Soft Computing in the Design of Nontoxic Chemicals

Les Sztandera,*,† Mendel Trachtman,‡ Charles Bock,‡ Janardhan Velga,§ and Ashish Garg§

Computer Information Systems Department, Chemistry Department, and School of Textiles and Materials Technology, Philadelphia University, Philadelphia, Pennsylvania 19144

Received September 6, 2002

This research focuses on the use of soft computing to aid in the development of novel, state-of-the-art, nontoxic dyes which are of commercial importance to the U.S. textile industry. Where appropriate, modern molecular orbital (MO) and density functional (DF) techniques are employed to establish the necessary databases of molecular properties to be used in conjunction with the neural network approach. In this research, we focused on the following: (1) using molecular modeling to establish databases of various molecular properties of azo dyes required as input for our neural network approach; (2) designing and implementing a neural network architecture suitable to process these databases; and (3) investigating combinations of molecular descriptors needed to predict various properties of the azo dyes.

1. INTRODUCTION

This research involves the integration of fuzzy entropies (used in the context of measuring uncertainty and information) with computational neural networks. An algorithm for the creation and manipulation of fuzzy entropies, extracted by a neural network from a data set, is designed and implemented. The neural network is used to find patterns in terms of structural features and properties that correspond to a desired level of activity in various azo dyes. Each molecule is described by a set of structural features, a set of physical properties, and the strength of some activity under consideration. After developing an appropriate set of input parameters, the neural network is trained with selected molecules, and then a search is carried out for compounds that exhibit the desired level of activity. High level molecular orbital and density functional techniques are employed to establish databases of various molecular properties required by the neural network approach.

The structural and electronic properties of the positional isomers of monomethoxy-4-aminoazobenzene (n-OMe-AAB) have been investigated using density functional theory with a basis set that includes polarization functions on all the atoms. These aminoazo dyes are of interest because their carcinogenic activities depend dramatically on the position (n) of the methoxy group, e.g. 3-OMe-AAB is a potent hepatocarcinogen in the rat, whereas 2-OMe-AAB is a noncarcinogen. Although the various isomers of OMe-AAB require metabolic activation via N-hydroxylation prior to reaction with cellular macromolecules, we have shown that there are structural and electronic features present in these isomers that correlate with their carcinogenic behavior.

3-Methoxy-4-aminoazobenzene (3-OMe-AAB) is a potent hepatocarcinogen in the rat.[1] This aminoazo dye requires metabolic activation to N-hydroxy-3-methoxy-4-aminoazobenzene (N-OH-3-OMe-AAB) prior to reaction with cellular macromolecules.[2] This conclusion is in accord with the observation that 3-OMe-AAB is mutagenic on the Ames' *Salmonella* system only after treatment with S-9, the 9000 g supernatant fraction of liver homogenate, whereas N-OH-3-OMe-AAB is strongly mutagenic without S-9 treatment[3,4] Interestingly, changing the position of the methoxy group on the phenyl rings dramatically influences the carcinogenic behavior of the resulting compound.[5] For example, 2-OMe-AAB is noncarcinogenic in rats, whereas 4′-OMe-AAB is carcinogenic but to a lesser degree than 3-OMe-AAB. This carcinogenic potency of 2- and 4′-OMe-AAB correlates well with their mutagenic activity in the Ames' *Salmonella* test, where neither 2-OMe-AAB nor its N-hydroxy derivative, N-OH-2-OMe-AAB are mutagenic even after treatment with S-9; 4′-OMe-AAB is very slightly mutagenic on *Salmonella* (TA98) and N-OH-4′-OMe-AAB is definitely mutagenic without S-9 treatment.[6] Unfortunately, the carcinogen/mutagenic activities of the remaining monomethoxy derivatives of 4-aminoazobenzene or their N-hydroxy analogues have not been reported.[7] (It is known that *N*,*N*-dimethyl-3′-OMe-AAB is carcinogenic.) For comparison, we note that the parent compound, 4-aminoazobenzene is only weakly carcinogenic in rats, nonmutagenic on *Salmonella* (TA98) with or without S-9 treatment and mutagenic on *Salmonella* (TA100) only in the presence of S-9.[6]

Although it is not entirely clear why there is such a radical difference in the carcinogenic behavior of 2- and 3-OMe-AAB, Kojima et al.[1] have determined that N-OH-3-OMe-AAB has a significantly greater effect than N-OH-2-OMe-AAB on DNA synthesis *in vivo*. This suggests that the observed differences in the carcinogenic activity of 2-OMe-AAB and 3-OMe-AAB may be linked to the differences in the inhibitory effects of their N-hydroxy derivatives on DNA replication. Hashimoto et al.[8] have established that the cytochrome P-450 enzymes efficiently catalyze the mutagenic activation of 3-OMe-AAB, and, in contrast to other carcinogenic aromatic amines, the activation is mediated by phenobarbital-P-450 rather than by 3-methylcholanthrene-P-450.

* Corresponding author e-mail: SztanderaL@PhilaU.edu.
† Computer Information Systems Department.
‡ Chemistry Department.
§ School of Textiles and Materials Technology.

10.1021/ci025600r CCC: $25.00
Published on Web 01/08/2003

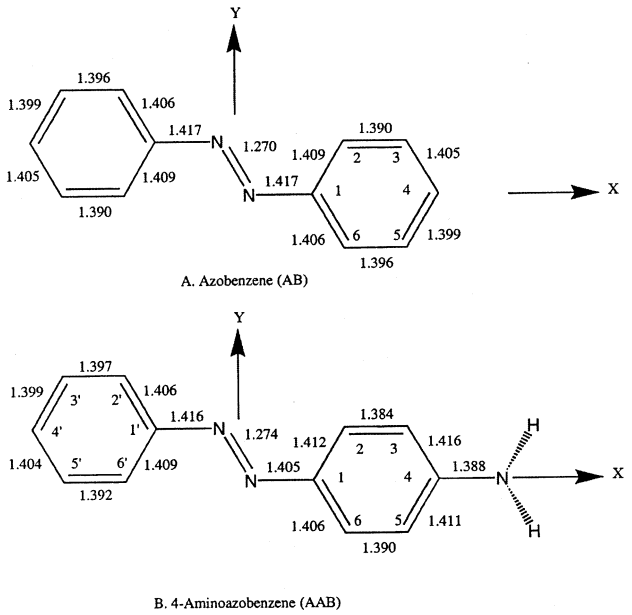

Figure 1. The structures, coordinate system, and numbering conventions for A. azobenzene (AB) and B. 4-aminoazobenzene (AAB). The bond lengths (Å) shown were calculated at the BP/DN**//BP/DN** computational level.

Despite significant interest in the carcinogenic behavior of the various positional isomers of OMe-AAB, relatively little is known about their structural or electronic properties. No experimental results from X-ray or electron diffraction studies have been reported for any of the OMe-AAB isomers[9] (Only a few experimental structures of azo dyes have been reported: O-Aminoazotoluene (X-ray) Kurosaki, S.; Kashino, S.; Haisa, M. *Acta Crystallogr*. **1976**, *B32*, 3160; Disperse Red 167 (X-ray) Freeman, H. S.; Posey, J. C., Jr.; Singh, P. *Dyes Pigm.* **1992**, *20*, 279; C.I. Disperse Yellow 86 (X-ray), Lye, J.; Hink, D.; Freeman, H. S. Computational Chemistry Applied to Synthetic Dyes 1997). Furthermore, no high-level computational results that compare the various OMe-AAB isomers using either molecular orbital or density functional theory calculations are currently available in the literature. It is important to note that substitution at the 2- and 6-positions or at the 3- and 5-positions in 4-aminoazobenzene are not equivalent, see Figure 1.

However, it is not evident that distinctions of this type were considered in the carcinogenic/mutagenic studies involving 2- and 3-OMe-AAB.[1−5] Thus, it is probably more appropriate to describe these studies as involving methoxy substitution at the meta and ortho positions, respectively. The purpose of this chapter is to describe the results of an extensive computational study using density functional theory (DFT) to establish the conformational preferences and relative energies of the positional isomers of OMe-AAB. Our goal is to identify any electronic and/or structural features that may be present among these positional isomers that can be correlated with their diverse carcinogenic behaviors and lead to a better understanding of the underlying molecular mechanism(s) involved.

2. COMPUTATIONAL METHODS

Density functional calculations were performed at the BP/DN** computational level with SPARTAN v5.0 on Silicon Graphics computers.[10] This level uses the nonlocal Becke-Perdew (BP) 86 functional and employs the numerically defined DN** basis set which includes polarization functions on all the atoms.[11,12] Complete optimizations for a variety of conformers of each OMe-AAB derivative were carried out; no symmetry constraints were employed in order to minimize the likelihood of optimizing to a transition state. In a few cases frequency analyses were performed to ensure that the optimized structures were local minima on the potential energy surfaces (PESs). The graphics utilities of SPARTAN were used to examine the electron densities, electrostatic potentials, and various Kohn−Sham orbitals for each conformer. Mulliken and electrostatic charges were also calculated.

2.1. Molecular Modeling Results and Discussion. Since no experimental structural data are available even for the parent compounds azobenzene (AB) and 4-aminoazobenzene (AAB), we first optimized these molecules at the BP/DN** computational level. The initial structures of both AB and AAB were taken as nearly trans about the azo linkage[13] (A second conformer of AB, with the phenyl rings twisted some 80°, was found to be 2.3 kcal/mol higher in energy at the BP/DN** computational level.), and, in the case of AAB, the amino group was taken as pyramidal with both hydrogen atoms on the same side of the ring.[14] (We also optimized a conformer of AAB in which the hydrogen atoms bonded to the amine nitrogen atom are on opposite sides of the ring but otherwise the structure is planar. This conformer is 8.0 kcal/mol higher in energy than the form shown in Figure 1. The optimized structure of a completely planar form of AAB is a transition state that is 0.15 kcal/mol higher in energy than the lowest energy form shown in Figure 1). The optimized structure of AB is found to be planar and a frequency analysis confirms that this is a local minimum on the PES. As can be seen in Figure 1, the azo linkage in AB significantly distorts the carbon−carbon bond lengths in the phenyl rings compared to their values in benzene, where the carbon−carbon bond distances are 1.403 Å at this computational level. The length of the N═N bond in AB, 1.270 Å, suggests considerable electron delocalization; the N═N bond lengths in $CH_3-N{=}N-CH_3$ and $CH_3-N{=}N-C_6H_5$ are shorter, 1.250 Å and 1.258 Å respectively. In the optimized structure of AAB, the two phenyl rings are practically planar and nearly coplanar with each other, see Figure 1. A frequency analysis confirms that this is a local minimum on the PES. To a large extent, the calculated structural parameters of AAB show that the amine group at the 4-position reinforces the geometrical changes already induced by the azo linkage. This is a consequence of electron delocalization using the lone pair of electrons from the amine nitrogen atom to give the C_4-N bond partial double bond character, which results in predictable adjustments of the bond lengths in the remainder of the molecule. For comparison, we note that the lengths of the C−N bonds in CH_3-NH_2 and $C_6H_5-NH_2$ are 1.478 Å and 1.408 Å, respectively, considerably longer than that found in AAB, 1.388 Å. Nevertheless, the structure at the amine nitrogen in AAB remains pyramidal—the sum of the three bond angles is 346.7° at this computational level, compared to 318.4° and 325.6° for NH_3 and NH_2-CH_3, respectively.

It is of interest to compare a few of the Kohn−Sham molecular orbitals of AB and AAB. The highest occupied molecular orbital (HOMO) in AB is a lone-pair orbital

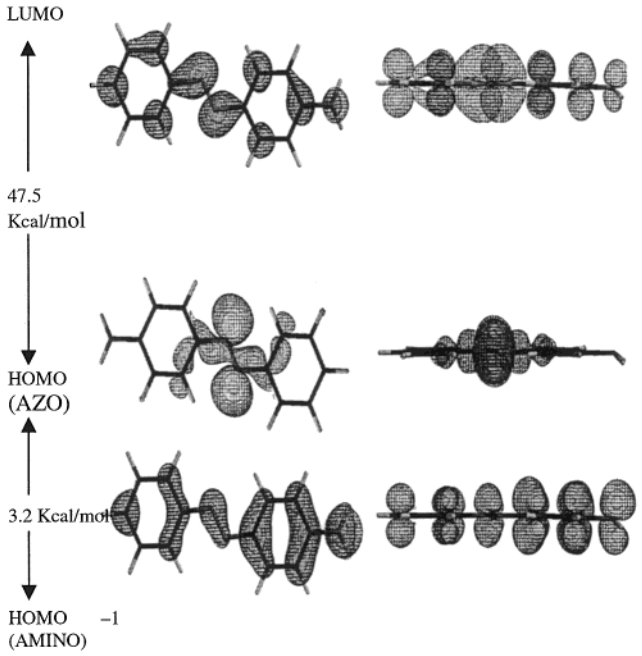

Figure 2. The HOMO{-1}, HOMO, and LUMO of 4-aminoazobenzene (AAB) calculated at the BP/DN**//BP/DN** computational level.

localized primarily in the vicinity of the azo linkage, which is 14.4 kcal/mol above the next highest occupied orbital (HOMO{-1}), a delocalized *pi*-bonding orbital. The lowest unoccupied molecular orbital (LUMO) in AB is a *pi*-antibonding orbital, some 46.7 kcal/mol above the HOMO. In AAB the HOMO also involves the azo lone-pair electrons, see Figure 2. It is nearly identical in shape to the HOMO found in AB, but it is 7.4 kcal/mol higher in energy. The HOMO{-1} in AAB involves the lone pair of electrons on the amine nitrogen atom, see Figure 2, but otherwise it is similar in shape to the HOMO{-1} (*pi*-bonding orbital) in AB. However, this HOMO{-1} is 18.6 kcal/mol above its counterpart in AB, reducing the energy gap between the two highest occupied orbitals in AAB to only 3.2 kcal/mol. The LUMO in AAB is similar in shape to the *pi*-antibonding LUMO in AB except that it includes a contribution from the amine nitrogen atom, see Figure 2. The energy gap between the HOMO and LUMO is 47.5 kcal/mol.

Two general types of conformers were considered for each of the positional isomers of monomethoxy AAB: one in which the O−Me bond is essentially in the nominal plane of the phenyl ring to which it is bonded and the other in which the O−Me bond is nearly perpendicular to this ring. For all nine isomers, the conformers in which the O−Me bond lies essentially in the ring plane are found to be *lower* in energy at the BP/DN**//BP/DN** computational level; the energies for these conformers are listed in Table 1 along with those for AB and AAB. Selected geometrical parameters and properties of the various OMe-AAB isomers calculated at the BP/DN** level are listed in Tables 2 and 3.

Since several orientations of the methoxy methyl group are possible, their positions for the lowest energy conformers at the BP/DN** level are shown in Figure 3. It should be noted that energy differences between some of the conformers of these monomethoxy isomers can be quite small. For example, rotating 180° about the C_ϕ-O bond in 4′-OMe-AAB and reoptimizing yields a structure only 0.3 kcal/mol higher in energy, whereas a conformer with the methyl group nearly perpendicular to the ring is 3.8 kcal/mol higher in energy.

Table 1. Total Molecular Energies (a.u.) of n-Methoxy-4-aminoazobenzene Calculated at the BP/DN**//BP/DN** Computational Level

n	total molecular energies (BP/DN**//BP/DN**)	relative energie
2	−742.931442	+6.7
3	−742.942450	0.0
5	−742.941042	+0.9
6	−742.936054	+4.0
2′	−742.935475	+4.4
3′	−742.940108	+1.5
4′	−742.940449	+1.3
5′	−742.941441	+0.6
6′	−742.930842	+7.3
AAB	−628.365269	
AB	−572.974339	

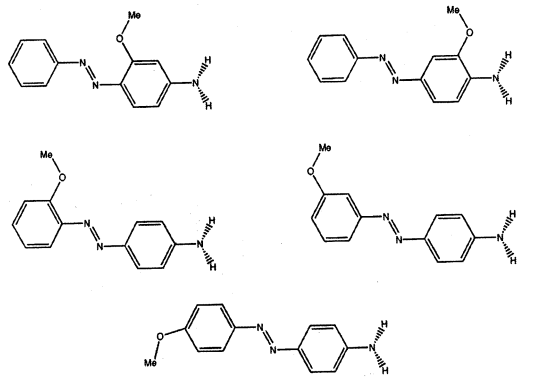

Figure 3. The orientation of the methyl group for the positional isomers of monomethoxy-4-aminobenzene calculated at the BP/DN**//BP/DN** computational level. (The orientation of the methyl group in 6-OMe-AAB is analogous to that for 2-OMe-AAB, etc.)

2.1.1. 3- and 5-OMe-AAB. As can be seen from Table 1, 3-OMe-AAB has the lowest total molecular energy among all the positional isomers at the BP/DN**//BP/DN** computational level. However, the other ortho derivative, 5-OMe-AAB, is less than 1 kcal/mol higher in energy. As expected, the presence of a methoxy group ortho to the amine group perturbs the pattern of carbon−carbon bond lengths in the phenyl rings compared to those in AAB. The most prominent changes occur at the point of attachment, e.g. in the case of 3-OMe-AAB, the length of the shorter bond (C_2−C_3) decreases while the length of the longer bond (C_3−C_4) increases, see Table 2. As might be expected, the lengths of the bonds in the unsubstituted phenyl ring are not significantly altered by the presence of the methoxy group. In both 3- and 5-OMe-AAB the length of C_4−N bond is shorter than that found in AAB and the amine group is less pyramidal, see Table 2. This suggests a further delocalization of the lone pair electron density on the amine nitrogen atom; the calculated Mulliken and electrostatic charges on this nitrogen are less negative than those found in AAB. The C_ϕ−O bond lengths in the 3- and 5-isomers, 1.378 Å and 1.380 Å, respectively, are *longer* than those calculated for the other positional isomers, see Table 2. To examine the effect of the amine group on the length of the C_ϕ−O bond, we replaced the amine group in 3-OMe-AAB with a hydrogen atom and reoptimized the structure. In the resulting 3-OMe-

Table 2. Structural Parameters (Bond Lengths (Å) and Bond Angles (deg)) of AB, AAB, and n-OMe-AAB

n	C_1-C_2	C_2-C_3	C_3-C_4	C_4-C_5	C_5-C_6	C_6-C_1	C_4-N	C_1-N	N=N
2	1.436	1.398	1.409	1.409	1.386	1.410	1.391	1.396	1.278
3	1.416	1.382	1.428	1.404	1.393	1.403	1.380	1.402	1.278
5	1.407	1.388	1.409	1.421	1.388	1.412	1.384	1.404	1.275
6	1.409	1.384	1.415	1.408	1.400	1.427	1.390	1.396	1.277
2′	1.411	1.384	1.415	1.410	1.392	1.407	1.392	1.406	1.277
3′	1.411	1.383	1.416	1.411	1.390	1.408	1.385	1.405	1.274
4′	1.412	1.385	1.415	1.410	1.390	1.407	1.390	1.404	1.276
5′	1.411	1.384	1.416	1.410	1.389	1.408	1.388	1.401	1.274
6′	1.412	1.383	1.414	1.410	1.392	1.406	1.392	1.412	1.277
AAB	1.412	1.384	1.416	1.411	1.390	1.406	1.388	1.405	1.274
AB	1.409	1.390	1.405	1.399	1.396	1.406		1.417	1.270

n	$C_{1'}-N$	$C_{1'}-C_{2'}$	$C_{2'}-C_{3'}$	$C_{3'}-C_{4'}$	$C_{4'}-C_{5'}$	$C_{5'}-C_{6'}$	$C_{6'}-C_{1'}$	$C_{\varnothing}-O$	O−C	Σ angles[a]
2	1.414	1.406	1.395	1.399	1.404	1.392	1.408	1.356	1.430	344.5
3	1.413	1.405	1.396	1.399	1.404	1.392	1.409	1.378	1.433	348.9
5	1.417	1.406	1.396	1.398	1.403	1.392	1.410	1.380	1.431	346.8
6	1.416	1.406	1.397	1.399	1.404	1.392	1.410	1.362	1.430	346.0
2′	1.404	1.425	1.404	1.398	1.401	1.392	1.406	1.364	1.431	344.6
3′	1.416	1.408	1.398	1.406	1.398	1.395	1.405	1.374	1.433	349.2
4′	1.411	1.408	1.389	1.406	1.408	1.391	1.406	1.370	1.432	345.7
5′	1.415	1.402	1.399	1.393	1.411	1.394	1.411	1.373	1.432	347.2
6′	1.401	1.410	1.390	1.397	1.398	1.404	1.432	1.359	1.433	345.2
AAB	1.416	1.406	1.397	1.399	1.404	1.392	1.409			346.7
AB	1.417	1.406	1.396	1.399	1.405	1.390	1.404			

[a] Sum of three bond angles at the amine nitrogen atom.

Table 3. Selected Properties of AB, AAB, and n-OMe-AAB Calculated at the BP/DN**//BP/DN** Computational Level

n	HOMO-{−1} (a.u.)	HOMO (a.u.)	LUMO (a.u.)	Log P[d]	electrostatic charge on amine nitrogen	dipole moment (D)
2	−0.186080[a]	−0.171712[b]	−0.101479	2.25	−0.72	4.85
3	−0.186278[b]	−0.184263[a]	−0.109082	2.13	−0.67	3.52
5	−0.185290[b]	−0.183553[a]	−0.109506	2.19	−0.67	3.54
6	−0.185048[a]	−0.179274[b]	−0.106652	2.26	−0.72	5.08
2′	−0.183749[a]	−0.179501[b]	−0.107565	2.29	−0.72	2.77
3′	−0.188844[a]	−0.186038[b]	−0.110120	2.24	−0.73	4.54
4′	−0.183084[c]	−0.182205[c]	−0.104728	2.34	−0.72	2.55
5′	−0.189970[a]	−0.187243[b]	−0.110805	2.53	−0.71	4.50
6′	−0.183052[a]	−0.173803[b]	−0.102735	2.30	−0.72	2.04
AAB	−0.192206[a]	−0.187648[b]	−0.111910	2.47	−0.72	3.61
AB	−0.222337[a]	−0.199380[b]	−0.126611	3.30		0.07

[a] Orbital involves amine lone pair. [b] Orbital involves azo lone pairs. [c] Orbital is mixed, see text. [d] Log P is the logarithm of the octanol−water partition coefficient calculated using the Dixon-Hehre algorithm in Spartan 5.0[10]. This involves explicit evaluation of $AM1_{oct}$ and $AM1_{aq}$ solvation models. The Ghose-Crippen approach gives Log P = 3.54 for all the OMe-AAB isomers (*J. Comput. Chem.* **1988**, *9*, 80).

AB compound (as well as in $MeO-C_6H_5$) the $C_\phi-O$ bond length, 1.373 Å, is slightly shorter than that found in 3-OMe-AAB. Thus, the amine group at the 4-position tends to impede the delocalization of lone-pair density on the methoxy oxygen atom in 3-OMe-AAB.

The presence of a methoxy group ortho to the amine group in AAB has an interesting effect on the two highest occupied Kohn−Sham molecular orbitals. The orbital localized at the azo linkage in 3(5)-OMe-AAB is similar in shape and only 0.9(1.5)kcal/mol higher in energy than the corresponding orbital in AAB; the in-plane lone pair on the methoxy oxygen is represented in this orbital but only to a small extent. The orbital involving the amine lone pair of electrons, which includes a significant contribution from the out-of-plane oxygen lone pair, is 5.3(5.7) kcal/mol higher in energy than the corresponding orbital in AAB. These relatively large increases in energy make this orbital the HOMO in both 3- and 5-OMe-AAB.

It is important to note that methoxy substitution at each of the positions on the phenyl rings increases the energy of this orbital, but the increases for the 3- and 5-isomers are more than triple the smallest increase we observed, 1.7 kcal/mol for 5′-OMe-AAB. The relatively large increase in the energy of this orbital appears to be a result of electron overcrowding involving the proximate lone pairs on the methoxy oxygen and amine nitrogen atoms. The shape of the LUMO in 3(5)-OMe-AAB is quite similar to the LUMO in AAB, having relatively little contribution from the out-of-plane lone-pair orbital on the methoxy oxygen atom. Furthermore, the energy of these LUMO in both the 3- and 5-isomers is less than 2 kcal/mol above the LUMO in AAB. The energies separating the HOMO and LUMO in 3- and 5-OMe-AAB, 47.2 and 46.5 kcal/mol, are just slightly lower than that found in AAB.

2.1.2. 2- and 6-OMe-AAB. The positional isomers 2- and 6-OMe-AAB are 6.7 and 4.0 kcal/mol higher in energy than the lowest energy isomer, 3-OMe-AAB, at the BP/DN**//BP/DN** computational level, see Table 1. The length of the C_4-N bond in both of these isomers is slightly longer than that found in AAB and the amine group is more pyramidal, see Table 2. The calculated Mulliken and electrostatic charges on the amine nitrogen atom in the 2- and 6-isomers are nearly the same as those in AAB. However, the $C_\phi-O$ bond lengths in 2(6)-OMe-AAB, 1.356 Å(1.362 Å), are some 0.02 Å shorter than the corresponding bond lengths in 3(5)-OMe-AAB. This indicates further delocalization involving the oxygen out-of-plane lone pair, which gives the $C_\phi-O$ bond additional double bond character. The Mulliken charge on the oxygen atom in 2-OMe-AAB is not as negative as that found in 3-OMe-AAB. To examine the effect of the amine group on the length of the $C_\phi-O$ bond in 2-OMe-AAB, we optimized the structure of the 2-OMe-AB. The length of the $C_\phi-O$ bond increases, but

only by about 0.002 Å; this change, however, is in a direction opposite to what we observed in going from 3-OMe-AAB to 3-OMe-AB. The shorter C_ϕ−O bond in 2(6)-OMe-AAB results in an elongation of both carbon−carbon bonds in the ring at the point of methoxy attachment when compared to that in AAB, see Table 2. Again, there are no significant changes in the carbon−carbon bond lengths in the unsubstituted phenyl ring compared to those in AAB.

The presence of a methoxy group meta to the amine group in AAB alters the two highest occupied Kohn−Sham orbitals differently than when the replacement occurs at an ortho position. In particular, the azo lone-pair orbital in 2(6)-OMe-AAB is 10.0(5.3) kcal/mol higher in energy than the corresponding orbital in AAB but only 0.9(1.5) kcal/mol higher for substitution at the 3(5)-position. This orbital involves contributions from the azo nitrogen lone pairs and from the in-plane methoxy oxygen lone-pair; its relatively large increase in energy is clearly the result of adverse lone-pair interactions in the region. The particular geometrical arrangement of atoms in the vicinity of the trans azo linkage causes the electron overcrowding in this region to be more severe for methoxy substitution at the 2-position than at the 6-position. This leads to a greater increase in the energy of the azo lone pair orbital in 2-OMe-AAB and results in the largest energy gap between the two highest occupied orbitals we observed in this study, 9.0 kcal/mol. The orbital involved with the amine nitrogen lone pair in 2(6)-OMe-AAB is similar in shape to the corresponding orbital in AAB, although it includes a contribution from the out-of-plane oxygen lone pair. Its energy is raised to a slightly lesser extent than it is when the methoxy group is at an ortho position. Thus, for 2(6)-OMe-AAB and AAB the orbital involving the azo lone pair is higher in energy than the orbital involving the amine nitrogen lone pair, whereas for 3(5)-OMe-AAB the order of these two orbitals is reversed. It is also interesting to note that the energy gap between the two highest occupied molecular orbitals in 2(6)-OMe-AAB, 9.0-(3.6) kcal/mol, is greater than that in AAB, 3.2 kcal/mol, and in 3(5)-OMe-AAB, 1.3(1.1) kcal/mol. The LUMO in both the 2- and 6-isomers is similar in shape to that in AAB but involves a significant contribution from the out-of-plane lone-pair orbital on the methoxy oxygen atom. The LUMO energies of 2- and 6-OMe-AAB are higher than those of 3- and 5-OMe-AAB, whereas the separation in energy between the HOMO and LUMO is smaller, 44.1 and 45.6 kcal/mol, respectively.

2.1.3. 3′- and 5′-OMe-AAB. The positional isomers 3′-(5′)-OMe-AAB are only 1.5 (0.6) kcal/mol higher in energy than 3-OMe-AAB, see Table 1. Interestingly, methoxy substitution at the 3′- or 5′-position has very little effect on the carbon−carbon bond lengths in either of the phenyl rings when compared to those in AAB, see Table 2. The C_4−N bond length in 3′-OMe-AAB is slightly shorter than that found in AAB while that of 5′-OMe-AAB is nearly the same as in AAB. The small differences in the geometrical parameters and the lack of any severe lone-pair interactions in 3′(5′)-OMe-AAB are consistent with the observation that the two highest occupied Kohn−Sham orbitals of AAB are closer in energy to those of 3′(5′)-OMe-AAB than to those of the other positional isomers. The azo and amine lone-pair orbitals are only 1.0(0.3) and 2.4(1.7) kcal/mol higher in energy than the corresponding orbitals in AAB, leading to a small energy gap of 1.8(1.7) kcal/mol. The LUMOs are again similar in shape to that found in AAB, with relatively little contribution from the out-of-plane oxygen lone-pair orbital; the HOMO−LUMO energy gaps, 47.6 and 48.0 kcal/mol, are slightly greater than that found in AAB.

2.1.4. 2′- and 6′-OMe-AAB. The positional isomers 2′- and 6′-OMe-AAB are found to be 4.4 and 7.3 kcal/mol higher in energy than 3-OMe-AAB, see Table 1. As can be seen in Table 2, methoxy substitution at the 6′-position of AAB leads to greater changes in several of the geometrical parameters than those found with the other monomethoxy derivatives. The relatively short $C_{6'}$−O bond length, 1.359 Å, in 6′-OMe-AAB indicates significant electron donation from the out-of-plane lone pair on the methoxy oxygen. This short $C_{6'}$−O bond is compensated for by an elongation of both carbon−carbon bonds in the ring involving the $C_{6'}$ atom, a shortening of the $C_{1'}$−N bond and a slight elongation of the N═N bond. Analogous changes in the bond lengths occur for substitution at the 2-position, but the presence of the amine group buffers the magnitude of these changes somewhat, particularly at the azo linkage, see Table 2. The energies of the highest two Kohn−Sham molecular orbitals in AAB are both significantly increased by substitution at the 6′-position.

The lone-pair orbital localized at the azo linkage also includes a contribution from the in-plane oxygen lone pairs and remains higher in energy than the orbital involving the amine nitrogen lone pair; the energy separation, 5.8 kcal/mol, is the second highest we observed for any of the monomethoxy derivatives. The structure of the LUMO is similar to that in AAB but contains a contribution from the out-of-plane oxygen lone-pair, similar to that observed for the 2- and 6-isomers. The energy separations between the HOMO and LUMO are 45.1 and 44.6 kcal/mol, respectively.

2.1.5. 4′-OMe-AAB. The positional isomer 4′-OMe-AAB is only 1.3 kcal/mol higher in energy than 3-OMe-AAB, see Table 1. As can be seen in Table 2, the $C_{4'}$−O bond length, 1.370 Å, is intermediate between that found for the 3-, 3′-, 5-, and 5′-isomers and that found for the 2-, 2′-, 6-, and 6′-isomers. The pattern of carbon−carbon bond lengths in the phenyl ring to which the methoxy group is attached is generally enhanced above that already found in AAB. The structures of the two highest occupied Kohn−Sham molecular orbitals are radically different from those observed for the other positional isomers. They are nearly degenerate (only separated by 0.5 kcal/mol) and appear as a mixture of the orbitals involving the azo and the amine lone pairs that are found for the other positional isomers. For comparison, we optimized several other AAB derivatives with substitution at the 4′-position. Similar combination orbitals are obtained for the HOMO and HOMO{-1} of 4′-OH-AAB, but for 4′-F-AAB the HOMO is clearly an azo lone-pair type orbital, whereas for 4′-SMe-AAB the HOMO is an amine lone-pair type orbital. The structure of the LUMO in 4′-OMe-AAB is similar to that observed for AAB but with a contribution from the out-of-plane lone-pair orbital on the oxygen atom; the energy gap between the HOMO and LUMO is 48.6 kcal/mol.

2.2. Remarks. The methoxy azo dyes 2-OMe-AAB, 4′-OMe-AAB, and 3-OMe-AAB are noncarcinogenic, moderately carcinogenic, and strongly carcinogenic, respectively.[5] The studies that established these results, however, have not

made a clear distinction between methoxy substitution at the 2- and 6-position or at the 3- and 5-position. Ames' *Salmonella* mutagenicity tests suggest that none of these molecules are mutagenic per se but require activation to their N-hydroxy derivatives prior to reaction with cellular macromolecules. Nevertheless, there appear to be some differences in the structures and electronic properties of the monomethoxy-AAB compounds themselves that may provide a basis for understanding their diverse carcinogenic behavior.

Many of the structural features in the monomethoxy AAB derivatives are determined to a large extent by electron delocalization at the azo linkage, which establishes a pattern of carbon−carbon bond lengths in the phenyl rings of AB; this pattern is enhanced by electron delocalization at the amine nitrogen atom in AAB. The presence of a methoxy group, with its two lone pairs of electrons, provides yet another site where delocalization is an issue, but it also introduces the possibility of lone-pair interactions involving the azo and amine nitrogen lone pairs.

Comparing the structures of 3(5)-OMe-AAB with those of 2(6)-OMe-AAB suggests that there is competition to delocalize lone-pair electron density at the amine nitrogen and methoxy oxygen atoms. For the 3(5)-isomers, where the methoxy oxygen is in close proximity to the amine nitrogen, it is energetically favorable to delocalize at the (less-electronegative) amine nitrogen atom by further increasing the double bond character of the C_4-N bond; for these isomers the $C_\phi-O$ bond is relatively long. For the 2(6)-isomers, where the methoxy oxygen is now in close proximity to the azo linkage, it becomes favorable to delocalize more at the methoxy oxygen atom by further increasing the double bond character of the $C_\phi-O$ bond; for these isomers the C_4-N bond is relatively long.

The Kohn−Sham HOMO and HOMO{-1} of AAB involve the azo and amine lone pairs respectively and these orbitals are relatively close in energy. For most of the monomethoxy AAB derivatives the two highest occupied orbitals are similar in shape to those found in AAB but involve contributions from one of the two lone pairs on the methoxy oxygen atom. In these n-OMe-AAB compounds, the energies of the two highest occupied orbitals are sensitive to the position (n) of the methoxy group because there is the potential for its lone pairs to be forced into close proximity with those on the AAB backbone.

It is interesting to note that the HOMO of the strongest carcinogen, 3(5)-OMe-AAB, involves the amine lone pair, whereas the HOMO of the noncarcinogen, 2(6)-OMe-AAB, involves the azo lone pairs. In the case of AAB itself, which is weakly carcinogenic, the HOMO involves the azo lone pairs, but the separation in energy between the two highest orbitals is smaller than that for 6-OMe-AAB and much smaller than that for 2-OMe-AAB. The carcinogenic potency of 4′-OMe-AAB is between that of 2(6)-OMe-AAB and 3(5)-OMe-AAB, and its HOMO is a mixed orbital that includes a contribution from the amine nitrogen lone pair.

The results of our investigation suggest that the carcinogenic activity of an OMe-AAB isomer is increased as the energy of the orbital involving the amine nitrogen lone pair is raised relative to that of the orbital involving the azo nitrogen lone pairs. (The energies of the LUMOs do not seem to correlate well with the carcinogenic potency of the AAB compounds, e.g. the LUMO of the noncarcinogen 2-OMe-AAB is 4.4 kcal/mol above the LUMO of the strong carcinogen 3-OMe-AAB but 6.5 kcal/mol above the LUMO of the weak carcinogen AAB.) This correlation can be further tested by noting that *N*-methyl-AAB and *N*,*N*-dimethyl-AAB compounds are usually more carcinogenic than the corresponding AAB compounds.[6] The HOMOs of both AAB and *N*-methyl-AAB are localized at the azo linkage. However, the separation in energy between the HOMO and HOMO-{-1} in *N*-methyl-AAB is only about 50% of the corresponding separation in AAB. On the other hand, the HOMOs of both 3-OMe-AAB and *N*-methyl-3-OMe-AAB involve the amine lone pair, but for these compounds the energy gap between the two highest occupied orbitals is three times greater in the *N*-methyl compound. Furthermore, the HOMO and HOMO{-1} of *N*,*N*-dimethyl-AAB are of the mixed type we observed in 4′-OMe-AAB, where both orbitals involve the amine nitrogen lone pair. These results are consistent with an increase in the carcinogenic potency of a methoxy AAB derivative when the primary amine is monomethylated or dimethylated. It must be pointed out that a variety of effects can influence the carcinogenic activity of a particular compound.[13−15] For example, the HOMO of *N*,*N*-dimethyl-4′-OH-AAB involves the amine nitrogen lone pair, and it is 2.7 kcal/mol higher in energy than the orbital involving the azo lone pairs. Based on our results for 2- and 3-OMe-AAB, this would suggest that *N*,*N*-dimethyl-4′-OH-AAB was a strong hepatocarcinogen in the rat, but this is not the case.[16] It is likely that the hydroxy group provides a site for the metabolic breakdown of this dye before it can act as a carcinogen.[16] In fact, studies have shown that *N*,*N*-dimethyl-4′-OH-AAB is formed from *N*,*N*-dimethyl-AAB during its metabolism by rat homogenates[17] and that demethylated hydroxyazo derivatives are present in the urine of rats fed the dye.[18] Additional calculations and further experimental carcinogenic/mutagenic studies on AAB derivatives will be required to establish the extent to which knowing the relative energies of the orbitals involving the azo and amine lone pairs in these compounds can be used as a predictive tool of the carcinogenic behavior of azo dyes. These studies are currently in progress.

3. SOFT COMPUTING APPROACH

Artificial neural networks (ANNs) are rapidly becoming the method of choice for QSAR/QPAR development. ANNs are model-free mapping devices capable of capturing complex nonlinear relationships in data that may be missed by conventional multilinear regression techniques. The particular approach we adopted in this investigation integrates fuzzy logic with ANNs. These two methodologies effectively complement one other: neural networks supply the computational power necessary to process rapidly large quantities of data, while fuzzy logic provides a high-level reasoning capability that guides the overall construction of the network. The algorithm we employed generates a feed-forward network architecture for a given data set, and, after generating fuzzy entropies at each node of the network, it switches to fuzzy decision making based on those entropies. Nodes and hidden layers are added as needed until the learning task is accomplished; in this study we restricted the architecture to a single hidden layer.

In the last several years there has been a large and energetic upswing in research efforts aimed at synthesizing fuzzy logic with computational neural networks in the emerging field of soft computing in AI. The enormous success of commercial applications (primarily by Japanese companies), which are dependent to a large extent on soft computing technologies, has led to a surge of interest in these techniques for possible applications throughout the U.S. textile industry.

The marriage of fuzzy logic with computational neural networks has a sound technical basis, because these two approaches generally attack the design of "intelligent" systems from quite different angles. Neural networks are essentially low level, computational algorithms that offer good performance in dealing with large quantities of data often required in pattern recognition and control. Fuzzy logic, introduced in 1965 by Zadeh,[19] is a means for representing, manipulating, and utilizing data and information that possess nonstatistical uncertainty. Thus, fuzzy methods often deal with issues such as reasoning on a higher (i.e., on a semantic or linguistic) level than do neural networks. Consequently, the two technologies often complement each other: neural networks supply the brute force necessary to accommodate and interpret large amounts of data and fuzzy logic provides a structural framework that utilizes and exploits these low level results.

This research is concerned with the integration of fuzzy logic and computational neural networks. Therefore, an algorithm for the creation and manipulation of fuzzy membership functions, which have previously been learned by a neural network from the data set under consideration, is designed and implemented. In the opposite direction we are able to use fuzzy tree architecture to construct neural networks and take advantage of the learning capability of neural networks to manipulate those membership functions for classification and recognition processes. In this research, membership functions are used to calculate fuzzy entropies for measuring uncertainty and information. That is, the amount of uncertainty regarding some situation represents the total amount of potential information in this situation. The reduction of uncertainty by a certain amount (due to new evidence) indicates the gain of an equal amount of information.

3.1. Fuzzy Entropy Measures. In general, a fuzzy entropy measure is a function f: $P(X) \rightarrow R$, where P(X) denotes the set of all fuzzy subsets of X. That is, the function f assigns a value f(A) to each fuzzy subset A of X that characterizes the degree of fuzziness of A. Thus, f is a set-to-point map, or in other words, a fuzzy set defined on fuzzy sets.[20]

DeLuca and Termini[21] first axiomatized nonprobabilistic entropy. Their axioms are intuitive and have been widely accepted in the fuzzy literature. We adopt them here. To qualify as a meaningful measure of fuzziness, f must satisfy the following axiomatic requirements:

Axiom 1. $f(A) = 0$ if and only if A is a crisp (nonfuzzy) set.

Axiom 2. f(A) assumes the maximum if and only if A is maximally fuzzy.

Axiom 3. If A is less fuzzy than B, then $f(A) \leq f(B)$.

Axiom 4. $f(A) = f(A^C)$.

Only the first axiom is unique; axioms two and three depend on the meaning given to the concept of the degree of fuzziness. For example, assume that the "less fuzzy" relation is defined, after DeLuca and Termini,[21] as follows

$$\mu_A(x) \leq {}_B(x) \text{ for } \mu_B(x) \leq {}^1/_2$$

$$\mu_A(x) \geq {}_B(x) \text{ for } \mu_B(x) \geq {}^1/_2$$

where μ is a membership function, $\mu_A(x)$, $\mu_B(x)$ are membership functions for sets A and B, respectively, and the term maximally fuzzy is defined by the membership grade 0.5 for all $x \in X$.

Motivated by the classical Shannon entropy function DeLuca and Termini proposed the following fuzzy entropy function:[21]

$$f(A) = -\sum_{x \in X}[(\mu_A(x)\log_2\mu_A(x) + (1 - (\mu_A(x))\log_2(1 - \mu_A(x))]$$

Its normalized version is given by $f(A)/|X|$, where $|X|$ denotes the cardinality of the universal set X. Similarly, taking into account the distance from set A to its complement A^C another measure of fuzziness, referred to as an index of fuzziness,[22] can be introduced. If the "less fuzzy" relation of Axiom 3 is defined by

$$\mu_C(x) = 0 \text{ if } \mu_A(x) \leq {}^1/_2$$

$$\mu_C(x) = 1 \text{ if } \mu_A(x) > {}^1/_2$$

where C is the crisp set nearest to the fuzzy set A, then the measure of fuzziness is expressed by the function[22]

$$f(A) = \sum_{x \in X}|\mu_A(x) - \mu_C(x)|$$

when the Hamming distance is used, and by the function[22]

$$f(A) = (\sum_{x \in X}(\mu_A(x) - \mu_C(x))^2)^{1/2}$$

when the Euclidean distance is employed.

It is clear that other metric distances may be used as well.[23] For example, the Minkowski class of distances yields a class of fuzzy measures

$$f_w(A) = (\sum_{x \in X}(\mu_A(x) - \mu_C(x))^w)^{1/w}$$

where $w \in [1, \infty)$.

However, both DeLuca and Termini measure and Kaufmann measure are only special cases of measures suggested by Knopfmacher[24] and Loo,[25] expressed in the form[23]

$$f(A) = h(\sum_{x \in X} g_x(\mu_A(x)))$$

where g_x ($\mu_A(x)$) are functions

$$g_x: [0,1] \rightarrow R^+$$

which are all monotonically increasing in [0, 0.5], monotonically decreasing in [0.5, 1], and satisfy the requirements that $g_x(0) = g_x(1) = 0$, and that $g_x(0.5)$ is the unique maximum of g_x, and h is a monotonically increasing function. It has

been shown that the degree of fuzziness of a fuzzy set can be expressed in terms of the lack of distinction between the set and its complement.[26−28]

It has been also established that a general class of measures of fuzziness based on this lack of distinction is exactly the same as the class of measures of fuzziness expressed in terms of a metric distance based on some form of aggregating the individual differences:[23]

$$f_C(A) = |X| - \sum_{x \in X} [\mu_A(x) - c(\mu_A(x))]$$

To obtain the normalized version of fuzzy entropy the above expression is divided by the cardinality of a fuzzy set. The previous definitions can also be extended to infinite sets.[23]

Another fuzzy entropy measure was proposed and investigated by Kosko.[20,29] He established that

$$f(A) = (\text{sigma count } (A \wedge A^C))/(\text{sigma count } (A \vee A^C))$$

where sigma count is a fuzzy cardinality.[30,31]

Kosko[29] claims that his entropy measure and corresponding fuzzy entropy theorem does not hold when we substitute Zadeh's operations[19] with any other generalized fuzzy operations. If any of the generalized Dombi's operations[32] are used, the resulting measure is an entropy measure, it is maximized; however, it does not equal unity at the midpoints.[33]

The generalized Dombi's operations proved to do well in different applications and were used by Sztandera[34] for detecting coronary artery disease and were suggested for image analysis by Sztandera.[35] However, we still have to use Zadeh's complement,[19] since the Kosko's theorem does not hold for any other class of fuzzy complements.

3.2. A New Concept of Fuzzy Entropy. In our experiments we used fuzzy entropy suggested by Kosko[29] and generalized fuzzy operations introduced by Dombi.[32]

Generalized Dombi's operations form one of the several classes of functions, which possess appropriate axiomatic properties of fuzzy unions and intersections. The operations are defined below. From our experience the parameter $\lambda = 4$ gives the best results.[33]

3.2.1. Dombi's Fuzzy Union

$$\mu_{A \vee B}(x) = \{1 + [(1/\mu_A(x) - 1)^{-\lambda} + [(1/\mu_B(x) - 1)^{-}]^{-1/\lambda}]\}^{-1}$$

where λ is a parameter by which different unions are distinguished, and $\lambda \in (0,\infty)$.

3.2.2. Dombi's Fuzzy Intersection

$$\mu_{A \wedge B}(x) = \{1 + [(1/\mu_A(x) - 1)^{\lambda} + [(1/\mu_B(x) - 1)]^{1/\lambda}]\}^{-1}$$

where λ is a parameter by which different intersections are distinguished, and $\lambda \in (0,\infty)$.

It is interesting to examine the properties of these operations. By definition, generalized fuzzy union and intersection operations are commutative, associative, and monotonic. It can be shown that they neither satisfy the law of the excluded middle nor the law of contradiction. They are also not idempotent, nor distributive. However, they are continuous and satisfy the de Morgan's laws (when the standard Zadeh's complement is used).[32] Zadeh's complement ($c(a) = 1 - a$) is by definition monotonic nonincreasing. It is also continuous and involutive. Other properties and the proofs can be found in Dombi's[32] and Zadeh's[19] papers.

3.3. Feed-Forward Neural Network Architecture. The proposed algorithm generates feed forward network architecture for a given data set, and after having generated fuzzy entropies at each node of the network, it switches to fuzzy decision making on those entropies. The nodes and hidden layers are added until a learning task is accomplished. The algorithm operates on numerical data and equates a decision tree with a hidden layer of a neural network.[33] A learning strategy used in this approach is based on achieving the optimal goodness function. This process of optimization of the goodness function translates into adding new nodes to the network until the desired values are achieved. When this is the case, then all training examples are regarded as correctly recognized. The incorporation of fuzzy entropies into the algorithm seems to result in a drastic reduction of the number of nodes in the network and in decrease of the convergence time. Connections between the nodes have a "cost" function being equal to the weights of a neural network. The directional vector of a hyperplane, which divides decision regions, is taken as the weight vector of a node.

The outline of the algorithm follows:

Step (i) For a given problem with N samples, choose a random initial weight vector.

Step (ii) Make use of learning rule

$$\Delta w_{ij} = -\rho \partial f(F)/\partial w_{ij}$$

where w_{ij} is a weight vector between i and j nodes, ρ is a learning rate, and f(F) is a fuzzy entropy function; and search for a hyperplane that minimizes the fuzzy entropy function

$$\min f(F) = \sum_r N_r/N \text{ entropy}(L, r)$$

where L is a level of a decision tree, R is total number of nodes in a layer, r is number of nodes, and f (F) is fuzzy entropy.

Step (iii) If the minimized fuzzy entropy is not zero, but it is smaller than the previous value compute a new node in the current layer and repeat the previous step. Otherwise go to the next step.

Step (iv) If there is more than one node in a layer compute a new layer with inputs from all previous nodes including the input data, then go to step (ii). Otherwise terminate.

4. AZO DYE DATABASE

We have conclusively demonstrated that density functional techniques can efficiently be used to investigate the structure and properties (charge distribution, band gap, logP, etc.) of a wide range of azo dyes. (Most prior calculations on dyes have used lower level semiempirical methods). We employed the gradient-corrected density functional (Becke-Perdew) method incorporated into the Spartan 5.0 molecular modeling package[10] using the polarized numerical DN** basis set (BP/DN**//BP/DN** level), which provides an exceptionally good description of the bonding in most organic molecules.

(This computational level can also be used with dyes that contain metals such as Cr, Co, Cu, etc.). The calculated structural and physicochemical properties of these dyes, augmented with experimental results (optical properties, toxicological activity, etc.), were incorporated into a database that was used to train the neural network.

Preliminary results from several trials suggest that, given a collection of dye molecules, each described by a set of structural features, a set of physical properties, and the strength of some activity under consideration, a neural network algorithm could be used to find patterns in terms of the structural features and properties that correspond to a desired level of activity.

It is well-known that ANNs with a single hidden layer, containing a sufficient number of neurons, can interpolate any multidimensional nonlinear function to a given accuracy and can implement exactly an arbitrary finite training set. This remarkable potential of ANNs to pattern nonlinear phenomena, however, can easily result in relatively poor predictive power of a trained ANN toward new data. To evaluate the generalization ability of our various networks, an n-fold cross-validation procedure was employed. To determine the effectiveness of the proposed algorithm, the performance was evaluated on a database of molecular properties involving 22 selected azo dyes (11 carcinogenic/mutagenic and 11 noncarcinogens). We used 80% of the database (18 molecules) for training purposes and 20% (four molecules) for testing. We repeated the process five times (20% Jacknife procedure).

We selected a pool of about 50 descriptors to construct our neural networks. Although an exhaustive search was not practical, a large variety of networks with different numbers and types of descriptors were generated and evaluated. After several trial-and-error approaches with different input sets, we opted for 11 input parameters (logP, number of benzene rings, XY shadow/XY rectangle, 1x gamma polarizability, Balaban index, YZ shadow/YZ rectangle, 1/2x beta polarizability, total dipole of the molecule, internal entropy/number of atoms, molecular volume/XYZ box, total molecular surface area). Using those parameters, in conjunction with experimental toxicological data, the network was able to learn and differentiate between mutagenic/carcinogenic and nonmutagenic/noncarcinogenic dyes. We expect the neural network to predict the mutagenic/carcinogenic nature of other chemical structures, provided very large databases exist.

We are currently looking into using so-called topological indices (modified Wiener's index, modified Balaban's index, modified Schultz's index, etc.)[36,37] that have been used successfully in the pharmaceutical industry in QSPR and QSAR studies. We plan to use one of these topological indices or develop one ourselves if none of these are adequate, in conjunction with logP and selected electronic properties from our density functional calculations as descriptors in our soft computing approach.

5. CONCLUDING REMARKS

Molecular modeling has allowed us to investigate the properties of a large number of azobenzene derivatives in a short period of time. It is clear that there are correlations between our calculated properties and their toxicological behavior. We are also certain that such correlations exist between molecular properties and various textile parameters such as light fastness, for example. Often these correlations are not evident until calculations on a sufficiently large number of related structures have been performed and the data carefully analyzed. Then appropriate molecular descriptors can more readily be identified and used as input into a neural network.

The special interest in ANNs for QSAR/QPAR development arises from their ability to cope with nonlinear relationships in the absence of an underlying model. Indeed, we have shown that an 11-descriptor ANN, developed with the described approach, can account for the variation in the mutagenic activity of the compounds in this data set. Furthermore, the predictive power of this network, as assessed by cross-validation, is exceptionally good, $(R_{CV})^2 = 0.98$.

From the soft computing point of view, the proposed approach shows a way in which neural network technology can be used as a "tool" within the framework of a fuzzy set theory. Generating membership functions with the aid of a neural network has been shown to be an extremely powerful and promising technology. In this research, membership functions are used to calculate fuzzy entropies for measuring uncertainty and information. The proposed neural network is a building block toward combining the two soft computing paradigms. It allows for a self-generation of a neural network architecture suited for a particular problem.

The main features and advantages of the proposed approach are as follows: (1) it is a general method of how to use numerical information, via neural networks, to provide good approximations to the membership functions; (2) it is a simple and straightforward quick-pass build-up procedure, where no time-consuming iterative training is required, resulting in much shorter design time than most neural networks; (3) there is a lot of freedom in choosing the membership functions and corresponding fuzzy entropies; this provides flexibility for designing systems satisfying different requirements; and (4) it performs successfully on data where neither a pure neural network nor a fuzzy system would work perfectly.

ACKNOWLEDGMENT

The authors would like to acknowledge the U.S. Department of Commerce, National Textile Center (Grant #I98-P01) for financial support of this research.

REFERENCES AND NOTES

(1) Kojima, M.; Degawa, M.; Hashimoto, Y.; Tada, M. *Biochem. Biophy. Res. Commun.* **1991**, *179*, 817.
(2) Hashimoto, Y.; Degawa, M.; Watanabe, H. K.; Tada, M. *Gann* **1981**, *72*, 937.
(3) Degawa, M.; Miyairi, S.; Hashimoto, Y. *Gann* **1978**, *69*, 367.
(4) Degawa, M.; Shoji, Y.; Masuko, K.; Hashimoto, Y. *Cancer Lett.* **1979**, *8*, 71.
(5) Miller, J. A.; Miller, E. C. *Cancer Res.* **1961**, *21*, 1068.
(6) Hashimoto, Y.; Watanabe, H. K.; Degawa, M. *Gann* **1981**, *72*, 921.
(7) Freeman, H. S.; Posey, J. C., Jr.; Singh, P. *Dyes Pigm.* **1992**, *20*, 279.
(8) Degawa, M.; Kojima, M.; Hashimoto, Y. *Mutation Res.* **1985**, *152*, 125.
(9) Lye, J.; Hink, D.; Freeman, H. S. In Computational Chemistry applied to synthetic dyes. *Computational chemistry and chemical engineering*; Cisneros, G., Cogordan, J. A., Castro, M., Wang, C., Eds.; Singapore World Scientific Publ: 1997.
(10) Spartan v.5.0, Wavefunction Inc., 18401 Von Karmen Avenue, Suite 370, Irvine, CA 92612.

(11) Perdew, J. P. *Phys. Rev.* **1986**, *B33*, 8822.
(12) Perdew, J. P. *Phys. Rev.* **1987**, *B34*, 7046.
(13) Chung, K. T.; Cerniglia, C. E. *Mutation Res* **1992**, *277*, 201.
(14) Ashby, J.; Paton, D.; Lefevre, P. A.; Styles, J. A.; Rose, F. L. *Carcinogenesis* **1982**, *3*, 1277.
(15) Cunningham, A. R.; Klopman, G.; Rosenkranz, H. S. *Mutation Res.* **1998**, *405*, 9.
(16) Miller, J. A.; Sapp, R. W.; Miller, E. C. *Cancer Res.* **1949**, *9*, 652.
(17) Mueller, G. C.; Miller, J. A. *J. Biol. Chem.* **1948**, *176*, 535.
(18) Miller, J. A.; Miller, E. C. *Cancer Res.* **1947**, *7*, 39.
(19) Zadeh, L. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338−353.
(20) Kosko, B. Fuzzy Entropy and Conditioning. *Inf. Sci.* **1986**, *40*, 165−174.
(21) DeLuca, A.; Termini, S. A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. *Inf. Control* **1972**, *20*, 301−312.
(22) Kaufmann, A. *Introduction to the Theory of Fuzzy Subsets*; Academic Press: New York, 1975.
(23) Klir, G. J.; Folger, T. A. *Fuzzy Sets, Uncertainty and Information*; Prentice Hall: Englewood Cliffs, 1988.
(24) Knopfmacher, J. On Measures of Fuzziness. *J. Math. Anal. Appl.* **1975**, *49*, 529−534.
(25) Loo, S. G. Measures of Fuzziness. *Cybernetica* **1977**, *20*, 201−210.
(26) Yager, R. R. On the Measure of Fuzziness and Negation. Part I: Membership in the Unit Interval. *Intl. J. General Systems* **1979**, *5*, 221−229.
(27) Yager, R. R. On the Measure of Fuzziness and Negation. Part II: Lattices. *Inf. Control* **1980**, *44*, 236−260.
(28) Higashi, M.; Klir, G. J. On Measures of Fuzziness and Fuzzy Complements. *Intl. J. General Systems* **1982**, *8*, 169−180.
(29) Kosko, B. *Neural Networks and Fuzzy Systems*; Prentice Hall: Englewood Cliffs, 1992.
(30) Zadeh, L. A Computational Approach to Fuzzy Quantifiers in Natural Languages. *Comput. Math. Appl.* **1983**, *9*, 149−184.
(31) Zadeh, L. The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems. *Fuzzy Sets Systems* **1983**, *11*, 199−227.
(32) Dombi, J. A General Class of Fuzzy Operators, the De Morgan Class of Fuzzy Operators and Fuzziness Measures. *Fuzzy Sets Systems* **1982**, *8*, 149−163.
(33) Cios, K. J.; Sztandera, L. M. Continuous ID3 Algorithm with Fuzzy Entropy Measures. In *Proceedings of the 1st International Conference on Fuzzy Systems and Neural Networks*; IEEE Press: San Diego, 1992; pp 469−476.
(34) Cios, K. J.; Goodenday, L. S.; Sztandera L. M. Hybrid Intelligence Systems for Diagnosing Coronary Stenosis. *IEEE Eng. Med. Biol.* **1994**, *13*, 723−729.
(35) Sztandera, L. M. Relative Position Among Fuzzy Subsets of an Image, M.S. Thesis, Computer Science and Engineering Department, University of Missouri-Columbia, Columbia, MO, 1990.
(36) Vedrina, M.; Markovic, S.; Medic-Saric, M.; Trinajstic, N. *Comput. Chem*. **1997**, *21*, 355−361.
(37) Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399−404.

CI025600R