Eur J Med Chem (1997) 32, 571–582

Quantitative structure–activity relationships in the 8-amino-6,7,8,9-tetrahydro-3*H*-benz[*e*]indole ring system. Analysis of serotonin 5-HT_{1A} effects in vivo and in vitro via partial least squares regression

LO Hansson[1], MD Ennis[2], P Stjernlöf[1]*

[1]*Institute of Physiology and Pharmacology, Department of Pharmacology, Medicinal Chemistry Unit, Göteborg University, Medicinaregatan 7, S-413 90 Sweden;*
[2]*Medicinal Chemistry Research, Upjohn Laboratories, Pharmacia & Upjohn Company, Kalamazoo, MI 49001, USA*

(Received 29 August 1996; accepted 16 December 1996)

Summary — The quantitative structure–activity relationship (QSAR) based on both in vivo activity and in vitro affinity data in a series of 51 differently substituted 8-amino-6,7,8,9-tetrahydro-3*H*-benz[*e*]indoles has been investigated by means of cross-validated partial least squares regression (PLS). The physicochemical description of the structures in the series was realized by utilizing variables derived from molecular mechanics and semi-empirical calculations. These descriptor variables were then used to model the affinity at the 5-HT_{1A} receptor and the effect of the drugs on 5-HTP synthesis in reserpine pretreated rats. An alternative strategy to the conventional alignment of energy-minimized structures, aimed at unravelling the biologically relevant directional influence of non-rigid structural elements, is proposed.

QSAR / PLS / physicochemical description / 5-HT_{1A} agonist / model evaluation

Introduction

In a series of articles we have reported on the central serotonergic (5-HT_{1A}, 5-HT_{1D}) and dopaminergic (D_2, D_3) activity of 8-amino-6,7,8,9-tetrahydro benzindoles [1–4]. Our interest in these compounds was focused on their serotonin 5-HT_{1A} activity, determined by Wikström et al, who reported that the *N,N*-dimethyl derivative **1** (fig 1) possessed considerable 5-HT_{1A} activity in addition to the previously known dopaminergic activity [5]. Later, it was reported that the *N,N*-dipropyl-1-formyl derivative **2** is a very potent, selective, but also a highly bioaccessible serotonin 5-HT_{1A} agonist [1, 2]. Even though the backbone of these compounds closely resembles that of the 5-HT_{1A} agonist prototype – the 8-substituted 2-*N,N*-dialkylaminotetralins (such as **8**-OH-DPAT) – there is no obvious similarity in the pattern by which the receptor affinities change between these structural classes when the aromatic substituent is changed. A striking example illustrating this is the difference in the results obtained upon introduction of an acetyl or a cyano group in these skeletons. In a paper by Liu et al, the effects of introducing these substituents into position 8 on the DPAT skeleton were described [6]. In this case, the acetylated compound (**8**-COMe-DPAT) was found to be more potent, displaying an affinity that was almost 30-fold higher than the corresponding 8-nitrile (**8**-CN-DPAT). In our recent articles, the activity at the 5-HT_{1A} receptor for a number of analogs of compound **2** is reported [3, 4]. In one of these studies, the 1-cyano substituted compound **3** was found to be among those compounds with the highest affinities, whereas the corresponding acetyl compound showed a 20-fold lower affinity [3]. Still, compound **2**, functionalized with a formyl group which has physicochemical features resembling those of an acetyl group, is one of the most potent compounds in this series. This dissimilarity in the structure–activity relationship (SAR) between these two structural families is intriguing, but may also shed some light on the processes underlying the drug–receptor interactions of these classes of compounds.

Several models of the drug–5-HT_{1A} receptor interactions explaining receptor affinity have been proposed [7–12]. Generally, these models were of a qualitative

*Correspondence and reprints

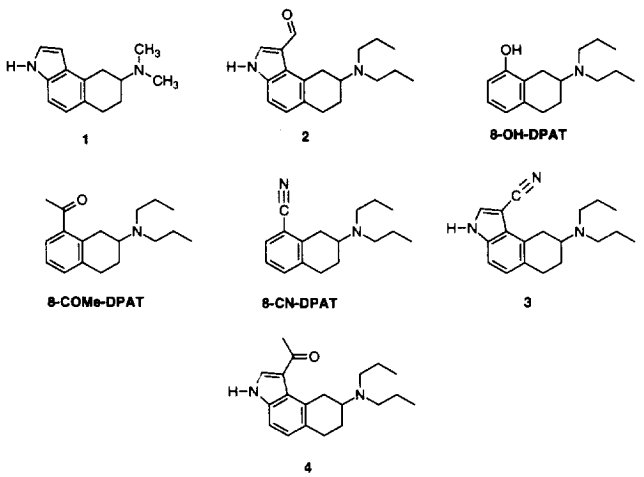


Fig 1. Structures **1**, **2**, **8**-OHDPAT, **8**-COMe-DPAT, **8**-CN-DPAT, **3** and **4** discussed in this study.

nature, and thus not useful for quantitative prediction purposes. We believed that the broad range of variation, along with the extensive biological testing reported in [3] and [4], could be useful in a partial least-squares regression – quantitative structure–activity relationship (PLS–QSAR) analysis of this class of 5-HT_{1A} active indoles. In this paper only racemic structures are discussed, but with a few exceptions. The reason for this is that only a limited number of enantiomerically pure 8-amino-6,7,8,9-tetrahydrobenzindoles has been reported. However, compounds within this class and within similar structural classes (ie, aminotetralins) generally show low or moderate stereoselectivity for 5-HT_{1A} receptors as long as no additional substitution in the cyclohexene part of the molecule is present [6, 10, 13, 14].

Partial least squares regression and related methods have gained increased use as statistical tools in QSAR analyses [15–19]. Compared to the more conventional methods, like multiple linear regression and stepwise multiple regression, PLS circumvents the problems of collinearity and multicollinearity [20, 21]. It also offers the obvious advantage of handling data sets where the number of independent variables is greater than the number of observations [22].

Until a few years ago, the conventional way to find variables representative of differences between the objects in a series of compounds has been to use tabulated data. Hammett constants (σ_p, σ_m), Verloop parameters and Hansch π-values are some examples of descriptors often used in this context. A problem with such sources of data is that tabulated data are invariably incomplete. If substances with uncommon substituents are a part of the series, it will generally be impossible to extract the necessary information for all substituents of interest. An alternative approach is to calculate variables with similar or perhaps higher informational content [23]. One method using this approach is the comparative molecular field analysis (CoMFA) [24], where information about directional influences, in contrast to conventional descriptors, is accounted for [17, 25–29]. However, CoMFA calculations are at present out of reach for most ordinary desktop computers (ie, IBM compatibles or Macintosh). For the purpose of generating descriptor variables carrying directional information without accessing programs and hardware handling CoMFA, we have used estimates of relative populations in different spatial directions occupied by rotatable groups (ie, energy profiles emerging from the analysis of dihedral torsional drivers, obtained by means of molecular mechanics calculations; see *Materials and methods*). This task was easily performed for small structural elements (chains containing a maximum of about five consecutive atoms) by ordinary molecular mechanics (MM) programs. However, longer chains did not lend themselves to this kind of treatment (see *Materials and methods*). During the course of this work, some papers dealing with dihedral torsional drivers in conjunction with PLS were published, although the implementations were somewhat different from those reported here [26, 30].

Another example of calculated descriptors included in this work is AM1 calculated proton affinity for the amino moiety of the drugs [31]. The estimation of this quantity allowed a judgement to be made on the relative basicity of the structures in an environment of low dielectric constant. This may be a more valid descriptor with respect to a central nervous system (CNS) receptor interaction, than the use of pK_a values (measured or calculated), because the bulk dielectric constant of CNS receptors is believed to be low [25, 32]. In an environment of high dielectric constant, ie, water or plasma, the amine basicity (ie, pK_a) is largely determined by the entropy of solvation. This results in quite different trends in basicity with structural changes, compared to that observed in a medium of low dielectric constant. Here we report on the use of calculated physicochemical variables as predictors for in vitro and in vivo serotonin 5-HT_{1A} biological activities by means of PLS.

Materials and methods

Biology

Synthesis and biological testing protocols both regarding in vivo and in vitro pharmacology have been reported elsewhere [3, 4]. In brief, the in vivo biological activities modelled were

the negative logarithm of the dose (mol/kg) required to obtain a half maximum lowering (pED_{50}) of the biosynthesis of 5-HTP (5-hydroxy tryptophan) in the limbic region of reserpine-pretreated rats. A few compounds (ie, **16**, **48** and **49**, table I) showed only partial agonism in the highest dose tested. For these compounds, approximated ED_{50} of 50, 20 and 20 μmol/kg respectively were used in the model. Compound **22** was previously reported to have partial effects at the highest dose tested [4]. The original data have now been reevaluated and an ED_{50} has been calculated after omission of a severe outlier. The in vitro measurement in this context was the negative logarithm of the inhibition constant (pK_i, M) for the drugs at the cloned 5-HT_{1A} receptor. The pK_i values were obtained by conventional binding experiments. A few values ($n = 3$) were obtained using calf hippocampus membranes instead of cloned cells. This obstacle was dealt with through by including this information as an independent indicator variable in the calculations (K_i values from cloned cell experiments were uniformly 3–5-fold lower where comparisons could be made). The compounds included in two different models (model 1, pK_i 5-HT_{1A} and model 2, pED_{50} 5-HTP levels) resulting from this work are listed in table I, together with data on their biological effects.

Molecular mechanics and semi-empirical calculations

General approach

The MM calculations performed employed MM+ force fields as implemented in the HyperChem™ [33] software. The semi-empirical (AM1) calculations were also performed with HyperChem, and were in all cases preceded by a molecular dynamics heating–cooling sequence (simulated annealing [32]), followed by a MM geometry optimization.

Dihedral drivers

The dihedral drivers run were in each case started from the lowest energy conformation found in a Random walk Monte Carlo conformational search [34, 35] (1000 iterations). All acyclic torsions in each molecule were included and ring torsional flexing (TFLEX) [36] was used for the cyclohexene part of the molecule in the conformational searches. In all cases identical skewed boat conformations of this part were found. The Metropolis criteria ([35] and references cited therein) were used for the acceptance of conformations. The ChemPlus™ 1.0a add-on module for HyperChem was used for both dihedral drivers and conformational searches. Structure refinements were in all cases obtained using a conjugate gradient minimizer according to the Polak–Ribiere method. The convergence criterion was always set to < 0.05 kJ/Å•mol.

PLS regression modelling

For the computation of PLS models we used the SIMCA-S software package [37]. Initially all variables were scaled to zero mean and unit variance (auto-scaling). The statistical significance of the models was judged by their cross-validated R^2 values (cumulated Q^2 or R^2_{cv} based on PRESS statistics) [21, 38, 39]. Evaluation of purely data fitting abilities of the models were performed by calculation of the standard deviation error of calculation (SDEC) [40]. To investigate the sensitivity of the regression coefficients to perturbations (exclusion or inclusion of data), observations (compounds) were excluded until each observation had been excluded exactly once from the original data set. The resulting N subsets, each consisting of N–1 observations, were then used to compute N regression submodels, always extracting the same number of PLS components as in the full model. The results from these submodels were then used to obtain estimates of the regression parameters and their errors. The method used here is known as the jackknife [41], and has been shown to yield unbiased estimates of a wide range of statistical parameters regardless of their distribution [42]. The jackknife estimate of a parameter θ is expressed as in eq (1) below, where the θ^* denotes the parameter derived from the full dataset and θ_i the parameter from the subset where the ith observation has been excluded. Eq (2) gives the jackknife estimate of the standard error $\sigma(\theta)$. In this formula, $\bar{\theta}_i$ denotes the average of the θ_is.

$$\theta = N(\theta^*) - \frac{N-1}{N}\sum_{i=1}^{N}\theta_i \qquad (1)$$

$$\sigma(\theta) = \sqrt{\frac{N-1}{N}\sum_{i=1}^{N}(\theta - \bar{\theta}_i)^2} \qquad (2)$$

The jackknife has been shown to reasonably well approximate the results obtained by the bootstrap [21, 43], another method commonly used in this context (ie, in the QSAR module of Sybyl, Tripos Associates Inc). However, we found the former easier to perform when using software not supporting such procedures for statistical evaluation of regression coefficients. The submodels obtained in the jackknife procedure were also used to make external predictions of the held-out observations from each submodel. In this way, external predictions of all compounds included in the full model were obtained.

The outcome of a PLS regression is scaling-dependent and the usual default scaling technique (auto-scaling) is not necessarily the optimum method [25, 44]. Therefore the scaling weights of the variables were optimized using a central composite design followed by evaluation of the resulting response surfaces [22, 45]. The different blocks of conceptually similar variables could then be assigned scaling weights offset from unit variance (S Wold, pers commun). The cross-validated R^2 was used as optimization criterion to maximize the predictive ability of the models rather than to improve the data-fitting ability. Using this procedure, an increased number of significant PLS components and some increase in the overall R^2_{cv} were obtained.

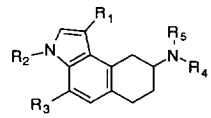
Physicochemical description of the molecules

The physicochemical description of the structures was obtained from MM and semi-empirical calculations on the different substructures of the drugs, as indicated in figure 2. Breaking up the structures into smaller subunits was found necessary to keep the computational times within reasonable limits. During this study, a variety of descriptors were calculated, some of which showed low or no contribution to the modelling of Y (pK_i or pED_{50}). The descriptors omitted at an early stage include partial and global clog$P$ values, HOMO and LUMO orbital energies, partial (Mulliken or Gasteiger) charges and dipole moments related to R_1. A summary of the variables finally used is given in table II.

Size variables (1, 31, 32, and 34–36; substructures A and C)

Van der Waals (VdW) volume (of R_1, R_4 and R_5) variables were calculated for the different substructures indicated in figure 2

Table I. Summary of structural variation of the compounds discussed and their observed and calculated activities[a].



Compound	Substitution					pK_i (M) 5-HT_{1A}			pED_{50} (mol/kg) 5-HTP_{limb}	
	R_1	R_2	R_3	R_4	R_5	Obs Model 1	Calc Model 1	Pred Model 1	Obs Model 2	Calc Model 2
1	-H	-H	-H	Me-	Me-	7.82	7.94	8.24	NT[b]	–
2	-CHO	-H	-H	Pr-	Pr-	9.70	9.29	9.23	6.96	6.90
3	-CN	-H	-H	Pr-	Pr-	9.40	8.86	8.56	6.77	6.43
4	$-COCH_3$	-H	-H	Pr-	Pr	7.85	8.07	8.28	4.81	5.39
5	-Me	-H	-H	Pr-	Pr-	7.89	8.38	8.43	NT	–
6	-*n*-Pr	-H	-F	Pr-	Pr-	7.96	7.85	7.15	NT	–
7	-Cl	-H	-H	Pr-	Pr-	8.70	8.41	8.38	5.85	6.24
8	-H	-H	-F	Pr-	Pr-	8.85	8.55	8.52	6.22	6.57
9	-H	-H	-H	Pr-	Pr-	8.48	8.59	8.60	7.02	6.65
10	-SMe	-H	-H	Pr-	Pr-	7.89	7.86	8.04	5.00	4.90
11	-CN	-H	-F	Pr-	Pr-	8.64	8.82	8.79	6.28	6.36
12	-CHO	-H	-F	Pr-	Pr-	9.30	9.25	9.24	7.10	6.81
13	$-SO_2Me$	-H	-H	Pr-	Pr-	7.96	7.90	7.86	NT	–
14	$-COC_2H_5$	-H	-F	Pr-	Pr	8.15	7.77	7.78	5.20	4.77
15	$-CH_2NMe_2$	-H	-H	Pr-	Pr	6.95	6.89	7.23	I[c]	–
16	$-COCONH_2$	-H	-H	Pr-	Pr	7.62	7.68	7.99	4.30[d]	4.35
17	$-COCONEt_2$	-H	-H	Pr-	Pr-	8.14	8.25	7.84	NT	–
18	$-CF_3CH_2$	-H	-H	Pr-	Pr-	7.75	7.90	7.91	I	–
19	$-COC_6H_6$	-H	-H	Pr-	Pr-	8.07	8.09	7.65	NT	–
20	$-COCF_3$	-H	-H	Pr-	Pr-	7.70	7.94	8.17	4.90	5.13
21	-H	-H	-H	2-Thioph$(CH_2)_2$-[e]	Pr-	9.22	8.85	8.80	NT	–
22	-H	-H	-H	Pr-	Allyl-	8.34	8.36	8.37	6.13[f]	6.30
23	-H	-H	-H	Pr-	Me-	7.85	8.10	8.27	NT	–
24	-H	-H	-H	*o*-MeOPh$(CH_2)_3$-[g]	Pr-	8.54	8.53	8.52	NT	–
25	-H	-H	-H	*i*-Bu[h]	Pr-	8.85	8.70	8.68	5.74	6.56
26	-H	-H	-H	*c*-PrCH_2-[i]	Pr-	8.60	8.53	8.53	6.96	6.41
27	-H	-H	-H	*m*-ClPhO$(CH_2)_3$-[j]	Pr-	7.92	8.36	8.64	NT	–
28	-H	-H	-H	$-CH_2-(CH_2)_4-CH_2-$		7.82	8.48	8.57	NT	–
29	-H	-H	-H	$-CH_2-(CH_2)_2-CH_2-$		9.10	9.10	8.92	7.04	6.43
30	-H	-H	-H	$PhCH_2$-	Pr-	8.72	8.91	8.92	NT	–
31	-H	-H	-H	$PhCH_2CH_2$-	Pr-	9.16	8.76	8.70	NT	–
32	-H	-H	-H	Di-MeGlu$(CH_2)_4$-[k]	Pr-	8.72	8.35	8.15	NT	–
33	-CHO	-H	-H	2-Thioph$(CH_2)_2$-	Pr-	10.00	9.56	9.49	7.09	6.69
34	-CHO	-H	-H	Pr-	Allyl-	9.52	9.06	8.98	6.72	6.55
35	-CHO	-H	-H	Pr-	Me-	9.16	8.80	8.71	5.98	6.07
36	-CHO	-H	-H	Pr-	H-	9.00	9.02	9.02	NT	–
37	-CHO	-H	-H	*o*-MeOPh$(CH_2)_3$-	Pr-	8.92	9.23	9.30	NT	–
38	-CHO	-H	-H	*i*-Bu-	Pr-	8.96	9.40	9.44	NT	–
39	-CHO	-H	-H	*c*-PrCH_2-	Pr-	9.22	9.24	9.23	6.64	6.66
40	-CHO	-H	-H	*m*-ClPhO$(CH_2)_3$-	Pr-	9.10	9.06	9.07	5.32	5.61
41	-CHO	-H	-H	$-CH_2-(CH_2)_4-CH_2-$		9.62	9.79	9.68	NT	–
42	-CHO	-H	-H	$-CH_2-(CH_2)_2-CH_2-$		10.16	10.42	10.02	6.24	6.68
43	-CHO	-H	-H	Me-	Me-	8.82	8.64	8.64	5.54	5.51
44	-CHO	-H	-H	$PhCH_2$-	Pr-	8.57	9.61	9.76	I	–
45	-CHO	-H	-H	$PhCH_2CH_2$-	Pr-	10.40	9.47	9.33	NT	–
46	-CHO	-H	-H	Ph$(CH_2)_3$-	Pr-	9.22	9.49	9.50	6.18	6.44
47	-CHO	-H	-H	Di-MeGlu$(CH_2)_4$-	Pr-	8.89	9.05	9.13	6.18	5.95
48	-CHO	-Me	-F	Pr-	Pr-	7.55	7.47	7.70	4.70[d]	4.55
49	-CHO	-Me	-H	Pr-	Pr-	7.40	7.49	7.81	4.70[d]	4.61
50	-H	-Me	-F	Pr-	Pr-	6.59	6.76	7.18	NT	–
51	-H	-H	-H	Pr-	H-	8.50	8.32	8.13	NT	–

[a]Statistical limits for all biological data are given in [3] and [4]; [b]NT: not tested; [c]I: inactive at the highest dose tested; [d]approximated values for compounds reported with partial effects; [e]thiofen-2-yl ethyl; [f]recalculated ED_{50}; [g]4-(2-methoxyphenyl)propyl; [h](*R*)-(+)-enantiomer; [i]cyclopropylmethyl; [j]3-(3-chlorophenoxy)propyl; [k]4-(4.4-dimethyl-2.6-dioxo-piperidin-1-yl)butyl.

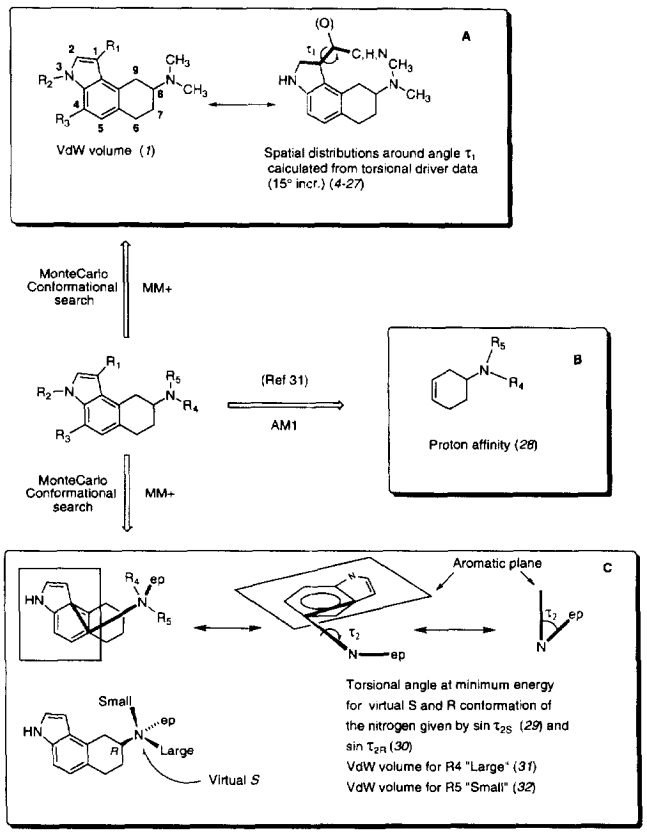


Fig 2. Graphical presentation of the calculated physicochemical descriptors.

using the method implemented in the ChemPlus 1.0a software. Because of the centering of data (zero mean) used in the subsequent calculations, the contribution to the volume by the unchanged parts will cancel out. These volumes were in all instances calculated without any substituents on the indole nitrogen. When the size of R_4 and R_5 differed, the R_4-substituent was always assigned as being the larger (see fig 2). The square and product terms *34–36* of *31* and *32* were also included.

Hydrogen bonding properties
Hydrogen bond acceptance ability at R_1 *(2)* was represented by an indicator variable (1/0). The ability of the indole NH moiety to function as a H-bond donor (substituted/unsubstituted) was similarly indicated *(3)*.

The torsional angle (τ_1) variables (4–27; substructure A)
The torsional angle τ_1 was defined as follows: eclipsing of the bulkiest group of R_1 with the indole C_2–C_1 double bond was defined as 0 degrees. However, an exception from this alignment was made for the formyl group, where the hydrogen (in this case the smaller part) was used as the point of reference. In the calculations on the oxalyl amides, the $-NH_2$ (**16**) and $-N(Et)_2$ (**17**) moieties as well as on the *n*-propyl (**6**) and the $-CH_2N(Me)_2$ (**15**) substituents, the next torsional angle was investigated by simultaneous rotation of both angles. The low energy trace for the first angle was then used in the final calculation of the relative populations in 24 angular directions, ie, using 15° increments. When R_1 was hydrogen, methyl, chloro, or cyano, it was assumed that they were equally populated in all directions, ie, ≈ 4.2% at each increment. The relative populations in the 24 different directions were then calculated from the torsional driver energy profiles, using Boltzmann's law of distribution at a temperature of 300 K, giving a 'Boltzmann-like' distribution. Thus, the basic approximation was made that these were the only possible states and that this calculation on non-relaxed structures did not distort the population estimates.

To further verify the validity of the τ_1 torsional driver variables, we performed a number of PLS regressions using only these as molecular descriptors and with the response data randomly scrambled. The results of these calculations where then compared with the results from a run using real response data (see *Results*).

Proton affinity (28; substructure B)
The amino nitrogen proton affinity (H^+_{aff}) was calculated exactly as described by Dewar et al [31] using AM1. This property is calculated as the difference between the sum of the ΔH_f of the free base and the proton and the ΔH_f of the protonated base (eq (3)). The experimentally determined ΔH_f of the proton (+ 367.2 kcal/mol) was used instead of the calculated value, since AM1 poorly approximated this quantity (+ 314.5 kcal/mol). These calculations were made on AM1 optimized structures both for protonated and unprotonated species with an *R* absolute configuration.

$$H^+_{aff} = \Delta H_{f,base} + \Delta H_{f,H} + - \Delta H_{f,protonated\ base} \quad (3)$$

Torsional angle τ_2 of the saturated amino moiety (29–30; substructure C)
To estimate the torsional angle τ_2 in a near global energy minimum conformation, all acyclic torsions in the N-ligands were included in a Monte Carlo conformational search performed as for τ_1. This procedure was performed on the C8 *R*-enantiomer separately for both invertomers (virtual *R* and *S*) of amine 'pseudo chirality'. The result from this modelling procedure gave rise to the title variables (see *Discussion*).

Additionally, an indicator variable (1/2) *(33)* was introduced to compensate for the differences between the two types of binding methods used (CHO cells or calf hippocampus membranes). All the above calculations were made using an IBM 486 or Pentium compatible computer.

Results

Variable scaling

After an initial investigation using unit variance scaling of all variables, the optimization of the scaling weights was performed using standard response surface techniques. The resulting optimized scaling weights are outlined in table III, expressed relative to the weights used in the unit variance scaling.

Table II. Summary of the calculated physicochemical descriptor variables used in the PLS regression models.

Variable No	*Identity*	*Substructure*	*Source*	*Units*
1	Volume of substructure A	A	MM+	Å^3
2	H-Acceptor ability of R_1	A	–	0/1
3	H-Donor ability of N-R_2	A	–	0/1
4–27	Relative population of the substituent R_1 at different torsional angles τ_1 (0–345° with 15° increments).	A	MM+	N
28	Calculated proton affinity for amine nitrogen (gaseous phase)	C	MM+/AM1	kcal/mol
29	Torsional angle at minimum energy for virtual *S* conformation of the amine nitrogen given by $\sin\tau_{2S}$	B	MM+ Monte Carlo conf search dihedr driv	–
30	Torsional angle at minimum energy for virtual *R* conformation of the amine nitrogen given by $\sin\tau_{2R}$	B	MM+ Monte Carlo conf search dihedr driv	–
31	Volume of substituent R_4	B	MM+	Å^3
32	Volume of substituent R_5	B	MM+	Å^3
33	Indicator variable (0/1) representing type of binding experiment (calf hippocampus or cloned cells)	–	–	0/1
34	Square term of *31*	–	–	Å^6
35	Square term of *32*	–	–	Å^6
36	Product term of *31* and *32*	–	–	Å^6

Table III. Optimized variable scaling of the two models.

Variable	*Scaling rel unit variance model 1*	*Scaling rel unit variance model 2*
1	1.20	1.20
2	1.00	1.00
3	0.50	1.00
4–27	0.75	0.75
28	1.30	1.60
29–30	1.00	1.00
31–32	1.30	1.30
33	0.50	Not included[a]
34–36	1.00	1.00

[a]Variable *33* (see table II) indicating the binding method is irrelevant in this context.

Model 1

Table IV shows the statistics for the 51-compound PLS model calculated using all the *X* variables listed in table II and the pK_i values at the serotonin 5-HT_{1A} receptor as the response variable *Y*. The observed, calculated and externally predicted values are given in table I and the relation between them is shown in figures 3a,b. The jackknife estimates of the regression coefficients with their standard errors are presented in figure 4. The approximate contribution to the explanation of the response variable variance by the different logical molecular entities in the model could be described via the VINFM (variable *in*fluence on the *m*odel) parameter [21]. This parameter reflects the proportion of variance explained in the dependent

Table IV. Summary statistics of the 51-compound cross-validated PLS model ($Y = pK_i$, 5-HT_{1A}).

	PLS						*Total*
	0	*1*	*2*	*3*	*4*	*5*	
Explained SS*X* (component)[a]	—	0.260	0.220	0.162	0.096	0.041	
Explained SS*X* (cumulated)	—	0.260	0.480	0.642	0.737	0.778	0.778
Explained SS*Y* (component)[b]	—	0.330	0.136	0.110	0.127	0.124	
Explained SS*Y* (cumulated)	—	0.330	0.467	0.577	0.703	0.827	0.827
Cross-validation R^2_{cv} (component)[c]	—	0.263	0.108	0.134	0.186	0.332	
Cross-validation R^2_{cv} (cumulated)	—	0.236	0.342	0.430	0.536	0.691	0.691
Standard deviation error of calculation (SDEC)[d]	0.792	0.648	0.579	0.515	0.432	0.329	0.329
Adj SDEC[e]	0.800	0.661	0.596	0.537	0.454	0.350	0.350

[a]Sum of squares explained by the component(s) in the independent (X) variable block; [b]sum of squares explained by the component(s) in the dependent (Y) variable block; [c]cross-validation performed with seven cross-validated (cv) groups, which gives approximately a leave-seven-out-at-a-time validation procedure; [d]SDEC is defined as: $\left[\frac{\sum_{i=1}^{N}(y_i - y_{i,calc})^2}{N}\right]^{1/2}$ where N is the number of objects, Y_i the observed and $Y_{i,calc}$ the calculated logK_i value [40]; [e]SDEC (adj) is defined as: $\left[\frac{\sum_{i=1}^{N}(y_i - y_{i,calc})^2}{N - A - 1}\right]^{1/2}$ where A is the number of PLS components extracted (adjusted for loss of degrees of freedom).

variable (pK_i, pED_{50}) by each of the independent variables. Thus, it was revealed that 9.4% of the response was related to variations in size and hydrogen bonding properties of R_1 (var *1–2*), whereas the relative population variables of R_{1A} (var *4–27*) accounted for 35.1%. Another 7.7% was explained by the indole nitrogen substitution (var *3*) and an additional 29.3% by the variables related to the saturated amine moiety (var *28–32, 34–36*). The variable related to binding method (*33*) was found to be relatively influential. However, when omitting the three compounds tested in calf hippocampus membranes, only a slight decline in performance of the model ($R^2_{cv} = 0.63$) was found.

Validation of torsional variables (4–27)

An indication of whether the importance of τ_1 in model 1 occurred by chance or not is given in figure 7. This figure shows a comparison of the resulting R^2_{cv} in the scrambled (n = 10) and real data runs respectively. Five PLS components were extracted for each case in this comparison. The results indicate that the importance of τ_1 is unlikely to be a chance occurrence. The jackknifed regression coefficients (obtained in the real data run) for each angular increment are shown with their standard errors in figure 8. It should be noted that the pattern of these coefficients is largely the same as that from the full model 1 (fig 4).

Model 2

Table V shows the statistics for the 27-compound in vivo PLS model. As is evident from figure 5, the predictive ability of 5-HT_{1A} receptor binding (pK_i) for the in vivo response (pED_{50} for the 5-HTP accumulation, limbic regions) is fair. The PLS regression model appeared, if judged by the data-fitting qualities (R^2 = 0.84; figure 6), as a superior predictor. However, when judged by the R^2_{cv} (see table V), the predictive capability was only slightly better to that of the pK_i values (R^2 = 0.61). Notable is, that adding the actual pK_i values to the physicochemical description did not result in any improvement of the model.

Discussion and conclusion

Similar variable influence patterns were found for both models. When performing simple linear regression analyses of the relation between the observed and the calculated Y's, the resulting equations showed slopes close to 1 and intercepts close to 0 for both models. This indicates that no systematic over- or underestimation is at hand over the entire ranges of the models.

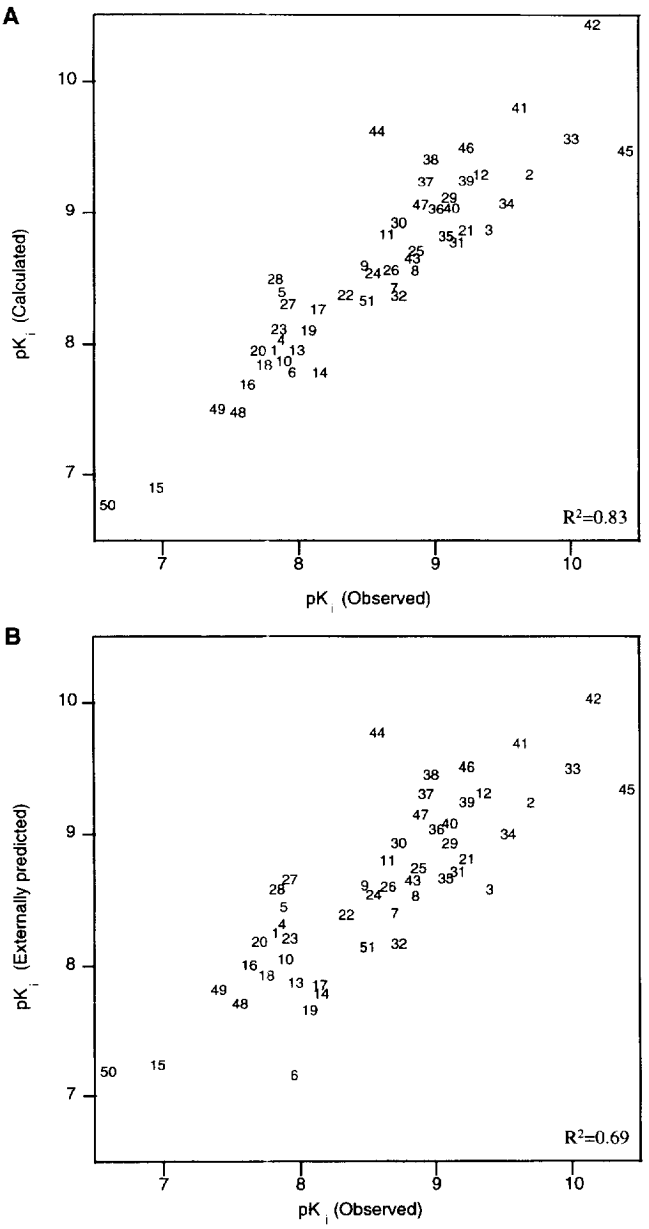


Fig 3. a. Observed vs calculated pK_i for 5-HT_{1A} binding affinities (model 1). **b.** Observed vs externally predicted pK_i for 5-HT_{1A} binding affinities (model 1).

Optimization of variable scaling weights

The optimization of the scaling weights used in the models was found to slightly improve the predictive ability of the model compared to when unit variance was used.

Variables related to R_1, R_4 and R_5 sizes (1, 31–32 and 34–36)

The size of the substituent R_1 should preferably be small, although this may be of minor importance. Judging the coefficients *31–32* and *34–36* jointly suggests that there might be an optimum size of R_4 around 500 Å^3, whereas the size requirement for R_5 is not easily deduced due to the limited variation (H, Me, Pr) of this property. However, in this range the variation in size of R_5 does not seem to influence the response.

H-bond acceptance ability R_1 (2)

The variation of the H-bond donor property of R_1 is not sufficiently large in the present series to draw any confident conclusions about effects related to this property (only one example of H-bond donors) and was therefore excluded. Nevertheless, the presence of H-bond acceptance properties at this position has a definite positive influence on affinity (see fig 4). This may, however, be a slight underestimate because several H-bond accepting structures show low affinities due to unfavourable conformations (τ_1 angle; see below).

H-bond properties of the indole nitrogen (3)

This variable accounts alone for the greatest influence on the model (7.7%). The effect of alkylation in this position has been pointed out earlier [3]. The low affinities of compounds **48**, **49** and **50** strongly suggest that the introduction of an indole *N*-methyl causes either a steric repulsive interaction with the receptor protein or the cancelling of an important H-bond donor possibility. An observation supporting the idea that the H-bond donor property is the more important one, is the comparison of the affinities of compound **50** and its indole *N*-benzylated analog (Stjernlöf and Ekman, unpubl res: 5-HT_{1A} receptor binding values (rat brain homogenate): K_i (**50**) = 280 nM, K_i (**7**) = 3.4 nM, K_i (**N–Bz**) = 880 nM). Here it was found that the benzyl compound showed a 3-fold lower affinity compared to **50**, whereas the indole *N*-methylation of **8** (yielding **50**) resulted in an 80-fold reduction in affinity.

Relative population at different dihedral angles of R_1 (4–27)

Another highly influential factor found in the present QSAR analysis was, when added together, the relative population of R_1 in different angular directions. Compounds with a dihedral angular segment ranging from about 165 to 195 ° are highly populated and generally display higher in vivo activities and in vitro affinities than compounds unable to adopt these conformations. With respect to the heterocyclic back-

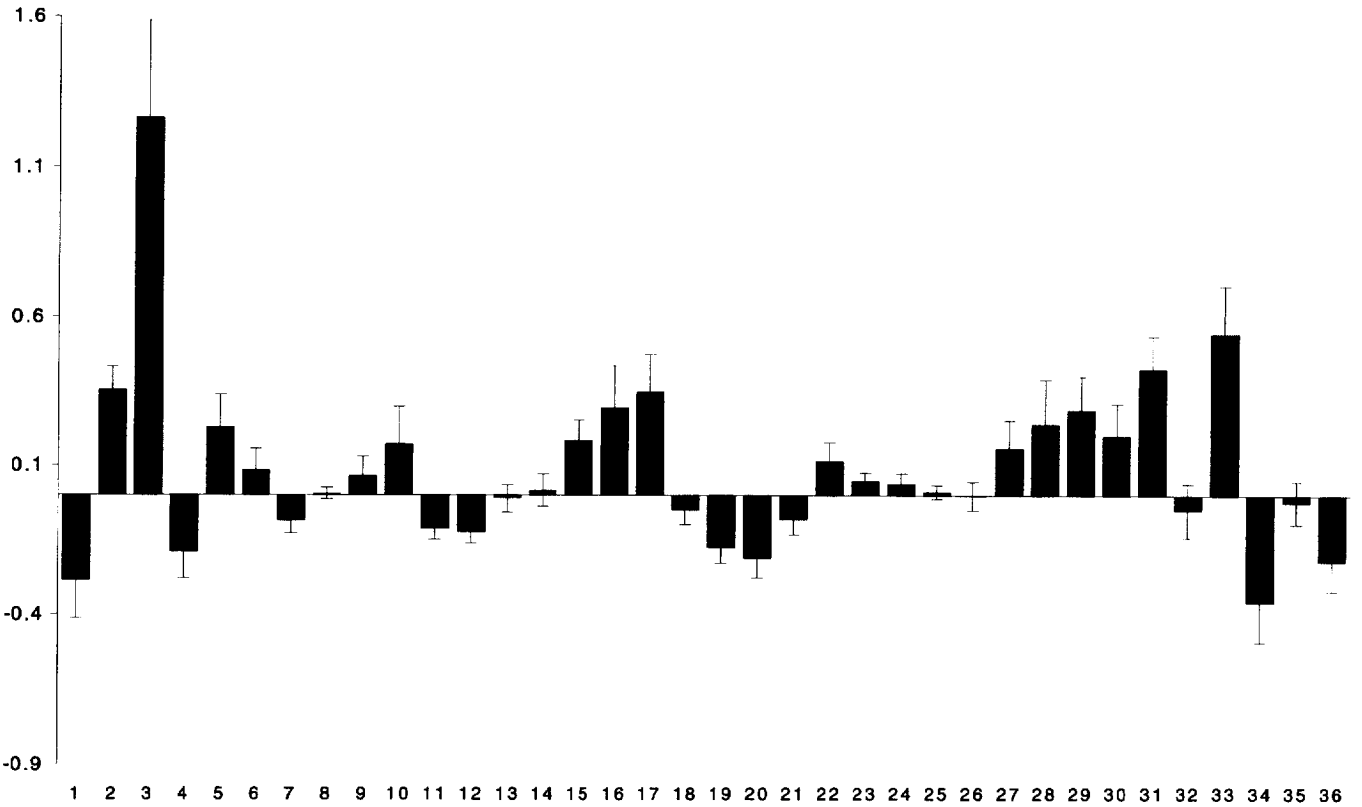


Fig 4. Jackknifed estimates of regression coefficients with standard deviations (model 1).

Table V. Summary statistics of the 27-compound cross-validated PLS model ($Y = pED_{50}$, limbic 5-HTP).

	PLS						*Total*
	0	*1*	*2*	*3*	*4*	*5*	
Explained SS*X* (component)[a]	–	0.377	0.208	0.147	0.081	0.061	
Explained SS*X* (cumulated)	–	0.377	0.585	0.732	0.814	0.875	0.875
Explained SS*Y* (component)[b]	–	0.344	0.227	0.145	0.090	0.030	
Explained SS*Y* (cumulated)	–	0.344	0.571	0.716	0.806	0.836	0.836
Cross-validation R^2_{cv} (component)[c]	—	0.215	0.204	0.273	0.156	0.064	
Cross-validation R^2_{cv} (cumulated)	—	0.215	0.375	0.546	0.617	0.641	0.641
Standard deviation error of calculation (SDEC)[d]	0.848	0.687	0.556	0.452	0.373	0.343	0.343
Adj SDEC[d]	0.864	0.714	0.589	0.537	0.490	0.413	0.389

[a,b,d]See table IV; [c]cross-validation performed with seven cv-groups which gives approximately a leave-four-out-at-a-time validation procedure.

bone, there are a variety of possible explanations. One of these is the obvious possibility that a high population at other torsional angles, beside those found to be reinforcing, would result in unfavourable sterical interactions at the binding site. However, the negative regression coefficients for angles around 230 ° show that effects that can be attributed to sterical crowding most likely exist, but are probably not the sole factor of importance. The observed impact of the VdW volumes of R_1 further supports the standpoint that the contribution of sterical effects is moderate within the range of substituent sizes investigated.

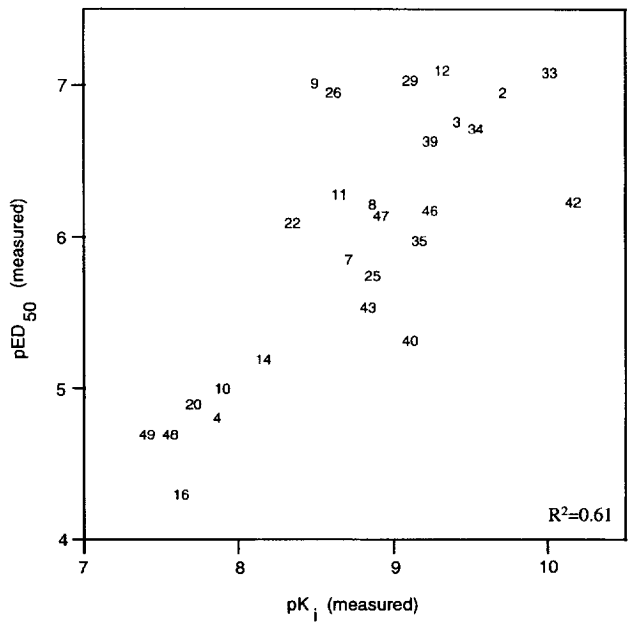


Fig 5. Observed pK_i for 5-HT_{1A} binding affinities vs observed pED_{50} for 5-HTP accumulation.

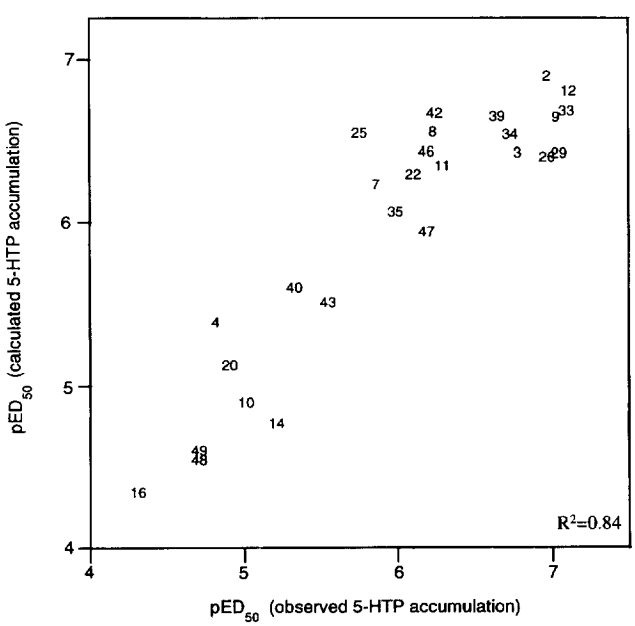


Fig 6. Observed vs calculated pED_{50} for 5-HTP accumulation (model 2).

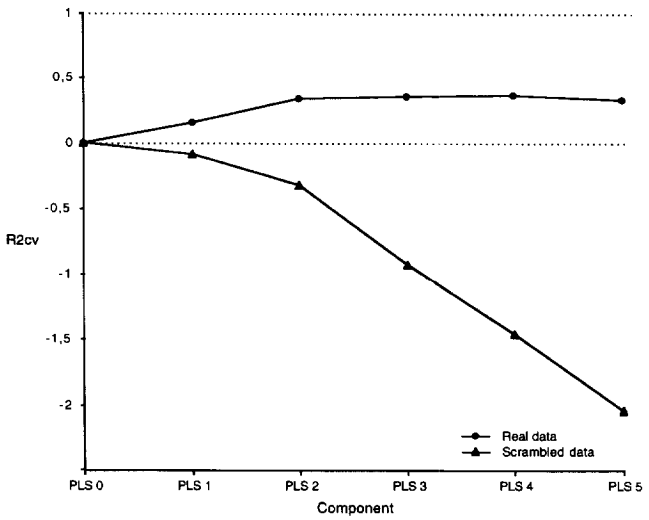


Fig 7. Comparison of R^2_{cv} for computation with true and scrambled Y data (n = 10). All models consisted of five PLS components regardless of significance.

Another aspect that may be important is the influence exerted by R_1 on the indole N–H bond properties, though mesomerically electron withdrawing groups in this position yield structures of a vinylogous amide type. Consequently, this will alter the hydrogen bond donor properties at this position (indole has a pK_a of ~ 17–18, whereas 3-formyl indole has a pK_a of ~ 12–13) [46]. This effect, of course, requires an overlap of the indole C3 and the R_1 carbonyl (or carbonyl like) atom p-orbitals. Efficient overlaps are possible when the R_1 dihedral angle is close to 0 or 180 °. Some structures, ie, oxalyl amides included in the analysis (compounds **16** and **17**) show high populations at angles between 165–195 °. Thus, these compounds would be expected to exert an influence on the N–H bond properties similar to that of formyl or cyano groups. However, in contrast to the high impact of highly populated states around 0 °, there is a negligible influence of a high relative population at dihedral angles around 180 °, which is shown by the small regression coefficients (fig 4, 8) in the PLS model (a positive sign reflects a positive contribution to the pK_i for high variable values). This is further substantiated by the fact that some compounds showing a large downfield ^{1}H-NMR shift for the indole N proton do not show very high affinities [3]. The only remaining reasonable explanation is that hydrogen bond acceptance in a relatively narrow directional range is causing the improvement in activity. This explanation is also compatible with the high affinity and activity reported for the 2-cyano substitu-

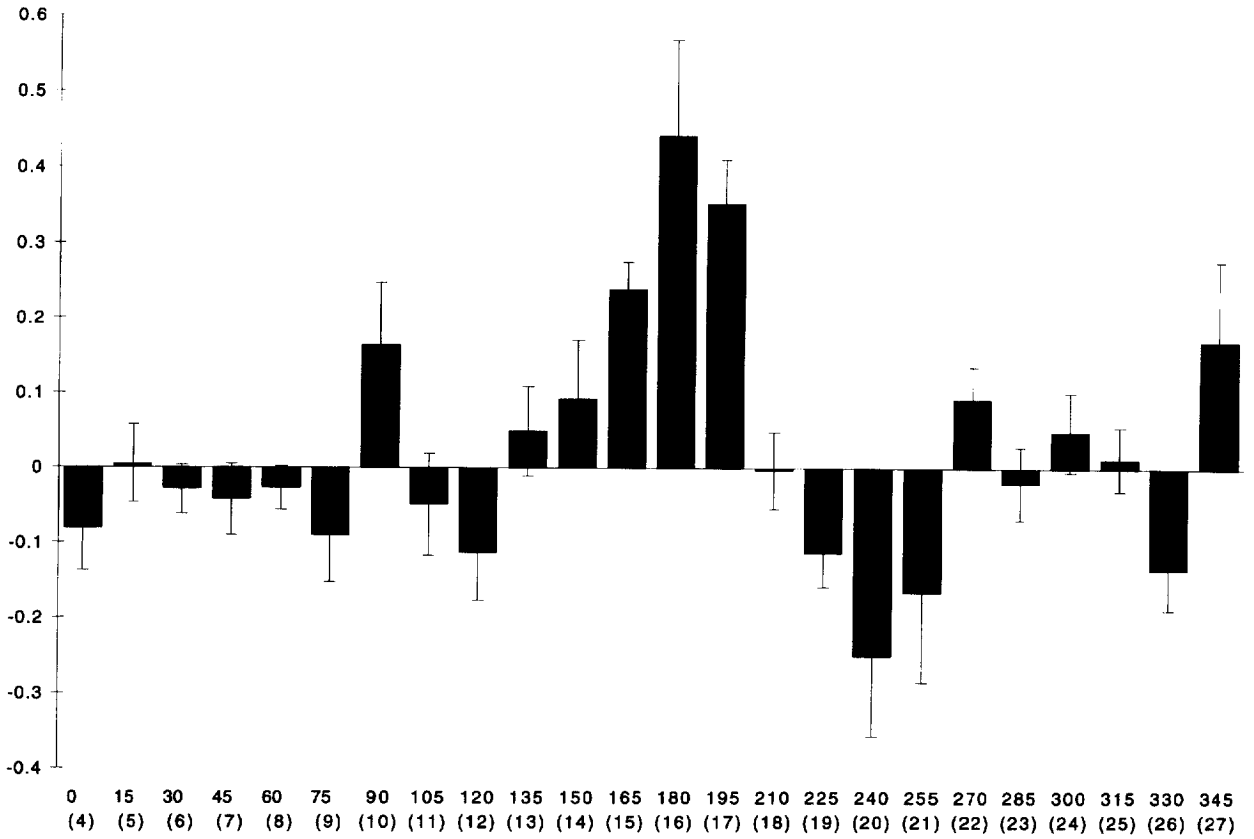


Fig 8. Jackknifed estimates of regression coefficients for each angular increment from true Y in the investigation of torsional angle τ_1 data. Below the angles are shown the respective variable number in parentheses.

ted analog [2, 47]. The maintained reinforcing effect on formylation of the indole *N*-methylated **50** yielding **48** (1 log unit enhancement) supports the above suggestion rather than any explanation founded on effects on the acidity of the indole N–H bond.

Proton affinity (28)

The PLS regression models clearly show that a high proton affinity of the amine nitrogen is favourable for both in vivo and in vitro activities. This is interesting because the calculations were made under in vacuo conditions (dielectric constant ~ 1.5), which may support the view that the environment at the ligand binding site of the receptor protein can be considered as hydrophobic. However, no comparison with calculations representing a high dielectric constant environment has been performed in this investigation.

Variables related to the torsion angle τ_2 (29 and 30)

The positive regression coefficients suggest that low energy for conformations corresponding to an antiperiplanar direction of the nitrogen lone pair, with respect to the α-hydrogen stabilizes the ligand–drug interaction. This would correspond to a position approximately perpendicular to the plane of the aromatic ring system. This finding appears to be valid for both of the possible invertomers, and also seems to be in agreement with previous observations [8–10, 12]. Initially, an attempt was made to treat this torsional angle in the same way as for τ_1 (relative population at different angles). This strategy was unsuccessful in this case, in as much as the dihedral drivers were unable to reproduce several of the local minima found in the Monte Carlo search. However, the torsion angle τ_2 recorded at the near global energy minimum conformations (those with the lowest energy found in the Monte Carlo search for both invertomers) was found to carry some additional information regarding directional influences. The observation that similar positive regression coefficient was found for both variable *29* and *30* implies that the large substituent can be pointing in two almost opposite directions without hampering the response. All of the above discussed observations are schematically summarized in figure 9.

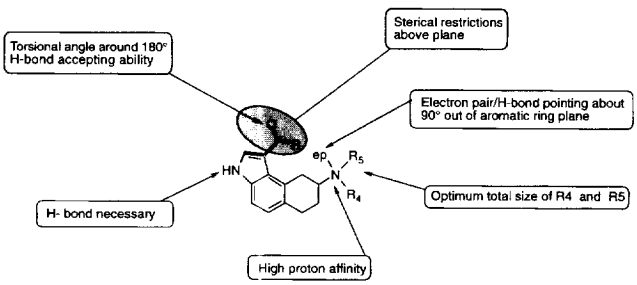


Fig 9. Schematic summary of the most significant QSAR findings.

In conclusion, we have been able to construct robust models, predictive for 5-HT_{1A} receptor binding affinity and *in* vivo activity for this set of compounds. The presented PLS models, behaved very well in all statistical respects, and have provided us with useful information regarding the impact of various physicochemical properties on 5-HT_{1A} receptor in vitro affinitiy and in vivo activity. The concept of optimisation of variable scaling weight in the PLS regression analysis appears to be useful, and may be worth further investigation. This study has also shown that it is possible to elucidate the directional influence of flexible molecular substructures, without any a priori assumptions about alignment, as long as a common point of reference can be established and the substructures considered in such contexts are of limited size. This approach may also be useful as a tool for unbiased analyses of molecular alignments prior to CoMFA investigations.

Acknowledgments

Financial support from The Upjohn Co, Kalamazoo, MI, USA is gratefully acknowledged. We also wish to express our gratitude to Prof Arvid Carlsson for providing excellent working environment.

References

1 Stjernlöf P, Gullme M, Elebring T et al (1993) *J Med Chem.* 36, 2059–2065
2 Stjernlöf P, Elebring T, Nilsson J et al (1994) *J Med Chem* 37, 3263–3273
3 Stjernlöf P, Ennis M, Hansson LO et al (1995) *J Med Chem* 38, 2202–2216
4 Ennis MD, Stjernlöf P, Hoffman RL et al (1995) *J Med Chem* 38, 2217–2230
5 Wikström H, Andersson B, Svensson A et al (1989) *J Med Chem* 32, 2273–2276
6 Liu Y, Yu H, Svensson BE, Cortizo L, Lewander T, Hacksell U (1993) *J Med Chem* 36, 4221–4229
7 Hibert MF, Gittos MW, Middlemiss DN, Mir AK, Fozard JR (1988) *J Med Chem* 31, 1087–1093
8 Hibert M, Midlemiss DN, Fozard JR (1987) *Brain 5-HT_{1A} Receptors.* Ellis Horwood Ltd, Chichester, UK, 27–33
9 Chidester CG (1993) *J Med Chem* 36, 1301–1315
10 Mellin C, Vallgårda J, Nelson DL et al (1991) *J Med Chem* 34, 497–510
11 Glennon RA, Westkaemper RB, Bartyzel P (1991) *In: Serotonin Receptor Subtypes: Basic and Clinical Aspects* (Peroutka SJ, ed) Wiley-Liss Inc, New York, 19–64
12 Hibert MF, McDermott I, Middlemiss DN, Mir AK, Fozard JR (1989) *Eur J Med Chem* 24, 31–37
13 Arvidsson LE, Johansson AM, Hacksell U et al (1987) *J Med Chem* 30, 2105–2109
14 Cornfield LJ, Lambert G, Arvidsson LE et al (1991) 39, 780–787
15 Norinder U (1993) *Eur J Med Chem* 28, 533–537
16 Hellberg S, Sjöström M, Skagerberg B, Wold S (1987) *J Med Chem* 30, 1126–1135
17 Norinder U (1993) *J Comput Aided Mol Des* 7, 671–682
18 Caldirola P, Coats E, Mannhold R, van der Goot H, Timmerman H (1993) *Eur J Med Chem* 28, 783–790
19 Waller CL, Oprea TI, Giolitti A, Marshall GR (1993) *J Med Chem* 36, 4152–4160
20 Topliss JG, Edwards RP (1979) *J Med Chem* 22, 1238–1244
21 Cramer III RD, Bunce JD, Patterson DE (1988) *Quant Struct–Act Relat* 7, 18–25
22 Wold S, Albano C, Dunn III WJ et al (1984) *In: NATO Adv Study in Chemometrics* (Kowalski B, ed) Reidel Publ Co, Dordrecht, The Netherlands, 17–95
23 Karelson M, Lobanov VS, Katritzky AR (1996) *Chem Rev* 96, 1027–1043
24 Cramer III RD, Patterson DE, Bunce JD (1988) *J Am Chem Soc* 110, 5959–5967
25 Davis AM, Gensmantel NP, Johansson E, Marriott DP (1994) *J Med Chem* 37, 963–972
26 Bureau R, Rault S, Robba M (1994) *Eur J Med Chem* 29, 487–494
27 Kim KH (1993) *J Comput Aided Mol Des* 7, 71–82
28 Waller CW, Marshall GR (1993) *J Med Chem* 36, 2390–2403
29 Debnath AK, Hansch C (1993) *J Med Chem* 36, 1007–1016
30 Oprea TI, Waller CL, Marshall GR (1994) *J Med Chem* 37, 2206–2215
31 Dewar MJS, Dieter KM (1986) *J Am Chem Soc* 108, 8075–8086
32 Mager PP (1994) *Eur J Med Chem* 29, 369–380
33 HyperChem™ 3.0. Autodesk Inc, Sausalito, CA, USA
34 Chang G, Guida WC, Still WC (1989) *J Am Chem Soc* 111, 4379–4386
35 von Freyberg B, Braun W (1991) *J Comput Chem* 12, 1065–1076
36 Kolossváry I, Guida WC (1993) *J Comput Chem* 14, 691–698
37 SIMCA-S 5.10. Umetri AB, Umeå, Sweden
38 Wold S (1978) *Technometrics* 20, 397–405
39 Clark M, Cramer III RD (1993) *Quant Struct–Act Relat* 12, 137–145
40 Cruciani G, Watson KA (1994) *J Med Chem* 37, 2589–2601
41 Miller RG (1974) *Biometrika* 61, 1–16
42 Miller RG (1964) *Ann Math Stat* 35, 1594–1605
43 Efron B (1979) *Ann Stat* 7, 1–26
44 Hansson LO, Waters N, Holm S, Sonesson C (1995) *J Med Chem* 38, 3121–3131
45 Box GEP, Hunter WG, Hunter JS (1978) *In: Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley & Sons Inc, New York
46 Remers WA, Brown RK (1972) *In: The Chemistry of Heterocyclic Compounds* (Houlihan WJ, ed) Wiley Interscience Inc, NY
47 Romero AG, Leiby JA, McCall RB, Piercy MF, Smith MW, Han F (1993) *J Med Chem* 36, 2066–2074