Estimation of Molecular Acidity via Electrostatic Potential at the Nucleus and Valence Natural Atomic Orbitals

Shubin Liu*,† and Lee G. Pedersen*,‡

Research Computing Center, University of North Carolina, Chapel Hill, North Carolina 27599-3420, and Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received: December 19, 2008; Revised Manuscript Received: February 26, 2009

An effective approach of estimating molecular pK_a values from simple density functional calculations is proposed in this work. Both the molecular electrostatic potential (MEP) at the nucleus of the acidic atom and the sum of valence natural atomic orbitals are employed for three categories of compounds, amines and anilines, carbonyl acids and alcohols, and sulfonic acids and thiols. A strong correlation between experimental pK_a values and each of these two quantities for each of the three categories has been discovered. Moreover, if the MEP is subtracted by the isolated atomic MEP for each category of compounds, we observe a single unique linear relationship between the resultant MEP difference and experimental pK_a data of amines, anilines, carbonyl acids, alcohols, sulfonic acids, thiols, and their substituents. These results can generally be utilized to simultaneously estimate pK_a values at multiple sites with a single calculation for either relatively small molecules in drug design or amino acids in proteins and macromolecules.

I. Introduction

Knowledge of pK_a values, the acid−base dissociation constant, as a measure of the strength of an acid or a base, is essential for the understanding and quantitative treatment of acid−base processes in solution, and is relevant in chemical synthesis, pharmacokinetics, drug design and metabolism, toxicology, and environmental protection. There has been an immense interest in the literature to develop new and reliable models to predict and estimate pK_a values with approaches using ab initio, density functional theory, molecular modeling, and statistical methods.[1−12]

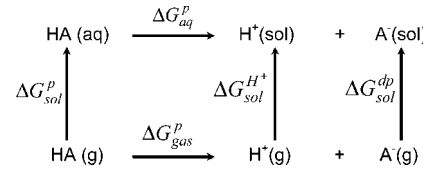
To compute accurate pK_a values according to the thermodynamic cycle (Scheme 1) using ab initio and DFT methods is a challenging task for large systems such as proteins and DNA because the simulations must be carried out in solution. According to the cycle, a number of free energy changes must be simulated:[1,13]

$$2.303RT\cdot \mathrm{p}K_a = \Delta G^{\mathrm{p}}_{\mathrm{aq}} = \Delta G^{\mathrm{dp}}_{\mathrm{sol}} + \Delta G^{\mathrm{H}^+}_{\mathrm{sol}} - \Delta G^{\mathrm{p}}_{\mathrm{sol}} + \Delta G^{\mathrm{p}}_{\mathrm{gas}} \quad (1)$$

where R is the Rydberg gas constant and T is the temperature. $\Delta G^{\mathrm{p}}_{\mathrm{aq}}$ is the sum of the free energy of deprotonation of the gas-phase species $\Delta G^{\mathrm{p}}_{\mathrm{gas}}$, the free energies of desolvation of the protonated form $-\Delta G^{\mathrm{p}}_{\mathrm{sol}}$, and solvation of the deprotonated form $\Delta G^{\mathrm{dp}}_{\mathrm{sol}}$, and the free energy of solvation for the proton $\Delta G^{\mathrm{H}}_{\mathrm{sol}}$. For large systems, ab initio simulations are still difficult even with the fastest software and hardware.

Much recent attention has been devoted to seeking statistical correlations of pK_a values with quantum descriptors such as highest occupied molecular orbital (HOMO) energies,[14] localized reactive orbital, frontier effective-for-reaction MOs (FERMO),[10] electrophilicity or group philicity,[15,16] etc. These relationships originated from the idea that proton or electron donor−acceptor reactions are driven by frontier molecular orbitals such as HOMO. However, the relations found were often only applicable within the same family of compounds like phenols, anilines, and azines.

SCHEME 1



It is our belief that molecular acidity is a property localized to the particular acidic atom and that the impact of the environment is reflected through the changes to that atom. The localized quantities that are relevant to the acidity of the given non-hydrogen acidic atom should be of either electrostatic or quantum nature, or both. In this work, we use two interdependent quantum descriptors to effectively and simultaneously estimate molecular pK_a values for amines, anilines, carbonyl acids, alcohols, sulfonic acids, thiols, and their substituents. The two quantum descriptors are molecular electrostatic potential (MEP) on the acidic atom, MEP at N, O, or S nucleus, and the sum of the valence p natural atomic orbitals, NAO, of the atom. Using MEP, or closely related quantities, to estimate pK_a values[3,17−20] and other properties[21,22] has a long history in the literature, and frontier orbitals such as FERMO have also been employed in predicting acidity.[10] To the best of our knowledge, however, this is the first time that quantum descriptors such as MEP at the acidic nucleus and NAO are introduced generally in pK_a estimation and that the interdependence of these two quantities is revealed. In addition, these descriptors are applied to simultaneously estimate pK_a values for more than one category of compounds at more than one atom type site.

* Corresponding authors. E-mail: shubin@email.unc.edu (S.L.); lee_pedersen@unc.edu (L.G.P.).
† Research Computing Center.
‡ Department of Chemistry.

10.1021/jp811250r CCC: $40.75
Published on Web 03/24/2009

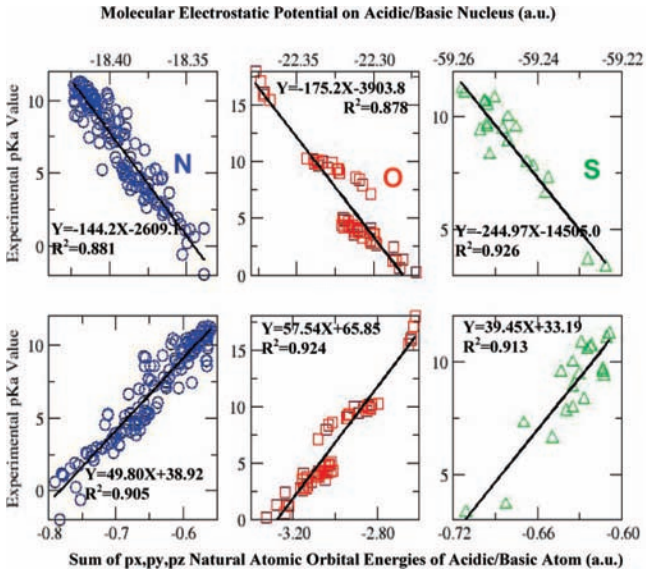


Figure 1. Linear relationships between molecular electrostatic potential on acidic nucleus and experimental pK_a values for amines (N), carboxylic acids and alcohols (O), and sulfonic acids and thiols (S) (upper panel); and linear relationships between the sum of three valence NAO 2p/3p orbitals and pK_a values (lower panel). See text for calculation details.

II. Computational Details

A total of 228 molecular systems (154 primary, secondary, and tertiary amines and anilines, 59 carboxylic acids and alcohols, and 15 sulfonic acids and thiols) have been investigated. A full structure optimization was first carried out at the DFT B3LYP/6-311+G(2d,2p) level. When a molecule has more than one stable conformation, all conformers will be examined, and the one with the lowest energy will be employed in the subsequent calculations. After structure optimization, single point calculations are performed to obtain the molecular electrostatic potential on each of the nuclei followed by a full NBO[23] analysis. We obtained the initial structure, and experimental pK_a values are from the literature.[24−31] To test the validity and applicability of the relationships presented in the text to other approaches, we also performed the same calculations with the Hartree−Fock method. We examined the results with the inclusion of the solvent effect in terms of the implicit PCM (Polarizable Continuum) model. All calculations are performed with the Gaussian 03 package[32] with tight self-consistent field convergence and Ultrafine integration grids.

III. Results and Discussion

Figure 1 exhibits linear relationships between experimental pK_a values and each of the two quantities for three categories of compounds, amines and anilines (N, blue color), carboxylic acids and alcohols (O, red color), and sulfonic acids and thiols (S, green color), as well as their derivatives. Their respective data of MEP and NAO are shown in Tables 1−3. It is seen that a reasonable linear relationship is obtained for each category of compounds for each of the two quantum descriptors, giving the correlation coefficient R^2 of 0.881, 0.878, and 0.926 for N-, O-, and S-containing compounds, respectively, from the MEP versus pK_a plot, and $R^2 = 0.905$, 0.924, and 0.913 for N-, O-, and S-containing compounds, respectively, from the NAO versus pK_a plot. An average correlation coefficient of 0.904 is observed from these correlations.

Moreover, if one given number, the MEP evaluated for the isolated neutral acidic atom, is subtracted from the MEP value

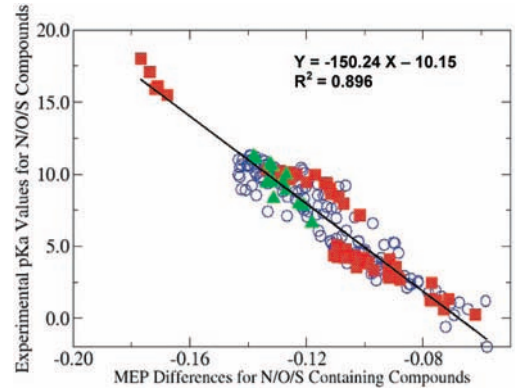


Figure 2. Linear relationship between the MEP difference and experimental pK_a values for all 228 data points. The MEP reference values for N, O, and S compounds are −18.28, −22.20, and −59.12 au, respectively. Symbols: N, blue ○; O, red ■; S, green ▲.

on the acidic nucleus for each of the three categories of compounds, and then all MEP differences of the three categories are plotted together against the experimental pK_a data, one single linear relationship, as shown Figure 2, is obtained with the correlation coefficient $R^2 = 0.896$. The aforementioned reference MEP value (isolated atoms of N, O, and S) employed in this work is −18.28 au for amine and aniline compounds, −22.20 au for carboxylic acids and alcohols, and −59.12 au for sulfonic acids and thiols.

The universality of the above linear relationship between the MEP difference [MEP (in molecule) − MEP (neutral isolated atom)] and pK_a values for different kinds of compounds can be understood in this manner. The molecular electrostatic potential on a nuclear R_A can be expressed as follows:

$$V_{R_A} = \sum_{i \neq A} \frac{Z_i}{|R_i - R_A|} - \int \frac{\rho(r)}{|r - R_A|} d\tau \qquad (2)$$

This quantity is system dependent because it is a function of $\{Z_i\}$. However, if one uses the sum of atomic electron densities as the zeroth-order approximation for the total molecular electron density, plus a local environment dependent correction:

$$\rho(r) = \sum_i \rho_i^0(r - R_i) + \sum_i f_i(|r - R_i|, \mathrm{NAO}_{v,i}) \qquad (3)$$

and inserts it into the MEP formula, the first term of the MEP can be arranged to cancel approximately, leaving the correction term dependent only on the local environment of the nucleus. To demonstrate, let us rewrite eq 3 as

$$\rho(r) = \rho_A^0(r - R_A) + \sum_{i \neq A} \rho_i^0(r - R_i) + \sum_i f_i(|r - R_i|, \mathrm{NAO}_{v,i}) \qquad (4)$$

With eq 4, we have

$$\int \frac{\rho(r)}{|r - R_A|} d\tau = \int \frac{\rho_A^0(r)}{|r - R_A|} d\tau + \sum_{i \neq A} \frac{Z_i}{|R_i - R_A|} + \sum_i \int \frac{g_i(|r - R_i|, \mathrm{NAO}_{v,i})}{|r - R_A|} d\tau \qquad (5)$$

TABLE 1: Molecular Electrostatic Potential on the Acidic Atom Nucleus, Experimental pK_a Data, and Valence NAO Energies for Amines and Anilines (N-Containing) Calculated at the B3LYP/6-311+G(2d,2p) Level of Theory[a]

compounds	MEP@N	exp. pK_a	NAO px	NAO py	NAO pz
Et_2NCH	−18.33812	−2.0	−0.2591	−0.2587	−0.2649
diethylcyanimide	−18.33841	1.2	−0.2576	−0.2577	−0.2647
acetanilide	−18.34462	0.61	−0.2602	−0.2556	−0.2744
$NCH_2CH_2CN_3$	−18.34761	1.1	−0.2553	−0.2538	−0.2699
$HNCH_2CN_2$	−18.34941	0.2	−0.2449	−0.2439	−0.2700
p-nitrobenzne	−18.35054	1.02	−0.2808	−0.2539	−0.2432
$EtNCH_2CN_2$	−18.35225	−0.6	−0.2439	−0.2488	−0.2598
3-methyl-4-nitrobenzene	−18.35365	1.5	−0.2495	−0.2400	−0.2782
4-chloro-3-nitrobenzene	−18.35567	1.9	−0.2450	−0.2396	−0.2758
p-cyanobenzene	−18.35809	1.74	−0.2734	−0.2456	−0.2358
35-dimethyl-4-nitrobenzene	−18.36144	2.59	−0.2415	−0.2323	−0.2698
m-nitrobenzene	−18.36192	2.5	−0.2369	−0.2361	−0.2698
m-cyanobenzene	−18.36429	2.76	−0.2345	−0.2335	−0.2673
3,5-dibromobenzene	−18.36433	2.34	−0.2303	−0.2386	−0.2671
3,5-dichloro-aniline	−18.36477	2.37	−0.2294	−0.2379	−0.2666
3-methoxy-5-nitrobenzene	−18.36538	2.11	−0.2288	−0.2367	−0.2664
4-methyl-3-nitrobenzene	−18.36640	2.96	−0.2328	−0.2294	−0.2657
3,5-dibromo-4-methoxybenzene	−18.36833	2.98	−0.2258	−0.2353	−0.2609
$EtNCH_2CH_2CN_2$	−18.36848	4.55	−0.2304	−0.2426	−0.2426
35-dibromo-4-methylbenzene	−18.36912	2.87	−0.2253	−0.2327	−0.2622
dicyanodiethylamine	−18.36961	5.2	−0.2247	−0.2261	−0.2614
$HNCH_2CH_2CN_2$	−18.37007	5.26	−0.2360	−0.2517	−0.2239
35-dibromo-4-hydroxybenzene	−18.37157	3.2	−0.2225	−0.2292	−0.2594
$PhNMe_2$	−18.37325	5.1	−0.2264	−0.2261	−0.2335
dimethylaminoacetonitrile	−18.37434	4.2	−0.2215	−0.2206	−0.2435
m-bromobenzene	−18.37435	3.51	−0.2240	−0.2243	−0.2574
3-chloro-aniline	−18.37477	3.52	−0.2242	−0.2229	−0.2570
m-chlorobenzene	−18.37477	3.34	−0.2242	−0.2229	−0.2570
p-bromobenzene	−18.37538	3.91	−0.2271	−0.2185	−0.2562
m-fluorobenzene	−18.37572	3.59	−0.2252	−0.2199	−0.2560
3-fluoro-aniline	−18.37572	3.58	−0.2252	−0.2199	−0.2560
2-chloro-aniline	−18.37606	2.64	−0.2229	−0.2240	−0.2558
4-chloro-aniline	−18.37636	3.99	−0.2260	−0.2175	−0.2553
p-chlorobenzene	−18.37636	3.98	−0.2552	−0.2261	−0.2175
2-fluoro-aniline	−18.37697	3.2	−0.2222	−0.2186	−0.2537
$PhNEt_2$	−18.37729	6.6	−0.2271	−0.2230	−0.2306
3-chloro-5-methoxybenzene	−18.37780	3.1	−0.2170	−0.2237	−0.2538
3-bromo-4-methylbenzene	−18.37862	3.98	−0.2199	−0.2188	−0.2529
3-chloro-4-methylbenzene	−18.37907	4.05	−0.2213	−0.2162	−0.2524
$CF_3CH_2N(CH_3)_2$	−18.38004	4.75	−0.2200	−0.2169	−0.2358
4-fluoro-aniline	−18.38037	4.65	−0.2214	−0.2134	−0.2512
p-fluorobenzene	−18.38037	4.65	−0.2510	−0.2216	−0.2134
diethylaminoacetonitrile	−18.38089	4.5	−0.2153	−0.2156	−0.2405
n-piperidine-CH_2CN	−18.38182	4.55	−0.2370	−0.2196	−0.2164
aminoacetonitrile	−18.38220	5.3	−0.2131	−0.2222	−0.2422
H_2NCH_2CN	−18.38220	5.34	−0.2131	−0.2222	−0.2422
3-bromo-4-methoxybenzene	−18.38280	4.08	−0.2160	−0.2129	−0.2484
Et_2NCH_2CN	−18.38315	4.55	−0.2150	−0.2213	−0.2302
β-dimethylaminopropionitrile	−18.38334	7.0	−0.2132	−0.2147	−0.2379
m-hydroxybenzene	−18.38508	4.17	−0.2158	−0.2104	−0.2468
aniline	−18.38581	4.58	−0.2460	−0.2171	−0.2078
3,5-dimethoxybenzene	−18.38610	3.82	−0.2072	−0.2164	−0.2456
$CF_3CH_2NHCH_3$	−18.38635	6.05	−0.2172	−0.2356	−0.2054
n-piperidine-CCH_3CN	−18.38648	9.22	−0.2131	−0.2111	−0.2345
$CF_3CH_2NH_2$	−18.38776	5.7	−0.2105	−0.2225	−0.2334
3-methoxyl-aniline	−18.38807	4.2	−0.2120	−0.2080	−0.2438
m-methoxybenzene	−18.38807	4.2	−0.2120	−0.2080	−0.2438
m-methylbenzene	−18.38808	4.69	−0.2118	−0.2078	−0.2442
$Et_2NC(CH_3)_2CN$	−18.38892	9.13	−0.2132	−0.2152	−0.2278
4-methyl-aninile	−18.38929	5.08	−0.2127	−0.2041	−0.2426
p-methylbenzene	−18.38929	5.12	−0.2127	−0.2041	−0.2426
n-methyleamphetamine-$(CH_2)_2CN$	−18.38970	6.95	−0.2235	−0.2194	−0.2149
p-hydroxybenzene	−18.38980	5.5	−0.2113	−0.2036	−0.2416
$Et_2N(CH_2)_2CN$	−18.38981	7.65	−0.2236	−0.2098	−0.2185
3,5-dimethylbenzene	−18.39012	4.91	−0.2032	−0.2132	−0.2408
m-aminobenzene	−18.39058	4.88	−0.2401	−0.2062	−0.2099
3,4-dimethylbenzene	−18.39133	5.17	−0.2101	−0.2025	−0.2404
β-diethylaminopropionitrile	−18.39143	7.6	−0.2083	−0.2145	−0.2247
2-amino-2-cyanopropane	−18.39167	5.3	−0.2083	−0.2261	−0.2214
4-methoxyl-aniline	−18.39174	5.36	−0.2105	−0.2037	−0.2363

TABLE 1: Continued

compounds	MEP@N	exp. pK_a	NAO px	NAO py	NAO pz
p-methoxybenzene	−18.39174	5.29	−0.2093	−0.2017	−0.2397
β-aminopropionitrile	−18.39304	7.7	−0.2045	−0.2237	−0.2235
phenyl_OHOHOHH	−18.39345	8.58	−0.2052	−0.2034	−0.2490
n-amphetamine-$(CH_2)_2CN$	−18.39407	7.23	−0.2031	−0.2009	−0.2390
epinephrine	−18.39415	8.55	−0.2037	−0.2082	−0.2311
3-amino-4-hydroxybenzene	−18.39512	5.7	−0.2066	−0.1988	−0.2352
p-aminobenzene	−18.39595	6.08	−0.2049	−0.1972	−0.2352
triethanolamine	−18.39601	7.77	−0.2013	−0.2013	−0.2265
arterenol	−18.39616	8.55	−0.2080	−0.2286	−0.2127
$Et_2N(CH_2)_3CN$	−18.39621	9.29	−0.2048	−0.2040	−0.2221
2-methyleanilne-Et_2	−18.39666	7.18	−0.2021	−0.2068	−0.2168
n-methylmorpholine	−18.39918	7.41	−0.2213	−0.2059	−0.1981
n-allylmorpholine	−18.39975	7.05	−0.2057	−0.1994	−0.2200
nn-dimethyl-2-2-aminoethoxyethanol	−18.40040	9.1	−0.1961	−0.1968	−0.2199
$Et_2N(CH_2)_4CN$	−18.40046	10.08	−0.2171	−0.2020	−0.1992
n-benzoylpiperazine	−18.40053	7.78	−0.1995	−0.1954	−0.2259
β-difluoroethylamine	−18.40078	7.52	−0.1947	−0.2375	−0.1934
triallylamine	−18.40090	8.31	−0.2141	−0.2036	−0.2028
dimethylethanolamine	−18.40128	10.3	−0.1984	−0.1966	−0.2157
n-ethylmorpholine	−18.40141	7.7	−0.1999	−0.1975	−0.2229
benzyldimethylamine	−18.40151	8.93	−0.1980	−0.2175	−0.1954
$Et_2N(CH_2)_5CN$	−18.40182	10.46	−0.2068	−0.2020	−0.2067
allyldimethylamine	−18.40272	8.72	−0.1945	−0.1943	−0.2155
diallylmethylamine	−18.40366	8.79	−0.1940	−0.1989	−0.2138
n-carbethoxypiperazine	−18.40371	8.28	−0.1972	−0.1896	−0.2246
$(CH_3)_3N$	−18.40449	9.76	−0.1920	−0.1920	−0.2165
morpholine	−18.40649	8.36	−0.2266	−0.1900	−0.1864
α-benzylpyrroline	−18.40658	7.08	−0.2027	−0.2096	−0.1946
n-allylpiperidine	−18.40716	9.69	−0.1915	−0.1915	−0.2156
triethylenediamine	−18.40816	8.8	−0.2232	−0.1849	−0.1849
benzyldiethylamine	−18.40854	9.48	−0.1925	−0.1982	−0.2000
ethanolamine	−18.40857	9.5	−0.2035	−0.1917	−0.2085
diallylamine	−18.40900	9.29	−0.1851	−0.2090	−0.1964
n-methylpiperidine	−18.40921	10.08	−0.1921	−0.1888	−0.2137
dimethyl-*n*-propylamine	−18.40975	9.99	−0.2059	−0.1935	−0.1897
dimethylethylamine	−18.40982	9.99	−0.2122	−0.1881	−0.1896
dimethyl-*n*-butylamine	−18.40987	10.02	−0.2058	−0.1931	−0.1894
benzylmethylamine	−18.40990	9.58	−0.1984	−0.1980	−0.1922
n-methylpyrrolidine	−18.41092	10.46	−0.1933	−0.1872	−0.2151
allylmethylamine	−18.41106	10.11	−0.1840	−0.1799	−0.2186
1*n*-propylpiperidine	−18.41122	10.48	−0.1882	−0.1880	−0.2124
n-methyltrimethyleneimine	−18.41175	10.4	−0.2146	−0.1886	−0.1822
nn-dimethylcyclohexylamine	−18.41175	10.0	−0.1880	−0.2009	−0.1949
2-2-aminoethoxyethanol	−18.41184	9.5	−0.2142	−0.1957	−0.1804
methyldiethylamine	−18.41187	10.29	−0.1904	−0.1890	−0.2055
12-dimethylpyrrolidine	−18.41198	10.26	−0.1905	−0.1899	−0.2141
benzylethylamine	−18.41266	9.68	−0.1935	−0.1817	−0.2097
$(CH_3)_2NH$	−18.41295	10.64	−0.1972	−0.1984	−0.1805
benzylamine	−18.41331	9.34	−0.1858	−0.2148	−0.1899
α-ethylpyrroline	−18.41335	7.43	−0.2103	−0.1905	−0.1864
$(C_2H_5)_3N$	−18.41377	10.65	−0.1880	−0.1898	−0.2031
n-ethylpiperidine	−18.41388	10.4	−0.2055	−0.1933	−0.1871
$(C_3H_7)_3N$	−18.41393	10.65	−0.1859	−0.1870	−0.2043
$(C_4H_9)_3N$	−18.41423	10.89	−0.1860	−0.1877	−0.2011
allylamine	−18.41470	9.49	−0.1821	−0.2101	−0.1925
quinuclidine	−18.41526	11.0	−0.1784	−0.1793	−0.2155
phenyl_HHHH	−18.41548	9.78	−0.1815	−0.1869	−0.2123
β-phenylethylamine	−18.41599	9.83	−0.1835	−0.2026	−0.1958
methoxypropylamine	−18.41712	10.1	−0.1786	−0.1910	−0.2078
phenyl_oOHOHOHCH$_3$	−18.41714	8.55	−0.1840	−0.1757	−0.2045
ethylenediamine	−18.41724	9.98	−0.2037	−0.1961	−0.1754
piperidine	−18.41752	11.22	−0.1781	−0.1762	−0.2156
γ-phenylpropylamine	−18.41761	10.2	−0.1779	−0.2052	−0.1929
diisobutylamine	−18.41775	10.5	−0.2177	−0.1766	−0.1763
i_$(C_3H_7)_3N$	−18.41776	11.05	−0.1827	−0.1822	−0.2029
CH_3NH_2	−18.41803	10.62	−0.2121	−0.1870	−0.1735
$(C_2H_5)_2NH$	−18.41862	10.98	−0.1815	−0.1868	−0.1949
NH_3	−18.41865	9.21	−0.1827	−0.1827	−0.2301
$(C_3H_7)_2NH$	−18.41886	11.0	−0.1898	−0.1843	−0.1897
pyrrolidine	−18.41906	11.27	−0.1752	−0.1790	−0.2161
$(C_4H_9)_2NH$	−18.41920	11.25	−0.1925	−0.1789	−0.1884

TABLE 1: Continued[a]

compounds	MEP@N	exp. pK_a	NAO px	NAO py	NAO pz
trimethyleneimine	−18.41950	11.29	−0.2115	−0.1780	−0.1745
1-ethylr-2-methylpyrrolidine	−18.41964	10.64	−0.1926	−0.1873	−0.2037
$C_2H_5NH_2$	−18.42005	10.63	−0.2052	−0.1899	−0.1730
$C_3H_7NH_2$	−18.42022	10.53	−0.2001	−0.1937	−0.1728
$C_4H_9NH_2$	−18.42049	10.59	−0.2073	−0.1857	−0.1724
phenyl_HHOHH	−18.42093	8.9	−0.1749	−0.1782	−0.2127
i_$(C_3H_7)_2$NH	−18.42188	11.0	−0.1761	−0.1941	−0.1906
i_$C_3H_7NH_2$	−18.42234	10.63	−0.1932	−0.1948	−0.1775
phenyl_HOHOHH	−18.42246	8.93	−0.1776	−0.1933	−0.1943
cyclohexylaime	−18.42277	9.82	−0.2120	−0.1806	−0.1712
cyclohexylamine	−18.42277	10.49	−0.2120	−0.1806	−0.1712
di-sec-butylamine	−18.42332	11.01	−0.1732	−0.2087	−0.1774
cycloheptylamine	−18.42339	9.99	−0.1740	−0.1709	−0.2184

[a] Atomic units.

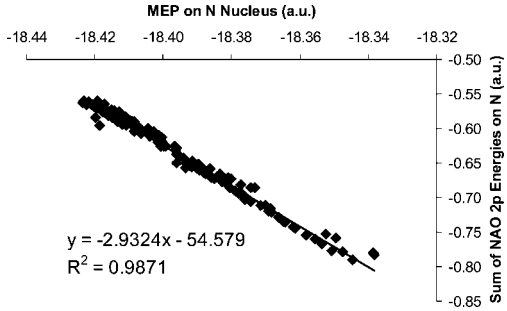
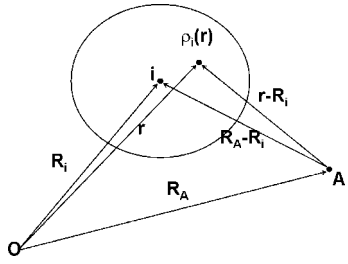


Figure 3. Strong linear relationship between MEP on N and the sum of nitrogen 2Px/2Py/2Pz NAO for N-containing compounds (amines and anilines) at the level of B3LYP/6-311+G(2d,2p). Atomic units.

SCHEME 2



To obtain the second term at the right-hand side of eq 5, we employed the approximation that R_i and R_A are separated (i.e., atoms A and i are not overlapped), so when calculating MEP at R_A from contributions of atoms R_i, we assume $r \approx R_i$ or $|r - R_A| \approx |R_i - R_A|$. That is:

$$\int \frac{\sum_{i\neq A} \rho_i^0(r - R_i)}{|r - R_A|} d\tau = \sum_{i\neq A} \int \frac{\rho_i^0(r - R_i)}{|r - R_A|} d\tau \approx \sum_{i\neq A} \int \frac{\rho_i^0(r - R_i)}{|R_i - R_A|} d\tau \approx \sum_{i\neq A} \frac{\int \rho_i^0(r - R_i)\, d\tau}{|R_i - R_A|} \approx \sum_{i\neq A} \frac{Z_i}{|R_i - R_A|} \quad (6)$$

The physical meaning of the above approximation is that the electrostatic potential at points A outside a spherical charge distribution $\rho_i(r)$ is equal to the electrostatic potential generated by the point charge Z_i from the center of the spherical atom i (Scheme 2). To get the last equality of eq 6, we used

$$\int \rho_i^0(r - R_i)\, d\tau = Z_i \quad (7)$$

The last term of eq 6 absorbed approximations from eq 7. Because

$$V_{R_A}^0 = \int \frac{\rho_A^0(r)}{|r - R_A|} d\tau \quad (8)$$

with eqs 2, 5, and 8, there arrives

$$V_{R_A} - V_{R_A}^0 = \sum_i \int \frac{g_i(|r - R_i|, \mathrm{NAO}_{v,i})}{|r - R_A|} d\tau \quad (9)$$

From the model density, eq 3, we know that the correction terms, $g_i(|r - R_i|,\mathrm{NAO}_{v,i})$, depend on differences in electron density between the atoms; these will be functions of the NAOs of the valence shells of the atoms. Because these local differences will be positive or negative, the rh's of eq 9 will thus be relatively small, see Figure 2, due to the significant cancelations in integration over the corrections. As seen in Figure 2, the rh's of eq 9 are indeed small for the large number of molecules studied; it is remarkable that these small numbers vary systematically with the pK_a values.

A strong correlation between the MEP on the acidic nucleus and the sum of the atom's valence natural atomic orbitals is observed. As an illustrative example, Figure 3 exhibits the relationship for amines and anilines. A similar correlation is seen for O- and S-containing compounds as well (not shown). Notice that the valence natural atomic orbitals employed in this study are 2p orbitals for nitrogen and oxygen and 3p orbitals for sulfur. We considered adding 2s/3s atomic orbitals in the summation, but no significantly different results were obtained. The strong correlation between the MEP on a nucleus and the valence NAO indicates that the correction term in eq 3, $f_i(|r - R_i|)$, is dominated by the contribution from the valence part of NAOs of the atom.

The MEP data are from DFT gas-phase calculations at the B3LYP/6-311+G(2d,2p) level. Taking the solvent effect into account does not destroy the correlation between MEP on the nucleus and experimental pK_a data. An example is illustrated in Figure 4 for the N-containing compounds, where one can

TABLE 2: Molecular Electrostatic Potential on the Acidic Atom Nucleus, Experimental pK_a Data, and Valence NAO Energies for Carbonyl Acids and Alcohols (O-Containing) Calculated at the B3LYP/6-311+G(2d,2p) Level of Theory[a]

compounds	MEP@O	exp. pK_a	NAO px	NAO py	NAO pz
2,2-dimethyl-propionic_acid	−22.31973	5.05	−0.3292	−0.3599	−0.3459
propionic_acid	−22.31848	4.87	−0.3342	−0.3574	−0.3475
butyric_acid	−22.31914	4.82	−0.3308	−0.3594	−0.3468
acetic_acid	−22.31648	4.76	−0.3327	−0.3633	−0.3495
p-methyl-benzoic_acid	−22.32054	4.37	−0.3292	−0.3575	−0.3449
vinyl-acetic_acid	−22.31360	4.35	−0.3384	−0.3628	−0.3520
phenyl-acetic_acid	−22.31572	4.31	−0.3433	−0.3513	−0.3529
m-methyl-benzoic_acid	−22.31932	4.27	−0.3302	−0.3588	−0.3459
succinic_acid	−22.31061	4.21	−0.3391	−0.3674	−0.3545
benzoic_acid	−22.31684	4.19	−0.3594	−0.3347	−0.3484
p-fluoro-benzoic_acid	−22.31179	4.14	−0.3383	−0.3661	−0.3534
3-chloro-propionic_acid	−22.30156	4.1	−0.3543	−0.3709	−0.3653
p-chloro-benzoic_acid	−22.31041	3.98	−0.3396	−0.3673	−0.3546
p-bromo-benzoic_acid	−22.31011	3.97	−0.3400	−0.3676	−0.3549
m-fluoro-benzoic_acid	−22.30922	3.87	−0.3400	−0.3690	−0.3555
m-chloro-benzoic_acid	−22.30871	3.83	−0.3405	−0.3696	−0.3560
glycolic_acid	−22.31175	3.83	−0.3401	−0.3651	−0.3545
m-bromo-benzoic_acid	−22.30859	3.81	−0.3454	−0.3650	−0.3562
formic_acid	−22.30199	3.75	−0.3755	−0.3448	−0.3622
m-cyano-benzoic_acid	−22.29973	3.6	−0.3518	−0.3764	−0.3647
p-cyano-benzoic_acid	−22.29935	3.55	−0.3776	−0.3512	−0.3651
methoxy-acetic_acid	−22.31280	3.54	−0.3392	−0.3575	−0.3493
3-butynoic_acid	−22.30687	3.32	−0.3485	−0.3665	−0.3586
fumaric_acid	−22.30203	3.05	−0.3461	−0.3749	−0.3614
bromo-acetic_acid	−22.30017	2.86	−0.3585	−0.3595	−0.3721
chloro-acetic_acid	−22.29858	2.81	−0.3549	−0.3761	−0.3666
2-chloro-propionic_acid	−22.30170	2.8	−0.3512	−0.3574	−0.3769
fluoro-acetic_acid	−22.29786	2.66	−0.3578	−0.3697	−0.3640
cyano-acetic_acid	−22.28694	2.44	−0.3750	−0.3751	−0.3781
nitro-acetic_acid	−22.28111	1.32	−0.3826	−0.3807	−0.3833
dichloro-acetic_acid	−22.28739	1.3	−0.3664	−0.3735	−0.3844
oxalic_acid	−22.28724	1.25	−0.3617	−0.3853	−0.3765
difluoro-acetic_acid	−22.28535	1.24	−0.3745	−0.3736	−0.3732
trichloro-acetic_acid	−22.28290	0.63	−0.3907	−0.3656	−0.3759
trifluoro-acetic_acid	−22.27212	0.23	−0.3770	−0.4020	−0.3881
t-butanol	−22.37672	18.0	−0.2766	−0.2942	−0.2904
isopropanol	−22.37374	17.1	−0.2817	−0.2867	−0.2970
n-propanol	−22.37101	16.1	−0.2748	−0.2956	−0.2993
ethanol	−22.37177	15.9	−0.2740	−0.2962	−0.2994
methanol	−22.36778	15.5	−0.2875	−0.2867	−0.3014
p-amino-phenol	−22.34360	10.3	−0.3147	−0.3091	−0.3188
p-methoxy-phenol	−22.33902	10.21	−0.3236	−0.3096	−0.3231
p-methyl-phenol	−22.33704	10.14	−0.3223	−0.3155	−0.3248
m-methyl-phenol	−22.33582	10.08	−0.3358	−0.3047	−0.3259
phenol	−22.33335	9.98	−0.3259	−0.3197	−0.3283
p-hydroxy-phenol	−22.33679	9.96	−0.3218	−0.3157	−0.3253
p-fluoro-phenol	−22.32713	9.95	−0.3317	−0.3257	−0.3344
m-amino-phenol	−22.33835	9.87	−0.3332	−0.3022	−0.3233
m-methoxy-phenol	−22.33569	9.65	−0.3359	−0.3049	−0.3257
m-hydroxy-phenol	−22.33002	9.44	−0.3410	−0.3106	−0.3312
p-chloro-phenol	−22.32374	9.38	−0.3354	−0.3289	−0.3374
p-bromo-phenol	−22.32300	9.36	−0.3363	−0.3296	−0.3381
m-fluoro-phenol	−22.32288	9.28	−0.3477	−0.3185	−0.3380
m-bromo-phenol	−22.32181	9.03	−0.3488	−0.3198	−0.3391
m-chloro-phenol	−22.32212	9.02	−0.3490	−0.3187	−0.3387
m-cyano-phenol	−22.31146	8.61	−0.3586	−0.3301	−0.3490
m-nitro-phenol	−22.30897	8.4	−0.3617	−0.3323	−0.3514
p-cyano-phenol	−22.30726	7.95	−0.3451	−0.3525	−0.3526
p-nitro-phenol	−22.30165	7.15	−0.3506	−0.3583	−0.3576

[a] Atomic units.

see that the correlation coefficient is similar to that of the gas-phase results. Also, we performed MEP calculations at other levels of theory, such as Hartree−Fock theory (Figure 5) or with different density functionals; no significant difference in the correlation was seen. In addition, for amines we also considered the protonated, conjugate species, but no statistically significant correlation between MEP at N and pK_a data is observed (results not shown).

One possible application of these results is to estimate pK_a values with a single DFT calculation for amino acids and peptides where different pK_a values at different atom sites are possible. As an illustrative example, we estimated pK_a values

TABLE 3: Molecular Electrostatic Potential on the Acidic Atom Nucleus, Experimental pK_a Data, and Valence NAO Energies for Sulfonic Acids and Thiols (S-Containing) Calculated at the B3LYP/6-311+G(2d,2p) Level of Theory[a]

compounds	MEP@S	exp. pK_a	NAO px	NAO py	NAO pz
methyl_thioglycolate	−59.24106	7.8	−0.1978	−0.2189	−0.2228
ethyl_mercaptan	−59.25179	10.5	−0.1780	−0.2011	−0.2457
o-aminothiophenol	−59.23813	6.59	−0.1853	−0.2519	−0.2122
$HOCH_2CH(OH)CH_2$-thiol	−59.25194	9.51	−0.1779	−0.1971	−0.2385
CH_2═$CHCH_2$-thiol	−59.24700	9.96	−0.1837	−0.2240	−0.2273
n-C_4H_9-thiol	−59.25272	10.66	−0.1765	−0.2004	−0.2445
t-C_5H_{11}-thiol	−59.25812	11.21	−0.2142	−0.1733	−0.2211
$C_2H_5OCOCH_2$-thiol	−59.24254	7.95	−0.1964	−0.2176	−0.2211
$C_2H_5OCH_2CH_2$-thiol	−59.25366	9.38	−0.1833	−0.2073	−0.2218
$HOCH_2CH(OH)CH_2$-thiol	−59.25194	9.66	−0.1779	−0.1971	−0.2385
n-C_3H_7-thiol	−59.25237	10.65	−0.1771	−0.2006	−0.2449
thioglycolic_acid	−59.22815	3.67	−0.2144	−0.2535	−0.2149
mercaptoethanol	−59.24517	9.5	−0.1849	−0.2072	−0.2513
cysteamine	−59.25026	10.81	−0.1798	−0.2027	−0.2466
thioacetic_acid	−59.22381	3.33	−0.2240	−0.2678	−0.2193

[a] Atomic units.

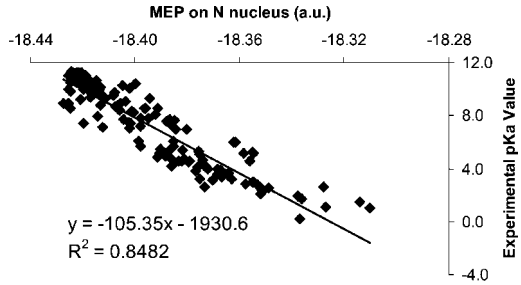


Figure 4. The impact of the solvent effect on the correlation between MEP on N and experimental pK_a data for N-containing compounds (amines and anilines). The implicit PCM (Polarizable Continuum model) and 6-311+G(2d,2p) basis set were used.

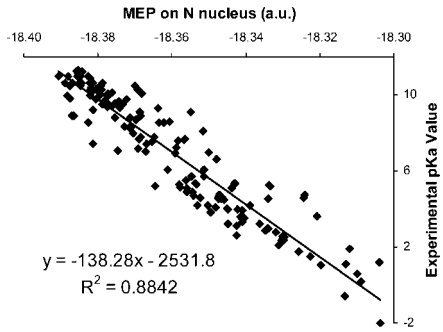


Figure 5. The strong linear relationship between MEP on N nucleus and experimental pK_a data for amines and anilines using the Hartree−Fock method and 6-311+G(2d,2p) basis set.

of cysteinylcysteine, which has four acidic sites, O, S1, S2, and N. Using the relationships in Figure 1, we obtained the pK_a values to be 3.5 (O), 6.9 (S1), 8.2 (S2), and 9.8, respectively, whereas experimental data give 2.7, 7.3, 9.4, and 10.9, respectively. Similar results are obtained when the relationship in Figure 2 is employed. In both cases, reasonable pK_a values are obtained, and the order of acidity of the four atoms is correctly predicted.

IV. Conclusions

An effective approach of estimating molecular pK_a values from simple gas-phase density functional calculations is proposed in this work, using either the molecular electrostatic potential on the nucleus of the acidic atom or the sum of valence natural atomic orbitals. A strong correlation between experimental pK_a values and each of these two quantities has been discovered. Moreover, if the MEP is subtracted by a given reference value for each category of compounds, we observe a single unique linear relationship between the MEP difference and experimental pK_a data of amines, anilines, carbonyl acids, alcohols, sulfonic acids, thiols, and their substituents. With a single DFT calculation, these results can conveniently be utilized to simultaneously estimate pK_a values at multiple sites of small molecules in drug design and of amino acids in proteins and macromolecules.

Acknowledgment. This work was supported in part by the National Institutes of Health (HL-06350) and NSF (FRG DMR-0804549). We acknowledge the use of the computational resources provided by the Research Computing Center at University of North Carolina at Chapel Hill and the Biomedical Unit of the Pittsburgh Supercomputer Center.

References and Notes

(1) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, *109*, 6857.
(2) Potter, M. J.; Gilson, M. K.; McCammon, J. A. *J. Am. Chem. Soc.* **1994**, *116*, 10298.
(3) Rajasekaran, E.; Jayaram, B.; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 8238.
(4) Alagona, G.; Ghio, C.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 9855.
(5) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190.
(6) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.
(7) Namazian, M.; Heidary, H. *J. Mol. Struct. (THEOCHEM)* **2003**, *620*, 257.
(8) Nielsen, J. E.; Mccammon, J. A. *Protein Sci.* **2003**, *12*, 1894.
(9) Fitch, C. A.; Garcia-Moreno, B. *Biophys. J.* **2004**, *86*, 86A.
(10) da Silva, G.; Kennedy, E. M.; Dlugogorski, B. Z. *J. Phys. Chem. A* **2006**, *110*, 11371.
(11) Ohno, K.; Sakurai, M. *J. Comput. Chem.* **2006**, *27*, 906.
(12) MacDermaid, C. M.; Kaminski, G. A. *J. Phys. Chem. B* **2007**, *111*, 9036.
(13) Pliego, J. R. *Chem. Phys. Lett.* **2003**, *367*, 145.
(14) De Proft, F.; Amira, S.; Choho, K.; Geerlings, P. *J. Phys Chem. B* **1995**, *98*, 5227. Machado, H. J. S.; Hinchliffe, A. *J. Mol. Struct. (THEOCHEM)* **1995**, *339*, 255.
(15) Deka, R. C.; Roy, R. K.; Hirao, K. *Chem. Phys. Lett.* **2004**, *389*, 186.
(16) Gupta, K.; Roy, D. R.; Subramanian, V.; Chattaraj, P. K. *J. Mol. Struct. (THEOCHEM)* **2007**, *812*, 13.
(17) Nagy, P.; Novak, K.; Szasz, G. *J. Mol. Struct. (THEOCHEM)* **1989**, *60*, 257.
(18) Brinck, T.; Murray, J. S.; Politzer, P.; Carter, R. E. *J. Org. Chem.* **1991**, *56*, 2934.
(19) Gross, K. C.; Seybold, P. G.; Peralta-Inga, Z.; Murray, J. S.; Politzer, P. *J. Org. Chem.* **2001**, *66*, 6919.

(20) Ma, Y.; Gross, K. C.; Hollingsworth, C. A.; Seybold, P. G.; Murray, J. S. *J. Mol. Model.* **2004**, *10*, 235.
(21) Politzer, P. *Theor. Chem. Acc.* **2004**, *111*, 395.
(22) Politzer, P.; Ma, Y. G.; Jalbout, A. F.; Murray, J. S. *Mol. Phys.* **2005**, *103*, 15.
(23) Glendening, E. D.; Badenhoop, J. K.; Reed, A. E.; Carpenter, J. E.; Bohmann, J. A.; Morales, C. M.; Weinhold, F. *NBO 5.0*; Theoretical Chemistry Institute: University of Wisconsin, Madison, 2001.
(24) Lu, H.; Chen, X.; Zhan, C.-G. *J. Phys. Chem. B* **2007**, *111*, 10599.
(25) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.
(26) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.
(27) Schruumann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.
(28) Gross, K. C.; Seybold, P. G. *J. Org. Chem.* **2001**, *66*, 6919.
(29) da Silva, R. R.; Ramalho, T. C.; Santos, J. M.; Figueroa-Villar, J. D. *J. Phys. Chem. A* **2006**, *110*, 1031.
(30) Chaudry, U. A.; Popelier, P. L. A. *J. Org. Chem.* **2004**, *69*, 233.
(31) Lide, D. R. *Handbook of Chemistry and Physics*, 88th ed.; CRC Press: Boca Raton, NY, 2007.
(32) 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 E.01; Gaussian, Inc.: Pittsburgh, PA, 2003.

JP811250R