Structure-Mutagenicity Study of 12 Trimethylimidazopyridine Isomers Using Orbital Energies and "Spectrum-like Representation" As Descriptors†

M. Vraćko,*,‡ A. Szymoszek,‡,§ and P. Barbieri‡,||

National Institute of Chemistry, Hajdrihova 19, Ljubljana, Slovenia, Faculty of Chemistry, University of Wroclaw, Poland, and Università degli Studi di Trieste, Via Giorgieri,1, 34127, Trieste, Italy

Received October 30, 2003

The set of 12 trimethylimidazopyridine isomers with mutagenic potency toward two strains of Salmonella was treated in this study. Ten isomers with known mutagenic properties were taken to build the models. Fifteen molecular orbital energies, or a "spectrum-like" representation of 3D structures, were taken as descriptors. As modeling techniques the multiple linear regression and the counter propagation neural network were applied. Models were tested with the recall ability test and the leave-one-out cross-validation tests. For two isomers, which have not been synthesized yet, we report predicted values for both mutagenic potencies obtained with different models. The best models were found when unoccupied molecular orbital energies are among the descriptors.

1. INTRODUCTION

The mutagenic and carginogenic potency of aromatic amines has been an object of several theoretical and experimental studies. Many monocyclic amines derived from azo dyes, which are widely used in the chemical and pharmaceutical industry, are toxic, mutagenic, and carcinogenic for humans and animals. A review of the structure-mutagenicity studies (QSAR) is reported by Chung et al.[1] Considering a series of studies it was generally recognized that metabolic activation is required for the mutagenicity and carcinogenicity of most aromatic amines. Metabolic activation may involve initial N-oxidation and a covalent bonding on biomolecules. In many QSAR studies a strong correlation between molecular orbitals (HOMO, LUMO) and mutagenic potency was found which implies that electron transfer mechanisms play an important role in the mutagenesis. Another important fact is that small modifications of molecular structures drastically affect their mutagenic potency. This implies that the mechanism of mutagenicity is sensitive on the 3D molecular structure, i.e., on the spatial arrangement of amino and other groups. In such a case other mechanisms such as hydrogen bonding could be involved. Zhang et al.[2] studied mutagenicity of heterocyclic amines using the CASE method.[3] They searched for the molecular fragments that can be responsible for mutagenic activity. On the basis of a large data set they found that the fragment $NH_2-C=N-C$ is responsible for mutagenicity (biophores). They also found that 1-methylbenzol[1,2-]imdazole, 3H-imidazo-[4,5-b]pyridine, and 3H-benzo[1,2-d]imidazole, which are structurally similar to compounds treated in this study (see below), are nonmutagenic (biophobes). In addition, they concluded that the mutagenic potency is directly related to LUMO energies. On the other hand, Chung et al.[4] found no correlation between orbital energies and mutagenic potency for a set of benzidine analogues. Benigni et al.[5] give a comprehensive review on QSAR studies of mutagenic and carcinogenic aromatic amines. Basak et al.[6,7] applied a hierarchical QSAR on a set of 95 aromatic amines. This data set was compiled by Debnath et al.[8] and was intensively studied by many authors.[9−16] The best correlations to mutagenic potency were found with topostructural and topochemical indices, which carry implicitly the information on molecular structures.[17] The inclusion of geometrical and quantum chemical indices, which carry information on precise 3D structure and orbital energies, did not improve the correlation. In addition, further reports[18−20] provide information on mutagenicity of cyclic amines.

In this study we treated the set of 12 trimethylimidazopyridine isomers,[21] which are described in the section *Data Set*. They belong to heterocyclic amines, which can be formed during the cooking processes. Some of them are mutagenic and carcinogenic. The treated isomers have similar structures, but they show a very different mutagenic potency. The main question is how the spatial arrangement of amino group and methyl groups influences the mutagenicity. Trying to attack this problem we applied a "spectrum-like" representation to describe 3D structures in models.[22,23] This representation is very sensitive on 3D structural parameters (see Section *Methods*). Since some of the previous studies show a strong correlation between LUMO energy and mutagenicity we also took several orbital energies as descriptors. One modeling experiment with molecular orbital energies as descriptors was made with Multiple Linear Regression (MLR), another with Counterpropagation Neural Network (CP NN).[24,25] Both methods are briefly described in the section *Methods*.

2. METHODS

2.1. Multiple Linear Regression. In the MLR technique the output variable (mutagenic potency) is expressed as a

* Corresponding author phone: +386 1 4760315; fax: +386 1 4760300; e-mail: marjan.vracko@ki.si.

† Dedicated to George W. A. Milne, former Editor-in-Chief of the *Journal of Chemical Information and Computer Sciences.*

‡ National Institute of Chemistry.

§ University of Wroclaw.

|| Università degli Studi di Trieste.

10.1021/ci030420i CCC: $27.50
Published on Web 02/12/2004

multilinear expansion of descriptors. The building of linear models is very computer effective, and therefore a large number of models with different combinations of descriptors can be analyzed. The method is suitable to select the descriptors that are correlated to the selected property. We applied the program package CODESSA using the *heuristic option* to select the descriptors,[26] which are mostly correlated to the mutagenic potency. The selection algorithm is as follows. The program calculates the one-parameter correlation equations and eliminates descriptors that do not fulfill the following criteria:

The F-test's value is below one.

The correlation coefficient is less than r_{min} ($r_{min} = 0.01$).

The t value is less than t_1 ($t_1 = 0.1$).

The descriptor is highly correlated with another descriptor, and this other descriptor has a higher correlation coefficient in the one-parameter correlation equation.

With the remaining descriptors all possible two-parameter correlations are calculated, and the procedure is extended to a given maximum number of descriptors.

2.2. Counter Propagation Neural Network. The CP NN has been described in detail in many textbooks and articles.[27−29] It consists of two layers of neurons the input (Kohonen) layer and the output layer, which are arranged in a two-dimensional rectangular matrix. Learning in the input layer is done in the same way as in a Kohonen network. This means a vector of input variables that represents an object is presented to all neurons and the program selects the neuron that has weights most closely to the input values (winning neuron). In the next step, the weights of the winning neuron and the neighboring neurons are modified. This procedure is repeated till the weights are stabilized and the network is considered to be trained. The most important feature of the trained network is that the objects with similar input vectors are located close to each other. This makes the counterpropagation neural network a suitable tool for clustering.

The output layer has the same arrangement of neurons as the input layer. During the learning process the output values are given to the output layer, and the weights in the output layer are changed in a way that they become similar to the given values. A trained counter propagation neural network is a suitable tool for prediction of properties for new objects. Such an object is, first, situated on a neuron which weights are most similar to its weights. Second, this position is transferred to the output layer, which gives the output (property) value. Analyzing the neighborhood of the selected neuron one can determinate the objects responsible for the prediction. Furthermore, it has been shown that with a proper definition of output layer the CP NN is a suitable tool for classification.[25] Generally, the CP NN is a nonlinear method, and thus it is a suitable tool for modeling when the relationship between structure and property is nonlinear. In comparison to MLR the CP NN can be applied also when the number of descriptors overwhelms the number of objects.[30] For example, different spectra can be used to represent the molecular structures.[31]

The models were tested with the recall ability test and with the leave-one-out cross-validation test. The recall ability test shows how good the model recognizes compounds of the training set. In the leave-one-out cross-validation test the objects one after another are selected out, whereas for every selected object the model is built up with remaining ones. Furthermore, this model is used to predict the value for the selected object. The test gives information on the prediction ability of models. For both tests we report the correlation coefficients for regression lines r_{ra} and r_{cv}, respectively. The r_{cv} was taken as criterion to optimize the technical parameters of models. The optimal dimension of network was found as 6×6, or 7×7, and the number of training epochs 500, or 1000.

2.3. Spectrum-like Representation of 3D Molecular Structures. Since the details of the representation have recently been published, only some basic ideas are given here.[22,23] A representation is constructed from a 3D molecular structure in three steps. First, a 3D structure is projected on the *xy*, *xz*, and *yz* planes. In the second step, each projection (figure) is treated separately. A figure is put into a circle with an arbitrary radius. A project beam from the center of the circle produces a pattern of points on the circle where each point represents a particular atom. In the third step, each point on a circle is taken as a center for the Lorenzian curve of the form:

$$s_i(\phi) = \frac{\rho_i}{(\phi - \phi_i)^2 + \sigma_i^2} \quad (1)$$

Here, ρ_i and ϕ_i are the distance between the origin of the coordinate system and the position of the ith atom and its polar angle, respectively. σ_i is a free parameter, which can be associated with any atomic property. If we considered only geometries, the σ_i's are set to one, otherwise we selected atomic charges as atomic property. The spectrum related to the figure is a sum of all atomic Lorenzians, and it is defined in the interval $(0,2\pi)$. The complete 3D structure is therefore represented with three spectra. It can be shown that the representation is uniform, unique, and reversible, but it is not invariant on translation and rotation. This problem can be solved by proper orientation of molecules. For molecules with a common substructure it is convenient to select an orientation rule to the superposition of the substructures.

In this study all three projections were taken to represent the structures. An example is shown in Figure 1. Since the molecules are close, planar higher resolution was taken in the *xy* plane as compared to the *xy* and *yz* planes. The resolution parameters were 50, 20, 20 points in the interval $(0,2\pi)$ for the *xy*, *xz*, and *yz* projections, respectively.

2.4. Molecular Orbital Energies and Atomic Charges as Descriptors. Energies of the highest occupied molecular orbitals (HOMO) and the lowest unoccupied orbitals (LUMO) are often used as molecular descriptors. According to the Koopmans' theorem the HOMO energies are related to the ionization potentials, the LUMO energies to the electron affinities, and the differences to the excitation energies.[32] A strong correlation between molecular orbital energies and property indicates a possible electron exchange mechanism. In this study the energies of 10 orbitals close to HOMO and five orbitals close to LUMO were checked as possible descriptors. Other parameters taken from the quantum chemical calculations were atomic charges, which were calculated according to the Mulliken population analysis. It is known that the Mulliken population analysis suffers several shortcomings such as the following: the partitioning scheme of

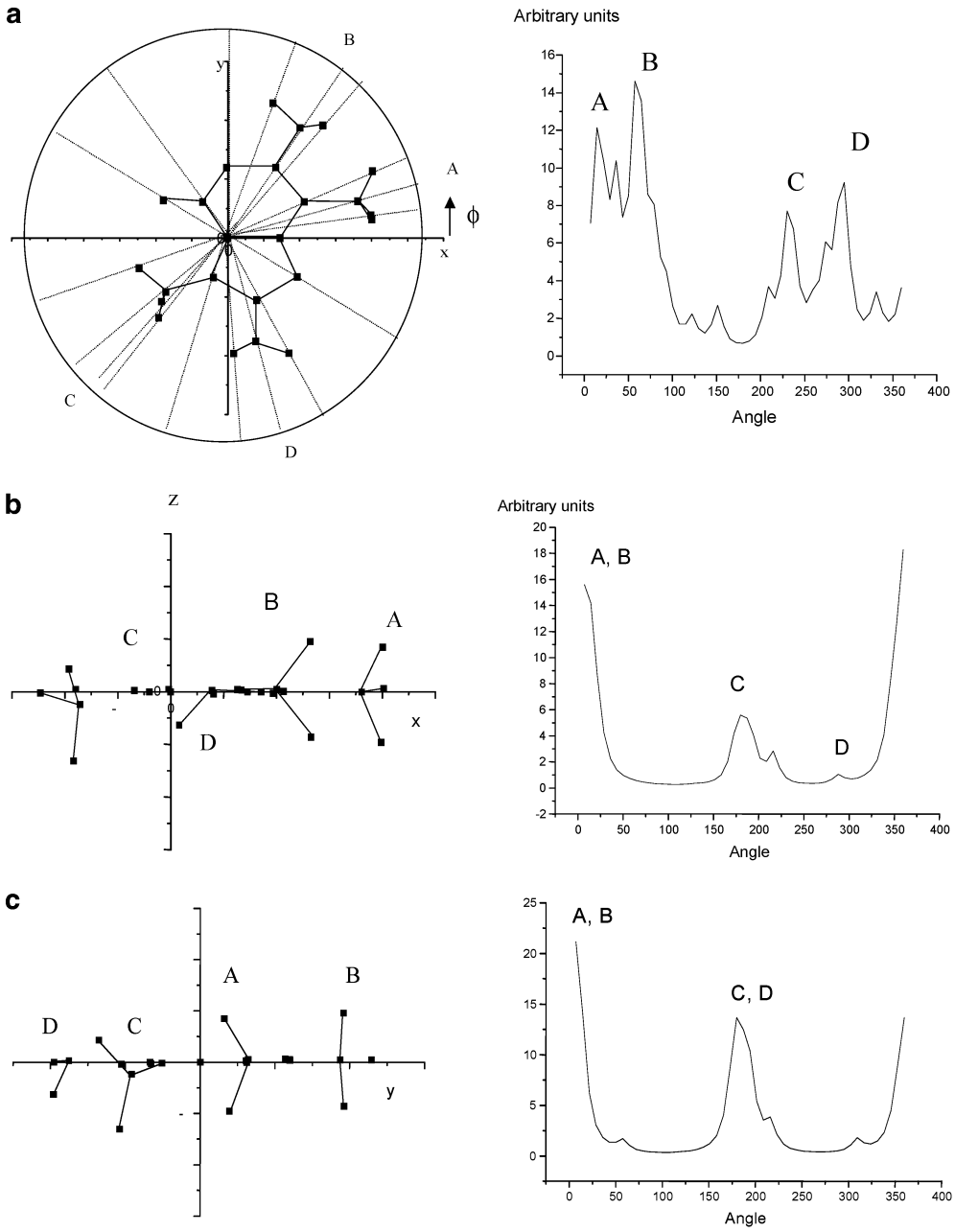


Figure 1. Construction of spectrum-like representation: a, *xy* projection; b, *xz* projection; and c, *yz* projection (A, B, C, D label the segments of molecule).

the total electron density is not unique, the calculated charges are strongly dependent upon the used basis set, and the atomic orbitals describing the charge on a particular atom are not always localized on this atom. In addition, there is a lack of experimental techniques for the direct measurements of atomic charges. Despite this, the comparison of charges calculated with the same basis set provides reasonable information on a charge distribution among molecules.[32] The atomic charges were incorporated into a spectrum-like representation according to eq 1. The quantum chemical calculations were performed with the program GAUSSIAN in the Hartree−Fock approximation using the 6-311G basis set.[33]

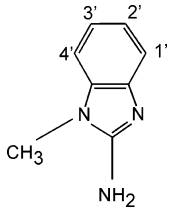
3. DATA SET

The molecules treated in this study are shown in Table 1. The mutagenic properties are reported by Felton et al.[21] The mutagenic potency expressed as a number of revertants per nanomole was measured for two strains of Salmonella, YG1024 and TA98. For 10 isomers the experimental data are available. Although the structures of isomers are similar, their mutagenic potencies span a range from <0.5 to 300 TA98 revertants per nanomole and from 6 to 870 revertants per nanomole in strain YG1024. For the modeling the natural logarithms (ln) of these numbers were taken. To avoid the negative numbers in the case of TA98 the numbers of revertants were multiplied by a factor of 10.

4. MODELS

The data set of selected isomers shows a wide range of mutagenic potency and provides a good example for the QSAR study. Due to the clear similarity of isomers it is expected that they follow the same mechanism of mutage-

Table 1. 12 Isomers and Their Mutagenic Potencies



molecule[a]	arrangement	YG1024(lnYG1024)	TA98(ln10*TA98)
1(a)	1′C−CH_3 2′C−CH_3 3′N 4′C−H	7.0(1.946)	0.5(1.609)
2(b)	1′C−H 2′N 3′C−CH_3 4′C−CH_3	6.7(1.902)	0.5(1.609)
3(c)	1′C−CH_3 2′C−H 3′N 4′C−CH_3		
4(d)	1′C−CH_3 2′N 3′C−H 4′C−CH_3	6.0(1.792)	0.9(1.792)
5(e)	1′C−H 2′C−CH_3 3′N 4′C−CH_3	9.2(2.219)	2.3(3.135)
6(f)	1′C−CH_3 2′N 3′C−CH_3 4′C−H	13.5(2.603)	1.6(2.773)
7(g)	1′C−CH_3 2′C−H 3′C−CH_3 4′N	103.0(4.635)	28.0(5.635)
8(h)	1′N 2′C−CH_3 3′C−H 4′C−CH_3	11.7(2.460)	5.7(4.043)
9(i)	1′C−CH_3 2′C−CH_3 3′C−H 4′N	555.0(6.319)	317.0(8.061)
10(j)	1′N 2′C−H 3′C−CH_3 4′C−CH_3		
11(k)	1′C−H 2′C−CH_3 3′C−CH_3 4′N	868.0(6.766)	317.0(8.061)
12(l)	1′N 2′C−CH_3 3′C−CH_3 4′C−H	15.6(2.747)	3.1(3.434)

[a] Characters a−l are labels of molecules in Kohonen networks.

Table 2. Statistical Results (r_{ra}, r_{cv}, F) and Descriptors (des) for the Best Linear Models (A − Best One-Parameter Model, B − Best Two-Parameter Model, C − Best Three-Parameter Model, D − Best Four-Parameter Model, E − Best Five-Parameter Model)

	YG1024				TA98			
	r_{ra}	r_{cv}	F	des	r_{ra}	r_{cv}	F	des
A	0.8488	0.7353	20.6099	LUMO+1	0.8415	0.7253	19.4199	LUMO+1
B	0.9421	0.8566	27.6137	LUMO+1	0.8987	0.7564	14.7042	LUMO+1
				LUMO+2				LUMO+2
C	0.9610	0.8401	24.1520	LUMO+1	0.9735	0.9140	36.2251	LUMO+1
				LUMO+2				HOMO-3
				HOMO-6				HOMO-4
D	0.9792	0.8915	29.1914	LUMO+1	0.9793	0.9035	29.2483	LUMO+1
				LUMO+2				HOMO-3
				LUMO+3				HOMO-4
				HOMO-3				HOMO-7
E	0.9953	0.9728	85.0032	LUMO+1	0.9851	0.9263	26.1943	LUMO+1
				LUMO+2				LUMO+2
				LUMO+3				HOMO-3
				HOMO-3				HOMO-4
				HOMO-7				HOMO-7

nicity. In this study we try to get an insight into this mechanism. The question is which molecular properties are more correlated to mutagenic potency, pure geometrical parameters, charges distribution, or orbital energies. The strong correlation between molecular geometry and mutagenic potency would imply the picture that the spatial arrangement of methyl and amino groups governs the binding on DNA. If the electrostatic properties play any role the representation with charge distribution must be correlated with mutagenic potency. The correlation to orbital energies implies the electron exchange mechanisms. Four cases were analyzed.

Case 1. Ten occupied and five unoccupied molecular orbital energies were taken as descriptors. Program CODESSA (Multiple Linear Regression)[26] was applied to select those energies that are mostly correlated with the mutagenic potency. Several hundreds of linear models were analyzed; we set three criteria to select the models: $r > 0.8950$, $F > 10.$ and $r_{cv} > 0.8366$. With the selected models the mutagenic potency for molecules c and j were predicted. The results are given in Table 2. In all selected models the LUMO+1 energy is the most important descriptor. This is in agreement with the previously reported results,[21] where a correlation to the LUMO energy is found.

Table 3. Neural Network Results (OE − 15 Orbital Energies as Descriptors, SLNC − Spectrum-like Representation without Atomic Charges, SLWC − Spectrum-like Representation with Atomic Charges)

	YG1024		TA98	
	r_{ra}	r_{cv}	r_{ra}	r_{cv}
OE	1.00	0.88	1.00	0.85
SLNC	1.00	0.80	1.00	0.68
SLWC	1.00	0.88	1.00	0.82

Case 2. Ten occupied and five unoccupied molecular orbital energies were taken as descriptors and the CP NN method was taken for modeling. The results of the leave-one-out tests are given in Figure 2a,d and in Table 3. The Kohonen map (Figure 3) shows three clusters: the first one contains the molecules k, i, and g, the second cluster contains the molecules h and l, and the third cluster contains the molecules a and e. This clustering matches well with the mutagenic potency. In the prediction procedure the molecule c is situated between molecules a and i, and the molecule j

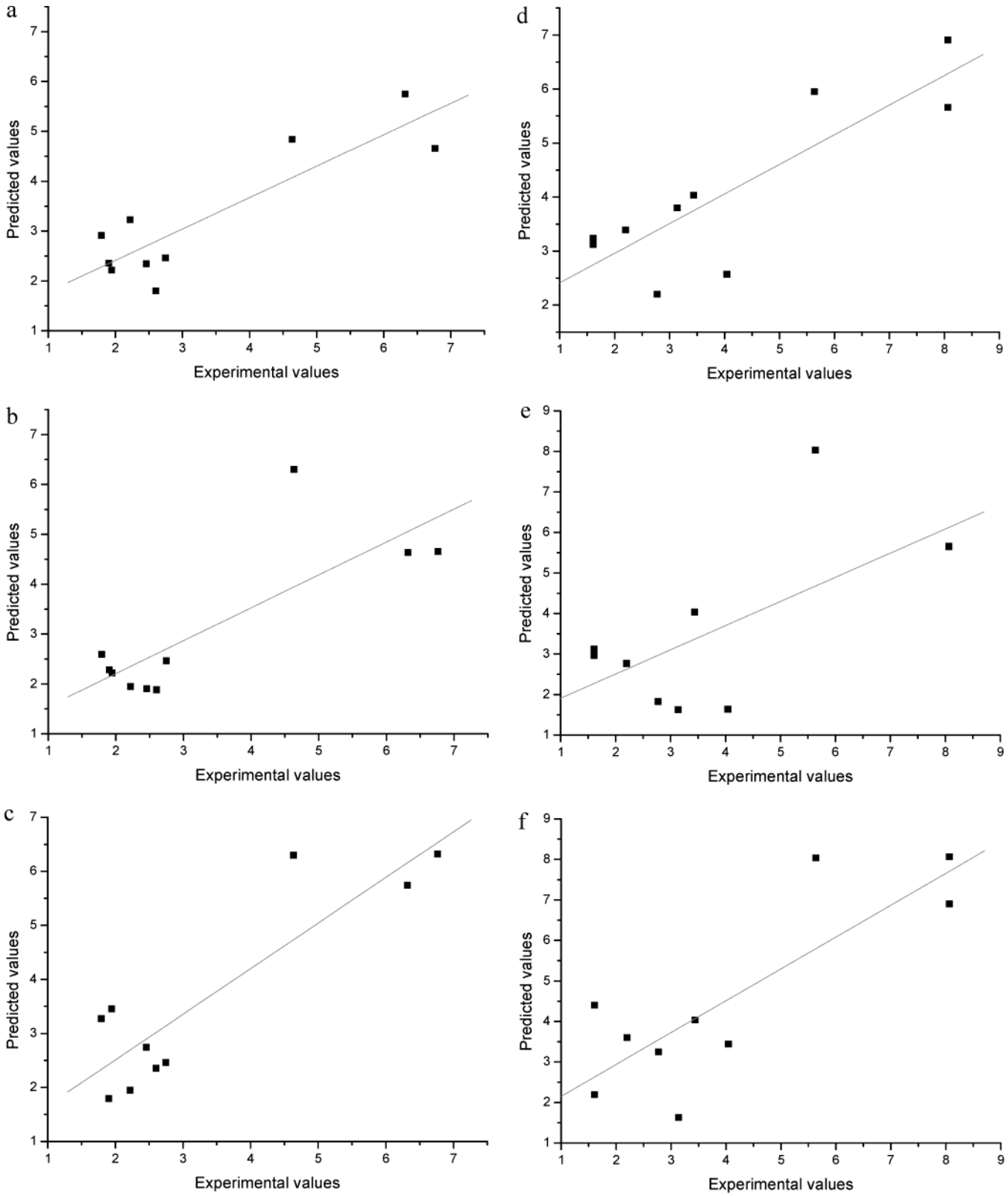


Figure 2. Leave-one-out cross-validation results for CP NN models: a, MOE yg1024; b, SLNC yg1024; c, SLWC yg1024; d, MOE TA98; e, SLNC TA98; f, SLWC Ta98.

in the cluster to the molecules l and h. The predicted values for both mutagenic potencies are given in Table 4.

Case 3. Spectrum-like representation without atomic charges was taken for descriptors and CP NN as the modeling technique. The Kohonen map and the results of the tests are given in Figure 2b,e and in Table 4. The Kohonen map (Figure 3) shows two clusters poorly correlated to mutagenicity. The clusters consist of molecules d, f, i, g and h, l, e, a, respectively. The predicted values are given in Table 4. The molecule c is situated between molecules a and e (similar to case 2), and the molecule j close to the molecule b. Indeed, the molecules j and b show very similar structures, the only difference is the position of one ring atom (nitrogen) and consequently the position of one hydrogen atom.

Table 4. Predicted Mutagenic Potencies for Two Isomers, in Parentheses Are the Natural Logarithms as Calculated from Models[a]

	molecule c				molecule j			
	YG1024		TA98		YG1024		TA98	
	mean[b]	s[b]	mean[b]	s[b]	mean[b]	s[b]	mean[b]	s[b]
OELM	10.60		1.39		8.00		0.70	
	(2.360)	(0.257)	(2.633)	(0.401)	(2.079)	(0.435[c])	(1.939)	(0.149[d])
OENN	7.99		1.04		13.59		4.15	
	(2.077)		(2.343)		(2.609)		(3.726)	
SLNC	8.06		1.10		6.71		0.50	
	(2.087)		(2.398)		(1.904)		(1.609)	
SLWC	9.20		2.30		15.60		3.10	
	(2.219)		(3.135)		(2.747)		(3.434)	

[a] OELM, orbital energies using a linear model; OENN, orbital energies using a neural network; SLNC, spectrum-like representation without atomic charges using a neural network; SLWC, spectrum-like representation with atomic charges using neural network. [b] Mean value and standard deviation *s* have been calculated only for the linear models, which meet the conditions described in the text. [c] One predicted value 216.8 (5.3791) was removed as outlier. [d] Two predicted values 0.3 (0.9509) and 17.9 (5.1907), were removed as outliers.

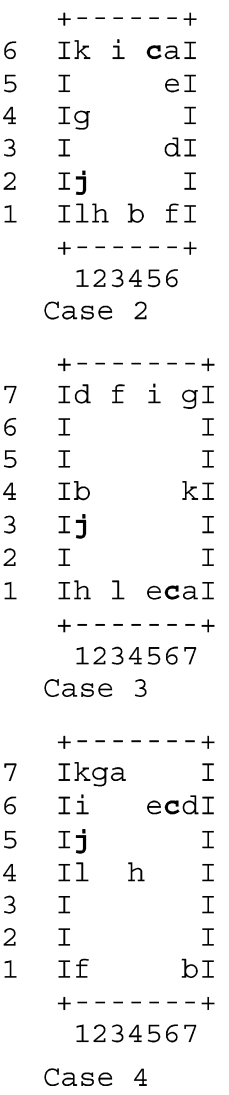


Figure 3. Kohonen maps for cases 2−4 (see text): case 2, OE; case 3, SLNC; case 4, SLWC characters a−l label the isomers, bold are predictions.

Case 4. Spectrum-like representation with atomic charges was taken for descriptors and CP NN as modeling technique. Atomic charges were calculated as described above. The Kohonen map (Figure 3) shows two clusters: one contains the molecules i, k, g, a, and the second cluster contains the molecules d and e. The first cluster correlates (beside the molecule a) well with mutagenicity. The prediction follows the similarities, c is between the molecules e and d, and j is between i and l.

5. DISCUSSION AND CONCLUSIONS

The study confirms the strong correlation between LUMO+*n* energies and mutagenic potency for this set of compounds.[21] MLR found three orbital energies close to the LUMO energy out of 15 orbital energies as the most correlated descriptors in the case of the YG1024 strain. In the case of the TA98 strain the most correlated is the LUMO+1 energy followed by HOMO-3, 4 energies (see Table 2). Several linear models gave comparable predictions for two isomers (see Table 4, OELM model). As a second method of modeling the CP NN was applied using three kinds of descriptors: molecular orbital energies, spectrum-like representation of 3D structures, and spectrum-like representation of 3D structures with incorporated atomic charges. The recall ability test shows that all three models recognized the data from the training set perfectly (Table 3). The results of the leave-one-out test are shown in Table 3 and in Figure 2a−f. The models with orbital energies and 3D representation with atomic charges overwhelm the model with pure 3D (geometrical) representation. The predicted values for the molecules c and j are shown in Table 4. For the molecule c the models predicted between 7 and 11 revertants per nanomole for strain YG1024 and between 1 and 2.3 revertants per nanomole for strain TA98. For the molecule j the situation is ambiguity: two models predicted a lower number of revertants and two models predicted a higher number. This is true for both strains of Salmonella.

ACKNOWLEDGMENT

M.V. thanks Ministry of Education, Science and Sport for the support of the work under the following contract: Program 034 507. A.S. and P.B. thank the EU for the grants in the framework of the project IMAGETOX (HPRN-CT-1999-00015).

REFERENCES AND NOTES

(1) Chung, K.-T.; Kirkovsky, L.: Kirkovsky, A.; Purcell, W.P. Review of mutagenicity of monocyclic aromatic amines: quantitative structure-activity relationships. *Mutat. Res.* **1997**, *387*, 1−16.
(2) Zhang, Y.P.; Klopman, G.; Rosenkranz, H.S. Structural basis of the mutagenicity of heterocyclic amines formed during the cooking processes. *Environ. Mol. Mutagen.* **1993**, *21*, 100−115.

(3) Klopman, G. Artificial intelligence approach to structure−activity studies: Computer automated structure evaluation of biological activity of organic molecules. *J. Am. Chem. Soc.* **1984**, *106*, 7315−7321.
(4) Chung, K.-T.; Chen, S.-C.; Wong, T.-Y.; Li, Y.-S.; Wei, C.-I.; Chou, M.W. Mutagenic studies of benzidine and its analogs: Structure−activity relationship. *Toxicol. Sci.* **2000**, *56*, 351−356.
(5) Benigni, R.; Giuliani, A.; Franke, R.; Gruska, A. Quantitative structure−activity relationships of mutagenic and carcinogenic aromatic amines. *Chem. Rev.* **2000**, *100*, 3697−3714.
(6) Basak, S. C.; Mills, D.; Balaban, A. T.; Gute, B. D. Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: A hierarchical QSAR approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 671−678.
(7) Basak, S. C.; Mills, D. Prediction of mutagenicity utilizing a hierarchical QSAR approach. *SAR QSAR Environ. Res.* **2001**, *12*, 481−496.
(8) Debnath, A. K.; Debnath, G.; Shusterman, A. J.; Hansch, C. A QSAR investigation of the role of hydrophobicity in regulating mutagenicity in the Ames test: 1. Mutagenicity of aromatic and heteroaromatic amines in Salmonella typhimurium TA98 and TA100. *Environ. Mol. Mutagen.* **1992**, *19*, 37−52.
(9) Benigni, R.; Andreoli, C.; Giuliani, A. QSAR models for both mutagenic potency and activity: Application to nitroarenes and aromatic amines. *Environ. Mol. Mutagen.* **1994**, *24*, 208−219.
(10) Benigni, R.; Passerini, L.; Gallo, G.; Giorgi, F.; Cotta-Ramusino, M. QSAR models for discriminating between mutagenic and non-mutagenic aromatic and heteroaromatic amines. *Environ. Mol. Mutagen.* **1998**, *32*, 75−83.
(11) Cash, G. G. Prediction of the genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2001**, *491*, 31−37.
(12) Frierson, M. R.; Klopman, G.; Rosenkranz, H. S. Structure-activity relationships (SARs) among mutagens and carcinogens: a review. *Environ. Mutagen.* **1986**, *8*, 283−327.
(13) Hawkins, D. M.; Basak, S. C.; Shi, X., F. QSAR with few compounds and many features. *Chem. Inf. Comput. Sci.* **2001**, *41*, 671−678.
(14) Kier, L. B.; Simons, R. J.; Hall, L. H. Structure activity studies on mutagenicity of nitrosamines using molecular connectivity. *J. Pharm. Sci.* **1978**, *67*, 725−726.
(15) Maran, U.; Karelson, M.; Katritzky, A. R. A comprehensive QSAR treatment of the genotoxicity of heteroaromatic and aromatic amines. *Quant. Struct.-Act. Relat.* **1999**, *18*, 3−10.
(16) Vracko, M.; Mills, D.; S. C. Basak, S. C. Structure-mutagenicity modelling using counter propagation neural network. *Environ. Toxicol. Pharmacol.* Accepted for publication.
(17) Randić, M. *Topological indices*. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer III, H. F., Schreiner, P. R., Eds.; Wiley: London, 1998; pp 3018−3032.
(18) Benigni, R.; Giuliani, A. Quantitative structure−activity relationship (QSAR) studies of mutagens and carcinogens. *Med. Res. Rev.* **1996**, *16*, 267−284.
(19) Hatch, F. T.; Lightstone, F. C.; Colvin, M. E. Quantitative structure-activity relationship of flavonoids for inhibition of heterocyclic amine mutagenicity. *Environ. Mol. Mutagen.* **2000**, *35*, 279−299.
(20) Karelson, M.; Sild, S.; Maran, U. Non-linear QSAR treatment of genotoxicity. *Mol. Simul.* **2000**, *24*, 229−242.
(21) Felton, J. S.; Knize, M. G.; Hatch, F. T.; Tanga, M. J.; Colvin, M. E. Heterocyclic amine formation and the impact of structure on their mutagenicity. *Cancer Lett.* **1999**, *143*, 127−134.
(22) Zupan, J.; Novic, M. General type of a uniform and reversible representation of chemical structures. *Anal. Chim. Acta* **1997**, *348*, 409−418.
(23) Zupan, J.; Vracko, M.; Novic, M. New uniform and reversible representation of 3D chemical structures. *Acta Chim. Slov.* **2000**, *47*, 19−37.
(24) Vracko, M. A study of structure−carcinogenic potency relationship with artificial neural networks. The using of descriptors related to geometrical and electronic structures. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1037−1043.
(25) Vracko, M.; Novic, M.; Zupan, J. Study of structure-toxicity relationship by a counterpropagation neural network. *Anal. Chim. Acta* **1999**, *384*, 319-332.
(26) Katritzky, A. R.; Karelson, M.; Lobanov, V. S. CODESSA, Version 2.20, University of Florida, 1994−1996.
(27) Zupan, J.; Gasteiger, J. *Neural networks in chemistry and drug design;* WILEY-VCH Verlag GmbH: Weinheim, 1999; p. 103.
(28) Hecht-Nielsen, R. Counter propagation networks. *Appl. Optics* **1987**, *26*, 4979-4984.
(29) Zupan, J.; Novic, M.; Gasteiger, J. Neural networks with counter-propagation learning strategy used for modelling. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 175−187.
(30) Arciniegas, F.; Bennett, K.; Breneman, C.; Embrechts, M. J. Molecular database mining using self-organizing maps for the design of novel pharmaceuticals. *Proceedings of the ANNIE conference*, St. Louis, MO, November 2001.
(31) Bursi, R.; Dao, T.; Wijk, T. V.; Gooyer, M. D.; Kellenbach, E.; Verwer, P. Comparative spectra analysis (CoSA): spectra as three-dimensional molecular descriptors for the prediction of biological activities. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 861-867.
(32) Hehre, J.; Random, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio molecular orbital theory*; Wiley: New York, 1986.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A.; GAUSSIAN 94; Gaussian Inc.: Pittsburgh, PA, 1995.

CI030420I