A QSAR for the Mutagenic Potencies of Twelve 2-Amino-trimethylimidazopyridine Isomers: Structural, Quantum Chemical, and Hydropathic Factors

M.G. Knize,[1]* F.T. Hatch,[2] M.J. Tanga,[3] E.Y. Lau,[1] and M.E. Colvin[4]

[1]*Biosciences Directorate, Lawrence Livermore National Laboratory, University of California, Livermore, California*
[2]*Meredith, New Hampshire*
[3]*Biosciences Division, SRI International, Menlo Park, California*
[4]*School of Natural Sciences, University of California, Merced, California*

An isomeric series of heterocyclic amines related to one found in heated muscle meats was investigated for properties that predict their measured mutagenic potency. Eleven of the 12 possible 2-amino-trimethylimidazopyridine (TMIP) isomers were tested for mutagenic potency in the Ames/Salmonella test with bacterial strain TA98, and resulted in a 600-fold range in potency. Structural, quantum chemical, and hydropathic data were calculated on the parent molecules and the corresponding nitrenium ions of all of the tested isomers to establish models for predicting the potency of the unknown isomer. The principal determinants of higher mutagenic potency in these amines are: (1) a small dipole moment, (2) the combination of *b*-face ring fusion and N3-methyl group, (3) a lower calculated energy of the π electron system, (4) a smaller energy gap between the amine HOMO and LUMO orbitals (Pearson ''softness''), and (5) a more stable nitrenium ion. Based on predicted potency from the average of six regression models, the isomer not yet synthesized and tested is expected to have a mutagenic potency of 0.77 revertants/μg in tester strain TA98, which is near the low end of the potency range of the isomers. Environ. Mol. Mutagen. 47:132–146, 2006.
© 2005 Wiley-Liss, Inc.

Key words: QSAR; TMIP; heterocyclic amine; cooking mutagen

INTRODUCTION

Aromatic and heterocyclic aromatic amines have been considered to be actual or potential mutagens and carcinogens for over a century [Rehn, 1895; Case et al., 1954]. For the past quarter-century, a variety of heterocyclic aromatic amines have been found in cooked foods, primarily well-done or over-cooked meats, and have been shown to be extensively genotoxic in many test systems and carcinogenic in rodents and monkeys [Adamson et al., 1995; Sugimura, 1995; Felton et al., 1995, 1999; Weisburger, 2002]. These amines are genotoxic only after activation by a series of enzymatic or biochemical reactions that convert the parent compound into an electrophilic derivative. Cytochrome P450 1A1 and 1A2 are catalysts for the first activation step [Guengrich, 1995; Guengrich et al., 1995; Turesky et al., 2002; Kim and Guengrich, 2004]. A quantum chemical computational study of the mechanism of this step, which is an oxidation involving removal of electrons from the parent amine has been carried out by Sasaki et al. [2002]. The first stable intermediate in the activation sequence is a hydroxylamine. This product is then esterified to an acetoxy or sulfate ester. Departure of the ester group is postulated to leave a nitrenium ion [Novak et al., 1998; Novak and Rajagopal, 2001], which is a reactive electrophile that covalently binds and damages DNA, primarily at guanine bases by forming a bulky adduct [Schut and Snyderwine, 1999]. A purely chemical production of putative nitrenium ions has been reported by Sabbioni and Wild [1992].

Since 1978, our laboratory has engaged in the identification, isolation, synthesis, and assessment of the genetic toxicology of the thermic food mutagens. We have published a series of quantitative structure-activity relationships (QSARs) directed at gaining insight into the chemical

mechanisms responsible for their genotoxic effects [Hatch et al., 1991; Hatch et al., 1992; Hatch et al., 1996; Hatch and Colvin, 1997; Colvin et al., 1998; Hatch et al., 2001]. This work has broadly covered mutagenic aromatic and heterocyclic amines whose structures bear a resemblance to the thermic mutagens.

An unknown mutagenic compound thought to be an amino-trimethylimidazopyridine (TMIP) was isolated from well-done beef [Felton, 1984]. Many years later, the one isomer found in a meat is the 2-amino-1,5,6-trimethylimidazo[4,5-b]pyridine (156bmip) in heated chicken breast, and in the meat drippings from beef, chicken thigh, and pork [Pais et al., 1999]. A model system based on the composition of meat formed the same 156bmip isomer [Borgen et al., 2001].

Interest in determining which of the 12 possible isomers of TMIP would be mutagenic or the most potent mutagen led to the present work to determine properties that could predict mutagenic potency. We have obtained 11 of the 12 possible isomers of TMIP (Table I) by chemical synthesis. The 12th isomer has eluded all synthesis efforts to date, giving only trace amounts for the several synthesis routes attempted. The isomers differ structurally only in the positions of methyl group substitution on the imidazole and pyridine rings and in the fusion face between the imidazole and pyridine rings. These molecules have structural correlations with an important cooking mutagen/carcinogen PhIP (2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine) and other studied congeners. These features seemed ideal for a QSAR study, although the mutagenic potency range (3 orders of magnitude) and sample size are rather limited for statistical analyses.

Three of the isomers are of moderate potency when compared with many other heterocyclic amines, and eight others are weak to very weak mutagens; the one not yet synthesized is of unknown potency. We have used structural, quantum chemical, and hydropathic data calculated on the parent molecules and their nitrenium ions (which are hypothesized to be the penultimate or ultimate mutagens after metabolic activation) of all the isomers to establish models for predicting the potency of the unknown. Limited data were also calculated for the imine tautomers and the hydroxylamines.

In an earlier article on the status of our food mutagen studies, we reported briefly [Felton et al., 1999] on the mutagenic potencies and preliminary semi-empirical and ab initio quantum chemical calculations on the TMIP isomers that are now being more fully reported. At that time, we showed that the strongest correlation with mutagenic potency came from the dipole moment, although this relationship was not significant in earlier studies on more diverse amines. Another difference from earlier work was that the lowest unoccupied molecular orbital (E_{lumo}) of the parent TMIPs showed only a borderline correlation with potency, whereas in other amine QSARs, this rela-

tionship is stronger. Finally, the earlier study found an extreme correlation ($r = 0.98$) between the highest occupied molecular orbital (E_{homo}) of the parent TMIPs and the energy to form the nitrenium ion, which reflects the oxidative loss of two electrons from the HOMO orbital of the parent amines. This is confirmed in the present study but not elaborated upon, since E_{homo} is not significantly correlated with mutagenic potency and was therefore not included in the predictor variables.

Vracko et al. [2004] have reported a QSAR study of the TMIP isomers. Mutagenic potency data in Salmonella strains TA98 and YG1024 were taken from Felton et al. [1999] for 10 isomers (the 11th was synthesized and assayed more recently). In their study, several machine learning techniques were applied to identify strong predictors of potency and the results pooled as a training set to predict the potency of the remaining two isomers. Specifically, the following four approaches were reported: (1) multiple linear regression was applied to series of orbital energies below HOMO and above LUMO and the best set of orbitals determined from the regression statistics. For unstated reasons the main HOMO and LUMO energies were not included; (2) counter Propagation Neural Network software was also applied to the orbital energy data; (3) a geometric "spectrum-like representation" of the 3D molecular structures of the isomers that was developed in the authors' laboratory was applied to produce the 3D position and polar angle for each atom leading to a Lorentzian curve, which could then be analyzed as a spectrum; (4) the same geometric method in method 3 was supplemented with Mulliken atomic charges for each atom.

The results of methods 2, 3, and 4 were subjected to cluster analysis producing Kohonen maps, which showed good relationships to mutagenic potency for methods 2 and 4, but not for method 3. Finally, all four methods were pooled to predict the potency of the two untested isomers. The prediction for 167bmip was in good agreement with our recently measured value. The prediction for the still untested isomer 347cmip was about one $\log_{10}$ higher than the prediction reached in this article.

In our article, we calculate a complete set of structural, quantum chemical, and hydropathic data on parent amines and nitrinium ions to establish models that correlate properties with potency, and use these to predict the potency of the one remaining not-yet synthesized isomer of the set of mutagenic TMIP.

MATERIALS AND METHODS

Abbreviations and definitions are presented in Table II.

Chemicals

Eleven of the 12 possible TMIP isomers were synthesized in the laboratory of co-author Dr. Mary Tanga at SRI International (Menlo Park,

TABLE I. Structure, CAS Number, Name, Synthesis Reference, Ring Fusion, N-Methyl Position, and Mutagenic Potency of TMIP Isomers
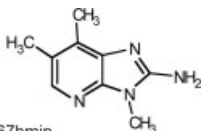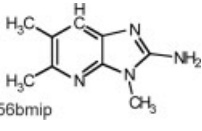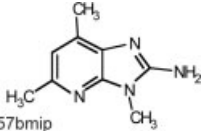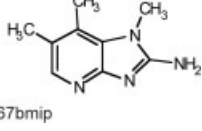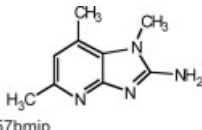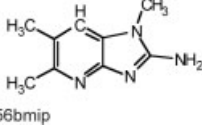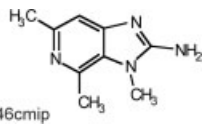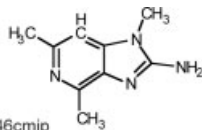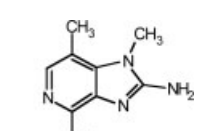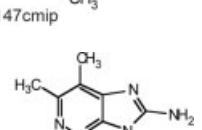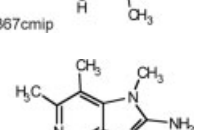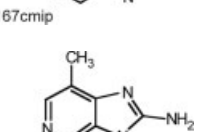
Structure abbreviation	CAS NO.	Name	Synthesis	b_c face	Nme	LogMP98	98 rev/μg	LogMP1024
367bmip	401560-72-3	2-Amino-3,6,7-trimethylimidazo[4,5-b]-pyridine	See Methods	b	3	1.7	317.0	2.0
356bmip	57667-51-3	2-Amino-3,5,6-trimethylimidazo[4,5-b]-pyridine	[Tanga et al., 1994]	b	3	1.6	233.0	2.2
357bmip	132898-06-7	2-Amino-3,5,7-trimethylimidazo[4,5-b]-pyridine	[Tanga et al., 1997]	b	3	0.7	28.0	1.3
167bmip		2-Amino-1,6,7-trimethylimidazo[4,5-b]-pyridine	[Ahiberg et al., 2000]	b	1	0.2	10.0	—
157bmip	401560-75-6	2-Amino-1,5,7-trimethylimidazo[4,5-b]-pyridine	[Tanga et al., 2003]	b	1	0.0	5.7	0.3
156bmip	161091-56-0	2-Amino-1,5,6-trimethylimidazo[4,5-b]-pyridine	[Tanga et al., 1994]	b	1	−0.3	3.1	0.4
346cmip	193690-74-3	2-Amino-3,4,6-trimethylimidazo[4,5-c]-pyridine	[Tanga et al., 1997]	c	3	−0.4	2.3	0.2
146cmip	401560-75-4	2-Amino-1,4,6-trimethylimidazo[4,5-b]-pyridine	[Tanga et al., 2003]	c	1	−0.5	1.6	0.4
147cmip	193690-65-2	2-Amino-1,4,7-trimethylimidazo[4,5-c]-pyridine	[Tanga et al., 1997]	c	1	−0.8	0.9	0.0
367cmip	401560-74-5	2-Amino-3,6,7-trimethylimidazo[4,5-c]-pyridine	See Methods	c	3	−1.0550	0.5	0.0911
167cmip	193690-71-0	2-Amino-1,6,7-trimethylimidazo[4,5-c]-pyridine	[Tanga et al., 1997]	c	1	−1.0550	0.5	0.0721
347cmip		2-Amino-3,4,7-trimethylimidazo[4,5-c]-pyridine	Not made	38c	3	—	—	—

TABLE II. Abbreviations and Definitions of Variables Used in the Tables and Text

Abbreviation	Meaning
Notation	
AHA80	Aromatic and heterocyclic amines studied previously [Hatch et al., 1992–2001]
nnn	Positions of first methyl group on the imidazole ring followed by the two methyl groups on the pyridine ring, shown in the abbreviation for each molecule
Suffixes	NH2, parent amine; NHOH, hydroxylamine derivative; NH^+, nitrenium ion; tau, imine tautomer of parent
Main dependent variable	
$LogMP98$	Log_{10} of revertants per nanomole in Ames/Salmonella strain TA98 or TA1538
LogYG1024	Log_{10} of revertants per nanomole in O-acetylase over-producing strain YG1024
Predictor variables:	
Structural factors	
Nme	Methyl group on imidazole ring N atom at position 1 or 3; in Table III, pos 3=1, pos 1=0
Face	Position of fusing imidazole and pyridine rings: pyridine face 2-3=b; face 3-4=c, shown in the name of each molecule; in Table III, face b=1, c=0
Quantum chemical factors	
RHF-	Restricted Hartree-Fock calculation of total energy in a.u. (Hartrees)
Ford-	Simulation of reaction energy calculation (Eproducts-Ereactants) vs aniline [Ford and Griffin, 1992] in kcal/mol
$\Delta E\_$	Difference in RHF energies between two suffixed derivatives
DipoleNH2	Dipole moment of parent amine from Hartree-Fock (ab initio) SpartanES calculation in Debyes
Dipoletau	Dipole moment of imine tautomer
ΔDipole	Difference: DipoleNH2–Dipoletau
EhomoNH2	Energy of Highest Occupied Molecular Orbital in electron volts
ElumoNH2	Energy of Lowest Unoccupied Molecular Orbital in electron volts
SoftNH2	Reciprocal of Pearson calculation (ElumoNH2-EhomoNH2)/2 of Hardness in reciprocal electron volts [Pearson, 1963; Pearson and Songstad, 1967; Jensen, 1999]
Ep_β	Total energy of π electrons in parent molecule calculated by simple Hückel theory with HMO software in β units; α units are constant and are omitted.
NPA	Natural Population Analysis charge on an atom, calculated by Spartan [Reed at al., 1985]
qNme_NH2	NPA charge on imidazole N atom containing the methyl substituent
Hydropathic factors	
LogP_GC	Ghose-Crippen calculation of LogP from constituent atoms and atomic connectivity [Ghose et al., 1988] by Spartan software
LogP_HL	LogP predicted by MMP software from molecular fragments [Hansch and Leo, 1979]

CA). All were greater than 98% purity as judged by HPLC using two different chromatographic systems. Their structures and references for synthesis are shown in Table I.

Synthetic routes for two of the isomers were not published. Eight steps were used to make the 367cmip isomer from 2,3-lutidine via nitration at the 4-position, reduction, bromination at the 5-position, dehalogenation with methylamine, and cyclization with cyanogen bromide (^{1}H NMR (CD$_3$O-d_4): δ2.40 (s, 3H), 2.53 (s, 3H), 3.61 (s, 3H), 8.15 (s, 1H)). The 367bmip isomer was made from 3,4-lutidine N-oxide reacted with N-methylchlorophenylmethanimine (made from N-methylbenzamide) to form 2-methylamino-4,5-dimethylpyridine, which was nitrated at the 2-position, reduced, and cyclized with cyanogen bromide (^{1}H NMR (DMSO-d_6): δ2.20 (s, 3H), 2.28 (s, 3H), 3.45 (s, 3H), 6.62 (br s, 2H), 7.64 (s, 1H)).

Salmonella Mutagenicity Assay

The mutagenic activity of the samples was determined using the standard plate incorporation assay described by Ames et al. [Ames et al., 1975; Maron and Ames, 1983], with *Salmonella typhimurium* strain TA98 (a gift of Professor Bruce Ames, University of California, Berkeley). Samples were also assayed with strain YG1024 [Watanabe et al., 1990; Einisto et al., 1991], an O-acetyltransferase overproducing derivative of TA98 (a gift of Dr. T. Nohmi, National Institute of Hygienic Sciences, Tokyo).

Two milligrams of Aroclor-induced rat liver S9 protein was added per plate for metabolic activation, and the compounds were tested in doses to give a linear response covering a limited range from 1 to 100 μg/plate. A positive control, 2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ), gave 1200–1500 revertants/5 ng dose. Dimethylsulfoxide was included in the negative controls (spontaneous revertant counts) and gave TA98 values of 25–60 revertant colonies per plate. Strain YG1024 gave higher spontaneous reversion rates of 100–150 colonies per plate.

A minimum of four dose points from duplicate platings was used, and the linear portion of the curve was used to calculate the number of revertant colonies per microgram [Moore and Felton, 1983]. The standard error of the linear fits for all samples was less than 15%.

It is important to point out that the Ames test was carried out with *S. typhimurium* strain TA98 bacteria. This strain and a closely related strain TA1538 are commonly used for studying aromatic and heterocyclic amines because they are especially sensitive to frame-shift mutations, which are caused by large planar molecules. The basis for the sensitivity of the assay are the mutations engineered into the strains by Dr. Ames as follows: *hisD*3052 mutation results in a strain that cannot grow without added histidine unless reverted by a frame-shift mutation [Fuscoe et al., 1988; Felton et al., 1995]; *uvrB* mutation results in a deficiency in DNA repair; and *rfa* mutation causes a partial loss of the lipopolysaccharide surface that increases permeability to large molecules [Ames et al., 1973].

TABLE III. TMIP Variables Database

Name	LogMP98	Log1024	Face	Nme	DipoleNH2	Dipoletau	ΔDipole	EhomoNH2	ElumoNH2	SoftNH2	LogP_GC	Ep_β	qNmeNH2
367bmip	1.7471	1.9503	1	1	3.030	2.380	0.650	−0.2742	0.1407	4.8208	2.159	17.667	−0.530
356bmip	1.6134	2.1845	1	1	2.990	2.474	0.516	−0.2681	0.1419	4.8784	1.873	17.665	−0.529
357bmip	0.6932	1.2588	1	1	2.960	2.444	0.516	−0.2688	0.1436	4.8490	1.873	17.695	−0.527
167bmip	0.2460		1	0	7.070	2.555	4.515	−0.2740	0.1361	4.8774	2.159	17.662	−0.522
157bmip	0.0019	0.3142	1	0	6.020	2.370	3.650	−0.2828	0.1359	4.7776	1.873	17.687	−0.517
158bmip	−0.2626	0.4391	1	0	6.390	2.348	4.042	−0.2689	0.1426	4.8613	1.873	17.661	−0.518
346cmip	−0.3923	0.2098	0	1	4.850	1.046	3.804	−0.2949	0.1409	4.5886	0.871	17.774	−0.517
146cmip	−0.5499	0.3763	0	0	6.450	0.450	6.000	−0.2717	0.1492	4.7516	0.871	17.781	−0.520
147cmip	−0.7998	0.0242	0	0	6.780	0.130	6.650	−0.2738	0.1530	4.6861	1.157	17.769	−0.520
367cmip	−1.0550	0.0911	0	1	5.670	0.390	5.280	−0.2924	0.1475	4.5462	1.157	17.755	−0.517
167cmip	−1.0550	0.0721	0	0	6.870	0.304	6.566	−0.2746	0.1451	4.7650	1.157	17.755	−0.522
347cmip			0	1	5.830	0.708	5.122	−0.2932	0.1522	4.4905	1.157	17.768	−0.515
avgtop3					2.993	2.433	0.561	−0.2704	0.1421	4.8494	1.968	17.676	−0.5287
avgnext8					6.263	1.199	5.063	−0.2791	0.1438	4.7317	1.390	17.731	−0.5191
T-testprob					0.0000	0.0129	0.0000	0.0548	0.4768	0.0294	0.0183	0.0260	0.0003
Correl. LogMP98	1.000	0.971	0.770	0.493	−0.799	0.785	−0.913	0.425	−0.484	0.637	0.741	−0.735	−0.836

The bacteria are added to a buffered suspension of microsomes (the 9000g supernatant of homogenized rat or hamster liver cells), which supply CYP 450 and other enzymes. Microsomes are tiny membranous vesicles derived from the endoplasmic reticulum of the mammalian cells. When the bacteria are incubated in this suspension with certain supplements, reactive molecules produced in the activation process diffuse or are transported through the bacterial membrane into the interior, where they may react with bacterial DNA, or with other enzymes or biochemicals that may divert them from genotoxic effects. Despite the *rfa* mutation, the hydrophobicity of the molecules could still affect their rate of transport. In a QSAR for mammalian systems, the activated molecules encounter several membrane layers on the way from the site of mutagen exposure to the DNA within the nuclear membrane. In the mammalian case, the consideration of hydrophobic properties and possible detoxification reactions becomes more complicated.

Computational System and Software

The computer used was an HP Vectra VL420 1.9 GHz Pentium 4 (Hewlett-Packard, Mountain View, CA) running Windows XP Pro (Microsoft, Redmond, WA). Database management was with Excel 97 (Microsoft) supplemented with Accord 3 for Excel (Accelrys, formerly Synopsys, San Diego, CA) and with [Excel 2002]. Structures for Table I were drawn with Isis/Draw (MDL, San Leandro, CA) and incorporated into the table with Accord. Graphics and statistical analysis were done with Excel 2002 and Small Stata 8 (Stata, College Station, TX); 3-D graphics were generated with PSI-Plot, ver. 7.8 (Poly Software Intl., Pearl River, NY). Forward stepwise multiple linear regression models were calculated with Small Stata ver. 8. Probability for a variable to enter a model was set at 0.20 and to be removed at 0.21. Ab initio quantum chemical calculations, made with SpartanES 04 (Wavefunction, Irvine, CA), begin with optimization of geometry using in sequence MMFF, AM1, and Hartree-Fock (basis set 6-31G**) levels of theory. Following optimization, the Restricted Hartree-Fock (RHF) energies, molecular orbital energies, dipole moment, LogP_Ghose-Crippen, and natural population atomic charges (NPA) on several atoms were calculated. Simple Hückel calculation of total π electron energy was performed with the HMO program ver. 3.0a (Trinity Software, Plymouth, NH). This program had an extra line added to the parameter file including parameters for a positively charged N-atom and auxiliary induction and methyl group correction factors were added to the HMO parameter file. LogP_HL was calculated on minimized structures with Molecular Modeling Pro ver. 5 (MMP, ChemSW, Fairfield, CA).

Dependent and Predictor Variables

The dependent variable LogMP98 was derived from the Ames test data by conversion of revertants/microgram of mutagen into $\log_{10}$ revertants/nanomoles. The predictor variables were derived from the software calculations described above and are defined in Table II.

RESULTS

Mutagenic Potency

The database of the main dependent variable of Ames/Salmonella mutagenic potency in strains TA98 and the independent or predictor variables are presented for the set of 12 TMIP isomers in Table III. Potency data for *O*-acetylase-overproducing strain YG1024 are also given but were not analyzed in detail. Below the data are the averages for the more potent three isomers, the next eight isomers, and the probability of a *t*-test for the significance of the difference, using unequal variances. A significant difference between the isomer subsets suggests a predictor that should be examined further for influence on potency. The bottom line of this table gives the correlation coefficient for each predictor variable with LogMP98.

Whereas our previously reported series of 80 aromatic and heterocyclic amines (AHA80) of heterogeneous structure covered a potency range of about 10 orders of magnitude [Hatch et al., 2001], the 11 measured TMIPs had a range of nearly three orders of magnitude between the 30th and 76th percentiles of the AHA80 series. The narrower potency range, as well as extensive multicollinearity among many pairs of predictor variables, place limita-

tions on our ability to produce reliable statistical models relating potency to predictor variables.

Predictor Variables

General

Independent or predictor variables in this study are separated into three types: structural, quantum chemical, and hydropathic, as described in Table II. Because there are many substantial correlations among the predictor variables, it is essential to avoid combining highly correlated variable pairs ($r > \sim 0.7$) in multivariate regression models to prevent artifacts created by multicollinearity [Glantz and Slinker, 1990] pp 181 ff.]. Carefully selected combinations of predictors were incorporated into models that are believed to be free of artifact.

The relatively small size of the data set poses a challenge in developing a statistically significant QSAR from a sizable number of potential predictor variables. Topliss and Edwards [Topliss, 1979] used Monte Carlo simulations to investigate the probability of significant correlations occurring by chance in models built from sets of random numbers. Although no simple rule of thumb emerges from that study, it suggests that to achieve a $P < 0.05$ level of confidence that a regression model is not spurious, 11 observations will support up to six or seven variables. Recently, Livingstone and Salt published a new analysis of the question of "selection bias" in multiple regression analysis [Livingstone and Salt, 2005]. They provide an approximate equation for Fmax that is corrected for this bias. The significance of the models in this article was tested using this adjusted Fmax (see later).

The database for the isomers is presented in Table III, showing mutagenic potency in Salmonella and only those predictor variables that contributed to satisfactory multivariate regression models or were of potential mechanistic interest. Other variables were calculated with the several software programs described under Methods, but did not appear to be significant for prediction of potency or for mechanistic insight; and these data are omitted.

Table IV presents the RHF energies of the amine, tautomer, hydroxylamine, and nitrenium ion forms of the TMIPs and for aniline derivatives to facilitate calculation of reaction energies. In addition, the RHF energy differences between the reaction stages and the relative energies for hydroxylamine and nitrenium ion formation are provided (calculated relative to aniline, following Ford and Griffin [1992]). Table V contains a correlation matrix for the dependent and all relevant independent variables of the isomers database (Tables III and IV). Values are printed only if their significance level (shown on the second line of each entry) is below 0.1. Clearly, all of the predictors are at least moderately well-correlated pairwise

TABLE IV. RHF Energies and Ford Calculations

Name	LogMP98	RHF_NH2	RHF_tau	RHF_NHOH	RHF_NH+	ΔEtau_NH2	ΔENHOH_NH2	ΔENH$^+$_NHOH	FordNHOH_NH2	FordNH$^+$_NHOH
367bmip	1.74707	−565.64641	−566.68556	−640.43652	−564.78991	0.01084	−74.79011	75.64661	−10.96762	−6.66353
356bmip	1.61336	−565.64519	−565.63537	−640.43584	−564.79708	0.00982	−74.79986	75.63876	−11.30836	−11.58885
357bmip	0.69316	−595.65023	−565.63926	−640.44147	−564.80193	0.01097	−74.79124	75.63954	−11.67608	−11.10128
167bmip	0.24600	−565.62959	−565.63540	−640.42596	−564.77416	0.00420	−74.79637	75.65180	−14.89269	−3.40926
157bmip	0.00186	−565.64009	−565.63425	−640.43007	−564.78149	0.00469	−74.78998	75.64858	−10.88228	−5.42859
156bmip	−0.26263	−565.63660	−565.62850	−640.42936	−564.78609	0.00235	−74.79276	75.64327	−12.62738	−8.76380
346cmip	−0.39227	−565.63763	−565.63512	−640.43327	−564.76357	0.00913	−74.79565	75.66970	−14.43901	7.82630
146cmip	−0.54987	−565.64085	−565.62606	−640.43605	−564.78662	0.00573	−74.79520	75.64943	−14.15976	−4.89521
147cmip	−0.79975	−565.63252	−565.62596	−640.43022	−564.77626	0.00646	−74.79770	75.65396	−15.72728	−2.05196
367cmip	−1.05502	−565.63786	−565.62596	−640.42898	−564.75993	0.01190	−74.79112	75.86905	−11.59952	7.41717
167cmip	−1.05502	−565.62915	−565.62389	−640.42641	−564.77006	0.00526	−74.79726	75.65635	−15.45306	−0.55095
347cmip	−1.05502	−565.63576	−565.62455	−640.43212	−564.76002	0.01121	−74.79636	75.67210	−14.88579	9.32856
aniline		−285.74760	−285.70953	−360.52024	−284.86300					
avgtop3		−565.64727	−565.63673	−640.43795	−564.79631	0.01054	−74.79067	75.64164	−11.31735	−9.78455
avgnxt8		−565.63554	−565.62932	−640.43004	−564.77477	0.00599	−74.79450	75.65527	−13.72262	−1.23204
T-test prob		0.00133	0.00792	0.02116	0.00437	0.00126	0.00684	0.01138	0.00684	0.01138
Correl. LogMP98	1.000	−0.684	−0.642	−0.584	−0.725	0.452	0.603	−0.376	0.602	−0.657

TABLE V. Correlation Matrix

|	LogMP98	Dipole NH2	ΔDipole	qNmeNH2	LogP_GC	Ep_β	FordNH+_NHOH
LogMP98	1.000						
DipoleNH2	−0.799	1.000					
	0.003						
ΔDipole	−0.913	0.918	1.000				
	0.000	0.000					
qNmeNH2	−0.836	0.707	0.721	1.000			
	0.001	0.010	0.008				
LogP_GC	0.741		−0.676	−0.577	1.000		
	0.009		0.016	0.050			
Ep_β	−0.735		0.686	0.549	−0.970	1.000	
	0.010		0.014	0.065	0.000		
FordNH+_NHOH	−0.657		0.561	0.700	−0.677	0.712	1.000
	0.028		0.058	0.011	0.016	0.009	
SoftNH2	0.637			−0.680	0.746	−0.788	−0.930
	0.035			0.015	0.005	0.002	0.000
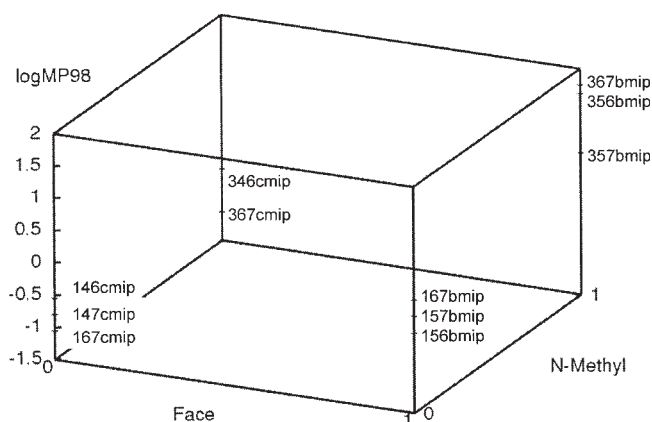
Fig. 1. 3-D graph showing the segregation of the mutagenic potencies of the TMIP isomers by the face of pyridine–imidazole ring fusion and the position of the imidazole N-methyl substituent.

with mutagenic potency and these form the source for regression models.

Structural Predictors

Two structural variables, Face and Nme, show extraordinary relationships to potency (Fig. 1). The six isomers of higher potency all have the pyridine ring fused to the imidazole ring at the *b*-face, with the pyridine ring nitrogen at position 4. The more potent three of these have the N-methyl group at position 3 and the less potent three at position 1. The five isomers of lower potency all have a *c*-face ring fusion, with the pyridine ring nitrogen at position 5. Three of these have the N-methyl at position 1 and two at position 3; the unsynthesized isomer is also at position 3.

Quantum Chemical Predictors

The quantum chemical variables found to be potential predictors of potency are ΔDipole, qNmeNH2, Dipole
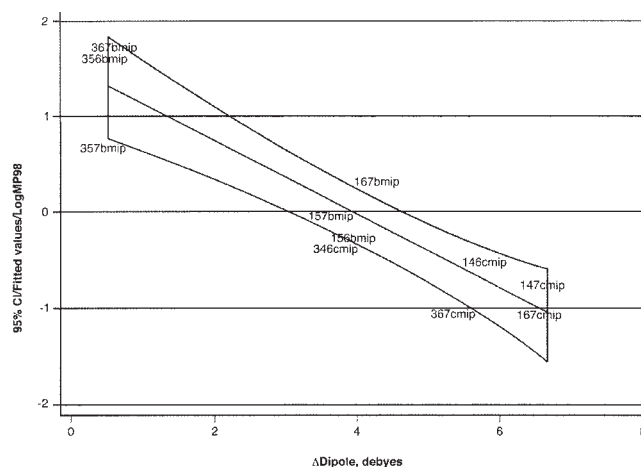

Fig. 2. Scatter graph of the relationship between LogMP98 and ΔDipole, which is the difference between the dipole moments of the parent amine and the imine tautomer. The correlation coefficient is highly significant at −0.913. Also shown are the linear predicted fit line and its 95% confidence interval bands. Actual data points lie under the "*b*" or "*c*" of the isomer names.

NH2, LogP_GC, Eπ_β, FordNH$^+$_NHOH, and SoftNH2 in descending order of correlation with LogMP98.

During study of the various activation intermediates, we became interested in the imine tautomer that could be formed by transfer of a hydrogen from the exocyclic N-atom of the parent amine to the nonmethylated nitrogen of the imidazole ring and shifting of the adjacent double bond to form an exocyclic imine. The RHF energies to form the tautomers from the parent amines are fairly low (1.5–7.0 kcal/mol), indicating that reactions involving tautomer formation could be quite fast at room temperature. There is little overall correlation between the energy of the tautomers and mutagenic potency, but the tautomer formation energy of the three most mutagenic isomers is much higher than that of most of the other isomers. The tautomer dipole moment, Dipoletau, is
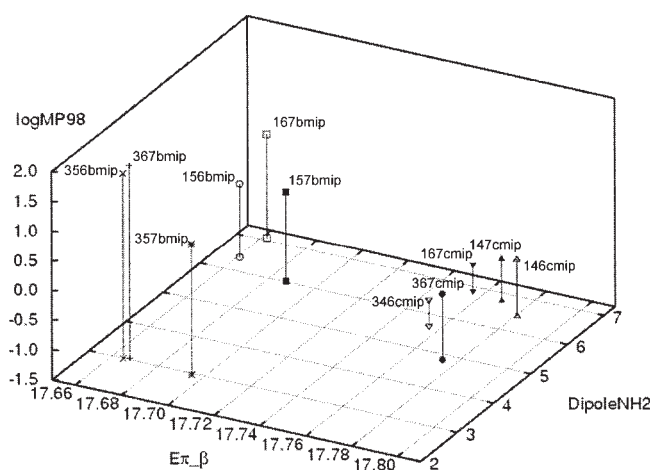
Fig. 3. 3-D graph showing the segregation of the mutagenic potencies of the TMIP isomers by the predictors Eπ_β and DipoleNH2. Eπ_β creates a major separation based on Face, and DipoleNH2 creates a modest separation based on Nme position.

well correlated with potency. The difference between the parent dipole and the tautomer dipole, ΔDipole, has the highest correlation with potency of the variables we have studied ($r = -0.91$) (Fig. 2).

The variable qNmeNH2 is the NPA charge on the imidazole N-atom containing the methyl substituent. This variable is strongly correlated ($r = -0.89$) with LogMP98.

DipoleNH2 exhibits a striking relation to potency; the three isomers of highest potency have the lowest dipole moments by a wide margin, and these are *b*-face and N3-methyl molecules (Fig. 3). Among the *c*-face isomers, the 3-Nmethyls also have lower dipole moments than the 1-Nmethyls, but by a lesser margin. DipoleNH2 is the main contributor to four of the regression models (Table VI).

Eπ_β is the total energy of π electrons calculated using simple Hückel theory. Although the number of π electrons is uniformly 12 in all isomers, the total energies show slightly smaller β values in the more potent isomers. In fact, the π-electronic energies are less negative for all *b*-face isomer energies than for all *c*-face isomers. This variable is moderately correlated with potency and makes a moderate contribution to one regression model (Table VI and Fig. 3).

Reaction energies for some steps in the activation process were calculated according to the method of Ford and Herman [1991, 1992]. This scheme, as applied to heterocyclic amine food mutagens by Ford and Griffin [1992], calculates the relative energy of formation of the TMIP nitrenium ions from the parent amines, using the analogous reaction for aniline as a reference. The energy to form the nitrenium ion may relate to the energy to form reactive intermediates that ultimately react with DNA to form adducts, as measured by LogMP98. Ford and Griffin, using the semi-empirical AM1 method and omitting the esterification and deesterification reaction steps because of

their complexity, found that only the reaction to form the nitrenium ion from the parent amine showed any correlation between reaction energy and potency [Ford and Griffin, 1992].

Our calculations, using ab initio (RHF) energies and omitting the ester intermediates, also show a slight correlation between the mutagenic potency and the energy to form the nitrenium ion from the parent amine ($r = 0.52$), as well as slightly stronger correlations with potency for both the formation of the hydroxylamine derivative from the parent amine ($r = 0.60$) and the formation of the nitrenium ion from the hydroxylamine ($r = -0.66$). Note that the energy to form the nitrenium ion from the parent amine correlates very strongly with the energy to form the nitrenium ion from the hydroxylamine ($r = 0.96$), but not with the energy to form the hydroxylamine from the parent ($r = -0.10$).

FordNHOH_NH2 is the energy ($E_{products} - E_{reactants}$) of the isodesmic reaction between AnilineNHOH and TMIPNH2, yielding anilineNH2 and TMIPNHOH (Fig. 4). The slope is positive, indicating that the more potent three isomers have less negative reaction energies, and therefore their hydroxylamine derivatives will require more energy to form from the parent amines (Note that all of these reaction energies are calculated relative to aniline, so their sign does not indicate whether the reaction is exothermic or endothermic in absolute terms).

FordNH$^+$_NHOH is the energy of the reaction between anilineNH$^+$ and TMIPNHOH, yielding anilineNHOH and TMIPNH$^+$ (Fig. 5). The slope is negative and the more potent three isomers have relatively large negative reaction energies for the formation of more stable nitrenium ions from the N-hydroxyl species.

SoftNH2 shows a modest correlation with potency, which signifies that the more potent isomers are generally softer in the Pearson sense (i.e., have narrower energy differences between HOMO and LUMO orbitals and are more readily polarizable). See discussion of softness as a QSAR predictor in Arulmozhiraja and Morita [2004].

Hydropathic Predictors

The hydropathic variable LogP_GC is moderately correlated with mutagenic potency ($r = 0.74$). This Ghose-Crippen calculation is atom-based, in contrast to the more familiar Hansch method that is based on the summation of values for molecular fragments. The Hansch calculation (LogP_HL) was somewhat correlated with potency but did not enter significant regression models (data not shown).

Regression Models

Satisfactory regression models, based on potency as the dependent variable, require care in design because of the number of predictor variables and the frequency of multi-collinearity among them. Six acceptable model equations

TABLE VI. Multiple Linear Regression Models

Model						
Model 1	LogMP98=	−0.3304	*ΔDipole	+2.1818	*SoftNH2	−1.8472
	Pcoeff.	0.000		0.106		
	Beta	−0.7906		0.2522		
	Variance %:	81.60%		3.70%		
	RMSE	0.3751				
	Adjusted R^2:	85.30%				
	Fmodel:	39.06				
	Pmodel:	0.0002				
Model 2	LogMP98=	−0.3624	*DipoleNH2	1.0049	*LogP_GC	+0.4086
	Pcoeff.	0.002		0.005		
	Beta	−0.6091		0.5178		
	Variance %:	59.80%		24.00%		
	RMSE	0.3946				
	Adjusted R^2:	83.80%				
	Fmodel:	26.78				
	Pmodel:	0.0003				
Model 3	LogMP98=	−0.364	*DipoleNH2	−9.8782	*Ep_β	+176.9349
	Pcoeff.	0.002		0.006		
	Beta	−0.6118		−0.511		
	Variance %:	59.80%		23.20%		
	RMSE	0.4034				
	Adjusted R^2:	83.00%				
	Fmodel:	25.44				
	Pmodel:	0.0003				
Model 4	LogMP98=	−0.4128	*DipoleNH2	+4.1743	*SoftNH2	−17.6514
	Pcoeff.	0.001		0.007		
	Beta	−0.6937		0.487		
	Variance %:	59.80%		23.20%		
	RMSE	0.4035				
	Adjusted R^2:	83.01%				
	Fmodel:	25.43				
	Pmodel:	0.0003				
Model 5	LogMP98=	−128.0435	*qNmeNH2	−8.2513	*Ep_β	+80.4331
	Pcoeff.	0.005		0.032		
	Beta	−0.6257		−0.427		
	Variance %:	66.50%		13.09%		
	RMSE	0.4423				
	Adjusted R^2:	79.59%				
	Fmodel:	20.49				
	Pmodel:	0.0007				
Model 6	LogMP98=	−0.3819	*DipoleNH2	−0.0619	*FordNH$^+$_NHOH	+1.8472
	Pcoeff.	0.007		0.049		
	Beta	−0.3417		−0.4121		
	Variance %:	59.80%		13.10%		
	RMSE	0.5097				
	Adjusted R^2:	72.90%				
	Fmodel:	14.44				
	Pmodel:	0.0022				

were found, involving LogMP98 as the dependent variable and a variety of predictor variables. The coefficients of determination, adjusted R^2, are equal to or greater than 80% for five of the models and slightly lower for the sixth model. In each case, no other variables deemed to be relevant could be entered into the models without violating multicollinearity criteria for a proper model. Thus, a large fraction of the total variance in potency is "explained" by the models.

There is a further limitation to the regression analysis of the TMIP isomers. This is based on the work of Topliss and Edwards [1979], who replaced either or both of the dependent and independent variables in test QSARs with random numbers and showed that stepwise multiple regressions produced significant responses by chance when the ratio of number of variables to number of observations exceeded certain values. Since our study has seven variables and 11 observations, it could be subjected to the foregoing problem. A recent article [Livingstone and Salt, 2005] extends the work of Topliss and Edwards by producing an algorithm to predict Fmax values from large numbers of simulations with random number input at various numbers of observations, total predictors considered, and predictors in the model. From a rank-ordered
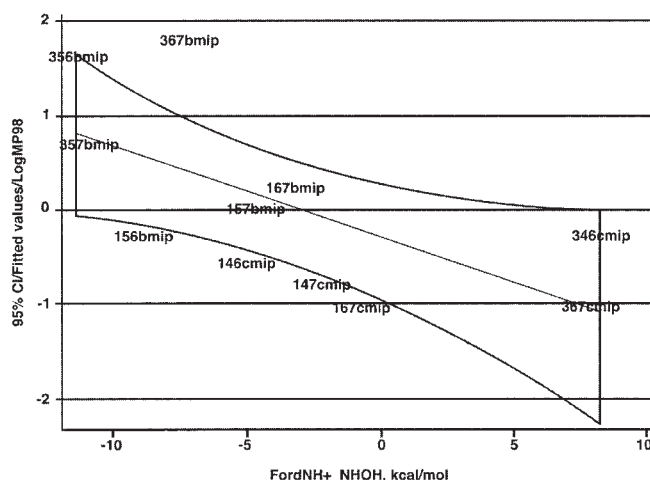
Fig. 4. Scatter graph of the relationship between LogMP98 and the value of FordNHOH_NH2, which is calculated as follows:

Reaction: TMIP_NH2 + PhNHOH = TMIP_NHOH + PhNH2

Calculation: ((RHF_NHOH + RHF_PhNH2) - (RHF_NH2 + RHF_PhNHOH))*627.51 kcal/mol. The correlation coefficient ($r = 0.646$) is significant. Also shown are the linear predicted fit line and its 95% confidence interval bands. Actual data points lie under the "b" or "c" of the isomer names. "Ph" equals Phenyl [Ford and Griffin, 1992].

Fig. 5. Scatter graph of the relationship between LogMP98 and the value of FordNH+_NHOH, which is calculated as follows:

Reaction: TMIP_NHOH + PhNH$^+$ = TMIP_NH$^+$ + PhNHOH

Calculation: ((RHF_NH$^+$ + RHF_PhNHOH) - (RHF_NHOH + RHF_PhNH$^+$))*627.51 kcal/mol. The correlation coefficient ($r = -0.657$) is significant. Also shown are the linear predicted fit line and its 95% confidence interval bands. Actual data points lie under the "b" or "c" of the isomer names. "Ph" equals Phenyl [Ford and Griffin, 1992]. Note that the slopes of Figures 4 and 5 are opposite.

list of the resulting Fmodels, Fmax values can be obtained for a series of probability levels. A power function equation was fitted to a large amount of simulation data. Both the simulation calculation and the equation output are freely available on the authors' website. When applied to our six regression models for 11 observations, seven total predictors, and two predictors per model, the Fmax for chance significance of the model at 5% probability by simulation was 11.8 and by equation was 13.0. For five of our regression models, the Fmodel ranged from 20 to 30, far above the chance result. For our weaker sixth model, Fmodel is 14.4, which gives a probability of a chance result of somewhat less than 5%; so chance is not completely ruled out.

Table VI details the regression models in descending order of the adjusted R^2. The first line of each model shows the equation derived from the data: the dependent variable is listed first, followed by the major predictor variable and other predictors, if any, and finally the constant. The second line gives the significance level of the coefficient above it; in several instances, significance greater than $P = 0.05$ was allowed for minor variables because the model was otherwise highly significant (Fmodel, Pmodel). The third line gives the β values for each coefficient, which are the coefficients that would be obtained if all variables had been normalized to a mean of 0 and a standard deviation of 1, thus correcting the coefficients for differences in scale among the calculated variables. This normalization was used by the software to
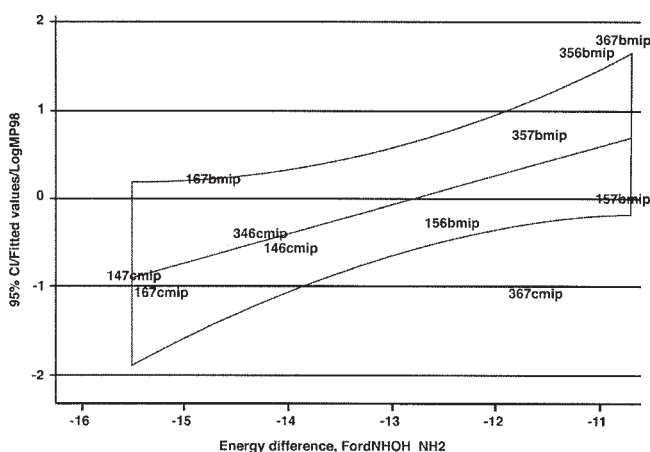
create all models presented. Thus, β gives the normalized value of each coefficient, or the relative weight of each predictor variable in the model. The fourth line shows the fraction of the total variance in the dependent variable that is "explained" by each predictor variable. The final three rows give the value of the coefficient of determination for the model, R^2 (adjusted for the number of predictor variables in the model); the value of the F-test for the significance of the model; and the Pmodel, which is the probability that the null hypothesis (i.e., no credible prediction) can be accepted.

The first five models give adjusted R^2 from 80 to 85% of total variance. Model 1 admits ΔDipole with a large contribution and a small addition from SoftNH2. In models 2–4, DipoleNH2 is the major contributor with additions of 23–24% from LogP_GC, Ep_β or SoftNH2. Only one other variable was allowed per model by forward stepwise regression. Model 5 admitted qNmeNH2 at 66.5% of variance and Ep_β at 13.1%. Model 6, at a lower adjusted R^2 of 72.9%, again had DipoleNH2 as the main predictor and FordNH$^+$_NHOH with a smaller contribution.

All six regression models were used to predict the potency of the not-yet synthesized isomer 347cmip by entering the values of the predictors for that molecule into the equation (Table VII). The predicted potencies from these model equations were averaged and gave a value of LogMP98 = −0.866 or 0.77 revertants/µg for 347cmip, which is near the bottom of the potency range of the 11 measured isomers.

TABLE VII. Regression Models for Predicting Mutagenic Potency (LogMP98) of 347cmip

						Predictions
Model 1	LogMP98=	−0.3304	*ΔDipole	+2.1618	*SoftNH2	−9.01425	−0.9990
		−0.3304	5.1223	+2.1618	4.49055	−9.01425	0.569rev/µg
Model 2	LogMP98=	−0.3624	*DipoleNH2	+1.0049	*LogP_GC	+0.4086	−0.5415
		−0.3624	5.83	+1.0049	1.157	+0.4086	1.631rev/µg
Model 3	LogMP98=	−0.364	*DipoleNH2	−9.8762	*Ep_β	+176.9349	−0.6675
		−0.364	5.83	−9.8762	17.768	+176.9349	1.22rev/µg
Model 4	LogMP98=	−0.4128	*DipoleNH2	+4.1743	*SoftNH2	−17.6514	−1.3131
		−0.4128	5.83	+4.1743	4.49055	−17.6514	0.276rev/µg
Model 5	LogMP98=	−126.0435	*qNmeNH2	−8.2513	*Ep_β	+80.4331	−1.2636
		−126.0435	−0.515	−8.2513	17.768	+80.4331	0.309rev/µg
Model 6	LogMP98=	−0.3819	*DipoleNH2	−0.0619	*FordNH+_NHOH	+1.8472	−0.957
		−0.3819	5.83	−0.0619	9.3286	+1.8472	0.627rev/µg
					Predicted LogMP98	−0.8664
					Average	0.772rev/µg

Relationships Among Predictor Variables

As shown in Table V, ΔDipole is moderately and inversely correlated with Face, LogP_GC, and Eπ_β; and it is positively correlated with qNme_NH2. DipoleNH2 makes a substantial contribution to determining the variance in mutagenic potency of the isomers. DipoleNH2 is negatively correlated with Nme, whereby the three highest potency isomers (all N3-methyl) have very low dipole moments and the remaining N3-methyl isomers are slightly less potent than the N1-methyl isomers. LogP_GC is correlated strongly with b-face over c-face, and inversely with Eπ_β. LogP_GC also correlates moderately with ElumoNH2 (inversely) and SoftNH2. Eπ_β is very strongly correlated with b-face over c-face and shows a sizeable gap between them (Fig. 3).

In summary, our results indicate that the most important determinants of relatively high mutagenic potency in the TMIP amines are: (1) a small dipole moment, (2) the combination of b-face and N3-methyl group, (3) a lower calculated energy of the π electron system, (4) a smaller energy gap between the amine HOMO and LUMO, and (5) a more stable nitrenium ion by the Ford and Griffin [1992] calculation.

DISCUSSION

Observations

An intriguing characteristic of the TMIP isomers is that the more potent six members have the pyridine nitrogen atom at the 4-position (b-face ring fusion) and the less potent six at the 5-position (c-face ring fusion). The potency variation is a continuum without a sharp distinction based on Face. Concomitantly, in the more potent three isomers, the imidazo-substituted methyl group is on the 3-position and on the 1-position of the next most potent three isomers; all of these are b-face. In the remaining less potent six isomers (all c-face), the N1- and N3-methyl substitutions are scattered, with the unsynthesized isomer being at N3-methyl and c-face. The relation between potency and methyl position on the imidazole ring (N1 vs. N3) contrasts with our previous AHA80 series, where N1-methyl was usually associated with higher potency than N3-methyl for several N-methyl-2-aminoimidazopyridine compounds [Hatch et al., 2001].

A low dipole moment is characteristic of the three more potent isomers, all of which have N3-methyl groups and b-face, and to a lesser extent of the remaining two less potent c-face N3-methyl isomers. This signifies a smaller asymmetry of atomic charge distribution. Separately including the three Cartesian components of the dipole moment (using a standard Cartesian reference frame for all isomers) in the regression did not provide any additional relation to potency; however, the TMIPs can be classified according to the direction of their dipole moment vector. The dipole moments of the N1-methyl-b-Face isomers average 6.51 D and lie on average 24.7° clockwise from a reference line drawn along the bond shared by the imidazole and pyridine rings (and oriented with the exocyclic amino group at the right [as in Table I]). The N1-methyl-c-Face isomers have an average dipole magnitude of 6.68 D and point 47.7° clockwise from this line. The most mutagenic N3-methyl-b-Face subset average 2.99 D and 124.7° clockwise, and the N3-methyl-c-Face isomers average 5.72 D and 120.0° clockwise. Thus, the positions of the N-methyl substituent

and the pyridine N-atom affect both the magnitude and orientation of the dipole moment. The position of the pyridine N-atom (Face) also obviously affects whether methyl substitutions can occur at positions 4 or 5.

The position of the pyridine ring N-atom at the 4 (*b*-face fusion) or 5 (*c*-face fusion) position is strongly correlated with potency, in that all six N4 isomers are more potent than the five measured N5 isomers. The N4 isomers correspond to the configuration found in all known thermic mutagens containing a pyridine ring that have been isolated from cooked foods. Furthermore, *b*- vs. *c*-face fusion is correlated positively with LogP_GC and SoftNH2, and inversely with Eπ_β. Thus, the fusion mode is related to a constellation of quantum chemical parameters that can contribute to an hypothesis of chemical mechanism.

The calculated LogP values produce interesting but disparate results in terms of relationships to each other and to other predictors. The classical fragment-hydrophobicity method LogP_HL of Hansch and Leo [1979] is not significantly correlated with potency ($r = 0.54$), but is significantly correlated with LogP_GC ($r = 0.68$) [Ghose et al., 1988]. However, LogP_HL is strongly correlated with SoftNH2 ($r = 0.91$), moderately with Eπ_β, and Face. LogP_GC is significantly correlated with potency ($r = 0.74$), strongly correlated with Eπ_β ($r = -0.97$) and Face ($r = 0.96$), and moderately with SoftNH2 ($r = 0.75$). This method is connectivity-based [Ghose et al., 1988], but upon examination appears to create many tiny fragments consistent with a refinement of the Hansch-Leo method. The modest correlation of LogP_GC with LogP_HL and the different correlation relationships of the two methods with other predictors signify important differences in what they actually measure, at least when applied to the TMIP isomers. These differences should be noted when evaluating the hydropathic properties of chemicals or drugs.

SoftNH2 is defined here as the reciprocal of Pearson's Hardness, which makes the correlation with mutagenic potency of the TMIPs positive (i.e., greater SoftNH2 corresponds to greater mutagenic potency). Although moderately correlated with potency ($r = 0.64$), SoftNH2 is rather well correlated with LogP_HL, Eπ_β, and Face; and moderately with LogP_GC; and strongly but inversely with FordNH$^+$_NHOH. A high SoftNH2 implies a narrow HOMO-LUMO gap in the parent amine. The further correlation of SoftNH2 with two of the LogP calculations is consistent with Softness, facilitating the crossing of the cell membrane and/or access to membrane-bound CYP450 enzymes.

Eπ_β represents the simple Hückel calculation of total π-electron energy, in β units. EB_β is correlated inversely strongly with Face and LogP_GC, moderately with potency ($r = -0.73$), and SoftNH2. EB_β completely segregates *b*-face and *c*-face isomers, with a sizable gap between them (Fig. 3). Thus, the shift of N4 to the N5 position increases the π-energy of the aromatic system significantly. A combination of Eπ_β and DipoleNH2 is capable of segregating the isomers into subgroups both by Face and Nme (Fig. 3).

Mechanistic Interpretations

This study is focused on finding relationships between the observed structural and calculated electronic properties of the TMIP mutagens and their measured mutagenic potencies. These properties presumably relate to the metabolic activation process that is required to convert the promutagenic TMIPs into active mutagens whose potency is measured by the Ames/Salmonella assay. However, one must be aware that the assay outcome includes several additional potential sources of variation for which data are not available: differential distribution or detoxification of the different isomers or their metabolic intermediates, interaction of the activated mutagen with DNA, structure-specific distortion of the DNA helix by bulky adducts and its effect on mutation, and any residual repair capacity of the repair-defective assay organism.

In chemical systems related to the thermic mutagens, there has been no isolation or direct spectroscopic observation of a nitrenium form of the thermic mutagens, although there is indirect evidence, such as reactant selectivity and reasonably long measured lifetimes for related species (e.g., up to 2×10^4 sec^{-1} for the 2-fluorenyl-nitrenium ion) (Anderson and Falvey [1993], and see Novak and Rajagopal [2001] for review). Such a product should be very reactive and prone to interaction with scavengers such as SH compounds or free radicals that would divert the ion from a mutagenic pathway. This would also be true for an N-ester with an excellent leaving group. Furthermore, the chemistry involved in the formation of the bulky DNA adduct may be equally amenable to a concerted S$_N$2 reaction (bimolecular nucleophilic substitution in which a DNA base nucleophile attacks an electrophilic center, forcing out a leaving group in a concerted reaction) or to the putative stepwise S$_N$1 pathway (unimolecular nucleophilic substitution in which a leaving ester group departs resulting in formation of an electrophilic nitrenium ion, followed by attachment of a DNA base nucleophile in a two-step reaction). This activation step may take place near the DNA molecule. For our purpose of developing a QSAR for mutagenicity, the use of the hypothetical nitrenium ion seems suitable, as this structure can either be an intermediate in an S$_N$1 pathway or a surrogate for a transition state in an S$_N$2 pathway. However, the conventional nitrenium ion structure with the positive charge at the exocyclic N-atom, which we used in our Ford calculations, is now believed to be correctly represented by a canonical imine structure with the charge distributed on the proximal ring [Novak and Lin, 1999].

Therefore, this type of calculation may be more appropriately done in accordance with Scheme 1 of that reference, which then produces an hydrated derivative in solution. However, the correct imine-type structures for the two-ring TMIP molecules (with a five-membered proximal ring) would have to be determined. [Note that the quantum chemical methods will automatically form the lowest energy electronic structure (nitrene, imine, etc.) for any given atomic connectivity].

All of the predictor variables analyzed here are well correlated with the mutagenic potency of the TMIP isomers: ΔDipole, qNmeNH2, DipoleNH2, Face, LogP_GC, Eπ_β, FordNH$^+$_NHOH, and SoftNH2 in descending order of pairwise correlation coefficients. Within this series of predictors, there are many significant pairwise correlations affecting multicollinearity, but the challenge remains to elucidate the mechanistic insight provided by these predictors. As in all studies of enzyme-mediated pathways, an important hypothesis is that the most potent compounds differentially bind into the active site based on idiosyncratic structural and electronic properties of the substrate. Clearly, several of the predictors of mutagenic potency, such as the dipole moment of the parent amine (DipoleNH2), the location of the pyridine endocyclic nitrogen (Face), and LogP_GC (affecting access to membrane-bound CYP450 enzymes), could plausibly affect the binding to the enzyme active site. The correlations to other properties, such as the energy of the π electron system (Eπ_β) and the reaction energy to form the nitrenium ion from the N-hydoxy species, are intriguing because they suggest that differences in the reactivity of the different isomers may also contribute to the observed differences in mutagenic potency. Hence, other compounds with the same binding affinity to the enzyme, but with different electronic properties, may exhibit widely varying mutagenic potency.

CONCLUSIONS

The objective of this article has been to apply QSAR technology to a series of mutagenic and potentially carcinogenic heterocyclic amine isomers. We have attempted to obtain some insights into the chemical features and calculable parameters that contribute to their significant variation in mutagenic potency. The findings may be limited in their applicability to the isomers and may not be directly relevant to molecules of greater structural diversity.

It seems amazing that a dozen isomers of nearly identical structure can display a 634-fold variation in mutagenic potency that is probably based on subtle differences in properties, most of which can only be calculated by use of quantum chemical theory. In addition to the calculated variables, unmeasured variables such as hyperconjugation from methyl substituents, variation in surface hydropathy, or details of molecular shape may influence the fitting of an isomer or its ion into the active site of a CYP450 1A enzyme catalyzing an early activation step, or the active site of an ester transferase enzyme catalyzing a late activation step. There is a strong suggestion from organic chemical mechanisms that the key properties tune each TMIP isomer to its own propensity for kinetic and/or thermodynamic reactivity in the activation reactions of the Ames/Salmonella assay. This study has presented regression models for the mutagenic potency of the 12 unique isomers of TMIP using a variety of structural, electronic, and hydropathic properties. Such a study is inherently limited by the choice and predictability of the regression properties, although this study has reduced one typical source of uncertainty in QSAR studies—the error in the predictor variables—by using accurate ab initio quantum chemical methods to predict most of the properties used in the regression. This study produced a number of regression models that reasonably account for the experimentally measured mutagenic potencies of the TMIP isomers. Finally, these regression models predict rather consistently the potency of the single unassayed isomer. We await the synthesis of sufficient material for the assay, and hope for confirmation of the QSAR.

ACKNOWLEDGMENTS

REFERENCES

Adamson RH, Farb A, Virmani R, Snyderwine EG, Thorgeirsson SS, Takayama S, Sugimura T, Dalgard DW, Thorgeirsson UP. 1995. Studies on the carcinogenic and myocardial effects of 2-amino-3-methylimidazo [4,5-f] quinoline (IQ) in nonhuman primates. In: Adamson RH, Gustafsson J-A, Ito N, Nagao M, Sugimura T, Wakabayashi K, Yamazoe Y, editors. Heterocyclic Amines in Cooked Foods: Possible Human Carcinogens. (23rd Intl Symp Princess Takamatsu Cancer Res Fund) Princeton, NJ: Princeton Scientific Publishing Co. Inc. pp 260–267.

Ahlberg LA, Danthi SN, Tanga MJ. 2000. Synthesis of the potential food mutagen 1,6,7-trimethyl-2-aminoimidazo[4,5-b]pyridine. American chemical Society 219:376 [Abstract]

Ames BN, Lee FD, Durston WE. 1973. An improved bacterial test system for the detection and classification of mutagens and carcinogens. Proc Natl Acad Sci USA 70:782–786.

Ames BN, McCann J, Yamasaki E. 1975. Methods for detecting carcinogens and mutagens with the Salmonella/mammalian microsomal mutagenicity test. Mutat Res 31:347–364.

Anderson GB, Falvey DE. 1993. Photogenerated arylnitrenium ions: absorption spectra and absolute rate constants for *tert*-butyl(4-halo-2-acetylphenyl)nitrenium ions measured by time-resolved laser spectroscopy. J Am Chem Soc 115:9870–9871.

Arulmozhiraja S, Morita M. 2004. Structure-activity relationships for the toxicity of polychlorinated dibenzofurans: approach through den-

sity functional theory-based descriptors. Chem Res Toxicol 17: 348–356.

Borgen E, Solyakov A, Skog K. 2001. Effects of precursor composition and water formation of heterocyclic amines in meat model systems. Food Chem 74:11–19.

Case RAM, Hosker M-E, McDonald DB, Pearson JT. 1954. Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry, Parts I, II. Br J Ind Med 11:75, 213

Colvin ME, Hatch FT, Felton JS. 1998. Chemical and biological factors affecting mutagen potency. Mutat Res 400:479–492.

Einisto P, Watanabe M, Ishidate M, Nohmi T. 1991. Mutagenicity of 30 chemicals in Salmonella typhimurium strains possessing different nitroreductase or O-acetyltransferase activities. Mutat Res 259: 95–102.

Felton JS, Knize MG, Wood C, Wuebbles BJ, Healy SK, Stuermer DH, Bjeldanes LF, Kimble BJ, Hatch FT. 1984. Isolation and characterization of new mutagens from fried ground beef. Carcinogenesis 5:95–102.

Felton JS, Wu R, Knize MG, Thompson LH, Hatch FT. 1995. Heterocyclic amine mutagenicity/carcinogenicity: influence of repair, metabolism, and structure. In: Adamson RH, Gustafsson J-A, Ito N, Nagao M, Sugimura T, Wakabayashi K, Yamazoe Y, editors. Heterocyclic Amines in Cooked Foods: Possible Human Carcinogens. (23rd Intl Symp Princess Takamatsu Cancer Res Fund) Princeton, NJ: Princeton Scientific Publishing Co. Inc. pp 50–58.

Felton JS, Knize MG, Hatch FT, Tanga MJ, Colvin ME. 1999. Heterocyclic amine formation and the impact of structure on their mutagenicity. Cancer Lett 143:127–134.

Ford GP, Herman PS. 1991. Comparison of the relative stabilities of polycyclic aryl nitrenium ions and arylmethyl cations: ab initio and semiempirical molecular orbital calculations. J Mol Struct (THEOCHEM) 236:269–282.

Ford GP, Griffin GR. 1992. Relative stabilities of nitrenium ions derived from heterocyclic amine food carcinogens: relationship to mutagenicity. Chem Biol Interact 81:19–33.

Ford GP, Herman PS. 1992. Relative stabilities of nitrenium ions derived from polycyclic aromatic amines: relationship to mutagenicity. Chem Biol Interact 81:1–18.

Fuscoe JC, Wu R, Shen NH, Healy SK, Felton JS. 1988. Base change analysis of Salmonella his gene revertant alleles. Mutat Res 201: 241–251.

Ghose AK, Pritchett A, Crippen GM. 1988. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: modeling hydrophobic interactions. J Comput Chem 9:80–90.

Glantz SA, Slinker BK. 1990. Primer of Applied Regression and Analysis of Variance. New York: McGraw-Hill. pp 181–236, 262–267.

Guengerich FP. 1995. Human cytochrome P450 enzymes. In: Ortiz de Montellano PR, editor. Cytochrome P450: Structure, Mechanism, and Biochemistry. 2nd edition. New York: Plenum. pp 473–535.

Guengerich FP, Humphreys WG, Yun C-H, Hammons GJ, Kadlubar FF, Seto Y, Okazaki O, Martin MV. 1995. Mechanisms of cytochrome P450 1A2-mediated formation of N-hydroxy arylamines and heterocyclic amines and their reaction with guanyl residues. In: Adamson RH, Gustafsson J-A, Ito N, Nagao M, Sugimura T, Wakabayashi K, Yamazoe Y, editors. Heterocyclic Amines in Cooked Foods: Possible Human Carcinogens. (23rd Intl Symp Princess Takamatsu Cancer Res Fund) Princeton, NJ: Princeton Scientific Publishing Co. Inc. pp 78–84.

Hansch C, Leo A. 1979. Substituent Constants for Correlation Analysis in Chemistry and Biology. New York: Wiley.

Hatch FT, Colvin ME. 1997. Quantitative structure-activity (QSAR) relationships of mutagenic aromatic and heterocyclic amines. Mutat Res 376:87–96.

Hatch FT, Knize MG, Felton JS. 1991. Quantitative structure-activity relationships of heterocyclic amine mutagens formed during the cooking of foods. Environ Mol Mutagen 17:4–19.

Hatch FT, Knize MG, Moore DH, II, Felton JS. 1992. Quantitative correlation of mutagenic and carcinogenic potencies for heterocyclic amines from cooked foods and additional aromatic amines. Mutat Res 271:269–287.

Hatch FT, Colvin ME, Seidl ET. 1996. Structural and quantum chemical factors affecting mutagenic potency of aminoimidazo-azaarenes. Environ Mol Mutagen 27:314–330.

Hatch FT, Knize MG, Colvin ME. 2001. Extended quantitative structure-activity relationships for 80 aromatic and heterocyclic amines: structural, electronic, and hydropathic factors affecting mutagenic potency. Environ Mol Mutagen 38:268–291.

Jensen F. 1999. Introduction to Computational Chemistry. New York: Wiley. p 353.

Kim D, Guengerich FP. 2004. Selection of human cytochrome P450 1A2 mutants with enhanced activity for heterocyclic amine N-hydroxylation. Biochemistry 43:981–988.

Livingstone DJ, Salt DW. 2005. Judging the significance of multiple linear regression models. J Med Chem 48:661–663.

Maron D, Ames BN. 1983. Revised methods for the Salmonella test. Mutat Res 113:173–215.

Moore D, Felton JS. 1983. A microcomputer program for analyzing Ames test data. Mutat Res 119:95–102.

Novak M, Lin J. 1999. Correlation of azido/solvent selectivities for nitrenium ions with ab initio hydration energies: understanding the kinetic lability of nitrenium ions in aqueous solution. J Org Chem 64:6032–6040.

Novak M, Rajagopal S. 2001. N-arylnitrenium ions. In: Tidwell TT, Richard JP, editors. Adv Phys Org Chem 36:167–253.

Novak M, Xu L, Wolf RA. 1998. Nitrenium ions from food-derived heterocyclic arylamine mutagens. J Am Chem Soc 120:1643–1644.

Pais P, Salmon CP, Knize MG, Felton JS. 1999. Formation of mutagenic/carcinogenic heterocyclic amines in dry-heated model systems, meats, and meat drippings. J Agric Food Chem 47:1098–1108.

Pearson RG. 1963. Hard and soft acids and bases. J Am Chem Soc 85:3533–3539.

Pearson RG, Songstad J. 1967. Application of the principle of hard and soft acids and bases to organic chemistry. J Am Chem Soc 89:1827.

Reed AE, Weinstock RB, Weinhold F. 1985. Natural population analysis. J Chem Phys 83:735–746.

Rehn L. 1895. Blasengeschwülste bei fuchsinarbeitern. Arch Klin Chir 50:588.

Sabbioni G, Wild D. 1992. Quantitative structure-activity relationships of mutagenic aromatic and heteroaromatic azides and amines. Carcinogenesis 13:709–713.

Sasaki JC, Fellers RS, Colvin ME. 2002. Metabolic oxidation of carcinogenic arylamines by P450 monooxygenases: theoretical support for the one-electron transfer mechanism. Mutat Res 506:79–89.

Schut HAJ, Snyderwine EG. 1999. DNA adducts of heterocyclic amine food mutagens: implications for mutagenesis and carcinogenesis. Carcinogenesis 20:353–368.

Sugimura T. 1995. History, present and future, of heterocyclic amines, cooked food mutagens. In: Adamson RH, Gustafsson J-A, Ito N, Nagao M, Sugimura T, Wakabayashi K, Yamazoe Y, editors. Heterocyclic Amines in Cooked Foods: Possible Human Carcinogens. (23rd Intl Symp Princess Takamatsu Cancer Res Fund) Princeton, NJ: Princeton Scientific Publishing Co. Inc. pp 214–231.

Tanga MJ, Bupp JE, Tochimoto TK. 1994. Syntheses of 1,5,6-trimethyl-2-aminoimidazo[4,5-b]pyridine and 3,5,6-trimethylimidazo[4,5-b]pyridine. J Heterocycl Chem 31:1641–1645.

Tanga MJ, Bupp JE, Tochimoto TK. 1997. Syntheses of five potential heterocyclic amine food mutagens. J Heterocycl Chem 34:717–727.

Tanga MJ, Bradford WW, Bupp JE, Kozocas JA. 2003. Syntheses of two potential food mutagens. J Heterocycl Chem 40:569–573.

Topliss JG, Edwards RP. 1979. Chance factors in studies of quantitative structure-activity relationships. J Med Chem 22:1238–1244.

Turesky RJ, Guengerich FP, Guillouzo A, Langouët. 2002. Metabolism of heterocyclic amines by human hepatocytes and cytochrome P4501A2. Mutat Res 506/507:187–195.

Vracko M, Szymoszek A, Barbieri P. 2004. Structure-mutagenicity study of 12 trimethylimidazopyridine isomers using orbital energies and "spectrum-like representation" as descriptors. J Chem Inf Comput Sci 44:352–358.

Watanabe M, Ishidate M, Jr, Nohmi T. 1990. Sensitive method for the detection of mutagenic nitroarenes and aromatic amines: new derivatives of Salmonella typhimurium tester strains possessing elevated O-acetyltransferase levels. Mutat Res 234: 337–348.

Weisburger JH. 2002. Comments on the history and importance of aromatic and heterocyclic amines in public health. Mutat Res 506/507:9–20.