

What are the pK_a values of C–H bonds in aromatic heterocyclic compounds in DMSO?

Kuang Shen, Yao Fu,^{*} Jia-Ning Li, Lei Liu and Qing-Xiang Guo^{*}

Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China

Received 15 October 2006; revised 6 December 2006; accepted 7 December 2006

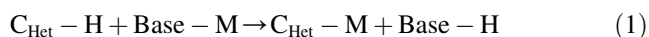
Available online 8 January 2007

Abstract—A first-principle method has been successfully developed for the prediction of pK_a values of aromatic heterocyclic compounds in DMSO solution with a precision of 1.1 pK_a units. Comparison of theoretical results and experimental data (where available) also shows excellent consistency. Armed with this useful approach, the pK_a values for a series of aromatic heterocycles were calculated in DMSO. Moreover, a discussion of the relationships between hydrogen acidities and molecular structures is conducted for the first time (determinants of C–H acidities, substituent effects, and some practical use of dehydrometalation). These statistics could be useful for synthetic chemists to design proper routes for introduction of aromatic heterocyclic moiety, especially when dehydrometalation reactions are used.
 © 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Aromatic heterocycles are present in all classes of organic compounds of interest in biology, pharmacology, optics, electronics, and material sciences.¹ Among all the possible ways of introducing an aromatic heterocyclic moiety into a complex molecule through carbon–carbon bond formation, the use of an organometallic reagent derived from direct metalation of an aromatic heterocycle is among the most versatile and efficient approaches.² Due to this reason, a variety of synthetic methods have been intensively studied for the preparation of metalated aromatic heterocycles during the past several decades. These methods include dehydrometalation, dehalometalation, transmetalation, oxidative addition, hydrometalation, carbometalation, and cross-coupling (see Fig. 1).² Of these approaches, dehydrometalation³ has attracted the most attention mainly due to the low cost and good accessibility of the starting materials (i.e., un-functionalized heterocycles) required by the reaction itself.

The general equation for the dehydrometalation reaction is shown in Eq. 1.



For dehydrometalation to proceed, a fundamental requirement is that the thermodynamic acidity of Base–H generated from Base–M (M=Li, MgX, etc.) should be lower than the acidity of $C_{\text{Het}}-H$. Thus, in order to facilitate the synthetic

studies concerning the dehydrometalation reactions of heterocycles, there is a strong need to know the acidities of the C–H bonds in various types of aromatic heterocyclic compounds in organic solutions. Unfortunately, due to the experimental difficulties (i.e., use of non-aqueous and strongly basic media, conversion of measured signals into the common pK_a scale by acidity functions, difficulty in deprotonation in specific positions), there are few direct experimental measurements of either the gas-phase or solution-phase acidities of the C–H bonds in the aromatic heterocycles.^{4,5}

Considering the difficulty in obtaining pK_a values for the C–H bonds in the aromatic heterocycles, it is necessary to develop some theoretical methods to predict these important quantities. These methods would be useful for researchers in a variety of traditional as well as emerging areas in organic chemistry ranging from base-promoted heterocycle dehydrometalation to aromatic C–H bond activation. In this

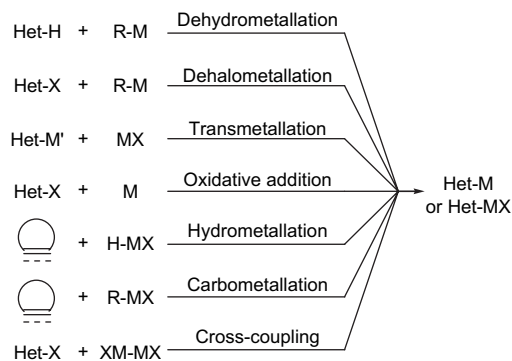


Figure 1. Approaches to produce metalated aromatic heterocycles.

^{*} Corresponding authors. Tel.: +86 551 360 7476; fax: +86 551 360 6689; e-mail addresses: fuyao@ustc.edu.cn; qxguo@ustc.edu.cn

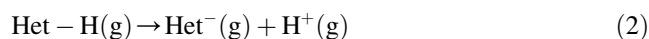
connection some attempts were made in the past to calculate the acidities of certain heterocycles in the gas phase with semi-empirical methods such as AM1 and MNDO.^{4a} However, because of the difficulty in handling the solvation effects, no theoretical method has ever been demonstrated to be able to accurately calculate the solution-phase pK_a values for the C–H bonds in aromatic heterocycles.

In the present study we develop an ab initio method that could reliably predict the pK_a values of various heterocyclic C–H bonds in DMSO.⁶ Our work took advantage of the recent developments of various continuum solvation models, which can be used to describe the solvation effects with increasing precision and efficiency.^{7,8} Previously we published a theoretical protocol, which could reach a precision of ca. ± 2.0 pK_a units in the calculation of pK_a values for over 100 organic acids in DMSO.⁹ Although this protocol could be directly used to handle aromatic heterocycles, our experience suggested that for each particular group of organic compounds the accuracy of the pK_a calculation could be significantly improved by using more strategically designed isodesmic reactions and more target-oriented calibration parameters in the DFT calculations and solvation treatments.

On the basis of the arguments above we collected available experimental data for aromatic heterocycles and used them to develop a method that would be specifically suitable for the calculation of pK_a 's of aromatic heterocycles. Our results demonstrated that by using this specifically calibrated method, the accuracy of pK_a calculation was improved to 1.1 pK_a units for the aromatic heterocycles (as compared to 2.0 pK_a units in our previous method developed for all groups of organic compounds⁹). Equipped with this protocol we systematically studied the pK_a values for many types of aromatic heterocycles for the first time. Knowledge of these pK_a values was expected to be valuable to synthetic chemists who need to design the experimental conditions for reactions that require the use of deprotonated heterocycles.













2. Gas-phase calculations

Before calculation of solution-phase pK_a values for C–H bonds in aromatic heterocyclic compounds, it is important to ascertain that the gas-phase acidity can be accurately calculated for the same bond, defined as the free energy change of the following reaction in the gas phase at 298 K, 1 atm.



To achieve this goal, we examined 12 representative compounds (see Table 1) and calculated their gas-phase C–H acidities using the standard B3LYP/6-311++G(2df,2p)//B3LYP/6-31G(d) method. Here, several aromatic compounds without heteroatoms were also included because of the similarity in aromaticity and lack of statistics. It was found that the theoretical results were systematically higher than the experimental values by 1.9 kcal/mol (see Fig. 2). This finding is not unexpected because the density functional theory (DFT) methods have been known to exhibit systematic underestimations or overestimations for many types of chemical properties.^{6j,10}

Table 1. Experimental and theoretical (B3LYP/6-311++G(2df,2p)//B3LYP/6-31G(d)) gas-phase acidities for 12 aromatic heterocycles (kcal/mol)

Compound	ΔG (exp) ⁴	ΔG (theor) before correction	ΔG (theor) ^a after correction
	380.0	383.3	381.4
	373.0	376.9	375.0
	392.9	394.1	392.2
	384.0	384.1	382.2
	373.4	374.8	372.9
	376.9	378.1	376.2
	383.1	385.2	383.3
	374.6	375.9	374.0
	378.6	381.7	379.8
	360.1	360.1	358.2
	383.8	387.1	385.2
	376.9	378.6	376.7

^a The correction method is shown in Eq. 3.

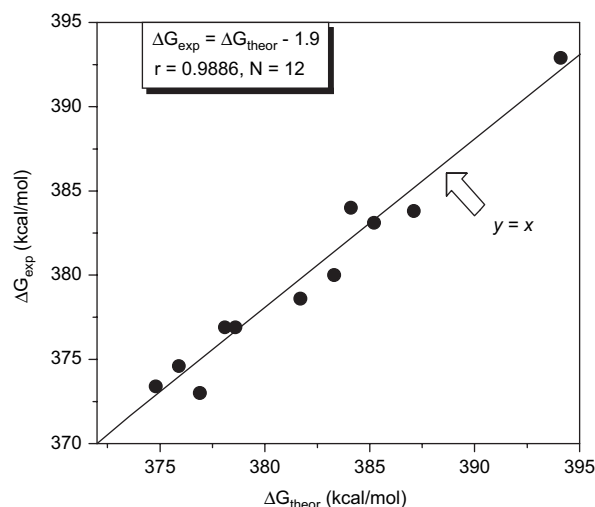


Figure 2. The correlation between the experimental and theoretical gas-phase acidities for the C–H bonds in aromatic heterocycles.

Evidently, the previously mentioned overestimation problem, if not rectified, will affect the quality of the subsequent pK_a calculations. As a result, we decide to use the following equation to calculate the gas-phase acidities:

$$\Delta G_{\text{gasphase}} = \Delta G_{\text{B3LYP}} - 1.9 \quad (3)$$

It can be seen that after the correction the predicted gas-phase acidities are in much better agreement with the experimental data. In fact, the root of mean square error (rmse) for the corrected theoretical gas-phase acidities is only 2.2 kcal/mol, which is close to the experimental error (ca. 2.0 kcal/mol) in the gas-phase acidity measurement.

3. Solution-phase calculations

With a reliable method to calculate the gas-phase acidities, we next needed to derive an equation for pK_a calculations. Despite the various solvents utilized in pK_a determination, there is an excellent linearity relationship among these different acidity scales,¹¹ with conversion of the pK_a values from one solvent to another given by Eq. 4.

$$pK_a(\text{THF}) = -0.963 + 1.046pK_a(\text{DMSO}) \quad (4)$$

Since it has been shown in the above section that the DFT methods tend to systematically overestimate or underestimate chemical properties, and the solvation energy of protons remains uncertain,¹² we decided to utilize an isodesmic reaction to improve the accuracy of the pK_a calculations. Thus, we consider the following proton-exchange reaction between furan (at its 2-position) and a target aromatic heterocyclic compound.









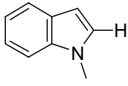
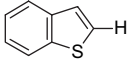
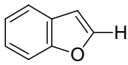
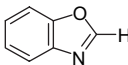
If the free energy change of the reaction above in DMSO is defined as $\Delta G_{\text{exchange}}$, the pK_a value of the target heterocycle can be calculated by Eq. 6,

$$pK_a(\text{Het-H}) = 35.0 + \frac{\Delta G_{\text{exchange}}}{2.303 \times RT} \quad (6)$$

where 35.0 is the experimental pK_a value for the C(2)–H bond of furan.⁵

According to the previous section, it is evident that the gas-phase contribution to $\Delta G_{\text{exchange}}$ can be reliably calculated, so whether Eq. 5 can successfully predict the pK_a values mainly relies on the quality of the solvation energy calculations. It is worth noticing that in previous studies we have examined the early versions of polarized continuum solvation (PCM) model in the pK_a calculations. Herein, to improve the accuracy we decided to test some more recent versions of PCM models including IEF-PCM (integral equation formalism model), C-PCM (polarizable conductor calculation model), and I-PCM (static isodensity surface polarized

Table 2. Experimental and theoretical (IEF-PCM, radii=Bondi, $f=1.20$) pK_a 's for 10 aromatic heterocycles

Compound	pK_a (exp)	pK_a (theor)	Ref.
	35.0	35.0	5a
	32.5	33.5	5a
	36.4	38.7	5a
	33.1	35.1	5a
	29.4	29.6	5b
	35.2	35.2	5a
	37.3	37.7	5a
	32.9	32.0	5a
	32.7	33.2	5b
	24.4	24.8	5b

continuum model). Two scales of atomic radii (i.e., the UA0 scale and the Bondi scale) have been tested for each PCM model. Furthermore, for the IEF-PCM/Bondi method, we have also tested different scale factors (f value ranging from 1.30 to 0.95) for the atomic radii.

Using each of the methods described above we have calculated the pK_a for 10 aromatic heterocycles whose experimental values have been reported (see Table 2). The results of these test calculations are shown in Figure 3. It is found that the IEF-PCM/Bondi method gives the lowest rmse error compared to the I-PCM, IEF-PCM/UA0, C-PCM/Bondi, and C-PCM/UA0 methods. Furthermore, for the IEF-PCM/Bondi method, the lowest rmse error is reached when the scale factor equals 1.20. As a result, the IEF-PCM/Bondi ($f=1.20$) method was concluded to be the most accurate for calculating the pK_a values of the C–H bonds in aromatic heterocycles. As shown in Figure 4, the correlation coefficient between the experimental pK_a values and the theoretical results provided by the IEF-PCM/Bondi ($f=1.20$) method is as high as 0.978. The mean error is $-0.6 pK_a$ unit and the rmse error is as low as 1.1 pK_a units.

4. Discussion

Through the above studies we have developed a first-principle method that can predict the pK_a values for C–H bonds in

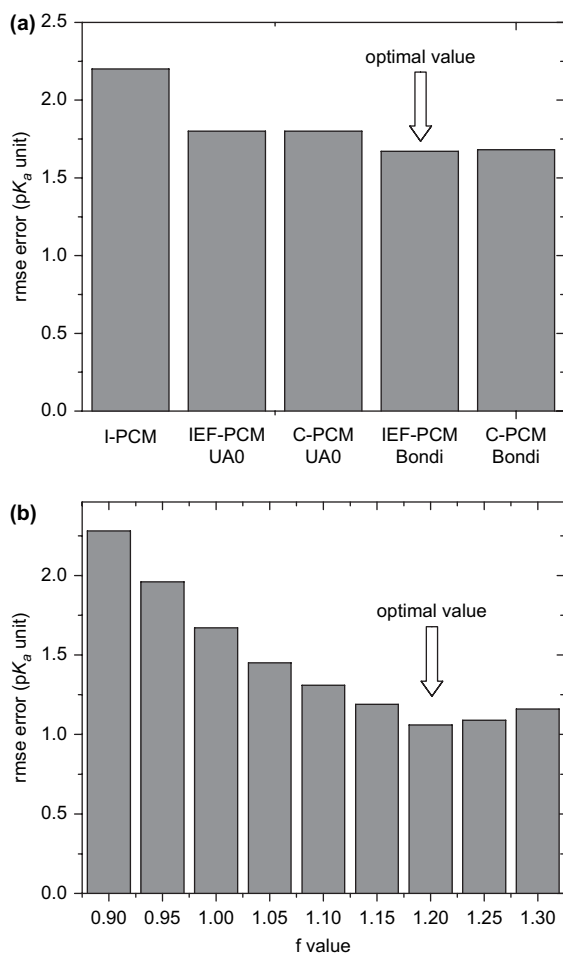


Figure 3. The dependence of the rmse error on different solvation models.

different aromatic heterocycles in DMSO. The error bar of the prediction is determined to be around 1.1 pKa units through comparison with experimental data available. Thus, the new method enabled us, for the first time, to estimate the pKa values of aromatic heterocycles in DMSO with a confident error bar.

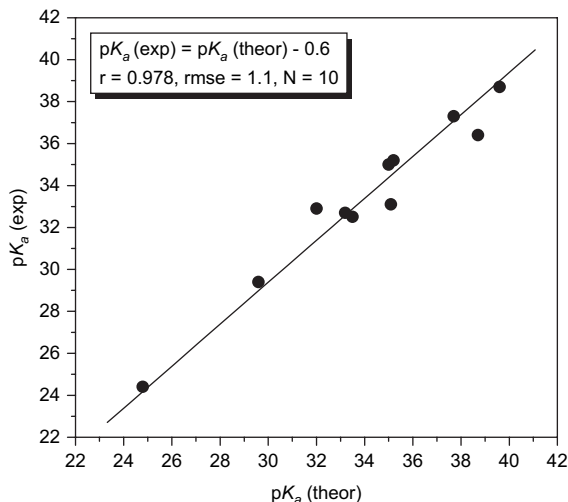


Figure 4. Correlation between the experimental and theoretical pKa values for 10 aromatic heterocycles.

Importantly, the pKa values for most C–H bonds in the aromatic heterocycles still remain unknown. As a result, the theoretical method developed here will be practically valuable for experimental chemists who need to estimate the pKa value of any heterocyclic C–H bond. Furthermore, the method developed above also allows for generation of some systematic data regarding the pKa values of heterocyclic C–H bonds in DMSO. On the basis of these data we can discuss, for the first time, some important yet unanswered questions concerning the structure–property relationships behind the acidities of the heterocyclic C–H bonds.

4.1. C–H bonds of five-membered heterocycles

Some most commonly seen five-membered heterocycles (including five-membered benzo-heterocycles) are furan, thiophene, pyrrole, pyrazole, imidazole, indole, benzofuran, and benzothiophene. These compounds are important intermediates in pharmaceutical research. Theoretical predictions of acidity are shown in Figures 5 and 6.

The acidities of five-membered heterocyclic compounds generally fall in the region from 24 to 45. The proton possessing the strongest acidity on the heterocycle is 2-H in benzo[d]-oxazole (24.8), and the weakest one is 3-H in *N*-methylpyrrole (44.8). Several overall rules can be proposed when we examine the results carefully. (1) Protons in α -position in single-heteroatom five-membered cycles usually possess stronger acidities than those in β -position by about 3–5 pKa units, thus α -Hs are more vulnerable when attacked by bases such as alkyl lithium, with formation of the corresponding carbanion. (2) Acidities of multi-heteroatom cycles are generally stronger than those of single-heteroatom ones, mainly due to the decrease of aromaticity and increase of ability in stabilizing negative charges. (3) Due to the conjugative effect of the benzene ring, benzo-coupled heterocycles generally have more acidic protons since negative electric charge can be more widely decentralized. (4) Hydrogens in benzene rings are much weaker acids than those in heterocycles.

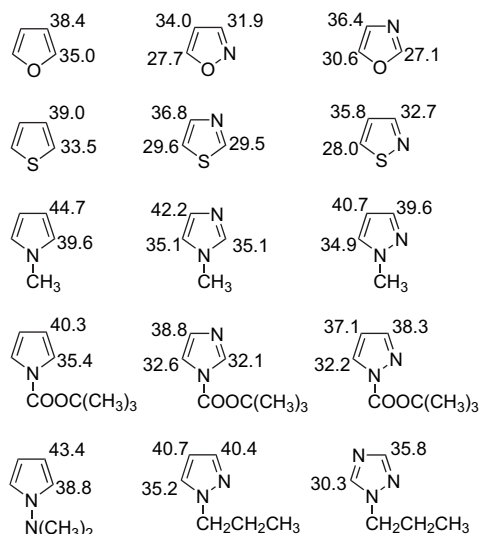


Figure 5. Theoretical prediction of acidities pKa of five-membered aromatic heterocyclic compounds.

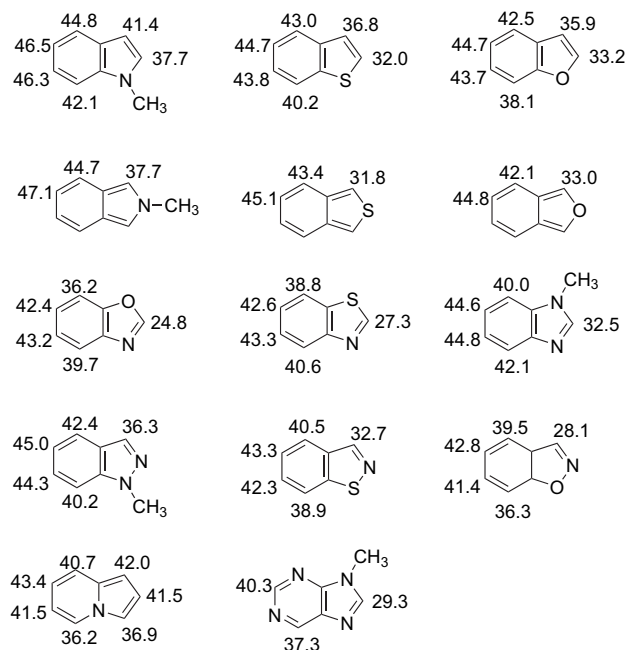


Figure 6. Theoretical prediction of acidities pK_a of five-membered aromatic benzo-heterocyclic compounds.

These rules are useful when trying to choose bases or to judge deprotonation points. For example, for protons with a pK_a of less than 30 such as 2-H in benzo[*d*]oxazole (pK_a (calcd)=24.8; pK_a (exp)=24.4), moderate bases such as lithium amide can be used, and there is no need to worry about the protons in the benzene ring since they are much weaker acids than those in heterocycles.

Since the present work deals with C–H bonds in heterocycles, nitrogen protons in pyrrole and its derivatives are substituted by a methyl group. This strategy can also be used in organic synthesis: the acidity of N–H is stronger than the C–H bond, so that *N*-alkyl-substitution must be utilized if we intend to prepare a carbanion reagent.

It is clear that the most important factor that affects acidities of C–H bonds in aromatic heterocycles is the polarity of heteroatom. Cycles containing oxygen or sulfur have significantly stronger acidities than those containing nitrogen atoms. If more than one heteroatom is included in the cycle, the proton adjacent to both heteroatoms generally possesses the strongest acidity, the one located near oxygen or sulfur atom is relatively weaker, and the protons located near the nitrogen atom the weakest. This rule can be explained by considering the different abilities of negative-charge stabilization of the different heteroatoms. According to the calculated Mulliken charges, the oxygen atom has a Mulliken charge of -0.394 and -0.470 before and after the deprotonation process (α -position in furan), while the corresponding parameter for sulfur is $+0.254$ and -0.034 (α -position in thiophene), and for nitrogen -0.361 and -0.364 (α -position in *N*-methylpyrrole). The negative charges in heterocycles are notably centralized on the oxygen and sulfur atoms, but not the nitrogen. Thus oxygen and sulfur are more able to stabilize negative charges and consequently carbanions, such that stronger acidities result.

Another rule that is worth noticing is that acidities of nitrogen-containing heterocycles will also be affected by substituent groups located on the nitrogen, especially when the electronic properties of the substituent change. For instance, if the methyl in *N*-methylpyrrole is replaced by an ester group (i.e., $-\text{COOC}(\text{CH}_3)_3$), the acidities of the corresponding sites will be enhanced greatly by about 4 pK_a units. This phenomenon can also be traced in other nitrogen-containing compounds such as *N*-substituted pyrazole and imidazole. It is generally observed that the greater the electron-withdrawing ability of the substituent group, the more acidic the heterocyclic protons will be. Yet this rule is expected when we come to the derivatives of benzene (discussed below in detail).

Combining experiments with theoretical predictions will allow for more meaningful discussion. Dehydrometalation by bases such as LDA, *t*-BuLi, etc. will result in the formation of carbanions on specific sites. The consequent product (carbanion) can react with electrophilic reagents such as aldehydes, ketones, esters, cycloxygen compounds, and so on. Some examples are as follows.

- (1) Synthesis of alcohol via direct lithiation of *N*-methylpyrrole. Here the deprotonation point supports the theoretical prediction that carbanion is formed in α -position (see Fig. 7).¹³
- (2) Application for the synthesis of bio-active molecules such as glycosides: the indole-modifying process of manose derivatives will take place according to the sequence shown in Figure 8. As expected, the lithiation will take place in nitrogen-adjacent site when R_1 is hydrogen, producing a kind of motif, which simulates a partial structure of RNase and has bio-activity.^{14,15}
- (3) The 5-position in thiazole is lithiated when the 2-position, which has the strongest acidity, has already been occupied. This further demonstrates our pK_a predictions (Fig. 9).¹⁶

A problem worth noting in geometry optimization is that some anions undergo ring-opening reactions, such as the

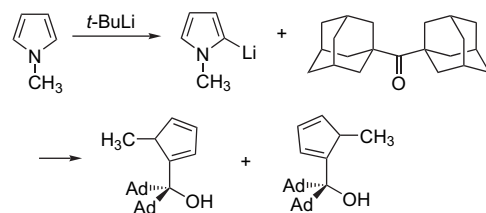


Figure 7. Synthesis of alcohol using lithiated *N*-methylpyrrole.

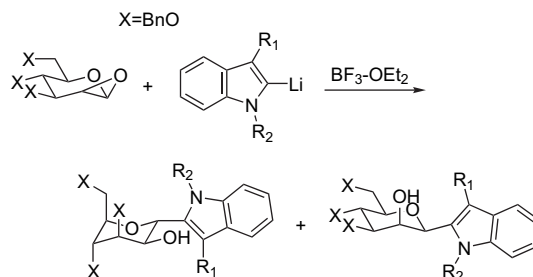


Figure 8. Indole-modification of manose.

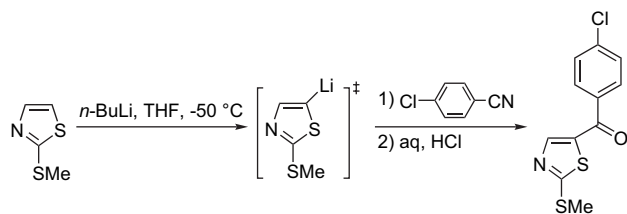


Figure 9. A secondary lithiation of thiazole.

2-anion in thiazole, 3-anion in isothiazole and isoxazole. Similarly, the corresponding benzo-heterocycle will also suffer from such problems. These compounds all contain one oxygen and one nitrogen atom. When the two heteroatoms are adjacent and the carbanion is generated in the carbon directly connected to nitrogen, the ring will open between the two heteroatoms, producing a cyanogen group and a sulfur anion (Fig. 10). The break up of isothiazole and isoxazole can be justified by experiments exactly.¹⁷ By comparison, five-membered heterocycles with more than one nitrogen atom are generally stable under similar conditions.

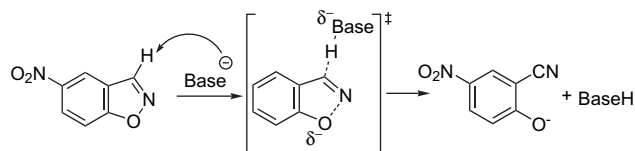


Figure 10. Ring-opening reaction of isoxazole.

4.2. C–H bonds of six-membered heterocycles

Since oxygen-family elements cannot take part in six-membered heterocycles, nitrogen and carbon are the only atoms on the ring. Theoretical calculations of pK_a of six-membered aromatic heterocycles are between 31 and 45 (see Figs. 11 and 12). The strongest acidity on heterocycles is the 3-position of pyridazine (31.1), and the weakest one is the 2-position of pyridine (43.6). The strongest acidity for the benzene ring is the 8-position of cinnoline (39.6), and the weakest one is that of benzene (44.7).

The acidities of hydrogens in benzene and pyridine is very weak (>40). Even a base as strong as sodium amide cannot deprotonate them directly, so that alkyl lithium must be used. Such a property is related to the highly-conjugated p orbitals in the ring. Consequently, if more nitrogen atoms are introduced into the ring, acidities of protons are enhanced due to the decrease in aromaticity. Calculation shows that pK_a of diazine generally falls in the region of 35–40, and that of triazine 31–35. Bases such as LiCHA or CsCHA ($pK_a(\text{HCHA}) \approx 40$) will be sufficient for deprotonation.

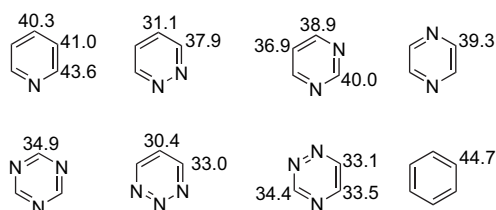


Figure 11. Theoretical prediction of acidities pK_a of six-membered aromatic heterocyclic compounds.

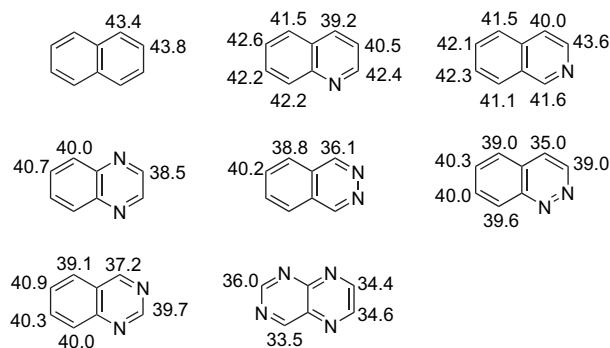


Figure 12. Theoretical prediction of acidities pK_a of six-membered aromatic benzo-heterocyclic compounds.

A crucial difference from five-membered heterocycles is that protons adjacent to heteroatom always have the weakest acidity, rather than the strongest. This is mainly because of the high repulsion effect between two electron clouds: the lone electron pair of nitrogen and the negative charges formed by carbanion. The angle between the two electron clouds in α -carbanions in six-membered rings ($\sim 60^\circ$) is smaller than that for five-membered carbanions ($\sim 72^\circ$). This means a higher static electricity repulsion will destabilize the carbanion, explaining the weaker acidity. Some experimental results support this reasoning: deprotonating pyridine with LDA or LTMP and without specific conditions will afford exclusive 3-position products.¹⁸ If 2- or 4-position anions are expected, halogen-exchange reaction must be employed.

The pK_a calculated for diazine agrees well with experimental observations.¹⁹ For instance, 1,3-diazine is deprotonated at the 5-position, which supports the calculation (Fig. 13).

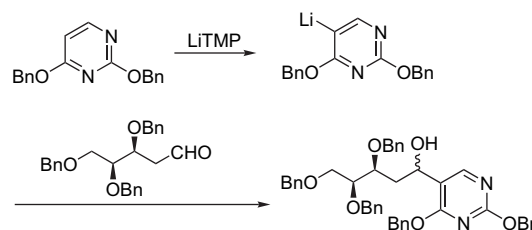


Figure 13. Theoretical prediction of deprotonation point in diazine agrees well with experiment.

4.3. Pyrimidine, purine, and pteridine

As the main structure of many bio-active molecules, pyrimidine, purine, and pteridine draw increasing interest from researchers. Here we discuss their properties (Fig. 14).

Pyrimidine is a diazine. The two sp^2 hybridized nitrogen atoms have little basicity. Acidities of C–H bonds in

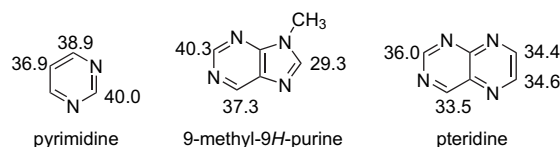


Figure 14. Theoretical prediction of acidities pK_a of pyrimidine, purine, and pteridine.

pyrimidine are stronger than benzene and pyridine because of deficiency of aromaticity. Purine similarly possesses acidities and basicities at the same time. If the nitrogen atom is not capped with methyl, the acidity of N–H is even stronger than phenols, but this is beyond our discussion. Bases such as LDA and NaNH_2 will be able to deprotonate the hydrogen in the five-membered ring without affecting those in the six-membered ring. Pteridine however, has C–H bonds with similar $\text{p}K_a$ (all between 33 and 36), and this will probably lead to a mixture when deprotonating.

4.4. The *ortho* effect

With the help of catalyst or specific groups substituted in heterocycle, *ortho* effect may occur, which means a direct anion formation on the carbon adjacent to the heteroatom regardless of the $\text{p}K_a$ calculated. Reactions using these kinds of effects are also called DoM (Direct *ortho*-Metalation) reaction, which enables us to introduce functional groups or substituents in aromatic heterocycles, sometimes even disobeying the common rules of metalation.

These effects show no contradiction to theoretical predictions for five-membered heterocycles in that hydrogens adjacent to heteroatoms are always the stronger acids, but in six-membered rings, such a reaction is especially useful because protons adjacent to the nitrogen possess the weakest acidities. For example, the *ortho*-proton has the weakest acidity in pyridine, and lithiation reaction with LDA and LTMP also supports this. However, if specific bases (BuLi – $\text{Me}_2\text{N}(\text{CH}_2)_2\text{OLi}$) are used, *ortho*-product will be obtained. The mechanism of this reaction predicts that a lithium complex is formed, which favors the *ortho*-position, such that the common rules are disobeyed (see Fig. 15).^{20b}

The same phenomenon can also be seen in reactions of azine. Protons having lower acidity but located adjacent to nitrogen are substituted (Fig. 16).^{20a,c}

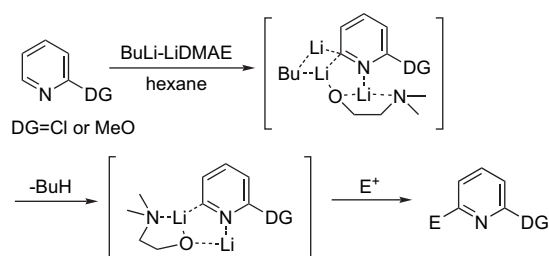


Figure 15. Mechanism of a DoM reaction.

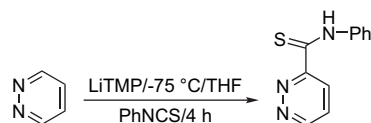


Figure 16. Another example of DoM reactions.

Therefore, a proper utilization of DoM reactions will simplify the process of synthesis when necessary.

4.5. The remote effect

Due to the similarity of aromatic heterocycles and benzene derivatives, we further calculated acidities of a series of

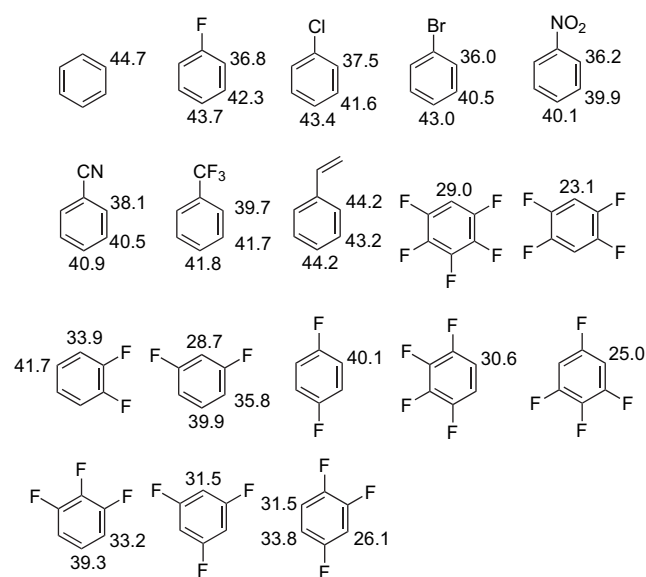


Figure 17. Theoretical prediction of acidities $\text{p}K_a$ of benzene derivatives.

benzene with different substituents. Using methods similar to Hammett's dealing with relationships between substituents and reaction rates or equilibrium constants, we obtained a quantitative result (Fig. 17).²¹

Since the parameter σ in Hammett equation is only determined by the properties and locations of substituents rather than specific reactions, and assuming that the dissociation constants of aromatic heterocycles also have the same relationship (most probably linear), we conducted the linear regression of $\text{p}K_a$ versus σ as below.

A clear trend of linearity between $\text{p}K_a$ and σ , although not perfect, can be seen from the graph with $r = -0.934$, and the large slope (-5.74) suggests a crucial effect of substituents on $\text{p}K_a$. Such linearity shows substituent effects on C–H acidities, where the relatively low regression coefficient may result from the difference between conditions in the Hammett equation and in our system (Fig. 18, Table 3).

We only discuss the *meta*- and *para*-substitutions of benzene rather than the *ortho*-ones because the substituent effect of *meta*- and *para*-groups are mainly determined by changes in

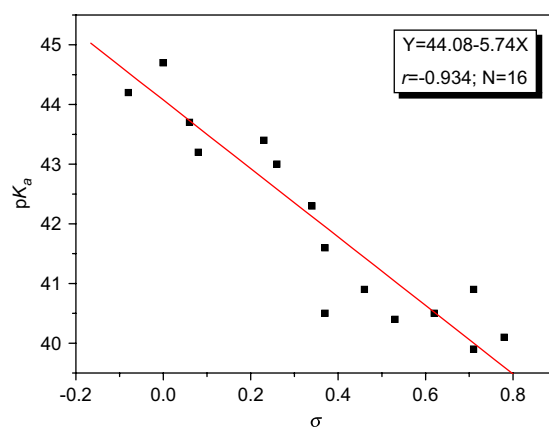


Figure 18. Linear regression of $\text{p}K_a$ versus σ .

Table 3. List of substituent effect constant σ and calculated pK_a 's

Substituent X	σ_X	pK_a	Substituent X	σ_X	pK_a
<i>m</i> -H	0	44.7	<i>p</i> -CH=CH ₂	-0.08	44.2
<i>m</i> -CH=CH ₂	0.08	43.2	<i>p</i> -H	0	44.7
<i>m</i> -F	0.34	42.3	<i>p</i> -F	0.06	43.7
<i>m</i> -Cl	0.37	41.6	<i>p</i> -Cl	0.23	43.4
<i>m</i> -Br	0.37	40.5	<i>p</i> -Br	0.26	43.0
<i>m</i> -CF ₃	0.46	40.9	<i>p</i> -CF ₃	0.53	40.4
<i>m</i> -CN	0.62	40.5	<i>p</i> -CN	0.71	40.9
<i>m</i> -NO ₂	0.71	39.9	<i>p</i> -NO ₂	0.78	40.1

enthalpy, but that of *ortho*- groups are simultaneously affected by enthalpy and entropy. According to the Gibbs free energy equation, ΔG (represented by pK_a) is linear to entropy when enthalpy remains constant. Therefore, *para*- and *meta*-substitutions may show this linearity, but not *ortho*- ones.

5. Summary

A first-principle theoretical protocol was developed successfully to predict the pK_a values of a number of aromatic heterocycles in DMSO with a precision of about 1.1 pK_a units. The theoretical predictions were found to be consistent with experimental data available. With this protocol, systematic studies of diverse types of aromatic heterocyclic compounds in DMSO were conducted. Thus, a scale of reliable pK_a values was constructed for the first time for aromatic heterocycles. These pK_a values will be helpful to synthetic chemists who need to design experimental conditions for handling deprotonated aromatic heterocycles. On the basis of these pK_a values a discussion on the relationships between hydrogen acidities and molecular structures was presented. Determination of C–H acidities, substituent effects, and some practical use of dehydrometalation were discussed.

6. Computational methodology

The theoretical calculations were conducted with the Gaussian 03 package.²² HF and B3LYP method with various basis sets (i.e., 6-31G(d), 6-31++G(2df,2p)), different solvation models (i.e., IEF-PCM, C-PCM, I-PCM, COSMO-RS), and varying cavity models (i.e., UA0, BONDI) and electrostatic scaling factors (0.90–1.30) were systematically utilized and compared. Finally the gas-phase energy calculations were conducted using the standard B3LYP/6-311++G(2df,2p)//B3LYP/6-31G(d) method. The PCM solvation model was used in its integral equation formalism (IEF-PCM) to calculate the solvation free energies in DMSO. All the IEF-PCM calculations were performed at B3LYP/6-311++G(2df,2p) level (Icomp=4, TSNUM=60, TSARE=0.4, radii=bondi, alpha=1.20).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 20602034). We also thank the USTC Supercomputer Center.

Supplementary data

The Cartesian coordinates of the optimized molecules, the calculated electronic energies, thermal corrections to Gibbs free energies, and solvation free energies. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.tet.2006.12.032.

References and notes

- (a) Joule, J. A.; Mills, K. *Heterocyclic Chemistry*; Blackwell Science: Oxford, 2000; (b) Katritzky, A. R.; Pozharskii, A. F. *Handbook of Heterocyclic Chemistry*; Pergamon: Oxford, 2003; (c) Xu, Y.; Guo, Q.-X. *Heterocycles* **2004**, *63*, 903–974.
- Chinchilla, R.; Najera, C.; Yus, M. *Chem. Rev.* **2004**, *104*, 2667–2722.
- (a) Mallan, J. M.; Bebb, R. L. *Chem. Rev.* **1969**, *69*, 693–755; (b) Marsais, F.; Queguiner, G. *Tetrahedron* **1983**, *39*, 2009–2011; (c) Snieckus, V. *Chem. Rev.* **1990**, *90*, 879–933; (d) Ramon, D. J.; Yus, M. *Eur. J. Org. Chem.* **2000**, *65*, 225–237; (e) Xu, B.; Ma, S.-M. *Chin. J. Org. Chem.* **2000**, *20*, 54–63; (f) Mongin, F.; Queguiner, G. *Tetrahedron* **2001**, *57*, 4059–4090; (g) Turck, A.; Ple, N.; Mongin, F.; Queguiner, G. *Tetrahedron* **2001**, *57*, 4489–4505.
- Gas-phase acidities of aromatic heterocycles: (a) DePuy, C. H.; Kass, S. R.; Bean, G. P. *J. Org. Chem.* **1988**, *53*, 4427–4433; (b) Meot-Ner, M.; Liebman, J. F.; Kafafi, S. A. *J. Am. Chem. Soc.* **1988**, *110*, 5937–5941; (c) Meot-Ner, M.; Kafafi, S. A. *J. Am. Chem. Soc.* **1988**, *110*, 6297–6303; (d) Davico, G. E.; Bierbaum, V. M.; Depuy, C. H.; Ellison, G. B.; Squires, R. R. *J. Am. Chem. Soc.* **1995**, *117*, 2590–2599; (e) Reed, D. R.; Kass, S. R. *J. Mass Spectrom.* **2000**, *35*, 534–539.
- pK_a of aromatic heterocycles: (a) Fraser, R. R.; Mansour, T. S.; Savard, S. *Can. J. Chem.* **1985**, *63*, 3505–3509; (b) Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 456–463 and references therein.
- Other theoretical studies: (a) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1991**, *95*, 10549–10551; (b) Smith, B. J.; Radom, L. *J. Am. Chem. Soc.* **1993**, *115*, 4885–4888; (c) Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1994**, *231*, 345–351; (d) Ochterski, J.; Petersson, G. A.; Wiberg, K. B. *J. Am. Chem. Soc.* **1995**, *117*, 11299–11308; (e) Bagno, A.; Scorrano, G. *J. Phys. Chem.* **1996**, *100*, 1536–1544; (f) Martin, J. M. L.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 136–143; (g) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764–7776; (h) Peterson, K. A.; Xantheas, S. S.; Dixon, D. A.; Dunning, T. H., Jr. *J. Phys. Chem. A* **1998**, *102*, 2449–2454; (i) Hammerum, S. *Chem. Phys. Lett.* **1999**, *300*, 529–532; (j) Burk, P.; Koppel, I. A.; Koppel, I.; Leito, I.; Travnikova, O. *Chem. Phys. Lett.* **2000**, *323*, 482–489; (k) Seo, Y.; Kim, Y.; Kim, Y. *Chem. Phys. Lett.* **2001**, *340*, 186–193; (l) Pokon, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, *105*, 10483–10487; (m) Ervin, K. M.; DeTuri, V. F. *J. Phys. Chem. A* **2002**, *106*, 9947–9956; (n) Wiberg, K. B. *J. Org. Chem.* **2002**, *67*, 4787–4794; (o) Remko, M. *J. Phys. Chem. A* **2002**, *106*, 5005–5010; (p) Yi, H.-B.; Duan, X.-H.; Lee, J.-Y.; Li, X.-Y.; Kim, K.-S. *J. Chem. Phys.* **2003**, *119*, 8854–8863; (q) Li, X.-Y.; Fu, K.-X. *J. Comput. Chem.* **2004**, *25*, 500–509; (r) Li, X.-Y.; Fu, K.-X.; Zhu, Q.; Shan, M.-H. *J. Comput. Chem.* **2004**, *25*, 835–842; (s) Zhu, Q.; Fu, K.-X.; Li, X.-Y.; Gong, Z.; Ma, J.-Y. *J. Comput. Chem.* **2005**, *26*, 399–409.

7. PCM-model theory: (a) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094; (b) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200; (c) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3093.
8. Application of PCM-model: (a) Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. *J. Phys. Chem. A* **2000**, *104*, 9619–9624; (b) Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A., III. *J. Phys. Chem. A* **2001**, *105*, 274–280; (c) Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327–1335; (d) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010–5018; (e) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963–11972.
9. Our previous study: (a) Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126*, 814–822; (b) Li, J.-N.; Liu, L.; Fu, Y.; Guo, Q.-X. *Tetrahedron* **2006**, *62*, 4453–4462; (c) Li, J.-N.; Fu, Y.; Liu, L.; Guo, Q.-X. *Tetrahedron* **2006**, *62*, 11801–11813.
10. (a) Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Chem. Soc., Perkin Trans. 2* **1999**, 2305–2313; (b) Henry, D. J.; Parkinson, C. J.; Mayer, P. M.; Radom, L. *J. Phys. Chem. A* **2001**, *105*, 6750–6756; (c) Song, K.-S.; Cheng, Y.-H.; Fu, Y.; Liu, L.; Guo, Q.-X. *J. Phys. Chem. A* **2002**, *106*, 6651–6658.
11. Streitwieser, A.; Wang, D. Z.; Stratakis, M.; Facchetti, A.; Gareyev, R.; Abbotto, A.; Krom, J. A.; Kilway, K. V. *Can. J. Chem.* **1998**, *76*, 765–769.
12. Uncertainty of solvation energy of H⁺: (a) Reiss, H.; Heller, A. *J. Phys. Chem.* **1985**, *89*, 4207–4213; (b) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610–5620; (c) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102*, 7787–7794.
13. Lomas, J. S.; Vaissermann, J. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1777–1781.
14. Manabe, S.; Ito, Y. *J. Am. Chem. Soc.* **1999**, *121*, 9754–9755.
15. Matsuzono, M.; Fukuda, T.; Iwao, M. *Tetrahedron Lett.* **2001**, *42*, 7621–7623.
16. Marcantonio, K. M.; Frey, L. F.; Murry, J. A.; Chen, C.-Y. *Tetrahedron Lett.* **2002**, *43*, 8845–8848.
17. Na, J.; Houk, K. N.; Hilvert, D. *J. Am. Chem. Soc.* **1996**, *118*, 6462–6471 and references therein.
18. (a) Comins, D.; La Munyon, D. *Tetrahedron Lett.* **1988**, *29*, 773–776; (b) Trécourt, F.; Mallet, M.; Marsais, F.; Quéguiner, G. *J. Org. Chem.* **1988**, *53*, 1367–1371; (c) Comins, D.; Bavevsky, M.; Hong, H. *J. Am. Chem. Soc.* **1992**, *114*, 10971–10972.
19. Kim, D. C.; Yoo, K. H.; Kim, D. J.; Chung, B. Y.; Park, S. W. *Tetrahedron Lett.* **1999**, *40*, 4825–4828.
20. (a) Gros, P.; Fort, Y.; Queguiner, G.; Caubere, P. *Tetrahedron Lett.* **1995**, *36*, 4791–4794; (b) Gros, P.; Choppin, S.; Fort, Y. *J. Org. Chem.* **2003**, *68*, 2243–2247; (c) Fruit, C.; Turck, A.; Ple, N.; Mojovic, L.; Queguiner, G. *Tetrahedron* **2002**, *52*, 2743–2753.
21. (a) Hammett, L. P. *Chem. Rev.* **1935**, *17*, 125–136; (b) Hansch, C.; Gao, H. *Chem. Rev.* **1997**, *97*, 2995–3060.
22. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C. 02*; Gaussian: Pittsburgh, PA, 2003.