Journal of Computer-Aided Molecular Design, 6 (1992) 349–383
ESCOM

J-CAMD 164

Applications of rule-induction in the derivation of quantitative structure-activity relationships

Mohammed A-Razzak[a] and Robert C. Glen[b,*]

[a]*Infolink Decision Services Ltd., 9–11 Grosvenor Gardens, London SW1 W0BD, U.K.*
[b]*Department of Physical Sciences, Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, U.K.*

Received 26 November 1991
Accepted 23 April 1992

Key words: Artificial intelligence, CNS; Computational chemistry; GABA, Inotrope; Malaria, QSAR, Rule-induction

SUMMARY

Recently, methods have been developed in the field of Artificial Intelligence (AI), specifically in the expert systems area using rule-induction, designed to extract rules from data. We have applied these methods to the analysis of molecular series with the objective of generating rules which are predictive and reliable.

The input to rule-induction consists of a number of examples with known outcomes (a training set) and the output is a tree-structured series of rules. Unlike most other analysis methods, the results of the analysis are in the form of simple statements which can be easily interpreted. These are readily applied to new data giving both a classification and a probability of correctness.

Rule-induction has been applied to in-house generated and published QSAR datasets and the methodology, application and results of these analyses are discussed.

The results imply that in some cases it would be advantageous to use rule-induction as a complementary technique in addition to conventional statistical and pattern-recognition methods.

INTRODUCTION

The derivation of useful relationships between computed and measured molecular properties of molecules and their biological effects has been attempted using many different types of data generation and analysis methods. The objective is usually to increase understanding of the processes involved and to establish predictive descriptions relating molecular properties to biological actions.

Classical QSAR analyses generally utilise physicochemical substituent constants describing properties of molecules; trends within series may be discovered using statistical methods, usually correlation analysis. Recently, more diverse types of data have been generated based on computational chemistry methods [1]. Because of the large number of possible molecular descriptors which

* To whom correspondence should be addressed.

can be calculated this approach may result in underdetermined problems in which there may be a high ratio of molecular descriptors to cases. In addition much of the generated data may be noise which although accurately describing the molecules in a series has no direct bearing on their biological activities.

To overcome some of these deficiencies nonparametric statistics and pattern recognition methods have been developed [2] and applied to these problems. For example principal components analysis, cluster analysis, nearest neighbour etc. These are generally described as being either supervised (in which a model is fitted e.g. linear regression) or unsupervised (in which the data is not fitted to a model e.g. a nonlinear map). Although these methods have been and continue to be applied successfully to many problems, there is often difficulty in interpreting the results of the analyses and in formulating the next step in drug synthesis.

Recently methods have been developed in the field of Artificial Intelligence (AI) specifically in the expert systems area designed to extract rules from data [3,4]. The input is usually a number of test cases and the output is a tree-structured series of rules (a class probability tree). This is an example of a supervised learning method.

Applications are diverse and include, for example, electronics [5], agriculture [6] and engine testing [7]. The attraction of these rule-induction methods is that the rules generated during the induction phase can be easily interpreted and in addition can be applied later to new cases. The performance of new examples can be tested and the reasons for success or failure easily deduced.

The implementation of rule-induction used in the examples analysed here is in the package EXTRAN [8]. This is based on the ID3 [9] algorithm which is modified to allow pruning of 'bushy' rule trees using a χ^2 criterion.

METHOD

It is helpful to define some of the vocabulary used in rule-induction:

Induction: the task of detecting rules in the example set.

Objects: each data point.

Attributes: the properties that describe the objects.

Decision tree: the set of rules (or one rule).

CX: a modification of ID3 to introduce a χ^2 test at each decision point.

ID3 attempts to construct a simple decision tree from a number of objects. It may not be the best tree (the method is iterative) but it is usually compact and extracts the essence of the information contained in the examples.

If C contains p objects of class P and n objects of class N then an object will belong to class P with probability $p/(p+n)$ and to class N with probability $n/(p+n)$.

The information in a decision tree is therefore:

$$I(p,n) = -p/(p+n)\log_2 p/(p+n) - n/(p+n)\log_2 n/(p+n)$$

If attribute A with values (A_1, A_{1+1}...) is used for the root of the decision tree then it will partition C into (C_1, C_{i+1}...) where C_i contains those objects in C that have value A_i of A. Let C_i contain p_i objects of class P and n_i objects of class N. The expected information for the subtree for C_i is $I(p_i, n_i)$. The expected information required for the tree with A as a root is then obtained as the

weighted average

$$E(A) = \sum_{i=1}^{v} (p_i + n_i)/(p+n) I(p_i, n_i)$$

The information gain on branching on A is therefore

$$\text{gain}(A) = I(p,n) - E(A)$$

ID3 examines all the attributes, then chooses the one that maximises the gain. The same process is followed recursively over the rest of the attributes or until a suitable level of reliability is achieved. The objects are then divided into subsets depending on their value of that attribute.

In the case of missing data, a null value is recorded and incorporated into the information entropy calculation. For noisy data a χ^2 test for stochastic independence has been found to be useful.

Suppose attribute A produces subsets $(C_1, C_2, \ldots C_n)$ of C where C_i contains p_i and n_i of class P and N. If the value of A is irrelevant to the class of an object in C the expected value p'_i of p_i should be

$$p'_i = p(p_i + n_i)/(p+n)$$

If n'_i is the corresponding expected value of n_i, the statistic

$$\sum_{i=1}^{v} (p_i - p'_i)^2/p'_i + (n_i - n'_i)^2/n'_i$$

is approximately χ^2 with $v-1$ degrees of freedom. The tree building algorithm can then be modified to reject attributes whose irrelevance cannot be rejected with high confidence.

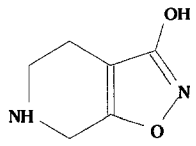
DATA ANALYSIS

Four datasets previously analysed by other methods were used to investigate the utility of rule-induction in the prediction of biological data. The datasets are typical of those analysed for structure–activity relationships and are composed of series of compounds with biological response results. The information content of some of the data (malaria data and CNS data) has been demonstrated by regression analysis, while pattern recognition methods have shown the relationships present in the GABA test set. The inotrope data shows clustering of compounds with similar biological effects in pattern recognition analyses but is a more difficult dataset for analysis (for example using regression) due to the embedded nature of the data.

Dataset 1. GABA data

GABA (4-aminobutyric acid) is an inhibitory neuro-transmitter concerned with the control of neuronal activity. A series of analogues of GABA were classified into agonist, weak agonist and inactives by Krogsgraad-Larsen et al. based on electrophysiological and binding studies [10].

In this study, the molecular structures were built using SYBYL [11] and conformationally restricted by least-squares fitting to THIP (4,5,6,7-tetrahydroisoxazolo[5,4-*c*]pyridin-3-ol),



THIP

a geometrically restricted active analogue, then geometry optimised using MOPAC [12] (AM1 PRECISE) on a VAX 85/50. The molecular properties were calculated using these optimised structures. Quantum-mechanically derived parameters were calculated from CNDO/2 calculations [13] on the neutral species. CNDO/2 was used to maintain compatibility with the previous pattern recognition analysis [14]. It should of course be borne in mind that in this example the zwitterions would certainly be present at neutral pH. (However, it is a much more complex and expensive task to calculate the zwitterionic species with solvent and counter ions.)

Principal component analysis of 7 computer-generated molecular properties for each of the 17 compounds showed [14] that a plot of the first two principal components, or a 2D nonlinear map clustered the molecules correctly into each class (for example the nonlinear map in Diagram 1).

We wished to see if rule-induction could achieve similar results in classification of the compounds. The data consists of 17 compounds, 24 descriptors (called attributes in AI research) and 3 classes (inactive, weak agonist and agonist) and is contained (along with chemical structures and a description of the attributes) in Appendix 1.

Each of the 17 examples is described by 24 numerical attributes in addition to the activity field.

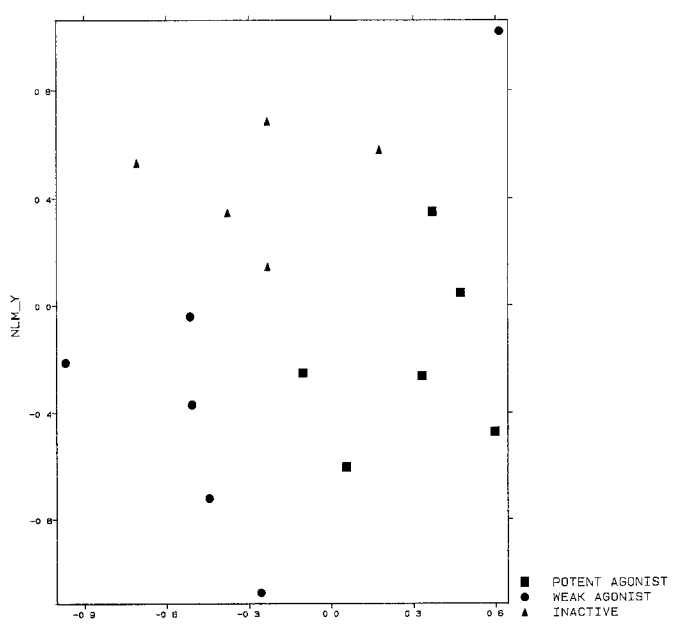


Diagram 1. GABA mimetics nonlinear map using dipole moment components, ellipsoid axes and surface area.

TABLE 1
CLASS PROBABILITY TREE USING AUTOMATIC INDUCTION ON THE GABA DATA

1)	[SURF_AR].			5,	6,	6.	17,	29 4 %
2)	|<162.15	[C_DIPMOM]:		0,	5,	6,	11,	0 0 %
3)	|	!<2.59:	1 0000	0,	5,	0,	5,	0.0 %
4)	|	|≥2 59	2.0000	0.	0,	6,	6,	0 0 %
5)	|≥162 15	[C_ATCH1]·		5,	1,	0,	6,	83.3 %
6)		|<−0.2241:	1 0000	0.	1,	0,	1,	0 0 %
7)		|≥−0.2241:	0 0000	5.	0,	0.	5,	100 %

Induction information

Node	Total Ex.	List of Examples																
1]	17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2]	11	1	2	3	4	5	6	8	9	10	11	12						
3]	5	8	9	10	11	12												
4]	6	1	2	3	4	5	6											
5]	6	7	13	14	15	16	17											
6]	1	7																
7]	5	13	14	15	16	17												

For the purposes of rule induction, the compounds were assigned the values 0 for inactive, 1 for weak agonist and 2 for agonist.

The result of automatic induction on the complete set of data gives an insight into rules embedded in the data. The class probability tree for automatic induction is shown in Table 1.

The rule structure is shown pictorially in Diagram 2.

Table 1 is the standard computer-generated output of the induction results and may be read by considering the first attribute in square brackets, [SURF_AR], which is the surface area in Angstroms (Å) of the molecules (the program has decided to 'split' first on surface area as this descrip-

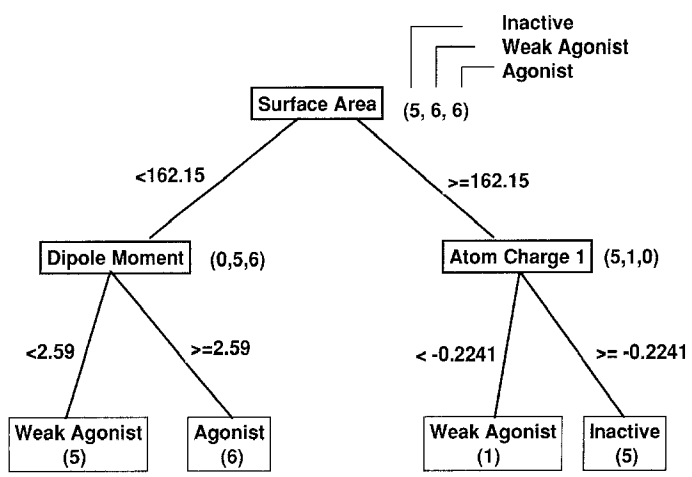


Diagram 2. Class probability tree for the GABA data.

tor contains the highest information entropy) and this is followed by three numbers. The first refers to the number in class '0', the second in class '1' and the third in class '2'. Next is the total number of examples, followed by the probability of classifying class '1' correctly.

On the next line, the numbers refer to the compounds left over after splitting on surface area. There are none of class '0', five of class '1' and 6 of class '2'. Following the other branch of the split on surface area to the bottom (line 5) we see that surface area $\geqslant 162.15$ Å^2 leaves 5 of class '0', one of class '1' and none of class '2'. This means that 'Compounds having a surface area greater than or equal to 162.15 Å^2 will probably be inactive, the probability, based on the training set, being 83% that this is true'. Similarly all the branches may be analysed to give the rule (which includes dipole moment in Debye (D)):

If SURF_AR $\geqslant 162.15$ Å^2 then probably inactive, possibly weak agonist. Otherwise, if C_DIPMOM $\leqslant 2.59$ D then weak agonist, otherwise agonist.

Based on the tree induced using the full example set (Table 1), it is clear that induction managed to classify the 3 classes very well and a clustering of each class appeared in the relevant branches of the induction tree. Only example 7 showed differences from the rest of the dataset. This is also seen in the principal component and nonlinear map analysis of the data, where this point appears as an outlier. The reason for this may be that the compound is not orientated correctly for this analysis or that the guanidinium group in this compound is not described well. It could even be that this compound has a different binding mode.

To investigate the stability of the induced rules on this very small dataset, 5 different training and test sets were generated randomly. Each training set was made up of 4 inactives, 4 weak agonists and 4 agonists. Obviously, we are attempting to generate rules from a very small set of data, however, in the early stages of drug research where we most often attempt to analyse this type of data, there are usually only a few tens of examples available.

Class probability trees were generated for the 5 training sets using the interactive CX induction algorithm with a confidence level of 90%. The test data were then filtered through the generated rules.

The experiments indicate that the predictive power of the final class probability trees was 64% while random guessing would yield 33% (for three classes).

The following points should be considered. With a small training set it is difficult to achieve stable prediction accuracy of the trees produced by induction. Also, the set of rules produced may have a different prediction success rate. Selecting a training set for induction may have an effect on the threshold values of attributes used in the tree. This may cause some examples in the training set to be misclassified because they fall just outside the threshold value. If compound 7 is not in the training set, then it will be misclassified.

Analysing the rules for some insight (speculative) into the role of molecular properties in recognition and binding of the molecules is possible. The surface area parameter is implying that there are size limitations (or steric restrictions) in fitting into the receptor. The dipole is separating agonists and weak agonists and suggests that smaller dipoles are more agonist-like. Perhaps it is easier for the receptor to orient those molecules with smaller dipoles for maximum agonist effect. The simple nature of the rules derived here is a reflection of the discriminatory power of a few parameters in the dataset. The other parameters are not required for classification in this case, which

does not mean that they are not important in describing important aspects of the molecules. As demonstrated in a later example, the influence of the other parameters can be investigated by forcing the induction algorithm to induce decision points on parameters in a user-defined sequence.

Dataset 2. Inotrope data

Molecules possessing inotropic properties increase the force of contraction of the heart muscle without increasing the rate, hence increasing the efficiency of the heart muscle. They may therefore be useful in cases of cardiac failure.

A previous analysis of a set of inotropic compounds using computational chemistry methods implied that potent compounds have particular structural and electronic requirements [15,16]. In particular, atom-centred charge density on the atom adjacent to the electrostatic potential minimum appeared to be related to inotropic potency of active molecules in the sulmazole and isomazole subsets.

In this analysis the molecular structures were built using SYBYL [11] and optimised using MOPAC [12] (AM1 PRECISE) on a Cray X-MP. They were all overlaid using the imidazo-pyridine ring system as a template.

As an example of analysis by pattern recognition methods, the 62 sulmazole and isomazole analogues displaying varying degrees of inotropic potency (pA_{50} measured as the log of the increase in the force of contraction of a guinea-pig papillary muscle upon administration of the drug) may be clustered reasonably well using a nonlinear map in 2 dimensions. This particular map was constructed from the complete parameter set and is shown in Diagram 3. (A more comprehensive pat-

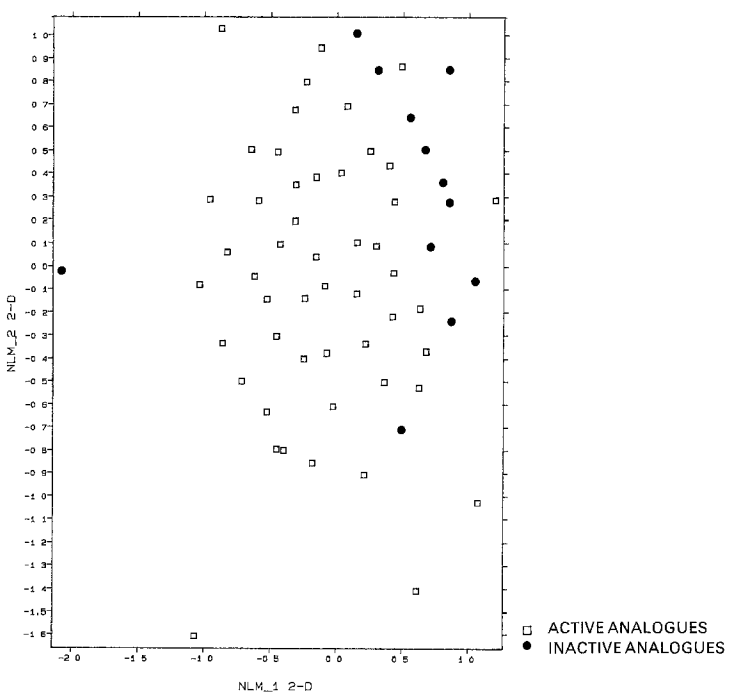


Diagram 3. 2D nonlinear map of potential inotropic compounds.

TABLE 2
CLASS PROBABILITY TREE OF THE INACTIVE/ACTIVE SULMAZOLE ANALOGUES

1)	[M_ATCH9]:					50,	12,	62,	19 4	%
2)	|< −0.10	[M_DIPMOM]				5,	9,	14,	64 3	%
3)	|	|<4 434	[M_ATCH1]:			5,	1,	6,	16 7	%
4)	|	|	|< −0.139.	ACTIVE		5,	0,	5,	0.0	%
5)	|	|	|⩾ −0 139	INACTIVE		0,	1,	1,	100	%
6)	|	|⩾4.434.	INACTIVE			0,	8,	8,	100	%
7)	|⩾ −0.10	[M_ATCH4]				45,	3,	48,	6 2	%
8)		|< −0.038:	ACTIVE			39,	0,	39,	0 0	%
9)		|⩾ −0 038	[P_E_AX_Y]			6,	3,	9,	33 3	%
10)			|<8.83:	[logP]		1,	3,	4,	75 0	%
11)			|	|<1.48	ACTIVE	1,	0,	1,	0.0	%
12)			|	|⩾1.48	INACTIVE	0,	3,	3,	100	%
13)			|⩾8.83	ACTIVE		5,	0,	5,	0.0	%

tern recognition analysis of this dataset using statistical and pattern recognition methods will be published elsewhere.)

This particular dataset is an example of embedded data in which one class of data is surrounded in parameter space by another class of data and this is particularly difficult to handle using for example, regression methods.

We wished to see if the rule-induction approach could give reliable results and offer insights into the role of different parameters in controlling inotropic potency.

The data consists of 62 compounds, 44 attributes and a numeric activity field and these are contained (along with chemical structures and a description of the attributes) in Appendix 1.

The data were divided into classes (potency of 0.0 to 2.0 and greater than 2.0). This was to differentiate between active and inactive molecules. The class probability tree for the full dataset using this threshold level is shown in Table 2 and pictorially in Diagram 4.

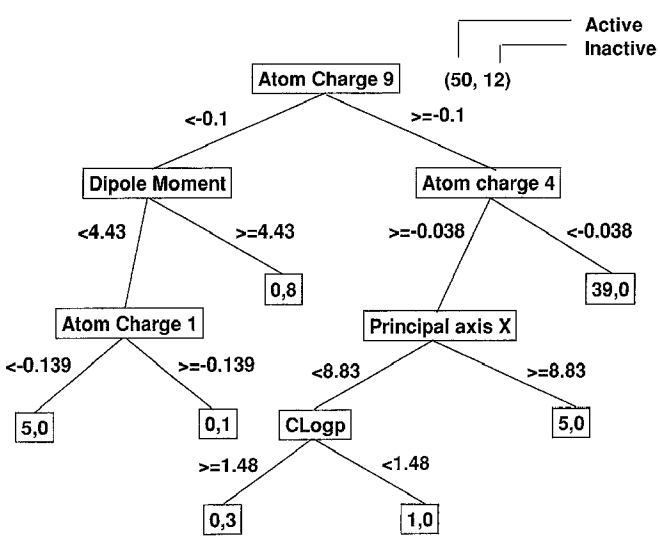


Diagram 4. Class probability tree for the active/inactive sulmazole analogues.

To test the performance of rule-induction in predicting active and inactive inotropes, 32 training and test sets were generated randomly by splitting the data as follows. The 50 actives were split into 34 for rule development and 16 for testing. The 12 inactives were split into 8 for rule development and 4 for testing. This gave 42 actives and inactives for rule development and 20 actives and inactives for testing in each of the 32 runs. Class probability trees were induced using the automatic CX procedure with a confidence level of 90%.

The average prediction rate over the 32 test runs showed that actives were correctly classified with 85% accuracy while inactives were correctly classified with 50% accuracy. This latter figure is much better than at first appears, as the random rate for correct prediction of inactives in this test series is calculated to be 19%. (No. of inactives/No. of compounds = 8/42 = 19%).

The nonlinear map shown earlier (Diagram 3) shows one representation of how the compounds are clustered in the parameter space. The map implied that this particular dataset did not clearly differentiate between actives and inactives, although distinct clustering of inactives is apparent. In applications where an unknown compound is classified using a graphical mapping of this type, the nearest-neighbour method is often used. In this approach, compounds may be classified according to the class membership of the nearest compounds to the unknown in the parameter space. This is obviously very difficult in this case, where compounds at the periphery of the 'inactive' volume are close to both actives and inactives. The performance of rule-induction is therefore good in view of the ambiguity in assigning compounds to one class or the other.

The inactives were removed from the dataset and threshold values of <3.5, 3.5–5.0 and >5.0 were chosen in an attempt at classifying levels of potency. The class probability tree for the actives in the dataset is shown in Table 3 and pictorially in Diagram 5.

Forty training and test sets were generated in a similar manner as before. The 50 actives were split into three groups of 15, 28 and 7 according to their potencies. The 15 were further split into 10 for rule-induction and 5 for testing. The 28 were split into 19 for rule-induction and 9 for test-

TABLE 3
CLASS PROBABILITY TREE OF THE ACTIVE INOTROPES

1)	[M_ATCH1]:							15,	28,	7,	50,	14 0%
2)	|< −0 17895	[M_HOMO_E]						0,	2,	7,	9,	77.8%
3)	|	<8.812.	gt 5 0					0,	0.	6,	6,	100. %
4)	|	|≥8 812	[logP]					0,	2,	1,	3,	33 3%
5)	|		|<1.697:	gt 5.0				0,	0,	1,	1,	100 %
6)	|		|≥1.697.	3.5–5.0				0,	2,	0,	2,	0 0%
7)	|≥ −0 17895	[M_LUMO_E]						15,	26,	0,	41,	0 0%
8)		|<0.6305:	lt 3.5					8,	0,	0,	8,	0 0%
9)		|≥0.6305:	[M_ESDL1]					7,	26,	0,	33,	0 0%
10)			|< −0 336·	3.5–5.0				0,	15,	0,	15,	0 0%
11)			|≥ −0.336:	[M_ATCH2]				7,	11,	0,	18,	0 0%
12)				|<0 0794:	[logP]:			4,	11,	0,	15,	0 0%
13)				|	|<1.09	3.5–5.0		0,	5,	0,	5,	0.0%
14)				|	|≥1.09:	[M_ESDL6]:		4,	6,	0,	10,	0.0%
15)				|		|< −0 248	lt 3 5	3,	2,	0,	5,	0 0%
16)				|		|≥ −0.248	3 5–5.0	1,	4,	0,	5,	0 0%
17)				|≥0.0794:	lt 3 5				3,	0,	0,	3,	0 0%

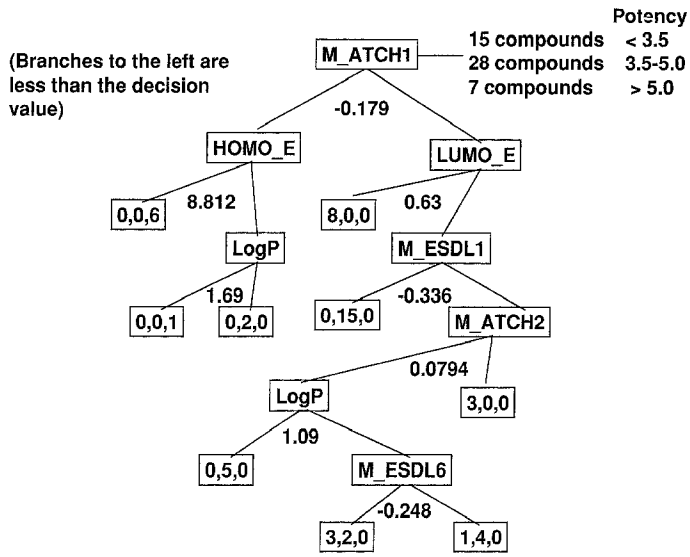


Diagram 5. Class probability tree for the active inotropes.

ing. The 7 were split into 4 for rule-induction and 3 for testing. This gave totals of 33 in the training set and 17 in the test set. Class probability trees were generated using automatic CX with a confidence level of 90% from each of the training sets and applied to their corresponding test sets. The prediction rates for correct classification averaged over the 40 runs were:

Potency <3.5	58%
3.5–5.0	62%
>5.0	79%

The most active compounds (the compounds which were of greatest interest in this case) appear to be predicted correctly with a high rate of success (79%). Clearly the rules induced from this analysis (Diagram 5) would serve as a very useful filter in deciding which compounds were most promising to synthesise.

Analysis of the class probability tree for potency classification (Table 3) shows that the level of potency is highly dependent on the charge on atom 1, the formally sp^2 nitrogen (in most structures) of the sulmazole ring. Structures that have an atom-centred charge of less than -0.179 electrons on atom 1 belong to the most active group. This is directly comparable to a previous QSAR analysis in which the formally sp^2 nitrogen atom-centred charge density was found to relate to inotropic potency as demonstrated by regression analysis [15,16].

Also the rule-induction analysis indicates that the calculated LUMO energy and the electrophilic superdelocalizability on atom 1 are important in separating the low- and medium-activity compounds. LogP also appears and is of less importance. However, logP does classify some of the remaining compounds. This was also seen in the QSAR analysis [15,16] in which low values of logP were associated with inactive compounds.

The class probability tree for the active/inactive set (Diagram 4) indicates that atom charge 1 is not important for classification here. This implies that, for example, compounds having the re-

quired electronic distribution for high potency (a sufficiently large negative charge on atom 1) may be rendered inactive by other factors. This was also seen in regression analysis [15,16] in which inactive compounds were removed as they appeared interspersed with actives in a plot of measured against predicted potencies.

The rule-induction analysis implies that if atom charge 9 is less than -0.10 electrons and the dipole moment is greater than 4.4 D then they are predicted to be inactive. The parameters selected by the induction algorithm are those giving the best classification. The dipole moment and the atom-charge distribution are influenced by the substitution pattern of the ring so it is possible that the selected parameters offer not the cause of inactivity, but may be reflections of the chemical modifications of the ring system.

Dataset 3. Malaria data

Sulfones and sulfonamides have been reported to act as inhibitors of dihydropteric acid synthase of various parasites including plasmodia. A systematic analysis using multiple regression and principal-component analysis was carried out by Weise et al. [17]. Observed biological data determined in cell-free and whole cell systems was tabulated (I_{50}, micromolar (μM)) along with several physicochemical descriptors. We examined the tabulated data pertinent to the cell-free folate synthesizing extracts of the mycobacterial strain *M. lufu*, as almost all the biological measurements were present (an explanation of the attributes is given in Appendix 1). A regression equation was derived by Weise et al.:

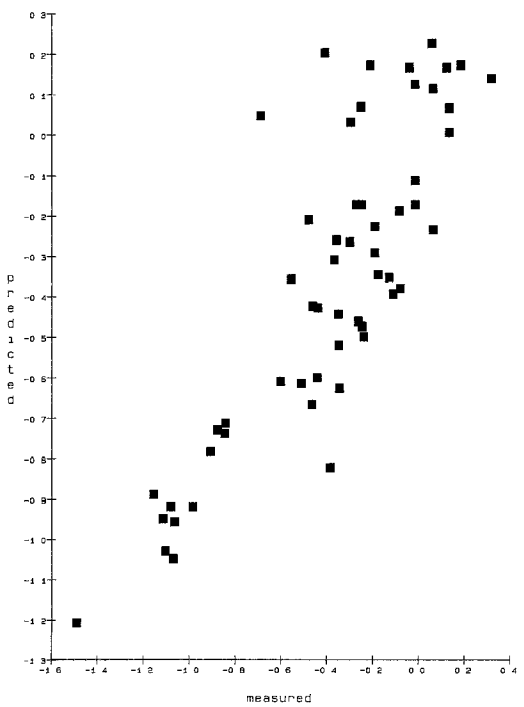


Diagram 6. Regression analysis on the *M. lufu* data.

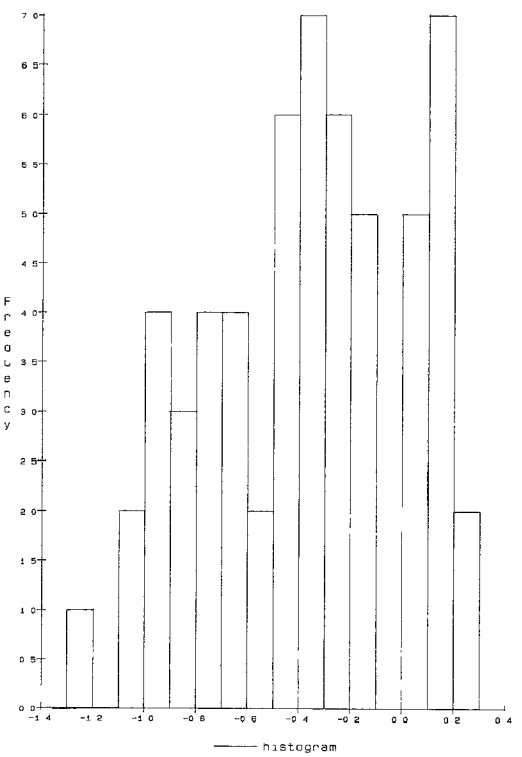


Diagram 7. Distribution of the *M. lufu* data.

$$\log 1/I_{50} = -4.98(0.42)D_{2/6}H + 0.098(0.022)\ MR - 0.0081(0.0017)\ MR^2 + 0.67(0.11)f_{ion} + 0.39(0.091)\ I_{OH} + 0.3(0.12)\ I_{Cl} - 1.01(0.059)$$
$$(n = 54, r = 0.905, s = 0.188, F = 25.67)$$

(n = number of examples, s is the standard deviation, r is the correlation coefficient, F is the ratio of explained to unexplained variance). A plot of calculated against observed activity for *M. lufu* is shown in Diagram 6. (This only contains points having all the data present.)

The major fraction of biological activity is explained by electronic substituent effects ($D_{2/6}H$) (changes in NMR chemical shifts) and a nonlinear dependence on MR (the molar refractivity) of the substituents in the 2′ position. The data are not scaled, so care must be taken in comparing substituents. We were interested in comparing the information obtained from rule-induction analyses with that from regression analysis.

Fiftyfour molecules were used in this analysis. Each was described by eight attributes in addition to the decision field (M_LUFU). These data (and a description of the attributes) are contained in Appendix 1. The classification was based on log(1/M_LUFU) with threshold values of −0.42 for the first set of experiments (splitting at the median point in the dataset) and −0.69 for the second set (classification of inactives) and at −0.05 (classification of most actives). A histogram of the distribution of the data is given in Diagram 7.

The objective was to classify molecules into the upper and lower bounds. Table 4 shows the class probability tree generated on the complete dataset by automatic induction with a threshold value of −0.42.

TABLE 4
CLASS PROBABILITY TREE USING AUTOMATIC INDUCTION ON THE *M. LUFU* DATA (THRESHOLD −0.42)

1)	[D2/6H] :							33,	21,	54, 61.1%
2)	| < −0 0695:	[VDWVOL]:						25,	5,	30, 83.3%
3)	|	| < 23 56	[logK]:					25,	2,	27, 92.6%
4)	|	|	| < −0.43815:	le −0 42				0,	1,	1, 0.0%
5)	|	|	| ⩾ −0.43815	[D2/6H]:				25,	1,	26, 96.2%
5)	|	|		| < −0.1015:	gt −0.42			18,	0,	18, 100. %
7)	|	|		| ⩾ −0.1015:	[D2/6H]:			7,	1,	8, 87.5%
8)	|	|			| < −0.0965:	le −0.42		0,	1,	1, 0.0%
9)	|	|			| ⩾ −0 0965:	gt −0 42		7,	0,	7, 100 %
10)	|	| ⩾ 23.56:	le −0.42					0,	3,	3, 0 0%
11)	| ⩾ −0 0695.	[logK]						8,	16,	24, 33 3%
12)		| < 1.34145:	[D2/6H]:					8,	6,	14, 57.1%
13)		|	| < 0 0205	[DNH2]·				8,	3,	11, 72.7%
14)		|	|	| < −0 231:	le −0 42			0,	2,	2, 0 0%
15)		|	|	| ⩾ −0.231:	[VDWVOL]·			8,	1,	9, 88.9%
16)		|	|		| < 5.68:	[DNH2]·		1,	1,	2, 50 0%
17)		|	|		|	| < −0 135	gt −0 42	1,	0,	1, 100. %
18)		|	|		|	| ⩾ −0 135	le −0.42	0,	1,	1, 0.0%
19)		|	|		| ⩾ 5.68	gt −0.42		7,	0.	7, 100. %
20)		|	| ⩾ 0 0205:	le −0.42				0,	3,	3, 0.0%
21)		| ⩾ 1.34145:	le −0.42					0,	10,	10, 0 0%

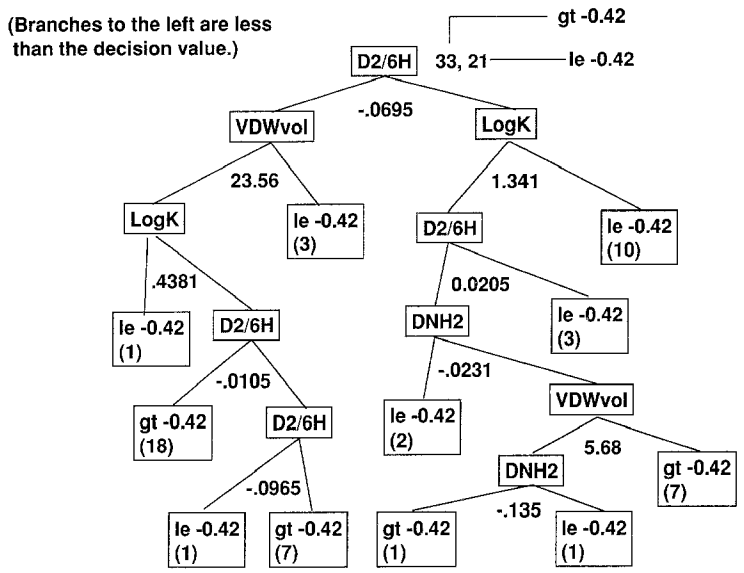


Diagram 8. Class probability tree of the *M. lufu* data (threshold −0.42).

Diagram 8 shows the tree pictorially. We can examine the rules to determine the main parameters classifying the more active set, for example. The principal split is on $D_{2/6}H$ (changes in NMR chemical shifts) followed by van der Waals volume (in Å^3) and logK (HPLC capacity factor). A large proportion of the actives are filtered through $D_{2/6}H$ having values of less than −0.0695 for $D_{2/6}H$. Further filtering by van der Waals volume (< 23.6 Å^3) and logK (≥ -0.438) gives 25 in the more active class to one in the less active class. This is a strong rule with a high classification probability. Analyses of this sort can offer insights into which parameters are most important in classification and also the threshold values separating classes.

In the first trial (−0.42 threshold) the data was broken randomly into 6 sets of 2/3 training set and 1/3 test set. The decision trees produced are in Table 5.

Table 6 contains the results of the test sets. The prediction rate was good (although variable), correctly predicting membership of the upper and lower areas at an average 80% and 74% success rate.

Analysis of the decision trees shows that the best descriptor (the most frequently selected) to split on is $D_{2/6}H$. Small negative values of $D_{2/6}H$ are selected as threshold values below which more active compounds are classified. This may be compared with the regression analysis in which the most influential contributor to this equation is also $D_{2/6}H$, with a large negative coefficient. This implies that small positive or negative values of $D_{2/6}H$ give more active compounds (higher log $1/I_{50}$).

Rule-induction is extracting similar information from the data as the regression equation. This may be interpreted in a structure-activity sense. Since ($D_{2/6}H = X -$ reference), $D_{2/6}H$ is negative for more shielding (nearer TMS) substituents. Shielded usually means more electron density around the proton, therefore this is a measure of the electron donation to this proton. More electron-donating substituents therefore give higher activity showing that the electronegativity of the substituents in the 2′ position is important.

TABLE 5
CLASS PROBABILITY TREES GENERATED FROM TEST SETS OF THE *M.LUFU* DATA (THRESHOLD −0.42)

1)	[D2/6H]				22,	14,	36,	61 1%
2)	| < −0 076	[VDWVOL]			18,	2,	20,	90 0%
3)	|	| < 22.07:	gt −0.42		18,	0,	18,	100 %
4)	|	| ⩾ 22.07:	le −0 42		0,	2,	2,	0 0%
5)	| ⩾ −0 076	[logK]:			4,	12,	16,	25.0%
6)		| < 1 34145:	gt −0 42		4,	4,	8,	50 0%
7)		| ⩾ 1 34145	le −0.42		0,	8,	8,	0 0%
1)	[VDWVOL]				22,	14,	36,	61.1%
2)	| < 5.68	[DNH2]:			4,	10,	14,	28 6%
3)	|	| < −0.134	gt −0 42		4,	3,	7,	57 1%
4)	|	| ⩾ −0 134:	le −0 42		0,	7,	7,	0 0%
5)	| ⩾ 5 68	[VDWVOL]:			18,	4,	22,	81 8%
6)		| < 21 785	gt −0 42		18,	0,	18,	100. %
7)		| ⩾ 21 785	le −0 42		0,	4,	4,	0.0%
1)	[D2/6H]				22,	14,	36,	61 1%
2)	| < 0.0205	[VDWVOL]			22,	9,	31,	71 0%
3)	|	| < 21 785:	[VDWVOL]:		21,	5,	26,	80 8%
4)	|	|	| < 5 68:	le −0 42	4,	5,	9,	44 4%
5)	|	|	| ⩾ 5 68:	gt −0 42	17,	0,	17,	100 %
6)	|	| ⩾ 21 785	le −0.42		1,	4,	5,	20 0%
7)	| ⩾ 0.0205	le −0 42			0,	5,	5,	0.0%
1)	[D2/6H]				22,	14,	36,	61.1%
2)	| < 0.0205	[VDWVOL]			22,	10,	32,	68 8%
3)	|	< 5.68:	[D2/6H]:		5,	7,	12,	41 7%
4)	|	|	| < −0.0795.	gt −0.42	5,	2,	7,	71 4%
5)	|	|	| ⩾ −0 0795	le −0.42	0,	5,	5,	0.0%
6)	|	| ⩾ 5.68:	[VDWVOL]:		17,	3,	20,	85 0%
7)	|		| < 21 785.	gt −0.42	16,	0,	16,	100 %
8)	|		| ⩾ 21 785	le −0.42	1,	3,	4,	25 0%
9)	| ⩾ 0.0205	le −0 42			0,	4,	4,	0.0%
1)	[VDWVOL]				22,	14,	36,	61 1%
2)	| < 5 68	[D2/6H]:			5,	10,	15,	33.3%
3)		| < −0 102:	gt −0.42		4,	0,	4,	100 %
4)	|	| ⩾ −0.102	le −0.42		1,	10,	11,	9.1%
5)	| ⩾ 5 68	[VDWVOL]			17,	4,	21,	81 0%
6)		| < 16 435:	gt −0 42		13,	0,	13,	100 %
7)		| ⩾ 16.435	gt −0.42		4,	4,	8,	50.0%
1)	[D2/6H]				22,	14,	36,	61 1%
2)	| < −0 0695	[VDWVOL]			17,	3,	20,	85.0%
3)	|	| < 23.63:	[logK]:		17,	1,	18,	94 4%
4)	|	|	| < −0.4139	le −0.42	0,	1,	1,	0 0%
5)	|	|	| ⩾ −0.4139	gt −0 42	17,	0,	17,	100 %
6)	|	| ⩾ 23 63.	le −0 42		0,	2,	2,	0 0%
7)	| ⩾ −0.0695	le −0 42			5,	11,	16,	31 3%

TABLE 6
TEST RESULTS ON THE *M. LUFU* DATA (THRESHOLD −0.42)

Experiment number	gt −0.42	le −0.42
1	11/11	3/7
2	10/11	3/7
3	6/11	7/7
4	10/11	7/7
5	8/11	5/7
6	8/11	6/7
	80%	74%

The next most important coefficient in the regression equation is for f_{ion}. However, in the induction analysis the next split is on van der Waals volume for substituents in the 2′ position. f_{ion} does not appear in any of the induced rules in the training sets. Van der Waals volumes less than 23.6 $Å^3$ are classified as more active in the cases where $D_{2/6}H$ is not enough for complete classification.

TABLE 7
CLASS PROBABILITY TREES GENERATED FROM TEST SETS OF THE *M. LUFU* DATA (THRESHOLD −0.69)

1)	[D2/6H] .			27,	9,	36,	75 0%
2)	|< −0.0385·	[DNH2]·		25.	1,	26,	96 2%
3)	|	|< −0 3645.	le −0.69	0.	1,	1,	0 0%
4)	|	⩾ −0 3645:	gt −0.69	25,	0,	25,	100 %
5)	|⩾ −0.0385·	le −0 69		2.	8,	10,	20 0%
1)	[DNH2] .			27,	9,	36.	75 0%
2)	|< −0 1275	[DNH2].		25.	1,	26.	96 2%
3)	|	|< −0 3625	le −0 69	0.	1,	1.	0 0%
4)	|	|⩾ −0 3625	gt −0.69	25,	0,	25,	100. %
5)	|⩾ −0 1275	le −0 69		2,	8,	10.	20 0%
1)	[DNH2]			27,	9,	36.	75 0%
2)	|< −0 1275	gt −0 69		25.	0.	25,	100. %
3)	|⩾ −0 1275.	le −0 69		2,	9,	11.	18.2%
1)	[D2/6H]			27,	9,	36.	75.0%
2)	|< −0.045 .	gt −69.0		21.	0,	21,	100 %
3)	|⩾ −0.045	[logK]·		6,	9.	15.	40 0%
4)	|	'< 1 34145·	gt −0 69	6,	4,	10,	60 0%
5)	|	|⩾ 1.34145	le −0 69	0,	5,	5.	0 0%
1)	[DNH2]			27.	9,	36.	75 0%
2)	|< −0.1275:	[VDWVOL]·		24.	2,	26.	92 3%
3)	|	|< 20.505	gt −0.69	22.	0,	22,	100 %
4)	|	|⩾ 20.505	gt −0 69	2,	2,	4,	50 0%
5)	|⩾ −0 1275	le −0.69		3.	7,	10,	30 0%

TABLE 8
TEST RESULTS ON THE *M. LUFU* DATA (THRESHOLD −0.69)

Experiment number	gt −0.69	le −0.69
1	10/14	4/4
2	11/14	3/4
3	11/14	2/4
4	12/14	2/4
5	11/14	3/4
	78%	70%

This implies a size restriction for substituents in this position.

The data was also split at −0.69 threshold value (near the lowest activity molecules) to examine the rules obtained for classification of low-activity molecules. The resulting decision trees of five random training sets are given in Table 7.

The results on 6 generated test sets are in Table 8.

Again, the results are encouraging, predicting compounds with activity values of greater than −0.69 78% of the time and less than −0.69 70% of the time.

In this case, the principal split is again on $D_{2/6}H$. However, the next split in these decision trees is usually on D_{NH}. For values of D_{NH} greater than about −0.36 (more positive) the more active molecules are selected (and therefore have electron-withdrawing substituents influencing this group).

Splitting at −0.05 (near the most active molecules) gives the class probability tree in Table 9. From this tree, if $D_{2/6}H$ is less than −0.092 and the van der Waals volume is in the range 5.68–17.42 $Å^3$ the probability is that they are in the most active set (75%), otherwise they are inactive.

In this example, many of the conclusions derived from regression analysis can also be inferred from rule-induction. However, in some respects rule-induction differs from regression analysis, in this case in the significance of less influential parameters.

TABLE 9
CLASS PROBABILITY TREE USING AUTOMATIC INDUCTION ON THE *M. LUFU* DATA (THRESHOLD −0.05)

1)	[D2/6H] .					43, 11, 54, 20 4 %
2)	| < −0.092·	[VDWVOL]:				16, 11, 27, 40.7 %
3)	|	| < 17 42.	[VDWVOL]·			10, 11, 21, 52 4 %
4)	|	|	| < 5.68:	gt −0 05		7, 2, 9, 22.2 %
5)	|	|	| ⩾ 5 68:	[VDWVOL]:		3, 9, 12, 75.0 %
6)	|	|		| < 14.87:	gt −0.05	0, 8, 8, 100. %
7)	|	|		| ⩾ 14.87	gt −0.05	3, 1, 4, 25 0 %
8)	|	| ⩾ 17 42.	le −0.05			6, 0, 6, 0.0 %
9)	| ⩾ −0.092·	le −0.05				27, 0, 27, 0.0 %

In this example, the actives have been classified (splitting = −0 05). From this tree, if D2/6H is less than −0 092 and the VDWVOL is in the range 5 68–17.42 the probability is that they are active (75%), otherwise inactive

Dataset 4. CNS data

The anticonvulsant and CNS-depressant activities of sixteen commercially available anti-epileptics were subject to regression analysis by Lien et al. [18]. We were interested in the analysis of the maximal electroshock data (MES). The dataset contains 16 compounds, 3 descriptors and the MES field. The data, compounds and a description of the attributes is contained in Appendix 1.

A simple correlation between the anticonvulsant activities and the physicochemical data was derived:

$$\log 1/C = 0.627(0.093)\log P + 2.58(0.16)$$
$$(n = 16, r = 0.76, s = 0.342)$$

A plot of calculated against observed datapoints is shown in Diagram 9.

Inclusion of additional parameters (for example a quadratic term and dipole moment) did not improve the fit to the data, for example:

$$\log 1/C = -0.03(0.135)\log P^2 + 0.714(0.32)\log P - 0.105(0.23)u + 2.727$$
$$(n = 16, r = 0.880, s = 0.362)$$

is not statistically significant.

When diazepam, clonazepam and carbamazepine were omitted (on reports that they interact

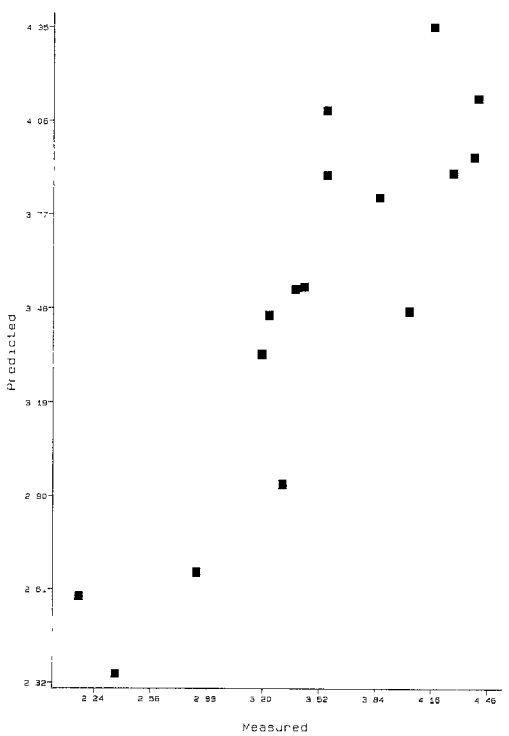


Diagram 9. Regression analysis 1 on the MES data.

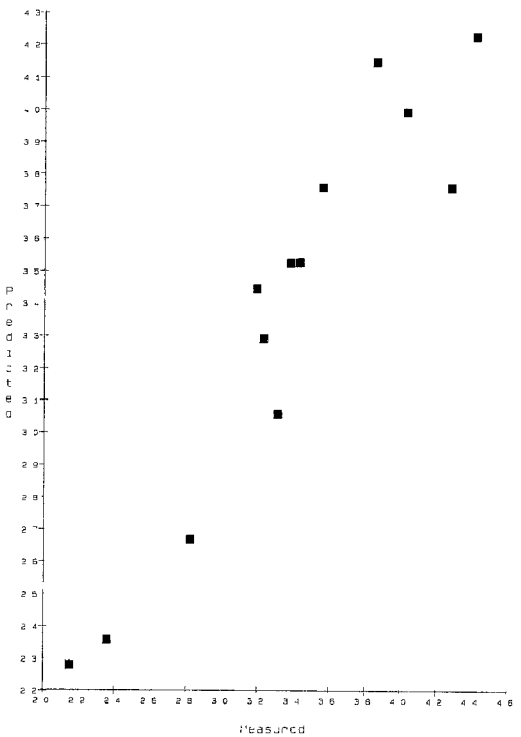


Diagram 10. Regression analysis 2 on the MES data

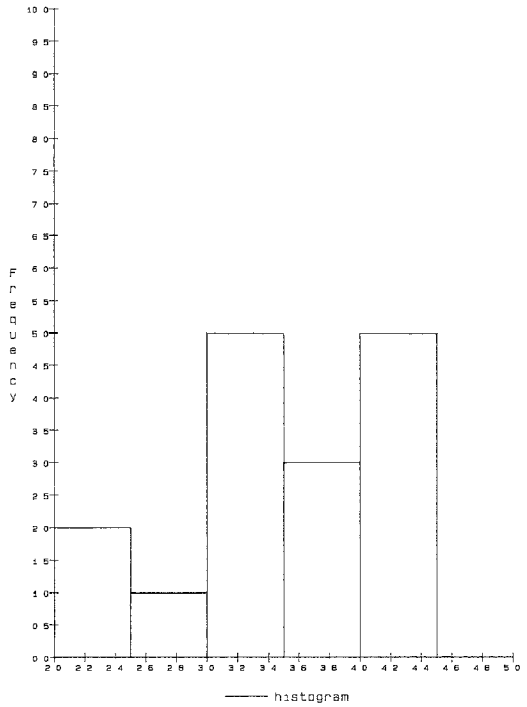


Diagram 11. Distribution of the MES data.

with different receptors) an improved fit of the data was obtained:

$$\log 1/C = 7.776(0.847)\log MW - 14.438(1.943)$$
$$(n = 13, r = 0.941, s = 0.241)$$

A plot of calculated against observed datapoints for this relationship is shown in Diagram 10.

The optimum lipophilicity was determined to be 1.42 from regression (derived from the parabolic equation) and the dependence on u (the dipole moment) was negative. We were interested to see if rule-induction could reduce the data in a similar fashion.

Initially the decision value for MES (maximal electroshock data) was selected as the mean, 3.5 (Diagram 11 shows a histogram of the data distribution).

Table 10 shows the decision tree obtained by automatic induction using all the data and Diagram 12 shows the class probability tree pictorially.

LogMW is the most significant descriptor. Using only logMW, 100% classification is obtained.

TABLE 10
CLASS PROBABILITY TREE USING AUTOMATIC INDUCTION ON THE MES DATA

1)	[logMW]		8.	8.	16.	50.0%
2)	|<2.325 :	lt3.5	8,	0,	8.	100. %
3)	|≥2.325 :	gt3.5	0.	8,	8.	0 0%

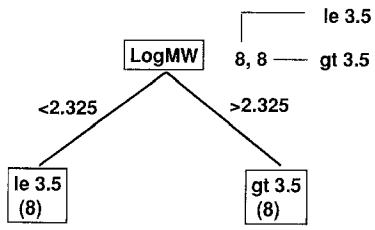


Diagram 12. Class probability tree of the MES data (threshold 3.5).

Compounds having a logMW value greater than 2.32 are classified as most active. This is interesting (but not surprising). The regression analysis of the full dataset indicated that logP and dipole moment were important but logMW was of lower significance, however, logP and logMW are highly correlated (r = 0.93).

When diazepam, clonazepam and carbamazepine were omitted from the regression analysis:

$$\log 1/C = 7.776(0.847)\text{logMW} - 14.438(1.943)$$
$$(n = 13, r = 0.941, s = 0.241)$$

The correlation coefficient has increased and now logMW has become the important descriptor in the regression equation, as was found in induction.

In order to obtain more information on a diverse set of parameters, or to view the data from a different viewpoint, induction may be forced to split on descriptors in a defined sequence. By forcing induction to split on logP first, the decision tree in Table 11 is obtained.

Interestingly, this gives about 90% classification of the full dataset assuming that the tree is pruned to remove nodes 3–4. This can be compared with the regression equation which used logP as a significant contributor (where the correlation coefficient was 0.88). The induction next splits

TABLE 11
CLASS PROBABILITY TREE USING FORCED INDUCTION ON LOGP FOR THE MES DATA

1)	[logP]			8,	8,	16,	50 0 %
2)	< 1 76	[logMW]:		8,	1,	9,	88.9 %
3)	|	|<2.34·	le 3 5	8,	0,	8,	100. %
4)	|	|≥2 34	gt 3.5	0.	1,	1,	0 0 %
5)	|≥1 76	gt 3 5		0.	7.	7,	0 0 %

TABLE 12
CLASS PROBABILITY TREE USING FORCED INDUCTION ON DIPOLE MOMENT FOR THE MES DATA

1)	[DIPOLE]			8,	8,	16,	50 0 %
2)	|<2 195	[logMW]:		8.	5.	13,	61 5 %
3)	'	|<2.325	le 3 5	8.	0,	0,	100 %
4)	,	|≥2 325.	gt 3.5	0.	5.	5,	0 0 %
5)	|≥2.195	gt 3 5		0.	3.	3.	0 0 %

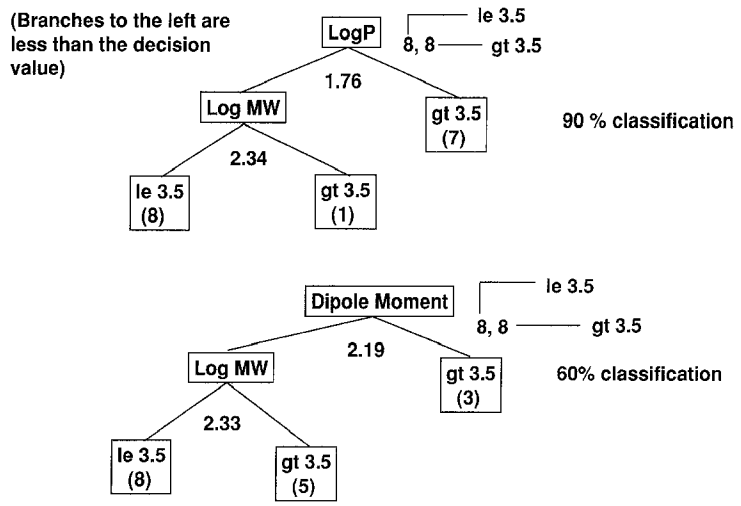


Diagram 13. Class probability trees for forced induction on the MES data.

on logMW while the next most significant contributor to the regression equation is dipole moment u.

The value for logP the induction algorithm split on was 1.76, which can be compared with the optimum value found for lipophilicity by regression which was 1.42. So, compounds with logP greater than 1.42 were in the more active set. The precise value of decision values is however, unstable in such small datasets and can change with the omission (or addition) of key compounds to the training set.

Splitting first on the dipole moment gives a lower classification (only 61%, induction tree shown in Table 12).

This is again comparable to regression in which the dipole was found to be less significant. The dependence on dipole was negative in the regression equation. This is not completely the case in rule-induction, where a number of compounds displaying a dipole greater than 2.19 D are classified more active using this descriptor. Perhaps this parameter only becomes influential in cases where other descriptors have values which enable the dipole to become significant. Diagram 13 shows pictorially the class probability trees from forced induction on this dataset.

Table 13 shows the decision tree obtained using automatic induction on all the data but classifying according to four ranges of MES (<3.0, 3.0–3.5, 3.5–4.0). The results are similar to the previous induced trees where splitting about the mean value was used. This shows that the more active compounds are classified, 3.5 and above dependent on logMW (>2.3) only.

This is also an interesting example in which to attempt exception programming [19]. This method is a modification of the induction algorithm and can be used to reveal combinations of physicochemical properties that are not present in the training set. It may therefore be useful in expanding the training set to cover more of the parameter space at minimum cost.

In this method, a decision tree is generated using the automatic CX induction algorithm to determine the decision points for each of the attributes. The values of the attributes are then converted from numeric to logical values, depending on threshold values appearing in the trees. A new class is created (undefined) for undefined compounds or combinations. The example sets are

TABLE 13
CLASS PROBABILITY TREE USING AUTOMATIC INDUCTION ON THE MES DATA (THRESHOLDS <3.0, 3.0–3.5, 3.5–4.0, >4 0)

1)	[logMW]·						3,	5,	3,	5,	16,	18.8%
2)	|<2.325 :	[logMW]·					3,	5,	0,	0,	8,	37 5%
3)	|	|<2.225:	lt 3				3,	0,	0,	0,	3,	100. %
4)	,	|≥2.225	3–3 5				0,	5,	0,	0,	5,	0.0%
5)	|≥2.325 ·	[logP]					0,	0,	3,	5,	8,	0 0%
6)		|<2 44	[logMW]				0,	0,	3,	3,	6,	0 0%
7)		|	|<2 38	[logMW]:			0,	0,	1,	3,	4,	0.0%
8)		!	!	|<2.355.	[logP]:		0,	0,	1,	1,	2,	0.0%
9)		!	|	|	|<2.095:	3.5–4	0,	0,	1,	0,	1,	0 0%
10)		|	|	|	|≥2.095·	gt 4	0,	0,	0,	1,	1,	0.0%
11)		|	|	|≥2 355	gt 4		0,	0,	0,	2,	2,	0 0%
12)		|	|≥2.38·	3 5–4			0,	0,	2,	0,	2,	0.0%
13)		|≥2.44:	gt 4				0,	0,	0,	2,	2,	0.0%

then entered as exceptions to the general case of undefined. Using automatic CX a new decision tree was generated (Table 14). The undefined points in the tree are cases which do not appear in the trial set (shown pictorially in Diagram 14). For example, there is no case of a compound with a dipole $\geqslant 2.195$ and a logP < 1.76 and a logMW < 3.325. This could be very useful in expanding a compound set to cover property space in a most efficient way.

CONCLUSIONS

Rule-induction using the CX algorithm can be an alternative method for extracting rules from data. The rules are compact and appear to offer insight into the role of molecular properties in the determination of biological activities. Induction appears to behave in some examples in a different way than, for example, regression and some pattern recognition methods in the determination of the relative contribution and importance of descriptors. The presentation of the results in the form of conditional statements is clear (from the point of view of deciding what molecule to make

TABLE 14
CLASS PROBABILITY TREE ON THE MES DATA SHOWING EXCEPTION PROGRAMMING

1)	[logMW]:			
2)	!<3 325 :	[logP]·		
3)	|	|<1 76	[DIPOLE]:	
4)	|	|	|<2.195	le 3 5
5)	|	|	|≥2 195	undef
6)	|	|≥1 76	undef	
7)	|≥3.325 :	[logP]:		
8)		|<1 76:	[DIPOLE]	
9)		|	|<2 195·	gt 3 5
10)		|	|≥2 195·	undef
11)		|≥1 76	gt 3 5	

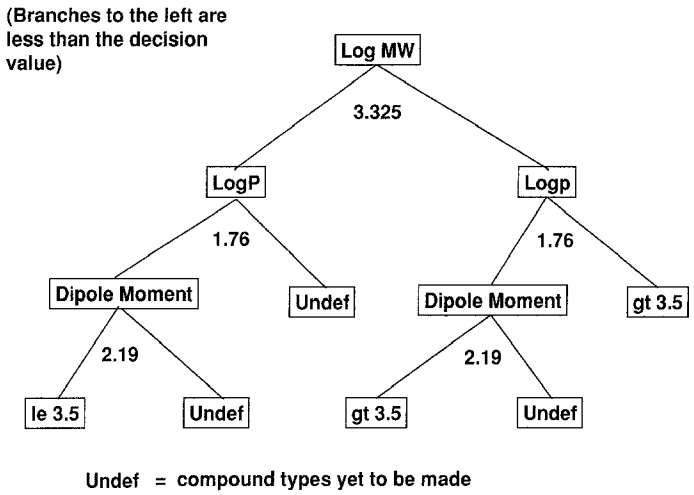


Diagram 14. Class probability tree on the MES data showing exception programming.

next) and may be compared to, for example, regression equations or nonlinear map plots which classify the objects of interest, but do not easily explain the reasons behind the classification.

The probabilities attached to the derived rules allow confidence to be attached to particular branches. Also, it is clear how the rules were derived and (if requested) which examples were used to determine the rules. In addition, the class probability trees can be used in exception programming to further explore property space economically.

Also, several trees may be induced from the same dataset by forcing induction on attributes of interest and so we can examine the data from different angles. A number of alternative production rules may thus be generated from the same dataset.

The results imply that in some cases it would be advantageous to use rule-induction as a complementary technique in addition to conventional statistical and pattern-recognition methods.

REFERENCES

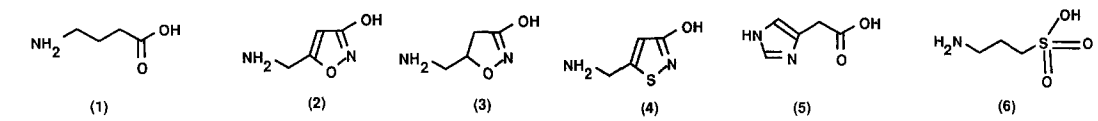
1 Hyde, R.M. and Livingstone, D.J., J. Comput.-Aided Mol. Design, 2 (1988) 145.
2 Chatfield, C. and Collins, A.J., Introduction to Multivariate Analysis, 3rd Ed., Chapman and Hall, 1986.
3 Hayes-Roth, F., Waterman, D.A. and Lenat, D.B. (Eds), Building Expert Systems, Addison-Wesley, London, 1984.
4 Winston, P.H., Artificial Intelligence, Addison-Wesley, London, 1984.
5 Gunhold, R. and Zettel, J., 'DIAESS: An expert system for diagnosis of faults in electronic circuit boards', In Proceedings of the Artificial Intelligence and Advanced Computer Technology Conference, 23–25 September 1986, Rhein-Main-Halle, Wiesbaden, 1987.
6 Austin, D.J., Blake, P.S., Fletcher, D.A. and Garrett, C.M.E., Chemometrics, 5 (1988) 53.
7 Modessitt, K., Experience with commercial tools involving induction on large databases for space shuttle main engine testing, 4th International Expert Systems Conference, London, June 1988.
8 A-Razzak, M., Hassan, T. and Ahmad, A., EXTRAN 7 User Manual, Infolink Decision Services Ltd., 9–11 Grosvenor Gardens, London SW1 W0BD, U.K.
9 Quinlan, J.R., Machine Learning, Kluwer, Boston, U.S.A., 1 (1986) 81–106.
10 Krogsgaard-Larsen, P., Jacobsen, P. and Falch, E. In S.J. Enna (Ed.), The GABA Receptors, The Humana Press, Clifton, NJ, U.S.A., 1983, pp. 149–176.
11 SYBYL molecular modelling package, Tripos Associates, St. Louis, MO, 1990.

12 MOPAC, Version 5.0, Quantum Chemical Program Exchange, Department of Chemistry, University of Indiana, Bloomington, IN.
13 CNDO/2. QCPE 91, Quantum Chemical Program Exchange, Department of Chemistry, University of Indiana, Bloomington, IN.
14 Glen, R.C. and Rose, V.S., J. Mol. Graph., 5 (1987) 79.
15 Barraclough, P., Beams, R.M., Black, J.W., Cambridge, D., Collard, D., Demaine, D.A., Firmin, D., Gerskowitch, V.P., Glen, R.C., Giles, H., Hill, A.P., Hull, R.A.D., Iyer, R., King, W.R., Livingstone, D.J. and Nobbs, M., Eur. J. Med. Chem., 25 (1990) 467.
16 Barraclough, P., Black, J.W., Cambridge, D., Collard, D., Firmin, D., Gerskowitch, V.P., Glen, R.C., Giles, H., Hill, A.P., Hull, R.A D., Iyer, R., King, W.R., Kneen, C.O., Lindon, J.C., Nobbs, M.S., Randall, P., Shah, G., Smith, S., Vine, S.J., Whiting, M.V. and Williams, J., J. Med. Chem., 33 (1990) 2231.
17 Weise, M., Seydel, J.K., Pieper, H., Kruger, G., Noll, K.R. and Keck, J., Quant. Struct.–Act. Relat., 6 (1987) 164.
18 Lien, E.J., Liao, C.H. and Shinouda, H.G., J. Pharm. Sci., 68 (1979) 463.
19 Hassan, T. and A-Razzak, M., Exception programming: A new approach to defining specification examples, Proceedings of the International Conference on Expert Systems, London, 1988, pp. 181–197.
20 Gomez-Jeria, J., Pharm. Sci., 71 (1982) 1423. (N.B. In this version of the calculation of superdelocalizability, the calculation as originally defined by Fukui et al. [21] is modified, as the Hückel parameter Lambda is replaced by the orbital energy E. E may tend to zero and is not scaled between molecules, however this parameter appears to act as a useful variable correlating well with biological activity in some cases.)
21 Fukui, K., Yonezawa, T. and Nagata, C., Bull. Chem. Soc. Jpn., 27 (1954) 423.
22 Edward, J.T., J. Chem. Ed., 47 (1970) 261.
23 Pearlman, R.S., SAREA, QCPE Bull., 1 (1981) 16.
24 Margenaux, H. and Murphy, G.M., The Mathematics of Physics and Chemistry, 2nd Ed., D. Van Nostrand Co. Inc., New York, NY, 1956.
25 CLOGP, Medchem software version 3.52 Pomona College Medicinal Chemistry Project, Claremont, CA, 1987
26 Weast, R.C. and Astle, M.J. (Eds), CRC Handbook of Chemistry and Physics, 60th Ed , 1980.

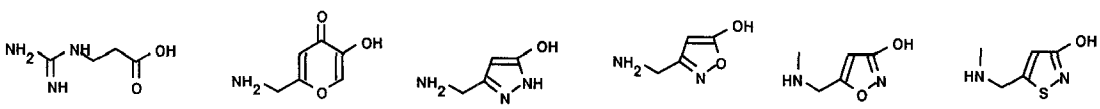
Appendix 1

GABA DATA [10,14]

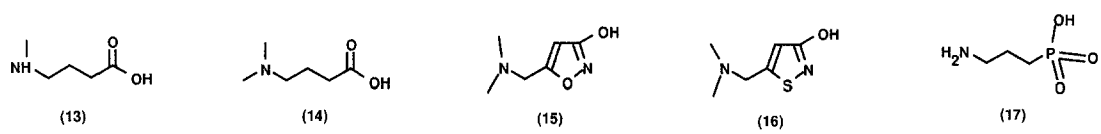
Potent GABA Agonists



Weak GABA agonists



No Intrinsic GABA activity



NAME, ACTIVITY, C_ATCH1, C_ATCH2, C_ATCH3, C_ATCH4, C_DIPV_X, C_DIPV_Y, C_DIPV_Z, C_DIPMOM, C_ESDL1, C_ESDL2, C_ESDL3, C_ESDL4, VDW_VOL, D_S_VOL, COL_DIAM, APP_DIAM, SURF_AR, M_OF_I_X, M_OF_I_Y, M_OF_I_Z, P_E_AX_X, P_E_AX_Y, P_E_AX_Z, MOL_WT

GAB1, 2, −0.1939, 0.3982, −0.2669, −0.3449, −2.066, 1.045, 1.1732, 2.5955, −11.3424, −15.825, −15.1215, −6.8112, 97.30002, 132.32803, 5.70651, 3.54235, 141.10939, 137.1953, 794.2288, 681.4996, 5.6999, 0.7706, 2.463, 103.12153

GAB2, 2, −0.184, 0.3294, −0.2572, −0.1984, −1.6785, 2.408, −2.188, 3.661, −0.0478, −11.9111, −14.7545, −23 209, 89.90002, 122.26402, 5.55802, 4.09777, 131.87355, 119.3546, 760.6179, 649.2339, 5.3198, 0.4181, 2.2495, 114.10422

GAB3, 2, −0.1882, 0.3568, −0.2579, −0.2795, −1.7156, 1.0285, −2.2313, 2.9966, −0.0953, −10.8412, −15.508, −7.6709, 103.80002, 141.16803, 5.83086, 4.10394, 139.38287, 202.2276, 882.1697, 687.9745, 5.1289, 0.393, 2.7611, 130.16481

GAB4, 2, −0.19, 0.2847, −0.2508, −0.1451, −0.1408, 2.2623, −1.6823, 2.8228, −0.0423, −14.9352, −14.9745, −30.1927, 95.50002, 129.88002, 5.67111, 4.53435, 142.56483, 135.6394, 808.6339, 703.8359, 5.4473, 0 8153, 2.2764, 116.12022

GAB5, 2, −0.0316, 0.4097, −0.2676, −0.3557, −2.2912, −1.7852, −2.0359, 3.547, −19.836, −4.2671, −0.0452, −0.1305, 104.50002, 142.12003, 5.84394, 4.37922, 143.70313, 205.7015, 887.0851, 805.53, 5.4316, 1.5695, 2.3874, 126.11522

GAB6, 2, −0.1956, 0.4212, −0.2411, −0.2645, −3.5008, 0.9008, −0.7745, 3.6968, −30.8645, −8.7232, −2.2284, −3.3491, 120.60003, 164.01604, 6.12984, 4.74089, 156.7924, 227.5761, 988.2823, 932.5047, 5.5185, 1.7578, 2.2575, 139.16986

GAB7, 1, −0.2563, 0.407, −0.268, −0.3429, −3.0017, 1.266, −4.4553, 5.5193, −25.9414, −13.3909, −4.0248, −6.8762, 118.40003, 161 02405, 6.09234, 4.7043, 163.78636, 182.8739, 1379.8918, 1312.61, 6.9205, 1.4852, 2.1849, 131.1349

GAB8, 1, −0.1897, 0.0941, −0.2456, −0.3706, −2.0894, 0.6066, −0.592, 2.2548, −0.067, −11.2709, −1.8885, −43.8357, 114.10002, 155.17603, 6.01767, 4.37805, 154.60587, 256.856, 1162.4065, 915.793, 5.6827, 0.4262, 2.9878, 141.12689

GAB9, 1, −0.188, 0.3046, −0.253, −0.1383, 1.8164, −0.1882, 1 8244, 2.5814, −0.0624, −13.5936, −15.1512, −19.674, 97.60002, 132.73602, 5.71237, 4.33011, 136.15846, 167.2367, 757.0826, 607.9944, 5.1479, 0.6337, 2.6454, 113.11952

GAB10, 1, −0.1884, 0.3844, −0.2545, −0.1718, −0.7567, 0.3028, 0.6563, 1.0464, −0.0582, −13.0808, −14.858, −29.1206, 89.90002, 122.26402, 5.55802, 4.1536, 131.95255, 164.6528, 737.6986, 590.8283, 5.0521, 0.6245, 2.6137, 114.10422

GAB11, 1, −0.1602, 0.3285, −0.258, −0.1979, −2.2803, −0.6702, 0.3044, 2.3962, −1.1246, −16.0404, −14.7507, −35.5146, 106.10002, 144.29604, 5.87361, 4.65464, 153.67284, 180.2722, 1147.9791, 985.6005, 6 1766, 0.5912, 2.587, 128.13123

GAB12, 1, −0.164, 0.3857, −0.2656, −0.2199, −0 126, −0.512, 0.442, 0.688, −0.0842, −10.9293, −14.894, −9.2505, 120.00002, 163.20003, 6.11966, 4.79428, 161.17467, 279.6127, 1285.0757, 1046.0752, 5 968, 0.8397, 3.0006, 144.1918

GAB13, 0, −0.1651, 0.4024, −0.3467, −0.268, −1.6528, −0.3633, −1.3448, 2.1615, −26.5051, −9.9168, −2.8427, −41.6255, 113.50002, 154 36003, 6.00711, 4.69316, 163.08752, 210.3617, 993.6855, 872.3694, 5.9479, 1.3794, 2.6622, 117.14854

GAB14, 0, −0.1481, 0.4008, −0.2689, −0.3475, −1.9597, 1.7686, −1.4977, 3.0351, −23.7139, −13.7375, −37.859, −38.116, 129.70003, 176.39204, 6.28029, 4.73681, 176.17818, 236 5967, 1238.6066, 1144.0458, 6.3996, 1.6464, 2.5139, 131.17552

GAB15, 0, −0.1409, 0.3283, −0.257, −0.1987, −2.1591, −0.7934, −0.4093, 2.3364, −3 9814, −14.9716, −14.3758, −31.8699, 122.30003, 166.32803, 6.15851, 4.93875, 171.34991, 263.4249, 1262.1583, 1145.2556, 6.1438, 1.6061, 2 5876, 142.15819

GAB16, 0, −0.1355, 0.3894, −0.266, −0.2235, −0.871, 2.1513, 0.3908, 2.3536, −0.0824, −14.7911, −14.7337, −33.3889, 136.20003, 185.23204, 6.3835, 5.19118, 182.87073, 321.8189, 1412.9445, 1256.6757, 6.095, 1.6185, 2.7504, 158.21877

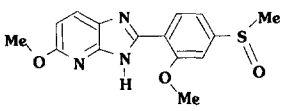
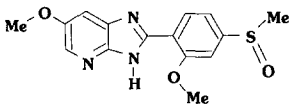
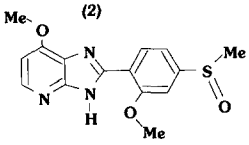
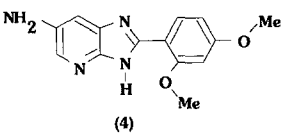
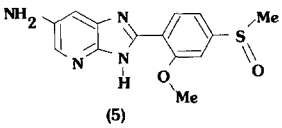
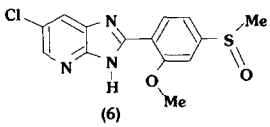
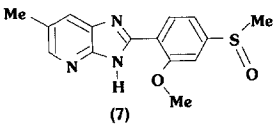
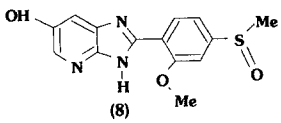
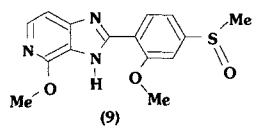
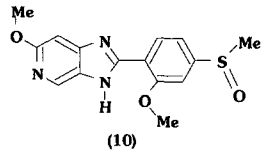
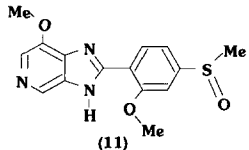
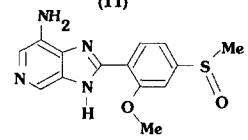
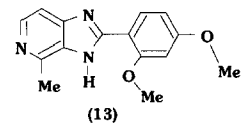
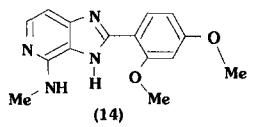
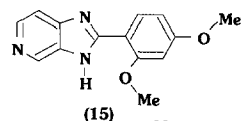
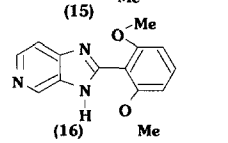
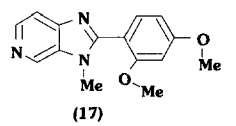
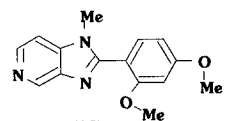
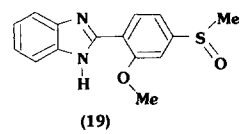
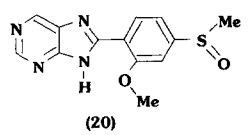
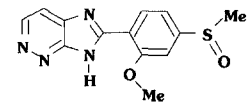
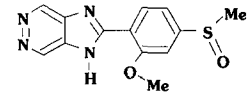
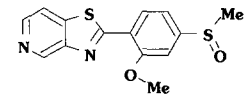
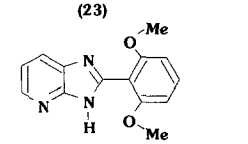
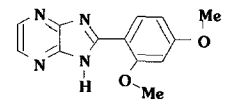
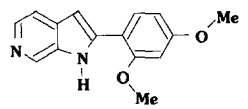
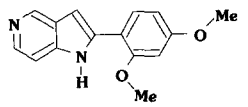
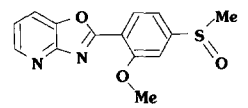
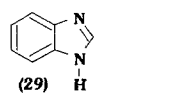
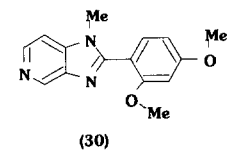
GAB17, 0, −0.1919, 0.4375, −0.2962, −0.3031, −1.6026, 1.5513, −2.6753, 3.4831, −22.5237, −15.0092, −33.4407, −36.0999, 117.90003, 160.34404, 6.08375, 4.54213, 164.02975, 241.1332, 1034.9072, 998.3727, 5.6979, 1.9252, 2.2428, 139.09166

Parameters

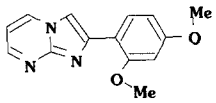
ACTIVITY. Classification based on binding and electrophysiological methods [10].
C_ATCH1-4, atom-centred charges, CNDO/2 (electrons) [13].
C_DIPV_X,Y,Z, dipole components in X,Y and Z (D) [13].
C_DIPMOM, total dipole moment (D) [13].
C_ESDL1-4, electrophilic superdelocalizability [20,21].
VDW_VOL, Van der Waals volume [22].
D_S_VOL, dead space volume [22].
COL_DIAM, collision diameter [22].
APP_DIAM, approach diameter [22].
SURF_AR, surface area [23].
M_OF_I_X,Y,Z, moments of inertia [24].
P_E_AX_X,Y,Z, principal axes [24].
MOL_WT, molecular weight (Daltons).

INOTROPE DATA

Chemical structures and inotropic potencies

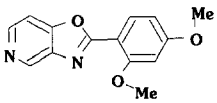
	Potency
 (1)	3.00
 (2)	3.63
 (3)	4.10
 (4)	4.80
 (5)	4.01
 (6)	3.61
 (7)	3.73
 (8)	3.75
 (9)	3.91
 (10)	3.76
 (11)	3.00
 (12)	3.70
 (13)	I
 (14)	I
 (15)	4.01
 (16)	I
 (17)	3.07
 (18)	2.77
 (19)	4.13
 (20)	3.52
 (21)	3.45
 (22)	3.58
 (23)	I
 (24)	2.93
 (25)	4.03
 (26)	I
 (27)	I
 (28)	3.23
 (29)	2.03
 (30)	2.77

Potency



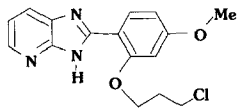
5.10

(31)



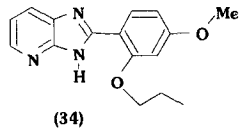
I

(32)



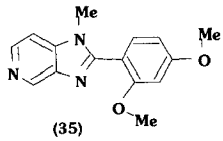
6.1

(33)



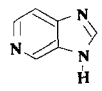
6.0

(34)



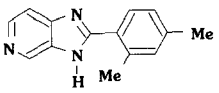
2.77

(35)



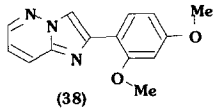
2.0

(36)



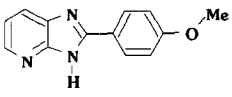
I

(37)



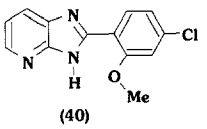
I

(38)



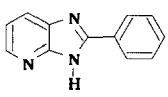
4.31

(39)



4.30

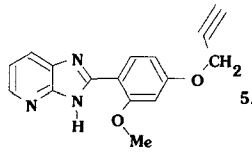
(40)



4.00

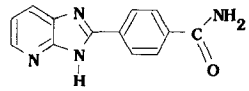
(41)

Potency



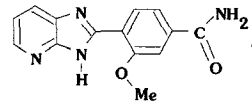
5.49

(42)



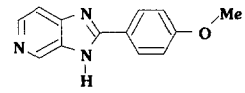
3.37

(43)



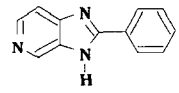
4.26

(44)



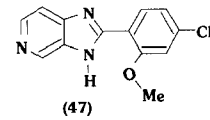
3.10

(45)



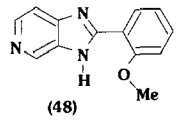
3.03

(46)



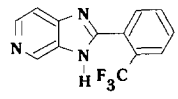
4.04

(47)



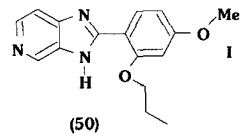
4.30

(48)



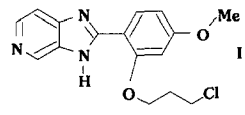
I

(49)



I

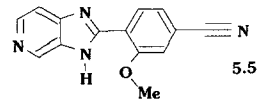
(50)



I

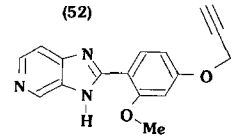
(51)

Potency



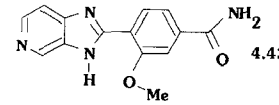
5.5

(52)



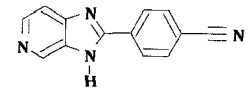
3.91

(53)



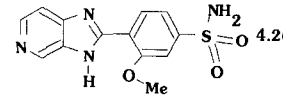
4.43

(54)



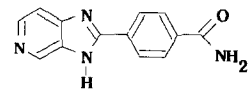
3.99

(55)



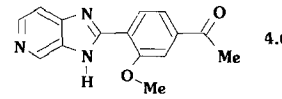
4.26

(56)



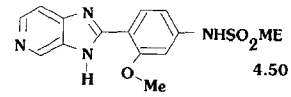
3.42

(57)



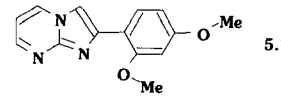
4.65

(58)



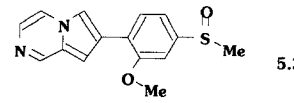
4.50

(59)



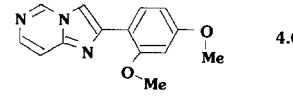
5.10

(60)



5.39

(61)



4.05

(62)

POTENCY, CATEGORY, CLOGP, M_ATCH1, M_ATCH2, M_ATCH3, M_ATCH4, M_ATCH5, M_ATCH6, M_ATCH7, M_ATCH8, M_ATCH9, M_DIPV_X, M_DIPV_Y, M_DIPV_Z, M_DIPMOM, M_LUMO_E, M_HOMO_E, VDW_VOL, SURF_AR, M_OF_I_X, M_OF_I_Y, M_OF_I_Z, P_E_AX_X, P_E_AX_Y, P_E_AX_Z, MOL_WT, COL_DIAM, M_ESDL1, M_ESDL2, M_ESDL3, M_ESDL4, M_ESDL5, M_ESDL6, M_ESDL7, M_ESDL8, M_ESDL9, M_NSDL1, M_NSDL2, M_NSDL3, M_NSDL4, M_NSDL5, M_NSDL6, M_NSDL7, M_NSDL8, M_NSDL9

ONEM, 3, 0, 1.128, −0.1482, 0.0678, −0.2011, −0.1342, 0.1147, −0.2143, −0.0181, −0.1189, 0.0207, −2.252, −1.1827, 0.4317, 2.5801, 0.992, 8.8442, 260.90005, 325.16028, 1001.1672, 8789.7842, 9623.833, 16.9309, 9.0645, 5.5085, 317.36322, 7.9279, −0.3338, −0.222, −0.2954, −0.3264, −0.2168, −0.2638, −0.2409, −0.2491, −0.2285, 0.6098, −5.1477, 0.7671, −10.9883, −8.6153, 2.5022, −14.3485, −3.9187, 3.8238

TWOM, 3.63, 0, 1.128, −0 1464, 0.0726, −0.1981, −0.1043, −0.1091, 0.0036, −0.0715, −0.0739, −0 0241, −1.1935, −0.0441, 0.2875, 1.2285, 1.0703, 8.7228, 260.90005, 320.65512, 989.2675, 8882.0742, 9765.7061, 16.9671, 8.8274, 5.0209, 317.36322, 7.9279, −0.3339, −0.2221, −0.2957, −0.3193, −0.2478, −0.2342, −0.2534, −0.2454, −0.2366, 0.1414, −25 4463, −3.3955, −34.1624, −39.0201, 2.3276, −41.4163, −25.489, 2.9097

THREEM, 4.1, 0, 1.128, −0.1225, 0.053, −0.2006, −0.1559, −0.0157, −0.2981, 0.182, −0.1312, 0.0299, −3.0246, 0.02, −0.3652, 3.0466, 0.884, 8.7705, 260.90005, 325.12833, 1382 8503, 7637.3027, 8894.5518, 14.6515, 10.0804, 5.1844, 317.36322, 7.9279, −0.3342, −0.2253, −0.2971, −0.3308, −0.2374, −0.2757, −0.2122, −0.2527, −0.2291, 0.9064, 2.4665, 1.2002, 3.1572, 3.6795, 2.6837, 5.5394, 2.6601, 2.8261

FOURM, 4.8, 0, 1 958, −0.1647, 0.0817, −0.2145, −0.1153, −0.0814, −0.0478, −0.092, −0.0718, −0.0131, 2.4748, −0.7057, −2.6943, 3.7259, 0.6609, 8.4256, 229.40005, 288.76575, 889.9173, 5632.8296, 6449.4692, 15.8026, 8.7023, 4.7525, 270.29156, 7.59506, −0.344, −0.2242, −0.3044, −0.326, −0.2487, −0.2422, −0.2579, −0.2483, −0.2386, 0.6232, 1.3296, 0.8448, 1.6824, 1.933, 2.1959, 2.2339, 1.8331, 2.0334

FIVEM, 4.01, 0, 0.575, −0.1557, 0.0642, −0.2065, −0.1082, −0.0799, −0.044, −0.0918, −0.076, −0.0145, −2.1216, −0.0519, −2.1664, 3.0327, 1.0485, 8.6055, 252.40005, 308.89001, 991.9869, 7496.624, 8404.3096, 15.7547, 8.9755, 4.7838, 302.35153, 7.84085, −0.3373, −0.2227, −0.2992, −0.3212, −0.2456, −0.2394, −0.2553, −0.2461, −0.2361, 0.6785, 1.088, 0.9169, 1.6341, 1.8487, 2.8028, 2.237, 1.9445, 2.6064

SIXM, 3.61, 0, 1.417, −0.1455, 0.0715, −0.2075, −0.1149, −0.0407, −0.1456, − 0.0212, −0.1025, 0.0067, 0.6354, −1 5207, −3 1869, 3.5879, 1.1278, 9.0479, 253.00005, 310.3851, 973.389, 9153.6953, 9929.3809, 16.2687, 8.6315, 5.5781, 321.78186, 7.84705, −0.3311, −0.2195, −0.295, −0.3178, −0.2372, −0.2466, −0.241, −0.2446, −0 2284, 1.5207, 1.4093, 4.4645, −1.2496, 10.454, 12.9682, −1.2495, 8.3757, 13.4325

SEVENM, 3.73, 0, 1.167, −0.1485, 0.0582, −0.2063, −0.1277, −0.0488, −0.145, −0.0296, −0.1062, 0.009, −0.9158, −0.4106, −3.2136, 3.3667, 0.9645, 8.8423, 254.70006, 319.013, 941.3352, 7495.8613, 8232.9141, 16.0309, 8.6999, 5.6941, 301.36383, 7.86459, −0.3365, −0.2238, −0.2988, −0.3254, −0.2431, −0.2573, −0.2474, −0.2496, −0.2325, 0.9678, 3.1692, 1.24, 4.2982, 4.3106, 3.2036, 6.4722, 3.071, 2.8074

EIGHTM, 3.75, 0, 1.114, −0.1591, 0.0678, −0.201, −0.1082, −0.0612, 0.011, −0.1329, −0.0668, −0.018, −1.6273, −2.2945, −0.054, 2.8135, 1.0401, 8.7825, 244.70004, 305.82007, 966.1109, 7543.7051, 8411.1504, 15.4656, 8.66, 4.9056, 303.33624, 7.76029, −0.3365, −0.2217, −0.297, −0.3196, −0.2432, −0.2288, −0.2552, −0.2435, −0.2353, 0.5943, −3.1812, 0.7214, −7.8011, −4.9182, 0.8801, −9.7804, −2.1221, 2.3533

NINEM, 3.91, 0, 1.128, −0.1632, 0.0667, −0.1875, 0.1624, −0.2213, −0.0593, −0.1533, −0.0145, −0.0989, −0.0801, −0.8489, −0.0134, 0.8528, 0.99, 8.8607, 260.90005, 327.26962, 1230.7749, 7669.8242, 8790.1016, 15.1776, 9.2551, 5.025, 317.36322, 7.9279, −0.3378, −0.2216, −0.2959, −0.2111, −0.3388, −0.2474, −0.2648, −0.2379, −0.2467, 0.9233, 2.6857, 1.289, 4.1072, 2.0988, 2.5425, 3.4178, 2.6826, 2.3751

TENM, 3.76, 0, 1.128, −0.154, 0.0781, −0.1984, −0.0233, −0.1824, 0.0966, −0.1419, −0.0132, −0.1228, 0.811, −0 9857, 0.0647, 1.278, 1.1011, 8.7246, 260.90005, 322.56335, 985.5779, 8940.2432, 9788.2207, 16.9216, 8.7928, 5.0805, 317.36322, 7.9279, −0.3344, −0.2202, −0.2943, −0.2379, −0.3296, −0.2229, −0.2658, −0.2374, −0.2494, −0.2377, −9.5755, −0.0351, −19.1413, −7.4517, 0.3163, −17.2903, −5.094, 1.7797

ELEVENM, 3, 0, 1.128, −0.1156, 0.0522, −0.1957, −0.0939, −0.0949, −0.1861, 0.0994, −0.0711, −0.0613, −0.9344, −1.3741, −0.0662, 1.663, 1.0352, 8.7514, 260.90005, 319.4274, 1332.0684, 7680.9326, 8882.4033, 14.677, 9.9079, 5.2625, 317.36322, 7.9279, −0.3306, −0.2242, −0 2929, −0.2493, −0.3225, −0.2626, −0.2249, −0.2449, −0 2411, 1 3336, 4.423, 1.4605, 7 1174, 3.4514, 3.542, 8.982, 3.3873, 3.5169

TWELVEM, 3.7, 0, 0.575, −0.1584, 0.0624, −0.2138, −0.093, −0.1046, −0.1626, 0.0584, −0.0825, −0.0503, −0.2586, −2 7633, −1.249, 3.0435, 1 0379, 8.5056, 251.00005, 311.19775, 1128.166, 6851.7798, 7868.772, 14.7101, 9.221, 5.1154, 302.35156, 7.82632, −0.3348, −0.222, −0.2981, −0.2504, −0.327, −0.2639, −0.2299, −0.2454, −0.2407, 1.0457, 3.0671, 1.2835, 4 1949, 2.5034, 2.4144, 5.0146, 2.6, 2.5567

THIRTEEM, 0, 1, 2.55, −0.1931, 0.0923, −0.2006, 0.0614, −0.1847, −0.0511, −0.1949, −0.0364, −0.084, −1.8802, −7.9085, −0.39, 8.1383, 0.166, 8.4415, 231.70006, 292.04529, 952.8539, 5068.8242, 5999.9795, 14.2749, 8.7823, 3.9246, 269.30386, 7.62036, −0 2917, −0.2188, −0.3463, −0.2391, −0.3454, −0.2503, −0.2708, −0.2377, −0.2486, 0.5329, 0.1563, 0.2141, 0.1429, 1.0083, 0.4355, 0.3992, 1.7281, 1.0166

FOURTEEM, 0, 1, 2.684, −0.2062, 0.093, −0.1869, 0.1971, −0.2475, −0.0322, −0.193, 0.0261, −0.1794, 6.1352, −0.9015, 1.877, 6.4789, 0.2121, 7.9984, 245.60007, 309.43112, 1146.52, 5673.2568, 6740.708, 14.7147, 9.5109, 4.5285, 284.31854, 7.76979, −0.3483, −0.2194, −0.2936, −0.2124, −0.3592, −0.2529, −0.2833, −0.2378, −0.2595, 0.3569, 0.8373, 0.5801, 1.4898, 0.7586, 1.5224, 1.1651, 1.2431, 1.3996

FIFTEENM, 4.01, 0, 2.051, −0.1652, 0.094, −0.2165, −0.0532, −0.1366, −0.1038, −0.1156, −0.0363, −0.0916, 5.3625, −1 7534, −0.9019, 5.7135, 0.6586, 8 9135, 215.50005, 276.00046, 808.1262, 4684.3779, 5428.0068, 14.9053, 8.7249, 4.9936, 255 27687, 7.43845, −0.34, −0.2206, −0.3005, −0.2437, −0.3318, −0.2539, −0.2601, −0.2415, −0.2457, 0.6547, 1.431, 0.879, 2.5548, 1.3726, 2.1603, 2.4774, 1.859, 2.2187

SIXTEENM, 0, 1, 1.491, −0.1856, 0.0623, −0.1367, 0.0085, −0.1727, −0.069, −0.1725, −0.0438, −0.1086, 0.4135, −6.6043, −0 0023, 6.6172, 0 5448, 8.3484, 215.50005, 268.30405, 1311.251, 3191.5369, 4481.4922, 12.8006, 10.6629, 4.1024, 255.27687, 7.43845, −0.2941, −0.2301, −0.3572, −0.2432, −0.3412, −0.2529, −0.2649, −0.2436, −0.2545, 0.7615, 1.1151, 0.5655, 1.7316, 1.1965, 1.4645, 1.8415, 1.8329, 1.4545

SEVENTEM, 3.07, 0, 2.09, −0.1476, 0.0951, −0.177, −0.1204, −0.0867, −0.1561, −0.0461, −0.1218, −0.0112, −0.287, −4.7959, 0.4918, 4.8295, 0.3171, 8.7917, 231.70006, 295.18063, 1026.495, 4666.7339, 5646.7036, 14.7448, 9.4222, 4.4806, 269.30386, 7.62036, −0.2865, −0.218, −0.3438, −0.2631, −0.2542, −0.3398, −0.2463, −0.2468, −0.2357, 0 8186, 0.8124, 0.4456, 1.4501, 1.9579, 0.7803, 1 5956, 2.0519, 1.2565

EIGHTEEM, 2.77, 0, 2.09, −0 1474, 0.0777, −0.1607, 0.012, −0.176, −0.0589, −0.1832, −0.03, −0.1039, 0.5409, −3.9921, 1.2007, 4.2037, 0.3608, 8.6529, 231.70006, 295.3046, 1008.6117, 4646.0107, 5538.123, 15.0678, 9.3576, 4.741, 269.30386, 7.62036, −0 2853, −0.2196, −0.3392, −0.2424, −0.3419, −0.2502, −0.2656, −0.2356, −0.2458, 0.9322, 1.1295, 0.5143, 2.0902, 1.4407, 1.5715, 2.1773, 2.2019, 1.4753

NINETEEM, 4.13, 0, 1.821, −0.1445, 0.054, −0.2116, −0.1337, −0.1121, −0.1536, −0.0695, −0.0721, −0.0285, −0.3132, 0.1457, −1.2381, 1.2854, 0.9368, 8.6414, 243.30005, 300.56342, 918.6081, 6201 7485, 7002.7471, 14.7275, 8.8155, 4.7049, 286 34915, 7.74546, −0.3383, −0.227, −0.3001, −0.2612, −0.2603, −0.2658, −0.2578, −0.2512, −0 2428, 0.8061, 2.0888, 0.9924, 2.51, 2.2336, 1.7538, 2.5366, 1.8778, 1.8827

TWENTYM, 3.52, 0, −0.338, −0.1388, 0.0646, −0.2008, −0.1796, 0.0022, −0.2244, 0.0494, −0.1588, 0.0349, 0.2802, −2.4843, −0.5138, 2.5523, 1.1661, 9.1168, 233.70004, 288.10791, 913.686, 6319.9097, 7123.5864, 14.4542, 8.6079, 4.8008, 288.32455, 7.64222, −0.3296, −0.2197, −0.292, −0.3248, −0.23, −0.3327, −0.2269, −0.2495, −0.2224, 1.1665, 0.4095, 1.4081, −0.5096, 0.0412, 3.322, 0.034, 1.185, 5.0189

TWENTYIM, 3.45, 0, −0.384, −0.1448, 0.0909, −0.1983, −0 0428, −0 0454, −0.135, −0.0953, −0.0804, −0.0429, −0.896, −4.199, −1.4665, 4.5371, 1.4255, 9.3682, 233.70004, 286.82062, 903.5471, 6293.354, 7075.9424, 14.6975, 8.6887, 4 6643, 288.32455, 7.64222, −0.3253, −0.215, −0.289, −0.3167, −0.3075, −0.24, −0.247, −0.2388, −0.2332, 1.0459, 1.316, 1.6963, 0.3096, 1.2521, 2.9329, 0.1671, 1.992, 4.0869

TWENTY2M, 3.58, 0, −0.384, −0.1307, 0.0728, −0.1913, −0.116, −0.0465, −0.0728, −0.0658, −0.0918, −0.0667, 2.5276, −3.7993, −1.5451, 4 8178, 1 3245, 9.3766, 233.70004, 287.94867, 907.4532, 6296.2319, 7082.8081, 14.2002, 8.7186, 4 5597, 288 32455, 7.64222, −0.3229, −0.2169, −0.2856, −0.2458, −0.308, −0.3116, −0.2435, −0.2402, −0 2342, 3.1116, 1.8631, 8.383, 1.8786, 7.6693, 19.6388, 3.676, 14.688, 32.8893

TWENTY3M, 0, 1, 1.005, −0.1316, −0.1631, 0.4708, −0.1405, −0.0382, −0.1598, 0 018, −0.0827, −0.3203, 0.1794, 0.7738, −2.5945, 2.7134, 1.6588, 9.0333, 242.50005, 292.05988, 935.7446, 6296.3892, 7106.916, 15.2067, 8.7757, 4.8303, 304.38214, 7.73696, −0.3121, −0.2271, −0.3803, −0.2533, −0.2344, −0.3138, −0.2305, −0.2315, −0.2522, 1.0967, 2.0224, 11.7074, 69.9037, 1.6358, 32.1346, 42.5603, 4.6197, 75.2954

TWENTY4M, 2.93, 0, 1.491, −0.1269, 0.0869, −0.2179, −0.1364, −0.0566, −0.2097, −0 0309, −0.1101, 0.0064, 2.7381, −2.7367, −0.7362, 3.9406, 0 4457, 8.513, 215.50005, 271.76608, 1314.0021, 3133.9302, 4433.2051, 12.8163, 10.5592, 3.9019, 255.27687, 7.43845, −0.3467, −0.2267, −0.3083, −0.3329, −0.2473, −0.2705, −0.2502, −0.2558, −0.2374, 0 4635, 0.957, 0.7294, 1.1506, 1 4053, 1.7286, 1.5999, 1.4473, 1.8405

TWENTY5M, 4.03, 0, 1.052, −0.1285, 0.0755, −0.2026, −0.0915, −0.1039, −0.1295, −0.0392, −0.0582, −0.0351, 0.377, −0.429, −1.1437, 1.2783, 1.1684, 9.1453, 233.70004, 289.418, 910.5193, 6264.6694, 7050.6543, 14.6663, 8.5215, 4.5513, 288.32455, 7.64222, −0.33, −0.2179, −0.2931, −0.311, −0.2415, −0.2445, −0.3111, −0.2372, −0.2315, 3 5831, 0.5548. 6.0438, 0.8521, 7.2715, 30.112, 6.4422, 9.9686, 28.3641

TWENTY6M, 0, 1, 1.806, −0.1959, 0.0289, −0.1787, −0.0104. −0.1847, −0.0567, −0.1967, 0.0227, −0.1403, 1.6001, −5.329, −0.2279, 5.5687, 0.4318, 8.0709, 220.30005, 271.20544, 827.3769, 4676.1328, 5483.5742, 14.9295, 8.7097, 3.9358, 254.28917, 7 49327, −0.3066, −0.2414, −0.2787, −0.2467, −0.3444, −0.2526, −0.2719, −0.2391, −0.2653, 0.6553, 0.8485, 0.5537. 1.2135, 0.9964, 1.1776, 1.3538, 1.6405, 1.3318

TWENTY7M, 0, 1, 2.557, −0.1959, 0.0285, −0.179, −0.0105, −0.1848, −0.0567, −0.1966, 0.0224, −0.1403, 1.6212, −5.3265, −0.1809, 5.5708, 0.4361, 8.0637, 220.30005, 271.56186, 827.1133, 4672.9985, 5486.3569, 14.9296, 8.7218, 3.8511, 254.28917, 7.49327, −0.3067, −0.2414, −0.2788, −0.2467, −0.3444, −0.2527, −0 2719, −0.2391, −0 2654, 0.6554, 0.8498, 0.5522. 1.2095, 0.9941, 1.1781, 1.3487, 1.6394, 1.3328

TWENTY8M, 3.23, 0, 0.376, −0.1337, 0.1026, −0.0731, −0.0663, −0.0821, −0.1653, −0.0734, −0.0605, −0.0287, −5.2048, −1.8409, −2.2367, 5 9567, 1.2456, 9.2346, 230.80003, 292.12427, 938.8732, 6110.0947, 6907.2319, 14 7871, 8.9057, 4.8664, 288.32153, 7.61048, −0.3225, −0.2116, −0.3168, −0.3152, −0.2414, −0.2548, −0.2409, −0.2343, −0.2317, −0.1492, 0.7113, 0.3811, −4.3699, −8.6919, −1.004, −5.7157, −8.8405, −1.1547

TWENTY9M, 2.03, 0, 1.547, −0.1523, −0.0303, −0.2389, −0.133, −0.0498, −0.2067, −0.0254, −0.1073, 0.0099, 1.2156, 0.1481, −0.0006, 1.2246, 0.1882, 9.3926, 100.90002, 136.38171, 212.194, 501.3243, 713.5096, 8.5624, 6.572, 2.5473, 119.12612, 5.77604, −0.3332, −0 2319, −0.2968, −0.3237, −0.2405, −0.2638, −0.2432, −0.2482, −0.2304, 0.4902, −0.6512, 0.8675, −0.9222, −0.5079, 2.8668, −1.217, 0.6505, 3.3564

THIRTYM, 2.77, 0, 2.09, −0.1557, 0.0906, −0.1576, 0.0155, −0.1802, −0.0556, −0.1907, −0.0303, −0.0964, −1.4734, −5.4411, −2.6387, 6.2241, 0.3924, 8.7601, 231.70006, 291.69229, 1025.3804, 4759.6821, 5718.5713, 14.7034, 9.5104, 4.2793, 269.30386, 7.62036, −0.2854, −0.2181, −0.3385, −0.2422, −0.3419, −0.2484, −0.2653, −0.2345, −0.245, 0.9271, 1.0134, 0.5053, 1.8478, 1.4917, 1.3564, 1.9325, 2.3319, 1.4164

THIRTYlM, 5.1, 0, 1.708, −0.1845, −0.0109, −0.0995, −0.0775, −0.0338, −0.2431, 0.0027, −0.1123, 0 0295, −4.1091, −1.7832, 1.0994, 4.6123, 0.5784, 8.3483, 215.50005, 272.93716, 828.5944, 4647.5322, 5446.2344, 14.5471, 8.7623, 4.1577, 255.27686, 7.43845, −0.2661, −0.2431, −0.3431, −0.3159, −0.2322, −0.2596, −0.2274, −0.279, −0.2289, −8.2695, −0.3223, −0.183, −2.8124, −8.7124, 0.0047, −0.2395, −5.5654, −1.8917

THIRTY2M, 0, 1, 1.767, 0.027, 0.0944, −0.181, 0.0241, −0.1657, −0.0415, −0.1687, −0.0504, −0.1167, −0.8736, −4 3616, 1.4032. 4.6643, 0.9212, 8.8011, 207.80003, 266.14459, 782.8915, 4386.4419, 5088.9063, 14.4991, 8.5386, 4.8693, 256.2616, 7.34878, −0.2893, −0.2045, −0 3222, −0.2323, −0.3187, −0.2377, −0.2606, −0.2245, −0.2366, 0.7896, 3.4906, 0.7722, 15.1051, 2.555, 8.672, 18.6444, 4.9528, 5.0666

THIRTY3M. 6.1, 0, 2.822, −0.2069, 0.0948, −0.1776, −0.1525, −0.0464, −0.2118, −0.033, −0.0842, −0.0052, 3.9835, −1.4882, −0.9014, 4.3469, 0.3346, 8.6828, 262.40005, 321.45673, 2126.3911, 4829.7539, 6734.2959, 14.3893, 10.4145, 5.7249, 317.77585, 7.94306, −0.3463, −0.2183, −0.2932, −0.3368, −0.2469, −0.2705, −0.2513, −0.2476, −0.2332, 0.4809, 1.009, 0.7237, 1.2276, 1.4408, 1.7402, 1.7584, 1.416, 1.7422

THIRTY4M, 6, 0, 3.109, −0.1835, 0.0884, −0.1641, −0.0269, −0.2122, −0.0419, −0.1587, −0.0152, −0.0839, 2.3523, −1.46, 0.4697, 2.8081, 0.3315, 8.7244, 247.90007, 317.09924, 1733.1324, 4730.2666, 6351.5078, 14.9466, 10.8205, 5.2478, 283.33084, 7.79397, −0.286, −0.2173, −0.3409, −0.2481, −0.2679, −0.2435, −0.3343, −0.2317, −0.2442, 0.8246, 0.8472, 0.4774, 1.6286, 2.0398, 1.2964, 1.1399, 2.1154, 1.3812

THIRTY5M, 2.77, 0, 2.09, −0.1525, 0.1017, −0.1661, 0.0137, −0.1795, −0.0565, −0.1945, −0.0297, −0.0915, −1.9523, −5.6689, −0.4738, 6.0143, 0.3352, 8.724, 231.70006, 291.84634, 1036.5201, 4784.0391, 5753.5024, 14.7066, 9.43, 4.5328, 269.30386, 7.62036, −0.2857, −0.2174, −0.3382, −0.2426, −0.3419, −0.2483, −0.2656, −0.2345, −0.2444, 1.0194, 1.2056, 0.4886, 1.2935, 1.4532, 1.0359, 1.4397, 2.3951, 1.2725

THIRTY6M, 2, 0, 0.414, −0.1613, −0.0148, −0.2505, −0.0568, −0.1326, −0.0998, −0.1115. −0 0337, −0.0871, 3.2429, −0.7751, 0.0064, 3.3343, 0.2196, 9.3734, 100.90002, 136.73784, 209.445, 511.9761, 721.4203, 8.5726, 7.0042, 2.5359, 119.12612, 5.77604. −0.3322, −0.2281, −0.2968, −0.2414, −0.3269, −0.25, −0.2557, −0.2384, −0.242, 0.5055. −0 4127, 0.9562, −0.7843, −0.099, 2.8137, −0.8921, 0.8998, 3.4329

THIRTY7M, 0, 1, 3.51, −0.1936, 0.0538, −0.1562, −0.0033, −0.1468, −0.0744, −0.1691, −0.0394, −0.1039, −1.0991, −5.1062, 0.2606, 5.2297, 0.5147, 8.7596, 203.10005, 256.47586, 524.3405, 3747.3855, 4123.9692, 13.5947, 7.4245, 4.9432, 223.27806, 7.29295, −0 2861, −0.2256, −0.3461, −0.2397, −0.3312, −0.2486, −0.2589, −0.2374, −0.2483, 0.9963, 1.1322, 0.7271, 2.4376, 1.8824, 1.8044, 2.5558, 2.9346, 1.8842

THIRTY8M, 0, 1, 1.68, −0.1253, −0.0559, −0.1428, −0.033, −0 2082, −0.0723, −0.0469, −0.0321, −0.0452, 0.6241, 1.8012, 0 7753, 2.0578, 0.3255, 8 4467, 215.50005, 274.32938, 786.7649, 4613.4014, 5375.1313, 14.8965, 8.7969, 4.5579, 255.27686, 7 43845, −0.2521, −0.252, −0.3408, −0.2387, −0.2561, −0.2312, −0.2964, −0.2656, −0.2242, 2.1043, 0.8975, 0.7781, −0.0593, 1.2921, 0.6356, −0.0738, 1.4095, 1.4485

THIRTY9M, 4.31, 0, 2.524, −0.1572, 0.0727, −0.2328, −0.1404, −0.0416, −0.2124, −0.0273, −0.0973, 0.0167, 0.9643, −0.2684, −1 6119, 1.8974, 0.7691, 8.8696, 193.10004, 249.39702, 391 5088, 4542.1826, 4882.5195, 14.5595, 6.8533, 4.4375, 225.25047, 7.17124, −0.3369, −0.222, −0.2981, −0.3254, −0.2402, −0.2646, −0.244, −0.2471, −0.2302, 0.7371, 1.5013, 1 1146, 2.305, 2.482, 3 1141, 3.4244, 2.3152, 3.1867

FORTYM, 4.3, 0, 2.778, −0.1617, 0.0803, −0.2073, −0.1466, −0.0475, −0.2059, −0.035, −0.0985, 0.0118, 2.0446, −0 2751, −0.1772, 2.0706, 0.7823, 8.7802, 207.60007, 255.25531, 823.2482, 4790.9517, 5594.9316, 14.0504, 8.8543, 3.6845, 259 6955, 7.34642, −0.3394, −0.2219, −0.298, −0.3283, −0.2428, −0.2655, −0.2461, −0.2484, −0.2314, 0.7174, 1.643, 1.0024, 2.0817, 2.4789, 2.518, 3.0177, 2.2212, 2.5495

FORTYlM, 4, 0, 2.512, −0.1493, 0.0734, −0.2337, −0.1475, −0.044, −0.2082, −0.0291, −0.0999, 0.0179, 1.2983, −0.6338, −0.0587, 1.4459, 0.6382, 8.9006, 170.70004, 216.8166, 359.3731, 2778.8918, 3138.1528, 12.8066, 6.722, 1.77, 195.22409, 6.88247, −0.3362, −0.2226, −0.2994, −0.3269, −0.2412, −0.2646, −0.2448, −0.2481, −0 23, 0.6691, 0.9995, 0.9633, 1.4867, 1.7949, 2.7155, 2.2153, 1.8929, 2.8168

FORTY2M, 5.49, 0, 1.742, −0.2141, 0.0982, −0.1699, −0.1529, −0.0404, −0.2121, −0.0324, −0.0929, 0.0026, 1.2424, −1.5711, 0.05, 2.0036, 0.3889, 8.7385, 242.30006, 302.29272, 1177.4409, 6553.207, 7704.3203, 17.3843, 9 8886, 3.9033, 279.29886, 7.73483, −0 3389, −0.2143, −0.2849, −0.332, −0.2442, −0.2679, −0.2474, −0.2429, −0.2274, 0.5176, 0.8376, 0.8138, 1.1209, 1.3658, 2.132, 1.5703, 1.4196, 2.1415

FORTY3M, 3.37, 0, 1.163, −0.1506, 0.0571, −0.2093, −0.1406, −0.0444, −0.2101, −0.0279, −0.1061, 0.0178, −0.6047, −0.4743, −2.4781, 2.5945, 1.1011, 8.9316, 226.40005, 279.48959, 858.0733, 5384.4194, 6216.2075, 14.834, 8.6015, 3.6162, 268.27557, 7.5618, −0.3358, −0.2232, −0.2988, −0.3267, −0.241, −0 2649, −0.2446, −0.2487, −0.2307, 2.4914, 18.5324, 3.4565, 25.8209, 25.1845, 5.8524, 37.0747, 14.718, 2.9506

FORTY4M, 4.26, 0, 0.78, −0.1506, 0.0571, −0.2093, −0.1406, −0.0444, −0.2101, −0.0279, −0.1061, 0.0178, −0 6047, −0.4743, −2.4781, 2.5945, 1.1011, 8.9316, 226.40005, 279.48959, 858.0733, 5384.4194, 6216.2075, 14.834, 8.6015, 3.6162, 268.27557, 7.5618, −0.3358, −0.2232, −0.2988, −0.3267, −0.241, −0.2649, −0.2446, −0.2487, −0.2307, 2.4914, 18.5324, 3.4565, 25.8209, 25.1845, 5.8524, 37.0747, 14.718, 2.9506

FORTY5M, 3.1, 0, 2.524, −0.1645, 0.091, −0.2396, −0.0422, −0.1528, −0.0917, −0.1158, −0.0338, −0.0876, 2.7026, −2.7082, −0.2428, 3.8337, 0.6228, 8.8945, 193.10004, 246.75613, 384.7461, 4576.1299, 4953.5649, 14.8753, 7.0122, 3.772, 225.25047, 7.17124, −0.3382, −0.2193, −0.2988, −0.241, −0.3322, −0.2512, −0.2582, −0.2399, −0.2444, 0.6166, 1.1194, 0 9394, 2.4781, 1.2057, 2.5983, 2.4325, 1.8223, 2.7661

FORTY6M, 3.03, 0, 2.512, −0.1561, 0.0807, −0.2382, −0.0478, −0.1526, −0.0916, −0.1137, −0.0345, −0.0843, 3.4071, −1.6181, 0.045, 3.7721, 0.6891, 9.1603, 170.70003, 212.69272, 356.9865, 2798.1802, 3155.002, 12.8492, 7.0212, 2.577, 195.22408, 6.88247, −0 3358, −0 22, −0.2975, −0.2409, −0.3318, −0.2507, −0.2574, −0.2393, −0.2422, 0.642, 1.1343, 0.9743, 2.5285, 1.241, 2.7388, 2.5071, 1.8938, 2.9086

FORTY7M, 4.04, 0, 2.778, −0.2058, 0.0885, −0.1802, −0 0405, −0.1564, −0.0811, −0.1225, −0.0107, −0.1235, 3.612, −1.2031, 0.5109, 3.8412, 0.7382, 9.2074, 191.40002, 238.67375, 506.2204, 4727.4902, 5232.6567, 14.1394, 7.6504, 2.4714, 245.66849, 7.15013, −0.3365, −0.2127, −0.2852, −0.2425, −0.3338, −0.2499, −0.2606, −0.2329, −0.2428, 0.8489, 2.9804, 1.0039, 8.6621, 2.7075, 4.0658, 8.8315, 2.6863, 3.3728

FORTY8M, 4.3, 0, 1.964, −0.1678, 0.0809, −0.2211, −0.041, −0.1601, −0.0918, −0.116, −0.036, −0.0879, 4.6905, −2.5497, 0.1662, 5.3413, 0.6484, 8.9582, 176.90002, 218.96803, 498.147, 2894.6042, 3392.6802, 12.903, 7.3693, 2.6259, 211.22348, 6.96481, −0 3417, −0.2224, −0.3018, −0.2428, −0.3363, −0.2537, −0.2605, −0.2423, −0.2465, 0.6073, 1 2401, 0.8402, 2.154, 1.163, 2.127, 2.1369, 1.6939, 2.1986

FORTY9M, 0, 1, 3.42, −0.1388, 0.1082, −0.2974, −0.0352, −0.134, −0.0821, −0.1214, 0.005, −0.0913, 5 5341, −0.5753, 0 5243, 5.5886, 1.3427, 9.373, 200.70001, 242.76047, 1078.4924, 3258.8757, 4164.6426, 12.1178, 8.1541, 4.9346, 263.22229, 7.26411, −0.3172, −0.2112, −0.3112, −0.2349, −0.3156, −0.2429, −0.2596, −0.2246, −0.2439, 0.8526, 5.3955, 1.4643, 19.4268, 3.0966, 10.8125, 23.872, 3.8203, 9.0644

FIFTYM, 0, 1, 3.19, −0.1845, 0.0732, −0.182, 0.0061, −0.1724, −0.067, −0.1715, −0.0453, −0.1002, −0 2922, −6.0796, 0.0711, 6.0871, 0.5147, 8.3959, 247.90007, 311.94223, 2082.8245, 4624.7866, 6668.5444, 14.398, 11.3794, 4.7426, 283.33084, 7.79397, −0.2926, −0.2264, −0.3544, −0.2421, −0.3396, −0.2515, −0.2637, −0.2423, −0.2517, 0.7496, 0.9463, 0.5571, 1 649, 1.2651, 1.4246, 1.7884, 1.9921, 1.5149

FIFTY1M, 0, 1, 2.822, −0.2145, 0.0909, −0.1769, −0.0418, −0.1613, −0.0843, −0.1262, −0.014, −0.123, 8.3136, −1.3773, 0.7082, 8.4566, 0.19, 8.7136, 262.40005, 323.4137, 2202.8647, 4858.3994, 6878.1543, 14.7787, 10.6184, 4.8386, 317.77585, 7.94306, −0.3478, −0.2172, −0.2912, −0.2483, −0.3419, −0.2553, −0.2668, −0.2387, −0.2487, 0.6664, 1.5328, 0.8479, 2.6388, 1.2886, 1.7582, 2.6055, 1.5159, 1.6478

FIFTY2M, 5.5, 0, 1.692, −0.2003, 0.1023, −0.1836, −0.0394, −0.1469, −0.0683, −0.1171, −0.0095, −0.132, 3.3633, 0.84, 0 1208, 3.4687, 1.0824, 9.3652, 212.50003, 264.49988, 793.5897, 4555.041, 5342.356, 14.6545, 8.7587, 3.602, 250.26018, 7.40377, −0.3316, −0.2097, −0 2828, −0.2421, −0.3333, −0.2422, −0.2539, −0.2309, −0.2398, 0.7058, 0.448, 1.0325, −1.7664, −0.072, 1.7262, −1.4134, 0.8434, 2.0927

FIFTY3M, 3.91, 0, 1.701, −0.1907, 0.0867, −0.1676, −0.1184, −0.0891, −0.1525, −0.0434, −0.133, −0.0106, 0.4354, −4.6769, 0.3919, 4.7134, 0.4004, 8.885, 242.30003, 300.87994, 904.0972, 6811.1748, 7665.5464, 17.3596, 9.0978, 4.1566, 279.29886, 7.73483, −0.2856, −0.2158, −0.3394, −0.2617, −0.2547, −0.3376, −0.2431, −0.2459, −0.2342, 0.8706, 1.1604, 0.555, 2.2939, 2.2429, 1.1489, 2.3365, 2.1681, 1.5252

FIFTY4M, 4.43, 0, 0.78, −0.1652, 0.0722, −0.2155, −0.0408, −0.1512, −0.0939, −0.1144, −0.0374, −0.0865, 1.5781, −1.5208, 0.5214, 2.2528, 1.1363, 9.1604, 226.40005, 276.02097, 863.9822, 5432.1084, 6281.9375, 14.8772, 8.8096, 3.5252, 268.27557, 7.5618, −0.337, −0.2203, −0.2977, −0.2408, −0.3316, −0.2513, −0.2577, −0.2399, −0.2442, 2.0418, 10.0434, 2.0517, 18.7306, 8.1373, 5.3749, 19.0349, 6.5717, 3.2188

FIFTY5M, 3.99, 0, 1.978, −0.1498, 0.0676, −0.2331, −0.0434, −0.1485, −0.09, −0.1111, −0.0376, −0.0838, 1.2641, 1.3919, −0.0406, 1.8807, 1.2251, 9.4306, 190.10002, 232.72774, 359.9007, 4427.0718, 4786.8984, 14.7398, 7.0289, 2.506, 220.23378, 7.1339, −0 3286, −0.2171, −0 2922, −0.2376, −0.3266, −0.2473, −0.2536, −0.2361, −0.2393, 0.7482, 0.2334, 1.6126, −2.2525, 0.3571, 3.1396, −2.3045, 1.9091, 4.4182

FIFTY6M, 4.26, 0, 0.453, −0.1721, 0.0651, −0.1726, −0.0192, −0.1592, −0.0572, −0.1212, −0.017, −0.1398, −0.825, 1.322, 0.1436, 1.5649, 1.9452, 9.3296, 246.10004, 297.72894, 987.5668, 7074.3008, 7881.8496, 14.2073, 8.903, 5.0166, 304.32394, 7.77506, −0.3372, −0.2176, −0.2815, −0.2358, −0.3322, −0.239, −0.2504, −0.2337, −0.2437, 1 0296, 0 7892, 1.4128, −0.0141, 0.6998, 3.2329, 0.2821, 2.1297, 3.8713

FIFTY7M, 3.42, 0, 1.163, −0.1534, 0.0692, −0.2315, −0.0425, −0.1496, −0.0901, −0.112, −0.0359, −0.0854, −0 1482, −0.896, −0.8366, 1.2348, 1.1643, 9.3939, 204.00005, 248.05879, 430 4654, 5355.0903, 5781.7241, 14.874, 7.0252, 3.2923, 238.24918, 7.30371, −0.3303, −0.2179, −0.293, −0.2381, −0.3276, −0.2479, −0.2543, −0.2363, −0.2402, 0.639, −0.4003, 1.3963, −3.7761, −0.3098, 2.1887, −3.8926, 1.2311, 3.5372

FIFTY8M, 4.65, 0, 1.691, −0.1646, 0.0744, −0.2188, −0.0415, −0.159, −0.0903, −0.1155, −0.0379, −0.0849, 2.4647, −1.6411, 0.9885, 3.1217, 1.0171, 9.0832, 228.70006, 284.68515, 880.4409, 5376.9619, 6207.249, 14.9729, 8.7932, 4.5181, 267.28787, 7.58732, −0.3379, −0.221, −0 2994, −0.2417, −0.3341, −0.2517, −0.259, −0.2411, −0.2443, 1.24, 4.6059, 1.4026, 7 7768, 3.6827, 3.2082, 8.0514, 3.5634, 2.5825

FIFTY9M, 4.5, 0, 0.946, −0.1744, 0.0946, −0.2185, −0.0422, −0.1611, −0.0906, − 0.1186, −0.0368, −0.0888, 3.2533, −6.2295, 0.4444, 7.0419, 2.3337, 8.7533, 263.70004, 322.19806, 939.4006, 8435.2578, 9179.1631, 16.8203, 8.8375, 5 2897, 318.35092, 7.95615, −0.3439, −0.2216, −0.3023, −0.2437, −0.3372, −0.2539, −0.2612, −0.243, −0.247, 0.4815, 0.86, 0.7424, 1.6733, 0.9396, 1.9399, 1.6017, 1.5151, 2.0879

SIXTYM, 5.1, 0, 1.708, −0.2, 0.0032, −0.1721, −0.0211, −0.2581, −0.0052, −0.1217, 0.0937, −0.0784, 2.8879, 2.4631, 0.6012, 3.843, 0.5965, 8.4598, 215.50005, 277.0029, 920.8661, 4619 9771, 5517.6655, 14.8613, 8.8769, 3.8591, 255 27687, 7.43845, −0.3492, −0.2458, −0.2572, −0.2297, −0.2637, −0.2277, −0.3161, −0.2066, −0.2668, 0.9875, 0.9045, 2.2799, 0.2514, 1.3607, 2.4025, 0.4733, 1.7128, 1.5065

SIXTY1M, 4.05, 0, 1.279, −0.1348, −0.0544, −0.214, −0.0318, −0.1309, −0.1454, −0.0992, −0.031, −0 0449, −3.1272, −2.8348, −0.7338, 4.2841, 0.6897, 8.3736, 243.30005, 300.14246, 910.797, 6399.2231, 7239.8262, 15.0724, 8.7579, 4.2907, 286.34912, 7.74546, −0.2597, −0.2503, −0.2801, −0.2376, −0.3358, −0.2574, −0.24, −0.271, −0.2292, 5.6952, 2.5348, 1.1675, −1.0829, 1.4467, 4.5366, 0.5882, 1.7904, 3.6123

SIXTY2M, 5.39, 0, 1.708, −0.1888, −0.0058, −0.1214, −0.1218, −0.0809, −0.1808, 0.0319, −0.1176, 0.0048, −0.9459, −2.1292, −0.0858, 2.3314, 0.4655, 8.3488, 215.50005, 271.61209, 867.6917, 4582.3896, 5438.1367, 14.7585, 8.8301, 4.0503, 255 27687, 7.43845, −0.267, −0.2418, −0.3429, −0.2522, −0.2447, −0.3227, −0.221, −0.2778, −0.231, −88.7792, −13.6899, −7.4579, −20.9946, −71.1708, −12.3053, −2.5168, −53.9695, −34.5461

Parameters

POTENCY, pA_{50}. The negative logarithm of the drug concentration required to give a 50% increase in basal contractile force of an isolated paced guinea-pig papillary muscle.

CLOGP [25], calculated octanol/water partition coefficient.
M_ATCH1-9 [12], atom-centred Coulson charges (electrons).
M_DIPV_X,Y,Z [12], dipole moment vectors (D)
M_DIPMOM [12], total dipole moment (D).
M_LUMO_E [12], energy of the lowest unoccupied molecular orbital (Hartree).
M_HOMO_E [12], energy of the highest occupied molecular orbital (Hartree).
VDW_VOL [22], van der Waals volume (Å^3).
D_S_VOL [22], dead space volume (Å^3).
COL_DIAM [22], collision diameter (Å).
SURF_AR [23], surface area (Å^2).
M_OF_I_X,Y,Z [24], moments of inertia (gM10^{-39}).
P_E_AX_X,Y,Z [24], principal elipsoid axes (Å).
MOL_WT [26], molecular weight (Daltons).
M_ESDL1-9 [20,21], electrophilic superdelocalizability (Hartree^{-1}).
M_NSDL1-9 [20,21], nucleophilic superdelocalizability (Hartree^{-1})

MALARIA DATA [17]

Observed biological activity and property data of 2′,4′-substituted 4-aminodiphenylsulfones

SULPHONES SULPHONAMIDES

Compound, *M.lufu*, D_{NH2}, $D_{2/6H}$, VDWVOL, MR, f_{ion}, I_{OH}, I_{Cl}, LogK

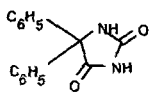
4′-NH_2, 1.2, −0.253, −0.108, 3.32, 1.03, 0, 0, 0, 0.339
4′-OCH_3, 7.55, −0.127, −0.038, 3.32, 1.03, 0, 0, 0, 1.5056
4′-NO_2, 31.03, 0.111, 0.058, 3.32, 1.03, 0, 0, 0, 1.5898
4′-H, 12.06, 0, 0, 3.32, 1.03, 0, 0, 0, 1.3749
4′-OH, 1.5, −0.156, −0.054, 3.32, 1.03, 0 451, 0, 0, 0 9645
4′-Cl, 12.99, 0.004, 0.006, 3.32, 1.03, 0, 0, 0, 2.267
4′-$NHCOCH_3$, 7.01, −0.114, −0.036, 3.32, 1.03, 0, 0, 0, 0.8954
4′-Br, 9 69, 0, 0, 3.32, 1.03, 0, 0, 0, 2.3804
4′-$NHCH_3$, 1.29, −0.253, −0.105, 3.32, 1.03, 0, 0, 0, 1.2032
4′-NHC_2H_5, 2.75, −0.25, −0.098, 3.32, 1.03, 0, 0, 0, 1.646
4′-CH_3, 8.08, −0.091, −0.027, 3.32, 1.03, 0, 0, 0, 1.8901
4′-$NH(CH_3)_2$, 1.76, −0.237, −0.089, 3.32, 1.03, 0, 0, 0, 1.7702
4′-$COOCH_3$, 11.75, 0.005, 0.026, 3.32, 1.03, 0, 0, 0, 1.7346
4′-COOH, 3.6, 0.03, 0.022, 3.32, 1.03, 1, 0, 0, −2.4711
4′-$CONHNH_2$, 12.72, −0.003, 0.022, 3.32, 1.03, 0, 0, 0, −0.1965
4′-$NHCH_2COOCH_3$, 2.24, −0 141, −0.095, 3.32, 1.03, 0, 0, 0, 0.7193
4′-$NHCH_2COOH$, 0.87, −0.141, −0.095, 3.32, 1.03, 1, 0, 0, −0.1716
2′-OH; 4′-NH_2, 0.73, −0.345, −0 095, 8.04, 2.85, 0, 1, 0, −0.0091
2′-OH, 4′-NHC_2H_5, 0.73, −0.338, −0.095, 8.04, 1.8318, 0, 1, 0, 1.2583
2′-OH; 4′-NHC_3H_7, 1.97, −0.341, −0.088, 8.04, 2.85, 0, 1, 0, 1.9318

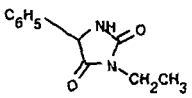
2′-OH; 5′-OH, 1.55, −0.185, −0.023, 8.04, 2.85, 0, 1, 0, 0.5562
2′-OH; 5′-NH_2, 1.55, −0.208, −0.036, 8.04, 2.85, 0, 1, 0, −0.1965
2′-OCH_3; 4′-NH_2, 1.87, −0.319, −0.114, 16.07, 7.87, 0, 0, 0, 0.0054
2′-OCH_3; 4′-NHC_2H_5, 1.77, −0.328, −0.114, 16.07, 7.87, 0, 0, 0, 1.2766
2′-OCH_3; 4′-NHC_3H_7, 1.21, −0.328, −0.111, 16.07, 7.87, 0, 0, 0, 1.8501
2′-OCH_3; 4′-NHC_4H_9, 1.03, −0.328, −0.114, 16.07, 7.87, 0, 0, 0, 2.4236
3′-OH, 4′-NH_2, 0.86, −0.247, −0.137, 3.32, 1.03, 0, 0, 0, 0.351
3′-OCH_3; 4′-NH_2, 2.77, −0.244, −0.064, 3.32, 1.03, 0, 0, 0, 0.7647
3′-OCH_3; 4′-NHC_2H_5, 3.24, −0.247, −0.061, 3.32, 1.03, 0, 0, 0, 2.03
3′-OCH_3; 4′-NHC_3H_7, 3.99, −0.25, −0.0621, 3.32, 1.03, 0, 0, 0, 2.604
2′-NO_2, 4′-NH_2, 1.34, −0.084, −0.075, 16.8, 7.36, 0, 0, 0, 0.8543
2′-NH_2; 4′-NO_2, 11.62, 0.001, 0.048, 10.54, 5.42, 0, 0, 0, 1.8199
2′-NO_2; 4′-NO_2, —, 0.29, 0.048, 16.8, 7.36, 0, 0, 0, 1.6384
2′-NO_2, 2.43, 0.131, 0.019, 16.8, 7.36, 0, 0, 0, 1.308
2′-NH_2, 2.22, −0.218, −0.018, 10.54, 5.42, 0, 0, 0, 1.0801
2′-NH_2; 4′-NH_2, 1.03, −0.312, −0 121, 10.54, 5.42, 0, 0, 0, −0.096
2′-Cl; 4′-NH_2, 0.48, −0.188, −0.111, 12, 6.02, 0, 0, 1, 1.0585
2′-CN; 4′-NH_2, 2.23, −0.084, −0.039, 14.7, 6.33, 0, 0, 0, 0.6815
2′-CF_3; 4′-NH_2, —, −0.198, −0.166, 21.33, 5.02, 0, 0, 0, 1.4095
2′-COOH; 4′-NH_2, 7.28, −0.5, 0.167, —, 6.93, 1, 0, 0, —
2′-SO_2NH_2; 4′-NH_2, 1.73, −0.211, −0.039, 30.84, 12.28, 0.5, 0, 0, −0.284
2′-NH_2; 4′-SO_2NH_2, 1.83, −0.068, 0.016, 10.54, 5.42, 0.5, 0, 0, 0.0227
2′-NO_2; 4′-SO_2NH_2, 53 46, 0.231, 0.035, 16.8, 7.36, 0.8, 0, 0, 0.7528
2′-CH_3; 4′-NH_2, 0.65, −0.244, −0.178, 13.67, 5.65, 0, 0, 0, 0.8327
2′-$NHCH_3$; 4′-NH_2, 2.32, −0.322, −0.111, 21.75, 10.33, 0, 0, 0, 0.6005
2′-$N(CH_3)_2$; 4′-$N(CH_3)_2$, —, −0.367, −0.137, 31.67, 15.55, 0, 0, 0, 1.925
2′-$COCH_3$; 4′-NH_2, 3.004, −0.335, −0.144, 25.37, 11.18, 0, 0, 0, 0.1923
2′-$NH(CH_3)_2$; 4′-NH_2, 6.96, −0.384, −0.147, 31.67, 15.55, 0, 0, 0, 0.8695
2′-Cl; 4′-NHC_3H_7, 0 86, −0.185, −0.106, 12, 6.02, 0, 0, 1, 2.5658
2′-Cl; 4′-NHC_2H_5, 1 03, −0.182, −0.108, 12, 6.02, 0, 0, 1, 2.2251
2′-CH_3; 4′-NHC_2H_5, 1.09, −0.247, −0.177, 13.67, 5.65, 0, 0, 0, 2.091
2′-CF_3; 4′-NHC_2H_5, 1.78, −0.188, −0.159, 21.33, 5.02, 0, 0, 0, 2.452
2′-$CONH_2$; 4′-NH_2, 14.29, −0.315, 0.012, 22.24, 9.81, 0, 0, 0, −1.5336
2′-$CH(OH)CH_3$, 4′-NH_2, 4.88, −0.257, −0.207, 28.49, 11.82, 0, 0, 0, 0.0022
2′-CH_2OH; 4′-NH_2, 2.55, −0.231, −0.186, 18.77, 7.19, 0, 0, 0, −0.1231
2′-CH_3; 4′-$NHCH_3$, 0.75, —, −0.177, 13.67, 5.65, 0, 0, 0, 1.659
2′-CH_3; 4′-NHC_3H_7, 1.62, —, −0.178, 13.67, 5.65, 0, 0, 0, 2.2182
2′-$COOCH_3$; 4′-NH_2, 2.92, —, −0.085, 29.07, 12.87, 0, 0, 0, 1.6255
4′-$NHCH_2CONHNH_2$, 2.9, −0.24, −0.099, 3.32, 1.03, 0, 0, 0, −0.7047
4-CN, —, 0.079, −0.035, 3.32, 1.03, 0, 0, 0, 1.068
'—' symbol means data is absent.

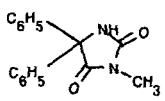
Parameters [17]

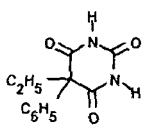
M. lufu, activity derived from bacterial growth kinetics of *M. lufu* using nonlinear regression.
$D_{2/6H}$, difference from reference NMR chemical shift of the 2,6 aromatic protons.
D_{NH2}, difference from reference NMR chemical shift of the 4-amino protons.
VDWVOL, van der Waals volume (taken from reference parameters [3]) for substituents in the 2′ position
MR, molar polarizability of substituents in the 2′ position.
f_{ion}, the fraction ionised in the 4′ position.
I_{OH}, indicator variable, is there an OH in the 2′ position.
I_{Cl}, indicator variable, is there a Cl in the 2′ position.
LogK, HPLC capacity factors.

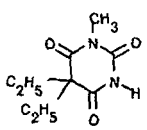
CNS DATA [18]

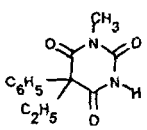

phenytoin

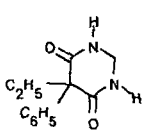

ethotoin

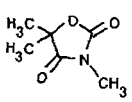

mephenytoin

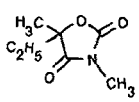

phenobarbital

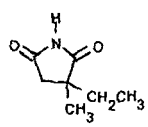

metharbital

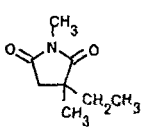

mephobarbital

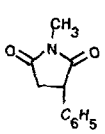

primidone

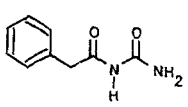

trimethadione

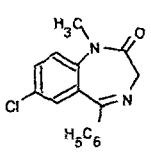

paramethadione

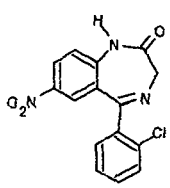

ethosuximide

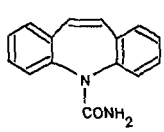

methsuximide


phensuximide


phenacemide


diazepam


clonazepam


carbamazepine

Number, Compound, MES, MET, logMW, logP, Dipole.

PHENYTOIN, 4.42, –, 2.4, 2.47, 1.74
ETHOTOIN, 3.38. 3 63, 2.31, *1.53*, 1.74
MEPHENYTOIN. 3.56. 3.86, 2.34, 2.09, 1.74
PHENOBARBITAL, 4.03, 4.25, 2.37. 1.42, 0.87
METHARBITAL, 3.19. 4.3, 2.3, 1.21, 1.13
MEPHOBARBITAL. 3.86. 4.02, 2.39, 1.98, 0.87
PRIMIDONE. 4.28, 3.57, 2.34, 2.1, 1.35
TRIMETHADIONE, 2.36, 2.68, 2.16, −0.37, 1.74
PARAMETHADIONE, 2.82, 3.4. 2.2, 0.13, 1.69
ETHOSUXIMIDE. 2.15. 3.04, 2.15, *0.01*, 1.47
METHSUXIMIDE. 3.43, 3.48, 2.31, 1.54, 1.61
PHENSUXIMIDE, 3.23, 3.58, 2.28, 1.4, 1.61
PHENACEMIDE, 3.31, 3.19, 2.25. 0.57, 2.06
DIAZEPAM, 4.17, 6 24. 2.45. 2.82, 2.65
CLONAZEPAM, 3.56, 7.55, 2.51. 2.41, 2.33
CARBAMAZEPINE, 4.4, –, 2.37, 2.18, 2.41

Parameters [18]

MES, Maximal Electroshock Seizure test. converted to log 1/C where C is the ED_{50} in mol/kg.
MET, the subcutaneous pentylene-tetrazol seizure test.
LogMW, log of molecular weight.
LogP. octanol/water partition coefficient.
Dipole. dipole moment (D).