A high-level theoretical study into the atmospheric phase hydration, bond dissociation enthalpies, and acidity of aldehydes

Sierra Rayne*a and Kaya Forestb

INTRODUCTION

The solution phase hydration of carbonyl compounds not only plays a fundamental role in organic chemistry and biochemistry owing to the mechanistic importance of tetrahedral intermediates,1,2 but also has substantial thermodynamic and kinetic relevance towards the environmental chemistry of natural and anthropogenically derived materials.3-4 Aqueous phase studies have shown that hydration equilibria can be substantial for some substituted aldehydes,5-10 and corresponding structure-reactivity investigations have correlated equilibrium constants with Hammett substituent parameters and applied other linear free energy approaches.10-16 Because carbonyl hydration is less favorable in the gas phase, theoretical studies have generally concentrated on aqueous phase reactivity and mechanistic/kinetic modeling and are typically limited to smaller compounds such as formaldehyde, with fewer studies focusing on larger databases of molecules and/or gas phase equilibria.17-31

Among the various classes of carbonyl derivatives, aldehydes play an important role in the chemistry of natural and polluted atmospheres.32-34 Recent findings suggest that some hydrated aldehydes may have significant populations in air samples35-38 warranting their consideration in environmental fate and toxicology modeling efforts, gas-particle and air-water partitioning studies, and understanding of formation pathways for secondary organic aerosols. Furthermore, given the complexity of atmospheric chemistry, and the ever increasing diversity of anthropogenic compounds being emitted into the troposphere, there is also a need to investigate the potential gas phase hydration behavior for a broad range of substituted aldehydes. An additional goal is the development of reliable structure-reactivity relationships, as these can be included in environmental models and be applied to systems that are difficult to measure experimentally or modeled theoretically. In some cases, experimental measurements of gas phase aldehyde hydration may not be possible because of the low value of an equilibrium constant or competitive side reactions for reactants and products.

In the current work, we employ high-level theoretical methods and established physical organic substituent constant/structure-property approaches to study equilibria for the gas phase hydration of a wide range of substituted aldehydes. In addition, we have examined the gas phase bond dissociation enthalpies.

Keywords: acidity; aldehydes; atmospheric chemistry; bond dissociation enthalpies; hydration

---

*a Correspondence to: S. Rayne, Chemologica Research, Moose Jaw, Saskatchewan, Canada.
E-mail: sierra.rayne@alumni.ubc.ca

b K. Forest
Department of Environmental Engineering Technology, Saskatchewan Polytechnic, Moose Jaw, Saskatchewan, Canada

---


Copyright © 2016 John Wiley & Sons, Ltd.
(BDEs) and aldehydic hydrogen acidities for various aldehydes with the intent of benchmarking the composite methods and examining possible structure–activity relationships. Gas phase BDEs are directly relevant for atmospheric chemistry, as they are strongly related to rates of degradation by oxidants such as the hydroxyl, nitrate, and chloride radicals. In addition, gas phase BDEs are often correlated with solution phase BDEs, and consequently offer insights into likely condensed phase reactivity trends. Gas phase acidities of these compounds are both of fundamental interest and often correlate with solution phase acidities in various solvents.

MATERIALS AND METHODS

Calculations were conducted using Gaussian 09. All calculations used the same gas phase starting geometries obtained with the PM6 method as implemented in MOPAC 2009 (http://www.openmopac.net; v. 9.281). Complete basis set (CBS) calculations used the CBS-Q//B3 method. Gaussian calculations used the G4 and G4(MP2) methods. All molecular enthalpies and free energies include zero point and thermal corrections. For hydrated aldehydes, no a priori assumptions were made regarding molecular conformations. All hydrated compounds were subjected to a molecular mechanics (MM) molecular dynamics (MD) conformational search with an unconstrained MD trajectory via the Verlet velocity algorithm and quasi-Newton gradient (maximum number of iterations = 10 000; ε = 10−4) using the following default parameters in Gabeedit v.2.2.12 (http://gabedit.sourceforge.net/). Amber MM model with inclusion of bond stretch, angle bend, torsion, non-bonded, and electrostatic charge–charge terms; post-processing identical structure removal with an energy tolerance of 0.01 kcal/mol and distance tolerance of 0.01 Å; and maximum 10 final lowest energy structures. Additional MM conformational studies were conducted using the MMFF94 force field and a systematic rotor search within Avogadro v.1.0.1 (http://avogadro.openmolecules.net/).

Depending on the number of conformers generated by each approach, between two and nine lowest energy MM conformations for each hydrated compound were subjected to calculations at each of the CBS-Q//B3, G4(MP2), and G4 levels of theory. For the aldehydes, acyl radicals, and acyl anions, multiple starting conformations were also considered where relevant. Only the lowest energy conformation for each compound from each composite method calculation was used for further analysis. Optimized structures were confirmed as true minima by vibrational analysis at the same level. Unless otherwise noted, all Hammett substituent constants and resonance and field parameters are from the compendium of Hansch et al. Statistical analyses were conducted using KryPlot v.2.0.b.15 (http://www.kyenslab.com/en/).

As described by Hansch et al., Hammett constants are widely employed in all branches of chemistry to describe a linear relationship between differing substituent identities and the equilibrium constant or rate of a reaction. For example, in the equation log (K/kR) = σΔ + pR, K is the equilibrium constant with substituent R, kR is the reference equilibrium constant where R is a hydrogen atom, σ is the substituent constant that depends only on the specific substituent R, and pR is the reaction constant that depends only on the type of reaction but not on the substituent used. Depending on the effect of interest (e.g., inductive, σi; resonance, σr; and field, σf), the positional relationship to the reaction center (e.g., meta, σm, and para, σp), and/or the charge/developing charge at the reaction center (e.g., σr′, σp′, σR′, and σR′), a range of substituent constants are available and have been determined experimentally and calculated theoretically. In the Swain–Lupton approach, field (F) and resonance (R) effects for a substituent are related to σR as follows: σR = FR + rR. A linear free energy relationship can be subsequently probed for equilibrium constants (or reaction rates) of the general form log (K/kR) = FR + rR + C (or log (k/kR) = FR + rR + C), where f, r, and C (a constant) are obtained via multiple linear regression of the dependent variables F and R against the independent variable log (K/kR) (or log (k/kR)) for a series of differing substituents and their corresponding equilibrium constants (or reaction rates). Analogous to the Hammett constants, the Swain–Lupton constants can reflect the charge/developing charge at the reaction center (e.g., R′ and R′).
Table 1. Substituent constants and estimated gas phase standard state (298.15 K, 1 atm) free energies of hydration ($\Delta_{\text{hydr}} G°^\circ(g)$) and corresponding hydration equilibrium constants (log $K_{\text{hydr,}}(g)$) for various aldehydes with electron withdrawing and electron releasing substituents at the CBS-Q/B3, G4(MP2), and G4 levels of theory

<table>
<thead>
<tr>
<th>Substituent</th>
<th>Substituent constants</th>
<th>$\Delta_{\text{hydr}} G°^\circ(g)$ (kcal/mol)</th>
<th>Log $K_{\text{hydr,}}(g)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(O)CH₃</td>
<td>0.33 0.17 n/a</td>
<td>0.9 2.1 1.0</td>
<td>CBS-Q/B3 G4(MP2) G4</td>
</tr>
<tr>
<td>NH₃</td>
<td>0.08 -0.74 -1.38</td>
<td>16.6 17.2 16.5</td>
<td>-12.1 -12.6 -12.1</td>
</tr>
<tr>
<td>C(CH₃)₃</td>
<td>-0.02 -0.18 -0.17</td>
<td>4.3 5.4 4.4</td>
<td>-3.2 -4.0 -3.2</td>
</tr>
<tr>
<td>C≡N</td>
<td>0.51 0.15 0.15</td>
<td>1.1 2.5 1.5</td>
<td>-0.8 -1.9 -1.1</td>
</tr>
<tr>
<td>N(CH₃)₂</td>
<td>0.15 -0.98 -1.85</td>
<td>20.7 21.2 20.3</td>
<td>-15.1 -15.5 -14.8</td>
</tr>
<tr>
<td>CH≡CH₂</td>
<td>0.13 -0.17 -0.29</td>
<td>7.1 8.0 7.1</td>
<td>-5.2 -5.9 -5.2</td>
</tr>
<tr>
<td>C≡CH</td>
<td>0.22 0.01 -0.04</td>
<td>4.6 5.7 4.8</td>
<td>-3.4 -4.2 -3.5</td>
</tr>
<tr>
<td>F</td>
<td>0.45 -0.39 -0.52</td>
<td>6.9 7.8 6.7</td>
<td>-3.0 -5.8 -4.9</td>
</tr>
<tr>
<td>H</td>
<td>0.03 0.00 0.00</td>
<td>1.3 2.3 1.2</td>
<td>-0.9 -1.6 -0.9</td>
</tr>
<tr>
<td>OH</td>
<td>0.33 -0.70 -1.25</td>
<td>14.4 15.3 14.3</td>
<td>-10.5 -11.2 -10.5</td>
</tr>
<tr>
<td>CH₃</td>
<td>0.01 -0.18 -0.32</td>
<td>2.8 3.8 2.7</td>
<td>-2.0 -2.7 -2.0</td>
</tr>
<tr>
<td>NO₂</td>
<td>0.65 0.13 0.14</td>
<td>-1.8 -0.1 -1.3</td>
<td>1.4 0.1 0.9</td>
</tr>
<tr>
<td>C₆H₅</td>
<td>0.12 -0.13 -0.30</td>
<td>7.0 8.0 7.1</td>
<td>-5.1 -5.9 -5.2</td>
</tr>
<tr>
<td>CF₃</td>
<td>0.38 0.16 0.23</td>
<td>-2.4 -0.9 -2.0</td>
<td>1.8 0.6 1.5</td>
</tr>
<tr>
<td>Si(CH₃)₃</td>
<td>0.01 -0.08 0.01</td>
<td>1.7 2.9 1.8</td>
<td>-1.3 -2.2 -1.3</td>
</tr>
<tr>
<td>Br</td>
<td>0.45 -0.22 -0.30</td>
<td>7.7 9.5 8.2</td>
<td>-5.6 -6.9 -6.0</td>
</tr>
<tr>
<td>Cl</td>
<td>0.42 -0.19 -0.31</td>
<td>8.0 9.4 8.3</td>
<td>-5.9 -6.9 -6.1</td>
</tr>
<tr>
<td>SiCl₃</td>
<td>0.44 0.12 0.13</td>
<td>-0.4 1.8 0.5</td>
<td>0.3 -1.3 -0.4</td>
</tr>
<tr>
<td>SiF₃</td>
<td>0.47 0.22 n/a</td>
<td>0.1 1.8 0.4</td>
<td>-0.1 -1.3 -0.3</td>
</tr>
<tr>
<td>SiH₃</td>
<td>0.06 0.04 0.08</td>
<td>3.2 4.4 3.2</td>
<td>-2.4 -3.2 -2.4</td>
</tr>
<tr>
<td>CBF₃</td>
<td>0.28 0.01 n/a</td>
<td>3.2 3.1 1.9</td>
<td>-1.7 -2.3 -1.4</td>
</tr>
<tr>
<td>CCl₃</td>
<td>0.38 0.09 n/a</td>
<td>0.5 2.1 0.9</td>
<td>-0.4 -1.5 -0.7</td>
</tr>
<tr>
<td>C≡CCF₃</td>
<td>0.37 0.14 n/a</td>
<td>2.4 3.6 2.6</td>
<td>-1.7 -2.6 -1.9</td>
</tr>
<tr>
<td>C≡CCH₃</td>
<td>0.29 -0.10 n/a</td>
<td>4.4 4.6 3.6</td>
<td>-3.2 -3.3 -2.7</td>
</tr>
<tr>
<td>C≡C≡N</td>
<td>0.92 0.04 n/a</td>
<td>-3.2 -1.6 -2.8</td>
<td>2.3 1.2 2.0</td>
</tr>
<tr>
<td>C₆H₅-4-Br</td>
<td>0.18 -0.06 -0.38</td>
<td>6.5 7.7 6.8</td>
<td>-4.8 -5.7 -5.0</td>
</tr>
<tr>
<td>C₆H₅-4-Cl</td>
<td>0.18 -0.06 -0.37</td>
<td>6.7 7.8 6.8</td>
<td>-4.9 -5.7 -5.0</td>
</tr>
<tr>
<td>C≡C≡F</td>
<td>0.17 -0.11 n/a</td>
<td>7.0 8.1 7.1</td>
<td>-5.1 -5.9 -5.2</td>
</tr>
<tr>
<td>C≡C≡CH₃</td>
<td>0.12 -0.15 n/a</td>
<td>7.7 8.8 7.9</td>
<td>-5.6 -6.4 -5.8</td>
</tr>
<tr>
<td>PF₄</td>
<td>0.54 0.26 n/a</td>
<td>-6.8 -4.3 -5.8</td>
<td>5.0 3.1 4.3</td>
</tr>
<tr>
<td>GeH₃</td>
<td>0.03 -0.02 n/a</td>
<td>3.1 4.8 3.6</td>
<td>-2.3 -3.5 -2.6</td>
</tr>
</tbody>
</table>

B3, G4(MP2), and G4 methods. The comparison suggests a systematic difference in the results between our two investigations, with qualitative agreement but modest quantitative disagreement. There is clearly substantial disagreement (on the order of several kcal/mol) between various levels of theory regarding the $\Delta_{\text{hydr}} G°^\circ(g)$ for even simple aldehydes. It is of interest that many of the prior semiempirical, Hartree-Fock (HF), density functional theory, and Møller-Plesset perturbation theory calculations predicted a $\Delta_{\text{hydr}} G°^\circ(g) < 0$ for formaldehyde. If such was the case, we would expect to see a dominant population of the hydrated form during gas phase studies of this compound, which is not observed experimentally.\[9,47]\n
More recently, Kent et al.\[48] have conducted a theoretical study of the conversion of gas phase methandiol to formaldehyde. Using the HF, MP2, MP4, CCSD, CCSD(T), and CC(Q) model chemistries with the cc-pVDZ and cc-pVTZ basis sets, these authors obtained $\Delta_{\text{hydr}} G°^\circ(g)$ ranging from 1.0 to 3.3 kcal/mol, with the highest level CCSD(T)/cc-pVQZ and QCI(T)/cc-pVTZ methods yielding $\Delta_{\text{hydr}} G°^\circ(g)$ of 1.0 and 1.1 kcal/mol, respectively. The values are in excellent agreement with our calculations, particularly our G4 $\Delta_{\text{hydr}} G°^\circ(g)$ of 1.2 kcal/mol. Consequently, the similar positive $\Delta_{\text{hydr}} G°^\circ(g)$ of 1 to 2 kcal/mol reported herein and by Kent et al.\[48] using high-level methods are likely the most accurate thermodynamic data for the gas phase hydration of formaldehyde.

Increasing EWG strength results in more favorable $\Delta_{\text{hydr}} G°^\circ(g)$ at the CBS-Q/B3, G4(MP2), and G4 levels of theory, as the electron withdrawing substituents increase the electrophilicity of the carbonyl carbon. To quantify this effect, we applied the Swain–Lupton approach\[49–51] using field effect (f) and resonance effect (r) parameters\[46] for each of the substituted aldehydes in the following equation, where $f$ and $r$ are coefficients obtained from multiple linear regression and C is a regression constant: log $K_{\text{hydr,}}(g) = f + rC$. During our initial statistical investigations, we observed that C=C=CH₃ ($r = -0.26$) was clustering separate from the other α-sp carbons (C=C-H, C=C-CCF₃, and C≡N).

R values are commonly derived from F values and Hammett substituent constants by way of equations such as $R = \sigma_p - aF$ (where $a = 0.921$).\[52] Using this equation and the $\sigma_p$ and F values given by Hansch et al.\[46] consistently yields R values for C=C-H (0.03 vs. 0.011), C≡N (0.19 vs. 0.15), C≡C=CH₃ (−0.24 vs. −0.26), C≡C≡F (0.17 vs. 0.14), and C≡C≡H₃ (0.03 vs. 0.01) within ±0.04 units of those provided in ref.[46] If one assumes the contribution of the α=C=C moiety to the $\sigma_p$ of C≡C-H, C≡C=CH₃, and C≡C=CH₂...
C≡C–CF₃, and C≡C–CH₂H₂ is constant, and that the σₚ are all accurate, linearly regressing the σᵣ of H (0.00), CH₂ (−0.17), CF₃ (0.54), and C₆H₅ (−0.01) against the σᵣ of C≡C–CH (0.23), C≡C–CH₂ (0.03), C≡C–CF₃ (0.51), and C≡C–CH₂H (0.16) should give a high correlation coefficient, which is the case (y = mx + b; r = 0.981, p = 0.02, m = 0.64, b = 0.17).

Similarly, assuming that the contribution of the α-C≡C moiety to the F value of C≡C–H, C≡C–CH₂, C≡C–CF₃, and C≡C–CH₂H is constant, and that the F-values are all accurate, linearly regressing the F of H (0.03), CH₂ (0.01), CF₃ (0.38), and C₆H₅ (0.12) against the F of C≡C–H (0.22), C≡C–CH₂ (0.29), C≡C–CF₃ (0.37), and C≡C–CH₂H (0.15) should also yield a straight line. This is not observed, and the regression is non-significant (p = 0.40). Omitting C≡C–CH₂ does not result in a significant regression (p = 0.36). F values can also be estimated from equations of the form $F = \alpha_m + \beta_r + \epsilon$, where constants of $a = 1.369/b = -0.373/\epsilon = -0.009^{[52]}$ and $a = 1.297 ± 0.147/b = -0.385 ± 0.089/\epsilon = 0.033 ± 0.026^{[46]}$ have been reported. Use of these constants and the σₘ and σᵣ values from ref.^[52] yields F values of 0.27 and 0.29, respectively, in agreement with the original F value of 0.29. To examine the potential inaccuracy of σₘ for C≡C–CH₂, σₘ for H (0.00), CH₂ (−0.07), CF₃ (0.43), and C₆H₅ (0.06) were linearly regressed against the σᵣ of C≡C–H (0.21), C≡C–CH₂ (0.21), C≡C–CF₃ (0.41), and C≡C–CH₂H (0.14). Neither regressions using all four paired compounds (p = 0.13) or C≡C–H, C≡C–CF₃, and C≡C–CH₂H (p = 0.24) yield significant relationships.

However, the available substituent constant dataset for α-sp carbons may be insufficient to evaluate the reliability of the C≡C–CH₂ R value. Thus, we looked at the correlation between the σₘ values (σₘ₁ + σₘ₂) for substituents on CH≡C(R₁R₂) functionalities and the σₘ values for the entire CH≡C(R₁R₂) moieties on the following groups: CH=CH₂, CH≡C(CH₃)₂, CH≡CH(NO₂)₃, CH=CH=CH(CF₃)₃, CH=CH=CH=CH(CH₃), CH=CH=CH=CH(CH₃), CH=CH=CH=CH(CH(CH₃)₃), CH=CH=CH=CH(CH₃)(CO₂CH₃). In this larger dataset, we found a significant linear relationship (r = 0.942, p = 0.025, n = 0.46, b = 0.02), suggesting that an analogous trend should exist for C≡C–R groups. Thus, approximating the trend between σᵣ for H, CH₃, and C₆H₅ against the σᵣ of C≡C–H, C≡C–CF₃, and C≡C–CH₂H (m = 0.56, b = 0.16), and intersecting the σᵣ of CH₂ (−0.07), yields a revised C≡C–CH₂ σᵣ of 0.12. With a σᵣ = 0.12 for C≡C–CH₂, F values of 0.14 and 0.18 are obtained with the equation $F = \alpha_m + \beta_r + \epsilon$ and the corresponding regression constants from ref.^[52] and^[46] respectively. Consequently, revised R values of 0.10 and 0.14 for C≡C–CH₂ arise from the relationship $R = \sigma_p - 0.921 F$. The use of $R = 0.10$ for C≡C–CH₂ places this substituent substantially closer to other α-sp carbons in plots of $\Delta_{\text{hydr}}G'(g)$ against R at each level of theory we examined, and we use this revised R value in subsequent analyses and discussions. However, future experimental clarification on this issue is required.

Using both the field and resonance effect substituent constants, we obtain reasonable training set correlation coefficients (n = 31) using multiple linear regression of the form log $K_{\text{hydr}}(g)$ = $F_R + R + C$ with corresponding values for $F$, $R$, and $C$ at the three levels of theory (error bars are standard errors throughout the manuscript): CBS-Q/B3, $r = 0.926, SE = 0.16, CV = -0.51, F = 2.79 ± 1.44$ (p = 0.06), $r = 13.1 ± 1.1 (p < 10^{-11}), C = -2.68 ± 0.54 (p < 10^{-4});$ G4 (MP2), $r = 0.918, SE = 0.16, CV = -0.40, F = 2.24 ± 1.44 (p = 0.13), r = 12.5 ± 1.1 (p < 10^{-10}), C = -3.51 ± 0.54 (p < 10^{-10});$ and G4, $r = 0.918, SE = 0.16, CV = -0.50, F = 2.40 ± 1.47 (p = 0.11), r = 12.7 ± 1.2 (p < 10^{-10}), C = -2.74 ± 0.55 (p < 10^{-4}).$ All three levels of theory yield statistically non-significant (p > 0.05) regression coefficients for the field effect constant $F$, suggesting a significant probability that $F = 0$ (and thus, negligible field effects) within the multiple regression equation. However, there is a highly significant (p = 0.01) individual relationship between $F$ and log $K_{\text{hydr}}(g)$. Although Hammett substituent constants ($\sigma_m, \sigma_p, \sigma_R, \sigma_a, \sigma_b, \sigma_R^\alpha, \sigma_R^\beta, \sigma_R^\gamma$) and $R^+$ values are not available for all substituents under consideration, the available subsets we examined did not provide as good correlations as the $F/R$ values and were not considered further.

Consequently, we omitted the field effect term from further analysis and investigated the theoretical gas phase hydration thermodynamics using only the resonance effect parameters R (Fig. 2) and $R^+$ (Fig. 3). Linearly regressing the CBS-Q/B3, G4 (MP2), and G4 log $K_{\text{hydr}}(g)$ against the corresponding R values gave the regression statistics in Table 2. Although strong

**Figure 2.** Correlations between the Swain–Lupton resonance effect parameter (R) and estimated gas phase standard state (298.15 K, 1 atm) hydration equilibrium constants (log $K_{\text{hydr}}(g)$) for various substituted aldehydes at the (a) CBS-Q/B3, (b) G4 (MP2), and (c) G4 levels of theory. A best fit regression for all datapoints (n = 31) is shown as a solid line. Identities of the substituents are given in Table 1, and summaries of the statistical analyses are provided in Table 2.
EWG C₆H₄-4-NO₂ substituent has a negative R⁺ value of -0.22, as do the parent halogen substituents (e.g., F = -0.52, Cl = -0.31, and Br = -0.30). The use of either R or R⁺-based resonance parameters for log K_{hydr,(g)} correlation analysis is theoretically defensible, because both the neutral and charge separated carbonyl resonance forms (C=O and C⁺-O⁻) are expected to contribute to the hydration reaction. Multiple linear regressions of the general form log K = rR + r'R + C (where r and r' potentially represent the relative contributions of the neutral and positively charged resonance forms) are precluded given the multicollinearity between R and R⁺ (r = 0.977 for the 20 substituents we considered).

Application of the R value-based regression equations towards 487 substituted aldehydes (including the 31 compounds in the training set) with available R values for uncharged substituents yields the estimated log K_{hydr,(g)} ranging between -16.0 (N(C₆H₅)₂) and +9.2 (N=NPO(OEt)₂) shown in Table S1. Analogous application of the R⁺ value-based regression equations towards 108 substituted aldehydes (including the 20 compounds in the training set) with available R⁺ values for uncharged substituents yields the estimated log K_{hydr,(g)} ranging between -17.7 (N(C₆H₅)₂) and +2.8 (P(C₆H₅)₂) shown in Table S2. Charged substituents (e.g., O⁻, CO₂⁻, and N⁻(CH₃)₃) were not considered in these gas phase log K_{hydr} estimates using either the R or R⁺ based approaches because these charged side groups are generally considered non-volatile and are not expected to have significant gas phase populations amenable to hydration reactions. Similarly, a number of substituents included in our analysis would also be expected to undergo hydrolysis in the presence of atmospheric water (e.g., acyl halides and transition metal complexes) at rates competitive with, or exceeding, that of corresponding aldehyde hydration reactions, but they are retained in the database for completeness. Regression-derived standard errors in the log K_{hydr,(g)} are about 1.0 to 1.5 units for most compounds. For the 108 substituents with both R and R⁺ values, good agreement was observed between log K_{hydr,(g)} obtained using both correlation approaches at each level of theory (r values ranging from +0.85 to +0.86, p < 10⁻⁵). The value of such correlations and resulting log K_{hydr,(g)} estimates resides in the conformational complexity and large size of many substituents for which R/R⁺ values are available. In many cases, high-level calculations on these moieties are either impossible (beyond the basis set limits) or impractical (large numbers of low-energy conformers requiring investigation and/or many heavy atoms present).

As a sample application of these results, experimental[35] and theoretical[36] evidence suggests perfluoroalkyl aldehydes (PFAlds) are likely present in the atmosphere in significant quantities as the hydrated form. These compounds are both derivatives and precursors of other well-known perfluorinated compounds and transition metal complexes) at rates competitive with, or exceeding, that of corresponding aldehyde hydration reactions, but they are retained in the database for completeness.

Regression-derived standard errors in the log K_{hydr,(g)} are about 1.0 to 1.5 units for most compounds. For the 108 substituents with both R and R⁺ values, good agreement was observed between log K_{hydr,(g)} obtained using both correlation approaches at each level of theory (r values ranging from +0.85 to +0.86, p < 10⁻⁵). The value of such correlations and resulting log K_{hydr,(g)} estimates resides in the conformational complexity and large size of many substituents for which R/R⁺ values are available. In many cases, high-level calculations on these moieties are either impossible (beyond the basis set limits) or impractical (large numbers of low-energy conformers requiring investigation and/or many heavy atoms present).

As a sample application of these results, experimental[35] and theoretical[36] evidence suggests perfluoroalkyl aldehydes (PFAlds) are likely present in the atmosphere in significant quantities as the hydrated form. These compounds are both derivatives and precursors of other well-known perfluorinated compounds and transition metal complexes) at rates competitive with, or exceeding, that of corresponding aldehyde hydration reactions, but they are retained in the database for completeness.
A HIGH-LEVEL THEORETICAL STUDY INTO ALDEHYDES

Table 2. Linear regression statistics for CBS-Q//B3, G4(MP2), and G4 estimated gas phase log \( K_{\text{hydr,(g)}} \) against Swain–Lupton resonance effect \( (R) \) and positive charged resonance effect \( (R^+ \) parameters for the substituted aldehydes in Figs. 2 and 3, respectively.

<table>
<thead>
<tr>
<th>Level of theory</th>
<th>( r )</th>
<th>( p )</th>
<th>SE</th>
<th>CV</th>
<th>m</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBS-Q//B3</td>
<td>0.923</td>
<td>&lt;10(^{-12})</td>
<td>1.62</td>
<td>−0.51</td>
<td>13.6 ± 1.1</td>
<td>−1.89 ± 0.31</td>
</tr>
<tr>
<td>G4(MP2)</td>
<td>0.921</td>
<td>&lt;10(^{-12})</td>
<td>1.55</td>
<td>−0.38</td>
<td>12.8 ± 1.0</td>
<td>−2.87 ± 0.29</td>
</tr>
<tr>
<td>G4</td>
<td>0.920</td>
<td>&lt;10(^{-12})</td>
<td>1.60</td>
<td>−0.49</td>
<td>13.1 ± 1.0</td>
<td>−2.06 ± 0.30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of theory</th>
<th>( r )</th>
<th>( p )</th>
<th>SE</th>
<th>CV</th>
<th>m</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBS-Q//B3</td>
<td>0.960</td>
<td>&lt;10(^{-10})</td>
<td>1.24</td>
<td>−0.29</td>
<td>7.55 ± 0.52</td>
<td>−1.69 ± 0.32</td>
</tr>
<tr>
<td>G4(MP2)</td>
<td>0.958</td>
<td>&lt;10(^{-10})</td>
<td>1.23</td>
<td>−0.24</td>
<td>7.20 ± 0.51</td>
<td>−2.70 ± 0.32</td>
</tr>
<tr>
<td>G4</td>
<td>0.959</td>
<td>&lt;10(^{-10})</td>
<td>1.23</td>
<td>−0.28</td>
<td>7.31 ± 0.51</td>
<td>−1.90 ± 0.32</td>
</tr>
</tbody>
</table>

log \( K_{\text{hydr,(g)}} \) for these compounds using correlations with \( R \) values are 0.1 CBS-Q//B3/−1.0 G4(MP2) (C\(_2\)), −0.2 CBS-Q//B3/−1.3 G4(MP2) (C\(_3\)), and 0.1 CBS-Q//B3/−1.0 G4(MP2) (C\(_4\)). \( R^+ \) substituent constants are not available for C\(_2\)F\(_2\), C\(_3\)F\(_2\), and C\(_4\)F\(_2\).

The results show that the longer chain (i.e., >C\(_4\)) PFAlds – which are of primary interest to the environmental chemistry community – may be present substantially as the hydrated form in the atmosphere. Similarly, other aldehydes with \( \alpha \)-EWGs such as nitro and variously substituted carbonyl moieties are also expected to have significant hydrated populations in gas phase atmospheric samples. Consequently, ambient atmospheric modeling efforts and associated risk assessments, as well as studies in engineered systems, will need to account for the gas phase hydration of such compounds. However, it is also important to note that various loss mechanisms (e.g., uptake on surfaces), the possible failure to achieve atmospheric equilibrium, and the relatively low equilibrium vapor pressure of water in the atmosphere will act collectively to reduce actual populations of the hydrates below their theoretical maxima. In another experimental study, Axson et al.[37] have recently provided an experimental log \( K_{\text{hydr,(g)}} \) of 2.2 ± 0.1 for methylglyoxal (H\(_2\)C(CO)(OH)). Our calculations at the CBS-Q//B3, G4(MP2), and G4 levels obtain log \( K_{\text{hydr,(g)}} \) of −0.6, −1.6, and −0.7, respectively. In their work, Axson et al.[37] noted that their reported log \( K_{\text{hydr,(g)}} \) is “significantly more favorable than theoretical predictions, even ones made in aqueous solution.” Our results are, however, consistent with those of Barsanti and Pankow, who estimated that gas phase hydration of methylglyoxal would not be thermodynamically favorable.\(^{[56]}\) Further clarification is required to better define whether the two theoretical estimates, or the experimental datapoint, accurately reflect the gas phase hydration thermodynamics of this model secondary organic aerosol precursor.

Our calculations were limited to gas phase asymmetric carbonyl substitutions of the general form \( R_1\text{C(O)R}_2 \) where \( R_2 = H \) (i.e., aldehydes). We sought to examine whether the findings could be extended into the aqueous phase and also towards symmetric and asymmetric carbonyl substitutions where \( R_2 \neq H \) and \( R_1 \neq R_2 \) (i.e., ketones and carboxylic acid derivatives). Gomez-Bombarelli et al.[28] have published a compilation of experimental log \( K_{\text{hydr,(aq)}} \) and conducted corresponding B3LYP/6-31++G(d,p), MP2/6-31++G(d,p), and CBS-Q log \( K_{\text{hydr,(aq)}} \) calculations directly using the integral equation formalism of the polarizable continuum model implicit solvation model. When the sum of the \( R \) values (\( \Sigma R = R_{\text{R1}} + R_{\text{R2}} \)) on either side of the carbonyl moiety from this compilation is plotted against the experimental log \( K_{\text{hydr,(aq)}} \) values, a strong correlation exists (\( n = 35, r = 0.947, p < 10^{-17}, \)) \( \text{SE} = 1.56, \text{CV} = -0.44, m = 14.43 ± 0.85, b = 1.70 ± 0.41 \) among all datapoints, although the three thioures (H\(_2\)OCS\(_2\)H\(_2\), H\(_3\)CC(O)SC\(_2\)H\(_2\), and F\(_2\)CC(O)SC\(_2\)H\(_2\)) are outliers (Fig. 4(a)). The use of \( R^+ \) (\( \Sigma R^+ = R^+_{\text{R1}} + R^+_{\text{R2}} \)) does not substantially change the summary statistics for the correlation (\( n = 28, r = 0.972, p < 10^{-17}, \)) \( \text{SE} = 1.20, \text{CV} = -0.28, m = 8.08 ± 0.38, b = 2.31 ± 0.38, \) but removes the outlying nature of various substituent groups (notably the thioures), thereby allowing greater confidence in applying the regression towards substituent functional group classes not included in the training set. SC\(_2\)H\(_2\) does not have a \( R^+ \) value in Hansch et al.[46], thus, a value of −0.81 was estimated based on the \( R^+ = -0.83 \) for SCH\(_3\) and the \( \Delta R^+ = 0.02 \) between CH\(_3\) (\( R^+ = -0.32 \)) and C\(_2\)H\(_5\) (\( R^+ = -0.30 \)).

The strong agreement for \( \Sigma R \) and \( \Sigma R^+ \)-based approaches towards log \( K_{\text{hydr,(aq)}} \) estimation across multiple functional group classes and over a log \( K_{\text{hydr,(aq)}} \) range of about 20 units suggests that not only can the gas phase \( R/R^+ \) value log \( K_{\text{hydr}} \) modeling approaches extend into aqueous systems but also that a simple summative approach to \( R/R^+ \) values about the carbonyl moiety can likely be used to extend our log \( K_{\text{hydr,(g)}} \) regression equations presented previously towards all possible functional group combinations of \( R_1\text{C(O)R}_2 \). A sample application of the \( \Sigma R \)-based approaches towards log \( K_{\text{hydr,(aq)}} \) estimation for various aldehydes is provided in Table S3. Using the regression equation shown in Fig. 4(b), log \( K_{\text{hydr,(aq)}} \) were estimated for 115 aldehydes having available \( R^+ \) substituent constants from ref.[46]. Predicted log \( K_{\text{hydr,(aq)}} \) range from −18.4 (S\(_5^2\) ) for strongly electron releasing substituents to +7.2 (P(C\(_6\)H\(_5\))\(_2\)) for strongly EWG groups. The estimated log \( K_{\text{hydr,(aq)}} \) end members are the symmetric ketones C(O)S\(_2\)\(^2\) (log \( K_{\text{hydr,(aq)}} \) = −39.0) and C(O)(P(C\(_6\)H\(_5\))\(_2\))\(_2\) (log \( K_{\text{hydr,(aq)}} \) = +12.0), well outside the current experimental range of −14.2 to +3.4. We note that many of the substituents are susceptible to rapid aqueous phase hydrolysis (e.g., various carboxylic acid derivatives) and would therefore hydrolyze having kinetic parameters competitive with – or exceeding – that of hydration reactions, but we retain these functional groups in the database for completeness.

Gas phase standard state (298.15 K, 1 atm) BDEs for the acyl C-H bond were also calculated at the CBS-Q//B3, G4(MP2), and G4 levels of theory on various substituted aldehydes (Table 3). Good agreement (i.e., generally within 1 kcal/mol) was found between the theoretical values and the limited available experimental dataset.
The radical species of several ω- and β-unsaturated members (R=C(O)CH₃, CH=CH₂, C=CH, NO₂, C≡C-CH₃, C≡C-CF₃, and C(N)₃), as well as R=PF₅, did not readily converge at all three levels of theory, respectively. The deprotonated anionic forms correspond to the currently known experimental data end members (e.g., OH, H), thereby substantially extending the range of lower quality of fit as compared with the analogous calculations were conducted to estimate the gas phase standard state (298.15K, 1 atm) enthalpies (ΔdG°(g)) and free energies (ΔdG°(g)) of aldehyde proton acid dissociation for these compounds (Table 4). As with the BDE calculations, good agreement was found between the theoretical values and the available experimental dataset. The deprotonated anionic forms did not converge at all levels of theory for the following aldehyde substituents: NH₂, C≡N, C≡C-H, C≡C-CH₃, C≡C-CF₃, NO₂, CF₃, Cl, Br, SiH₃, Si(CH₃)₃, SiF₃, SiCl₃, CCl₃, CBr₃, C(C≡N)₃, PF₅, and GeH₃.

Multiple linear regression of the σ₉/σ₁₋₉, σ₁₀/σ₁₋₉, and F/R pairs did not yield high quality correlations with statistical significance for both independent variables. Among the univariate correlations, σ₁ displayed an excellent quality of fit (n=9, r=-0.966, SE=3.1, CV=0.008, m=-60.3±6.1, b=387.1±1.4) against the G₄ ΔdG°(g) data in addition to the CBS-Q/B3 (n=9, r=-0.965, SE=3.1, CV=0.008, m=-59.7±5.9, b=386.7±1.4) datasets. Multiple regression of the general form aσ₁+bσ₁0=ΔdG°(g) gave insignificant (p>0.05) β-coefficients for σ₁₀, suggesting inductive influences dominate gas phase acidity variations for substituted aldehydes. Similar qualities of fit were obtained between σ₁ and the theoretical ΔdH°(g) (all n=9): CBS-Q/B3, r=-0.967, SE=3.1, CV=0.008, m=-60.6±6.0, b=395.4±1.4; G₄(MP2), r=-0.969, SE=2.9, CV=0.008, m=-59.8±5.7, b=394.8±1.3; and G₄, r=-0.969, SE=3.0, CV=0.008, m=-61.3±5.9, b=393.5±1.4. Using these correlations and σ₁ values for a range of substituents reported by Charton⁵⁷ yields the estimated ΔdH°(g) and ΔdG°(g) given in Table S5. The predicted ΔdH°(g) range within these substituents is from 351.8 to 396.0 kcal/mol, and a ΔdG°(g) range from 344.3 to 387.8 kcal/mol. These boundaries correspond to the currently known experimental data end members for fluorine and dimethylamino substituents, respectively. We also note that the Swain–Lupton F parameter yields a modestly lower quality of fit as compared with the σ₁ substituent constant (e.g., r of -0.924 against the G₄ ΔdG°(g) data, compared with a corresponding r of -0.966 for σ₁), but a larger parameter database is available for F, thereby substantially extending the range...
of possible substituents that can be modeled. In addition, $\sigma_I$ can be estimated from $F$ values using equations such as $F = 0.888 n_1 + 0.017$, further extending the scope of our $\Delta_d H^\circ(g)/\Delta_d G^\circ(g)$ models.

### CONCLUSIONS

Gas phase standard state (298.15 K, 1 atm) free energies of hydration ($\Delta_d H^\circ(g)$) and corresponding equilibrium constants (log $K_{\text{hydr}}(g)$), BDEs, and enthalpies ($\Delta_d H^\circ(g)$) and free energies ($\Delta_d G^\circ(g)$) of aldehyde proton acid dissociation were calculated using the composite method CBS-Q//B3, G4(MP2), and G4 levels of theory on a range of substituted aldehydes with various EWG and ERGs. More EWG substituents increasingly favor higher log $K_{\text{hydr}}(g)$, theoretical values display excellent correlation with the Swain-Lupton resonance effect parameters $R$ and $R'$, and no significant correlation with the corresponding field effect parameter ($F$). Using these correlations, log $K_{\text{hydr}}(g)$ was predicted for 487 substituted aldehydes having available $R$ values and 108 substituted aldehydes having available $R'$ values. The sum of $R$ and $R'$ on substituents attached to members of various carbonyl derivative classes also displays strong correlations with aqueous phase hydration equilibrium constants (log $K_{\text{hydr}}(g)$), reliably extending the modeling efforts into the condensed phase and suggesting the gas phase modeling approach can be used to predict log $K_{\text{hydr}}(g)$ for all possible symmetric and asymmetric carbonyl derivatives.

Both the computationally derived BDEs and $\Delta_d H^\circ(g)/\Delta_d G^\circ(g)$ exhibited good agreement with the limited experimental and theoretical datasets. No predictive univariate or multivariate correlations were found between the theoretical gas phase BDEs and the major Hammett substituent constants or Swain–Lupton parameters, suggesting that general structure–activity relationships across different types of functional groups may be absent for this combination of physical property and general compound class. High correlation coefficients were observed between gas phase acidities and the Hammett inductive substituent constant ($\sigma_I$), facilitating estimation of $\Delta_d H^\circ(g)/\Delta_d G^\circ(g)$ for other compounds with known $\sigma_I$ values. Good correlations were also found between $\Delta_d H^\circ(g)/\Delta_d G^\circ(g)$ and the corresponding $F$ values, allowing this substituent parameter (which is available for a large number of compounds, and can also be used to estimate $\sigma_I$ values) to also be employed for the estimation of gas phase aldehydic proton acidities.

### Acknowledgements

This work was made possible by the facilities of the Western Canada Research Grid (WestGrid:www.westgrid.ca; project 100185), the Shared Hierarchical Academic Research Computing Network (www.sharcnet.ca; project aqn-965), and Compute/Calcul Canada. We thank the anonymous reviewers whose comments greatly improved the quality of the manuscript.

### Table 4. Estimated gas phase standard state (298.15 K, 1 atm) enthalpies ($\Delta_d H^\circ(g)$) and free energies ($\Delta_d G^\circ(g)$) of aldehyde proton acid dissociation for various aldehydes with electron withdrawing and electron releasing substituents at the CBS-Q//B3, G4(MP2), and G4 levels of theory

<table>
<thead>
<tr>
<th>Substituent</th>
<th>$\Delta_d H^\circ(g)$</th>
<th>$\Delta_d G^\circ(g)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expt. CBS-Q//B3 G4(MP2) G4</td>
<td>Expt. CBS-Q//B3 G4(MP2) G4</td>
</tr>
<tr>
<td>C(O)CH$_3$</td>
<td>n/a</td>
<td>376.0</td>
</tr>
<tr>
<td>C(CH$_3$)$_3$</td>
<td>387.4 ± 4.1$^{[61]}$</td>
<td>392.1</td>
</tr>
<tr>
<td>N(CH$_3$)$_2$</td>
<td>399.1 ± 4.1$^{[62]}$</td>
<td>392.4</td>
</tr>
<tr>
<td>CH$_2$ = CH$_2$</td>
<td>n/a</td>
<td>392.2</td>
</tr>
<tr>
<td>F</td>
<td>349.7 ± 3.3$^{[63]}$</td>
<td>362.1</td>
</tr>
<tr>
<td>H</td>
<td>352.6 ± 4.6$^{[64]}$</td>
<td>394.5 ± 0.1$^{[65]}$</td>
</tr>
<tr>
<td>OH</td>
<td>n/a$^b$</td>
<td>379.2</td>
</tr>
<tr>
<td>CH$_3$</td>
<td>393.1 ± 0.9$^{[67]}$</td>
<td>393.3</td>
</tr>
<tr>
<td></td>
<td>391.0 ± 2.1$^{[68]}$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>387.0 ± 8.0$^{[69]}$</td>
<td>—</td>
</tr>
<tr>
<td>C$_6$H$_5$</td>
<td>n/a</td>
<td>388.1</td>
</tr>
<tr>
<td>C$_6$H$_4$-4-Br</td>
<td>n/a</td>
<td>382.3</td>
</tr>
<tr>
<td>C$_6$H$_4$-4-Cl</td>
<td>n/a</td>
<td>383.0</td>
</tr>
<tr>
<td>C$_6$H$_4$-4-F</td>
<td>n/a</td>
<td>384.9</td>
</tr>
<tr>
<td>C$_6$H$_4$-4-CH$_3$</td>
<td>n/a$^a$</td>
<td>389.2</td>
</tr>
</tbody>
</table>

Available values$^{[60]}$ are shown for comparison. Values are in kcal/mol.

$^a$Not available.

$^b$Experimental $\Delta_d H^\circ(g)$ and $\Delta_d G^\circ(g)$ are available for formic acid, but the values are for the more acidic carboxylic acid moiety ($\Delta_d H^\circ(g) \approx 340$ to 346 kcal/mol$^{[71–74]}$, $\Delta_d G^\circ(g) \approx 338$ to 339 kcal/mol$^{[71–75]}$).

$^c$Experimental $\Delta_d H^\circ(g)$ and $\Delta_d G^\circ(g)$ are available for 4-methylbenzaldehyde, but the values are for the more acidic methyl group ($\Delta_d H^\circ(g) = 358.2 ± 2.4$ kcal/mol$^{[76]}$, $\Delta_d G^\circ(g) = 352.6 ± 2.3$ kcal/mol$^{[76]}$).
REFERENCES


A HIGH-LEVEL THEORETICAL STUDY INTO ALDEHYDES


SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article at the publisher’s web site.