LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-211911

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

M. G. Knize, F. T. Hatch, M. J. Tanga, E. V. Lau, M. E. Colvin

May 3, 2005

Environmental and Molecular Mutagenesis

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.V. Lau[1], and M.E. Colvin[4]

[1]Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, University of California, Livermore, CA, Livermore, CA

[2]Meredith, New Hampshire

[3] Biosciences Division, SRI International, Menlo Park, CA

[4] School of Natural Sciences, University of California, Merced, CA

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

Running title: QSAR of TMIP isomers

Corresponding author: Mark Knize, Biology and Biotechnology Division, University of California, Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA 94550.

E-mail: knize1@llnl.gov

Abstract

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 regression model accounting for the largest fraction of the total variance in mutagenic potency contains four predictor variables: dipole moment, a measure of the gap between amine LUMO and HOMO energies, percent hydrophilic surface, and energy of amine LUMO. The most important determinants of high mutagenic potency in these amines are: 1) a small dipole moment, 2) the combination of b-face ring fusion and N3-methyl group, and 3) a lower calculated energy of the π electron system. Based on predicted potency from the average of five models, the isomer not yet synthesized and tested is expected to have a mutagenic potency of 0.84 revertants/μg in test strain TA98.

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 [Sugimura, 1995; Adamson et al.,

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 binds covalently to 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 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].

An unknown mutagenic compound thought to be an amino-trimethylimidazopyridine, 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 2-amino-

trimethylimidazopyridines 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 2-amino-trimethylimidazopyridine (Table I) by chemical synthesis. 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. The 12 isomers vary in mutagenic potency and structurally only in the positions of methyl group substitution on the imidazole and pyridine rings. These features seemed ideal for a QSAR study, although the potency range (3 orders of magnitude) and sample size are rather limited for statistical analyses.

Table I

Three of the isomers are of moderate potency compared to many other heterocyclic amines, and eight others are weak to very weak mutagens; one is not yet synthesized and thus, 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.

MATERIALS AND METHODS

Table II

Abbreviations and definitions are presented in Table II.

Chemicals

Eleven of the 12 TMIP isomers were synthesized in the laboratory of Dr. Mary Tanga at SRI International (Menlo Park, 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_3O-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 sample extracts was determined using the standard plate incorporation assay described by Ames et al. [Ames et al, 1975; Maron and Ames, 1983], with *S. 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 mg of Aroclor-induced rat liver S9 protein was added per plate for metabolic activation, and mutagens 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 per 5 nanogram 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 µg [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 with *Salmonella 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].

The bacteria are added to a buffered suspension of microsomes (the 9000 x g 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 plus 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, more complex, membranes 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 is an HP Vectra VL420 1.9GHz 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 Inc., 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 with PSI-Plot, ver. 7.8 (Poly Software Intl., Pearl River, NY). *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 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 done 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. LogP_HL and Hyphil were calculated on minimized structures with Molecular Modeling Pro ver. 5 (MMP, ChemSW, Fairfield, CA).

Dependent and Predictor Variables

The dependent variable LogMP98 is derived from the Ames test data by conversion of revertants/μg of mutagen into $\log_{10}$ revertants/nmol. The predictor variables are 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 TA 98 and the independent or predictor variables is 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.

Table III

Whereas our previously reported series of 80 aromatic and heterocyclic amines (AHA80) of heterogeneous structure covered a potency range of about ten orders of magnitude [Hatch et al., 2001], the 11 measured TMIPs had a range of nearly three orders of magnitude between the 30^{th} and 76^{th} percentiles of the AHA80 series. The narrower potency range, as well as extensive multicollinearity among many pairs of predictor variables, place limitations 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, 1993 pp. 181 ff.]. Carefully selected combinations of predictors were incorporated into models that are believed to be free of artifact.

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 after statistical analyses; and these data are omitted.

TableIIIA

able IIIA presents the Restricted Hartree-Fock (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 IV contains a correlation matrix for the dependent and all relevant independent variables of the isomers database (Tables III and IIIA). Values are printed only if their significance level (shown on the second line of each entry) is below 0.05. Clearly many of the predictors are moderately well correlated pairwise with mutagenic potency and these form the main source for regression models.

Table IV

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.

Fig. 1

Quantum Chemical Predictors. The quantum chemical variables found to be potential predictors of potency are ΔDipole, qNmeNH2, DipoleNH2, LogP_GC, Eπ_ß, and $FordNH^{+}$_NHOH in descending order of correlation with LogMP98. SoftNH2 and Hyphil are minor contributors to some regression models (Table V) and occasional sources of multicollinearity preventing their wider use in regressions. ElumoNH2 contributes to two models, although it does not correlate pairwise significantly with potency.

Table V

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 non-methylated 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 to 7.0 kcal/mole) 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 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).

Fig. 2

The variable qNmeNH2 is the NPA charge on the imidazole N atom containing the methyl substituent. This variable is strongly correlated 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 V).

Fig. 3

Eπ_ß 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 ß 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 two regression models (Table V 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 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 from the parent amine (r = 0.52) as well as slightly stronger correlation 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 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 (Eproducts - 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 the highest reaction energies, and therefore will require the most energy to form from the parent amines; (i.e. these reactions are the most endothermic.. (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). 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 the nitrenium ion from the N-hydroxyl species. Further quantum chemical analysis of the formation and stabilization of nitrenium ions is in Colvin et al. [1997] and discussion of the reactions of the nitrenium ions with guanine is in Parks et al. [2001]. Taken together, these results indicate that the most potent of the TMIP mutagens have N-hydoxyl forms that are relatively the most favorable to form, but

Fig. 4

Fig. 5

also release the most energy in reacting to form the corresponding nitrenium ion.

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 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. This Ghose-Crippen calculation is atom-based, in contrast to the more familiar Hansch method that is based on summation of values for molecular fragments. The Hansch calculation (LogP_HL) was somewhat correlated with potency but did not enter significant regression models. The hydropathic variable Hyphil does not correlate with potency, but enters one model with a small contribution. This is difficult to interpret since all the isomers are nearly planar and have three methyl groups protruding in different directions. Differing fractions of hydrophilic surface based on methyl group positions are not obvious; however, moving the pyridine ring N atom from position 4 to 5 may affect the geometry of surface hydrophilicity.

Regression Models

Satisfactory regression models, based on potency as the dependent variable, require care in design because of the large number of predictor variables and the frequency of multicollinearity among them. Nine acceptable model equations 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 percent for six of the models. In each case no other variables deemed to be relevant could be entered into the models without violating

multicollinearity criteria for a proper model.

Table V 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). The third line gives the Beta 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 Beta 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 model accounting for the largest fraction of the total variance in LogMP98 (88.7%) contains four predictor variables: DipoleNH2, SoftNH2, Hyphil, and ElumoNH2. The main predictor for this model, DipoleNH2, is highly correlated with potency of the isomers and appears in three additional regression models. It is clearly one of the valuable predictors (Fig. 3).

Another model accounting for a large fraction of the LogMP variation is Model 2 in

Table V that relates LogMP98 to the difference between DipoleNH2 and Dipoletau (i.e. ΔDipole), plus the NPA charge on the imidazole ring N atom containing the methyl substituent (qNmeNH2). The result shows a strong relationship to potency; the adjusted R^2 = 87.4%.

Models 3-5 combine Dipolenh2 with single other predictors: LogP_GC, Eπ_ß, and SoftNH2. Each of these other predictors contributes nearly one-quarter of the potency variance; but because of multicollinearity the predictors cannot be combined with each other.

Models 6 and 9 include FordNH^+_NHOH, indicating the reaction energy and potentially the reaction rate of the multistep formation (via an N-acetoxy or N-sulfonyl ester) of a nitrenium ion from the hydroxylamine. Ford and Herman (1992) and Ford and Griffin (1992) suggest that only the final deesterification step varies significantly in energy with the structure of the product and relates to mutagenic potency. In model 9 we find that the initial step forming the hydroxylamine also contributes a modest fraction of the potency variance.

Models 7 and 8 show the similar sizeable potency variance contributions by LogP_GC and Eπ_ß when treated individually. Note in Table IV that these two predictors are extremely well correlated inversely with each other (r = -0.97).

The five models with adjusted R^2 >= 83 percent were used to predict the potency of the not yet synthesized isomer 347cmip by entering the values of the predictors for that molecule (Table VI). The predicted potencies from these five model equations were averaged and gave a value (LogMP98 = -0.828 or 0.84 rev/g) for 347cmip that is near the bottom of the range of the 11 measured isomers.

Relationships Among Predictor Variables

As shown in Table IV, ΔDipole is moderately and inversely correlated with Face,

LogP_GC, and Eπ_ß; and it is positively correlated with qNmeNH2. 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_ß. LogP_GC also correlates moderately with ElumoNH2 (inversely) and SoftNH2. Eπ_ß is very strongly correlated with b-face over c-face and shows a sizeable gap between them (Fig. 3); and it is moderately inversely correlated with ElumoNH2.

Although LogP_HL has only a borderline correlation with potency, it shows some interesting contrasts with LogP_GC, with which it is correlated at r = 0.68. In comparison with LogP_GC, LogP_HL is more strongly correlated with SoftNH2, Hyphil and FordNH+_NHOH, moderately correlated with EhomoNH2, and less well correlated with EB_ß and Face.

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, and 3) a lower calculated energy of the π electron system.

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 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 degrees 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 degrees clockwise from this line. The most mutagenically potent N3-methyl-b-Face subset average 2.99 D and 124.7 degrees clockwise, and the N3-methyl-c-Face isomers average 5.72 D and 120.0 degrees 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, Eπ_ß, SoftNH2, and negatively with ElumoNH2. 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 the method of Ghose [Ghose et al., 1988], LogP_GC (r = 0.68). However, LogP_HL is strongly correlated with SoftNH2 (r = 0.91) and inversely with Hyphil, moderately with Eπ_ß, and Face. LogP_GC is significantly correlated with potency (r = 0.74), strongly correlated with Eπ_ß (r = -0.97) and Face (r = 0.96), moderately with SoftNH2 (r = 0.75), inversely with ElumoNH2 (r = -0.68), and slightly with Nme. 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. These differences should be noted when evaluating the hydropathic properties of chemicals or drugs. LogP_AV (Table II) was found to correlate strongly with DipoleNH2, but since the calculated dipole is one component of the expression for LogP_AV this is not an independent correlation. LogKOW (Syracuse Research Corp), available from the EPA website in EpiSuite311, did not distinguish among the TMIP isomers at all, and its author Dr. W. Meylan (personal communication) confirmed that this would be expected because of the LogKOW calculation method.

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 weakly correlated with potency, SoftNH2 is rather well correlated with LogP_HL, Eπ_ß, and Face; moderately with LogP_GC; inversely with Hyphil; and weakly inversely with ElumoNH2. A high SoftNH2 implies a narrow HOMO-LUMO gap in the parent amine. The further correlation of SoftNH2 with two of the LogP calculations and inversely with Hyphil is consistent with Softness facilitating the crossing of the cell membrane.

Eπ_ß represents the simple Hückel calculation of total π-electron energy, in ß units. EB_ß is correlated strongly with Face and LogP_GC, moderately with potency (r = 0.73), LogP_HL, SoftNH2, and ElumoNH2. EB_ß 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_ß and DipoleNH2 is capable of segregating the isomers into subgroups both by Face and Nme. The relation between Eπ_ß and LogMP98 is based primarily on the separation of the b- and c-face subsets; but within

subsets there is little correlation (i.e. between isomers with different N-methyl sites).

Mechanistic Interpretations

This study is focused on finding correlations 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: including 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 any residual repair capacity of the repair-defective assay organism.

We introduce the interpretation section with a comment from co-author F.H. regarding the prevailing dogma of formation of a heterocyclic nitrenium ion as the ultimate activation step toward a genotoxic agent. 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 $2x10^4$ sec^{-1} for the 2-fluorenyl-nitrenium ion) [Anderson and Falvey, 1993], and see Novak and Rajagopal [1993] 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 formation of the bulky DNA adduct may be equally amenable to a concerted S_N2 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_N1 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 path or a surrogate for a transition state in an S_N 2 path.

Most of the predictor variables analyzed here are well correlated with mutagenic potency of the TMIP isomers: ΔDipole, qNmeNH2, DipoleNH2, Face, LogP_GC, Eπ_ß, FordNH^+_NHOH, and SoftNH2 in descending order of correlation coefficients. Within this series of predictors there are many significant pairwise correlations, 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 discovered, such as the dipole moment of the parent amine (DipoleNH2) and the location of the pyridine endocyclic nitrogen (Face) could plausibly affect the binding to the enzyme active site. The correlations discovered to other properties, such as the energy of the π electron system (Eπ_ß) and the reaction energy to form the nitrenium 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.

CONCLUSION

. It seems amazing that a dozen isomers of nearly identical structure can display a 634-fold variation in 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 twelve unique isomers of TMIP using a very wide variety of structural and electronic 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, the regression models developed in this study 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:

The authors thank Rebekah Wu for performing the Ames/*Salmonella* tests, Dr. Felice Lightstone for helpful discussion of quantum chemistry, and Dr. John Figueras for modification of his HMO software. This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48 and supported by NCI grant CA55861.

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: R.H. Adamson, J-A. Gustafsson,N. Ito, M. Nagao, T. Sugimura, K. Wakabayashi, Y. Yamazoe (Eds.), 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.

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

Alkorta I, Villar HO. 1992. Quantum mechanical parameterization of a conformationally dependent hydrophobic index. Int J Quantum Chem 44:203-218.

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 density 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 Chemistry 74:11-19.

Case RAM, Hosker M-E, McDonald DB, Pearson J.T. 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 and II. Brit J Indust Med 11:75, 213

Colvin ME, Seidl ET, Nielsen IMB, Le Bui L, Hatch FT. 1997. Deprotonation and hydride shifts in nitrenium and iminium forms of aminoimidazo-azaarene mutagens. Chem-Biol Interact 108:39-66.

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, and 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: R.H. Adamson, J-A. Gustafsson,N. Ito, M. Nagao, T. Sugimura, K. Wakabayashi, Y. Yamazoe (Eds.), 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 Molec Struct (THEOCHEM) 236:269-282.

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

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.

Fuscoe JC, Wu R, Shen NH, Healy SK, Felton JS. 1988. Base change analysis of *Salmonella hisD* 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. Computational 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: P.R. Ortiz de Montellano (Ed.), Cytochrome P450: Structure, Mechanism, and Biochemistry (2nd Ed.), 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: R.H. Adamson, J-A. Gustafsson,N. Ito, M. Nagao, T. Sugimura, K. Wakabayashi, Y. Yamazoe (Eds.), 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.

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

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

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 & Sons. p353.

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

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, Kahley MJ, Eiger E, Helmick JS, Peters HE. 1993. Reactivity and selectivity of nitrenium ions derived from ester derivatives of carcinogenic N-(4-biphenylyl)hydroxylamine and the corresponding hydroxamic acid. J Amer Chem Soc 115:9453-9460.

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

Novak M, Rajagopal S. 2001. N-arylnitrenium ions. In: Tidwell TT, Richard JP, editors. Advances in Physical Organic Chemistry. Vol. 36:167-253.

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 Ag Fd

Chem 47:1098-1108.

Parks JM, Ford GP, Cramer CJ. 2001. Quantum chemical characterization of the reactions of guanine with the phenylnitrenium ion. J Org Chem 66:8997-9004.

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. Chemical Physics 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 Heterocyclic Chem 31:1641-1645.

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

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

Turesky RJ, Guengerich FP, Guillouzo A, Langouët. 2002. Metabolism of heterocyclic amines by human hepatocytes and cytochrome P4501A2. Mutat Res 506-7:187-195.

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.

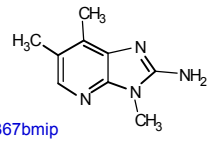
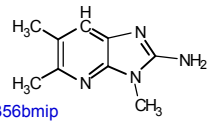
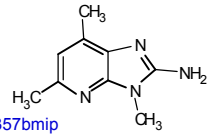
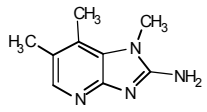
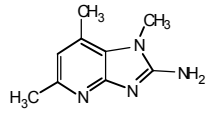
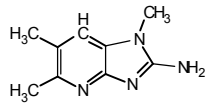
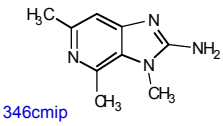
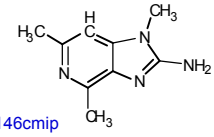
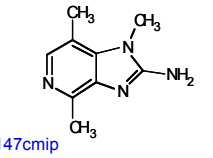
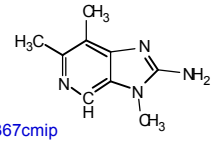
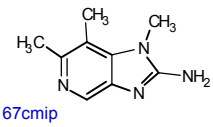
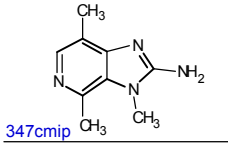
Structure/abbreviation	CAS NO.	Name	Synthesis	b_c face	Nme	LogMP98	98 rev/µg	LogMP1024
 367bmip	401560-72-3	2-Amino-3,6,7-trimethylimidazo[4,5-b]pyridine	See Methods	b	3	1.7	317.0	2.0
 356bmip	57667-51-3	2-Amino-3,5,6-trimethylimidazo[4,5-b]pyridine	[Tanga et al, 1994]	b	3	1.6	233.0	2.2
 357bmip	132898-06-7	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	[Ahlberg 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	161091-55-0	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	c	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
Face	Position of fusing imidazole and pyridine rings: face 2-3=b; face 3-4=c, shown in the name of each molecule
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
ΔE_	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]
qN_NH2	NPA charge on exocyclic N atom of parent molecule in electrons
qN_NH^+	NPA charge on exocyclic N atom of nitrenium ion in electrons
$\Delta qN\_NH^+\_NH2$	$qN(NH^+)$-qN(NH2), change of N atom charge after formation of nitrenium ion
qN3_NH2	NPA charge on N3 atom, regardless of methyl substitution
qNme_NH2	NPA charge on imidazole N atom containing the methyl substituent
Hydropathic factors	
Hyphil	Percent hydrophilic surface calculated by MMP software
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]
LogP_AV	Villar calculation of LogP from five-parameter equation 4 in [Alkorta and Villar, 1992]

Table III. TMIP Variables Database

Name	LogMP98	Log1024	Face	Nme	DipoleNH2	Dipoletau	ÆDipole	EhomoNH2	ElumoNH2	SoftNH2
367bmip	1.7471	1.9903	1	1	3.030	2.380	0.650	-0.2742	0.1407	2.4104
356bmip	1.6134	2.1845	1	1	2.990	2.474	0.516	-0.2681	0.1419	2.4392
357bmip	0.6932	1.2588	1	1	2.960	2.444	0.516	-0.2688	0.1436	2.4245
167bmip	0.2460		1	0	7.070	2.555	4.515	-0.2740	0.1361	2.4387
157bmip	0.0019	0.3142	1	0	6.020	2.370	3.650	-0.2828	0.1359	2.3888
156bmip	-0.2626	0.4391	1	0	6.390	2.348	4.042	-0.2689	0.1426	2.4307
346cmip	-0.3923	0.2098	0	1	4.850	1.046	3.804	-0.2949	0.1409	2.2943
146cmip	-0.5499	0.3763	0	0	6.450	0.450	6.000	-0.2717	0.1492	2.3758
147cmip	-0.7998	0.0242	0	0	6.780	0.130	6.650	-0.2738	0.1530	2.3431
367cmip	-1.0550	0.0911	0	1	5.670	0.390	5.280	-0.2924	0.1475	2.2731
167cmip	-1.0550	0.0721	0	0	6.870	0.304	6.566	-0.2746	0.1451	2.3825
347cmip			0	1	5.830	0.708	5.122	-0.2932	0.1522	2.2453
avgtop3					2.993	2.433	0.561	-0.2704	0.1421	2.4247
avgnext8					6.263	1.199	5.063	-0.2791	0.1438	2.3659
T-testprob					0.0000	0.0129	0.0000	0.0548	0.4768	0.0294
Correl. LogMP9	1.000	0.971	0.770	0.493	-0.799	0.785	-0.913	0.425	-0.484	0.637

Table III. TMIP Database continued

Name	Hyphil	LogP_GC	LogP_HL	Ep_§	qNnh2	qNnh+	ÆqNnh2_nh+	qNmeNH2	qNmenNH+	ÆqNmeNH+_N
367bmip	59.31	2.159	1.359	17.667	-0.9172	-0.5279	0.3893	-0.530	-0.4330	0.0970
356bmip	57.70	1.873	1.359	17.665	-0.9182	-0.5354	0.3828	-0.529	-0.4360	0.0930
357bmip	51.25	1.873	1.359	17.695	-0.9195	-0.5381	0.3815	-0.527	-0.4450	0.0820
167bmip	55.10	2.159	1.311	17.662	-0.9167	-0.5718	0.3450	-0.522	-0.4280	0.0940
157bmip	55.10	1.873	1.311	17.687	-0.9209	-0.5724	0.3485	-0.517	-0.4100	0.1070
156bmip	57.33	1.873	1.391	17.661	-0.9174	-0.5725	0.3449	-0.518	-0.4190	0.0990
346cmip	64.93	0.871	0.912	17.774	-0.9184	-0.5185	0.4000	-0.517	-0.3880	0.1290
146cmip	55.10	0.871	1.311	17.781	-0.9211	-0.5704	0.3507	-0.520	-0.4640	0.0560
147cmip	64.25	1.157	0.952	17.769	-0.9203	-0.5622	0.3581	-0.520	-0.4460	0.0740
367cmip	59.56	1.157	0.992	17.755	-0.9141	-0.5092	0.4049	-0.517	-0.3890	0.1280
167cmip	55.10	1.157	1.311	17.755	-0.9201	-0.5637	0.3564	-0.522	-0.4570	0.0650
347cmip	65.25	1.157	0.912	17.768	-0.9155	-0.5075	0.4080	-0.515	-0.3910	0.1240
avgtop3	56.087	1.968	1.359	17.676	-0.9183	-0.5338	0.3845	-0.5287	-0.43800	0.09067
avgnext8	58.310	1.390	1.186	17.731	-0.9186	-0.5551	0.3636	-0.5191	-0.42513	0.09400
T-testprob	0.4820	0.0183	0.0416	0.0260	0.7821	0.0583	0.0483	0.0003	0.2698	0.7621
Correl. LogMP9	-0.200	0.741	0.540	-0.735	0.099	0.217	0.219	-0.836	-0.112	0.053

Table IIIA. RHF Energies and Ford Calculations

Name	LogMP98	RHFnh2	RHFtau	RHFnhoh	RHFnh+	ÆEtau_nh2	ÆEnhoh_nh2	ÆEnh+_nhoh	Fordnhoh_nh2	Fordnh+_nhoh
367bmip	1.7471	-565.64641	-565.63556	-640.43652	-564.78991	0.01084	-74.79011	75.64661	-10.96762	-6.66353
356bmip	1.6134	-565.64519	-565.63537	-640.43584	-564.79708	0.00982	-74.79066	75.63876	-11.30836	-11.58885
357bmip	0.6932	-565.65023	-565.63926	-640.44147	-564.80193	0.01097	-74.79124	75.63954	-11.67608	-11.10128
167bmip	0.2460	-565.62959	-565.62539	-640.42596	-564.77416	0.00420	-74.79637	75.65180	-14.89269	-3.40926
157bmip	0.0019	-565.64009	-565.63540	-640.43007	-564.78149	0.00469	-74.78998	75.64858	-10.88228	-5.42859
156bmip	-0.2626	-565.63660	-565.63425	-640.42936	-564.78609	0.00235	-74.79276	75.64327	-12.62738	-8.76380
346cmip	-0.3923	-565.63763	-565.62850	-640.43327	-564.76357	0.00913	-74.79565	75.66970	-14.43901	7.82630
146cmip	-0.5499	-565.64085	-565.63512	-640.43605	-564.78662	0.00573	-74.79520	75.64943	-14.15976	-4.89521
147cmip	-0.7998	-565.63252	-565.62606	-640.43022	-564.77626	0.00646	-74.79770	75.65396	-15.72728	-2.05196
367cmip	-1.0550	-565.63786	-565.62596	-640.42898	-564.75993	0.01190	-74.79112	75.66905	-11.59952	7.41717
167cmip	-1.0550	-565.62915	-565.62389	-640.42641	-564.77006	0.00526	-74.79726	75.65635	-15.45306	-0.55095
347cmip		-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	75.65130	-11.31735	-9.78455
avgnxt8		-565.63554	-565.62932	-640.43004	-564.77477	0.00599	-74.79450	75.65901	-13.72262	-1.23204
T-test prob		0.0013	0.0079	0.0212	0.0044	0.00126	0.00684	0.23754	0.00684	0.01138
Correl. LogMP98	1.000	-0.684	-0.641	-0.585	-0.725	0.452	0.603	-0.376	0.602	-0.657

TMIP Table IV Correlation Matrix

_	LogMP98	Log1024	Face	Nme	DipoleNH2	Dipoletau	ÆDipole	EhomoNH2	ElumoNH2	SoftNH2	Hyphil	ÆEtau_NH2	ÆE_NHOH_NH2	ÆE_NH+_NHOH	FordNHOH_NH2	FordNH+_NHOH	LogP_GC	LogP_HL	Ep_§	qN_NH2	qN_NH+	ÆqN_NH2_NH+	qN3_NH2	qNme_NH2
LogMP98	1.0000																							
Log1024	0.9711	1.0000																						
	0.0000																							
Face	0.7701	0.7009	1.0000																					
	0.0056	0.0239																						
Nme				1.0000																				
DipoleNH2	-0.7987	-0.8876		-0.7879	1.0000																			
	0.0032	0.0006		0.0023																				
Dipoletau	0.7855	0.7113	0.9756			1.0000																		
	0.0042	0.0211	0.0000																					
ÆDipole	-0.9132	-0.8918	-0.7509	-0.5972	0.9180	-0.7953	1.0000																	
	0.0001	0.0005	0.0049	0.0403	0.0000	0.0020																		
EhomoNH2								1.0000																
ElumoNH2			-0.7354			-0.7905			1.0000															
			0.0064			0.0022																		
SoftNH2	0.6365	0.6414	0.8003			0.7240		0.8894	-0.5833	1.0000														
	0.0352	0.0456	0.0018			0.0078		0.0001	0.0465															
Hyphil								-0.6333		-0.7320	1.0000													
								0.0270		0.0068														
ÆEtau_NH2				0.9346	-0.7350							1.0000												
				0.0000	0.0065																			
ÆE_NHOH_NH2	0.6025		0.6448		-0.6713	0.6224	-0.7498						1.0000											
	0.0498		0.0236		0.0168	0.0307	0.0050																	
ÆE_NH+_NHOH			-0.6639					-0.8981		-0.8772	0.6926			1.0000										
			0.0186					0.0001		0.0002	0.0125													
FordNHOH_NH2	0.6022		0.6449		-0.6708	0.6226	-0.7496						1.0000		1.0000									
	0.0499		0.0236		0.0169	0.0306	0.0050						0.0000											
FordNH+_NHOH	-0.6566	-0.6517	-0.7694			-0.6696		-0.9177		-0.9301	0.6911			0.9082		1.0000								
	0.0282	0.0412	0.0034			0.0172		0.0000		0.0000	0.0128			0.0000										
LogP_GC	0.7408	0.7117	0.9585			0.9143	-0.6762		-0.6824	0.7458				-0.5921		-0.6768	1.0000							
	0.0091	0.0210	0.0000			0.0000	0.0158		0.0145	0.0054				0.0425		0.0156								
LogP_HL			0.7469			0.6699		0.8023		0.9140	-0.8487			-0.8489		-0.8680	0.6841	1.0000						
			0.0053			0.0172		0.0017		0.0000	0.0005			0.0005		0.0003	0.0141							
Ep_§	-0.7348	-0.7054	-0.9715			-0.9386	0.6862		0.7265	-0.7876			-0.5982	0.6090	-0.5985	0.7118	-0.9704	-0.7312	1.0000					
	0.0100	0.0227	0.0000			0.0000	0.0137		0.0075	0.0024			0.0399	0.0356	0.0398	0.0094	0.0000	0.0069						
qN_NH2																				1.0000				
qN_NH+				0.9319				-0.6430		-0.6473		0.8958		0.7758						0.6510	1.0000			
				0.0000				0.0241		0.0229		0.0001		0.0030						0.0219				
ÆqN_NH2_NH+				0.9362				-0.6377		-0.6484		0.9114		0.7697				-0.5776		0.5975	0.9977	1.0000		
				0.0000				0.0257		0.0226		0.0000		0.0034				0.0492		0.0402	0.0000			
qN3_NH2				0.9954	-0.7484							0.9353		0.5932							0.9510	0.9569	1.0000	
				0.0000	0.0051							0.0000		0.0420							0.0000	0.0000		
qNmeNH2	-0.8358	-0.9051			0.7065		0.7206	-0.6722		-0.6800						0.7001	-0.5768	-0.6117						1.0000
	0.0014	0.0003			0.0102		0.0082	0.0166		0.0150						0.0112	0.0496	0.0345						

Table V. Multiple Linear Regression Models

Model 1	**LogMP98=**	-0.3902	*DipoleNH2	+5.0146	*SoftNH2	+0.0646	*Hyphil	-35.1916	*ElumoNH2	-20.4626
Factors	Pcoeff.	0.001		0.007		0.095		0.164		
	Beta	-0.6557		0.5851		0.2722		-0.188		
	Variance Fractions:	59.8%		23.2%		3.3%		2.4%		
Adjusted R^2:	88.7%									
Fmodel:	20.75									
Pmodel:	0.0012									
Model 2	**LogMP98=**	-0.2705	*ÆDipole	-74.1787	*qNme_NH2	-37.6466				
Factors	Pcoeff.	0.004		0.053						
	Beta	-0.6472		-0.3682						
	Variance Fractions:	81.6%		5.8%						
Adjusted R^2:	87.4%									
Fmodel:	35.53									
Pmodel:	0.0001									
Model 3	**LogMP98=**	-0.3624	*DipoleNH2	+1.0049	*LogP_GC	+0.4086				
Factors	Pcoeff.	0.002		0.005						
	Beta	-0.6091		0.5178						
	Variance Fractions:	59.8%		24.0%						
Adjusted R^2:	83.8%									
Fmodel:	26.78									
Pmodel:	0.0003									
Model 4	**LogMP98=**	-0.3640	*DipoleNH2	-9.8762	*Ep_§	+176.9349				
Factors	Pcoeff.	0.002		0.006						
	Beta	-0.612		0.511						
	Variance Fractions:	59.8%		23.2%						
Adjusted R^2:	83.0%									
Fmodel:	25.44									
Pmodel:	0.0003									
Model 5	**LogMP98=**	-0.4128	*DipoleNH2	+4.1743	*SoftNH2	-17.6514				
Factors	Pcoeff.	0.001		0.007						
	Beta	-0.6937		0.487						
	Variance Fractions:	59.8%		23.2%						
Adjusted R^2:	83.0%									
Fmodel:	25.43									
Pmodel:	0.0003									
Model 6	**LogMP98=**	-0.1242	*FordNH+_NHOH	+25.5376	*ÆqN_NH+_NH2	-71.5655	*ElumoNH2	-0.3997		
Factors	Pcoeff.	0.001		0.007		0.034				
	Beta	-0.8269		0.5947		-0.3823				
	Variance Fractions:	36.8%		27.8%		15.0%				
Adjusted R^2:	79.7%									
Fmodel:	14.06									
Pmodel:	0.0024									
Model 7	**LogMP98=**	1.4378	*LogP_GC	-2.2081						
Factors	Pcoeff.	0.009								
	Beta									
	Variance Fractions:	49.9%								
Adjusted R^2:	49.9%									
Fmodel:	10.95									
Pmodel:	0.0091									
Model 8	**LogMP98=**	-14.201	*Ep_§	+251.5952						
Factors	Pcoeff.	0.010								
	Beta									
	Variance Fractions:	48.9%								
Adjusted R^2:	48.9%									
Fmodel:	10.56									
Pmodel:	0.0100									
Model 9	**LogMP98=**	-0.075	*FordNH+_NHOH	+0.2132	*FordNHOH_NH2	+2.5348				
Factors	Pcoeff.	0.08		0.137						
	Beta	-0.4995		0.4116						
	Variance Fractions:	36.8%		10.2%						
Adjusted R^2:	47.0%									
Fmodel:	5.43									
Pmodel:	0.0324									

Table VI. TMIP Regression Models for Predicting LogMP98 of 347cm

												Predictions
Model 1	**LogMP98=**	-0.3902	*DipoleNH2	+5.0146	*SoftNH2	+0.0646	*Hyphil	-35.1916	*ElumoNH2	-20.4626		-1.3602
		-0.3902	5.83	+5.0146	4.49055	+0.0646	65.25	-35.1916	0.1522	-20.4626		0.248rev/µg
Model 2	**LogMP98=**	-0.2705	*ÆDipole	-74.1787	*qNme_NH	-37.6466						-0.8302
		-0.2705	5.1223	-74.1787	-0.515	-37.6466						0.839rev/µg
Model 3	**LogMP98=**	-0.3624	*DipoleNH2	+1.0049	*LogP_GC	+0.4086						-0.5415
		-0.3624	5.83	+1.0049	1.157	+0.4086						1.63rev/µg
Model 4	**LogMP98=**	-0.3640	*DipoleNH2	-9.8762	*Ep_§	+176.9349						-0.6675
		-0.3640	5.83	-9.8762	17.768	+176.9349						1.22rev/µg
Model 5	**LogMP98=**	-0.4128	*DipoleNH2	+4.1743	*SoftNH2	-17.6514						-1.3134
		-0.4128	5.83	+4.1743	4.49055	-17.6514						0.28rev/µg
											Average	0.843rev/µg
											Pred. LogM	-0.8280

Table VI. TMIP Regression Models for Predicting LogMP98 of 347cmip

												Predictions
Model 1	**LogMP98=**	-0.3902	*DipoleNH2	+5.0146	*SoftNH2	+0.0646	*Hyphil	-35.1916	*ElumoNH2	-20.4626		-1.3602
		-0.3902	5.83	+5.0146	4.49055	+0.0646	65.25	-35.1916	0.1522	-20.4626		0.248rev/µg
Model 2	**LogMP98=**	-0.2705	*ÆDipole	-74.1787	*qNme_NH2	-37.6466						-0.8302
		-0.2705	5.1223	-74.1787	-0.515	-37.6466						0.839rev/µg
Model 3	**LogMP98=**	-0.3624	*DipoleNH2	+1.0049	*LogP_GC	+0.4086						-0.5415
		-0.3624	5.83	+1.0049	1.157	+0.4086						1.63rev/µg
Model 4	**LogMP98=**	-0.3640	*DipoleNH2	-9.8762	*Ep_§	+176.9349						-0.6675
		-0.3640	5.83	-9.8762	17.768	+176.9349						1.22rev/µg
Model 5	**LogMP98=**	-0.4128	*DipoleNH2	+4.1743	*SoftNH2	-17.6514						-1.3134
		-0.4128	5.83	+4.1743	4.49055	-17.6514						0.28rev/µg
											Average	0.843rev/µg
											Pred. LogMP	-0.8280

FIGURE LEGENDS

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.

Fig. 2. Scatter graph of the relationship between LogMP98 and ΔDipole, 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.

Fig. 3. 3-D graph showing the segregation of the mutagenic potencies of the TMIP isomers by the predictors Eπ_ß and DipoleNH2. Eπ_ß creates a major separation based on Face and DipoleNH2 creates a modest separation based on Nme position.

Fig. 4. Scatter graph of the relationship between LogMP98 and the value of FordNHOH_NH2, which is calculated as follows:

Reaction: RHF_NH2 + PhNHOH = RHF_NHOH + PhNH2

Calculation: ((RHF_NHOH + PhNH2) - (RHF_NH2 + PhNHOH))*627.51 kcal/mol

The correlation coefficient is significant at 0.646 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: RHF_NHOH + $PhNH^+$ = RHF_NH^+ + PhNHOH

Calculation: ((RHF_NH^+ + PhNHOH) - (RHF_NHOH + $PhNH^+$))*627.51 kcal/mol

The correlation coefficient is significant at 0.-657 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].

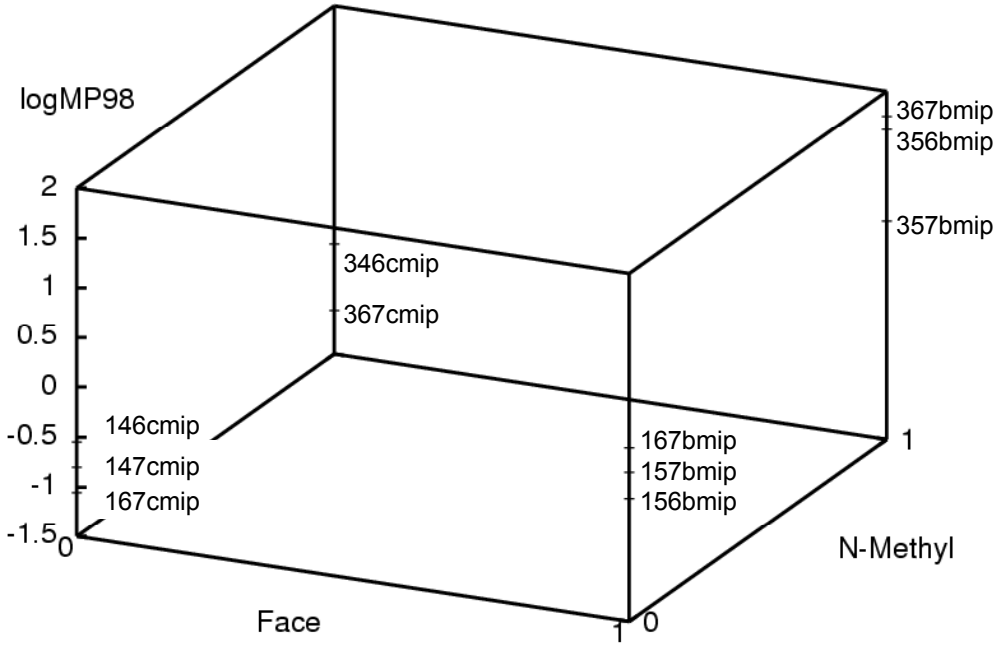

logMP98
2
1.5
1
0.5
0
-0.5
-1
-1.5
146cmip
147cmip
167cmip
346cmip
367cmip
367bmip
356bmip
357bmip
167bmip
157bmip
156bmip
0
Face
1
0
1
N-Methyl

Fig 1

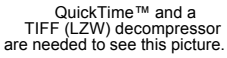

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Fig 2

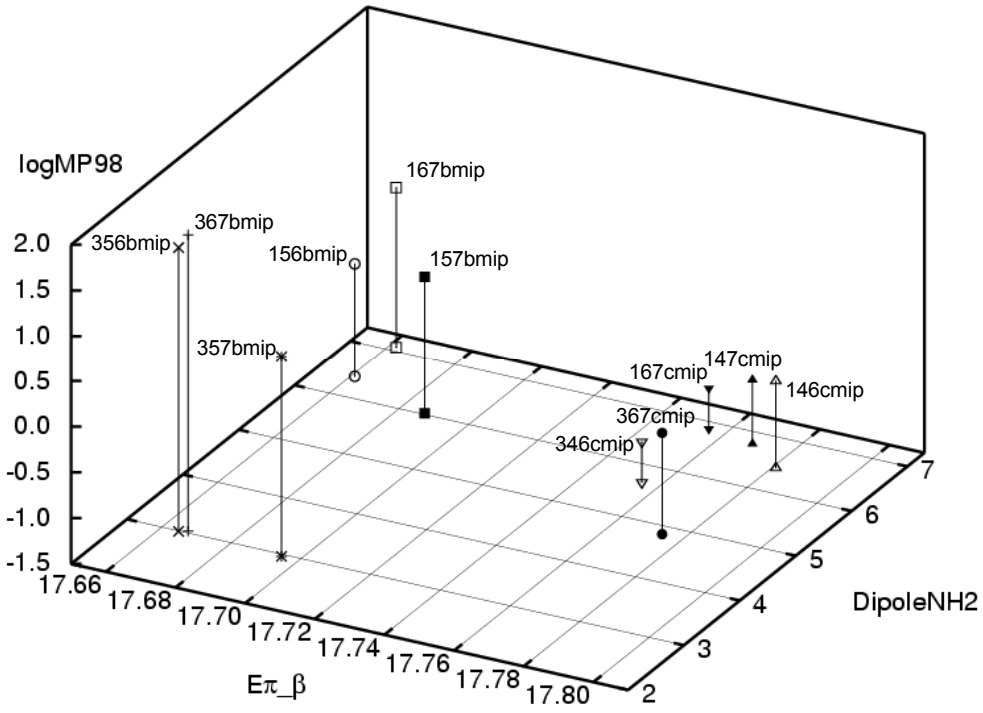

logMP98
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
356bmip
367bmip
156bmip
167bmip
157bmip
357bmip
167cmip
147cmip
146cmip
367cmip
346cmip
17.66
17.68
17.70
17.72
17.74
17.76
17.78
17.80
Eπ_β
2
3
4
5
6
7
DipoleNH2

Fig 3

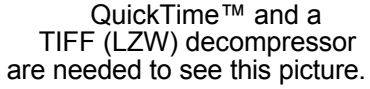
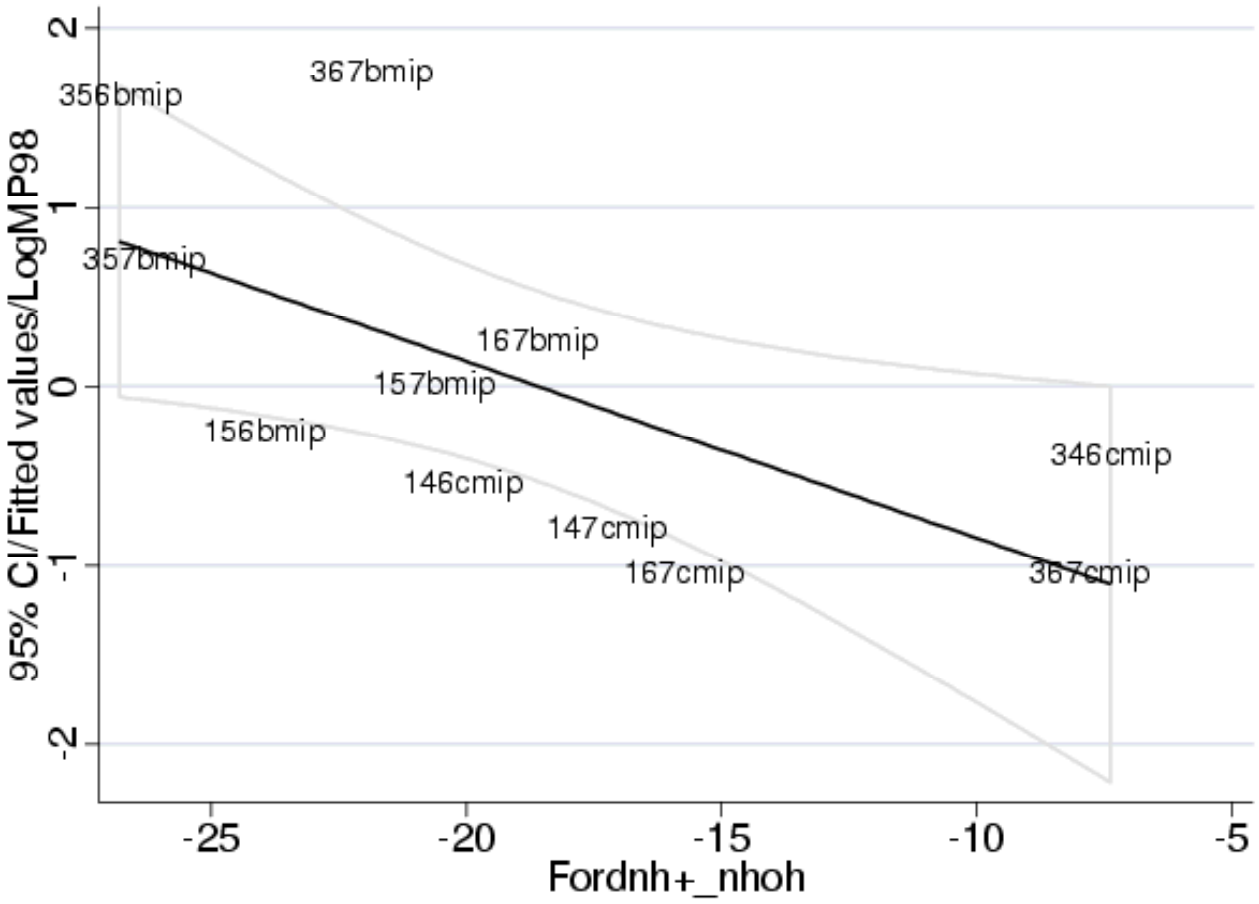

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

Fig 4

Fig 5


95% CI/Fitted values/LogMP98
2
1
0
-1
-2
-25
-20
-15
-10
-5
Fordnh+_nhoh
367bmip
356bmip
357bmip
167bmip
157bmip
156bmip
346cmip
146cmip
147cmip
167cmip
367cmip