JCTC Journal of Chemical Theory and Computation

Alchemical Variations of Intermolecular Energies According to Molecular Grand-Canonical Ensemble Density Functional Theory

O. Anatole von Lilienfeld*,† and M. E. Tuckerman†,‡

Department of Chemistry, New York University, New York, New York 10003, and Courant Institute of Mathematical Sciences, New York University, New York 10003

Received January 3, 2007

Abstract: Molecular grand-canonical density functional theory [*J. Chem. Phys.* **2006**, *125*, 154104] is employed for the alchemical variation of intermolecular energies due to changes in the chemical composition of small molecules. We investigate the interaction of a fixed binding target, formic acid, with a restricted chemical space, corresponding to an isoelectronic 10-proton system which includes molecules such as CH_4, NH_3, H_2O, and HF. Differential expressions involving the nuclear chemical potential are derived, numerically evaluated, tested with respect to finite difference results, and discussed regarding their suitability as gradients of the intermolecular energy with respect to compositional variations.

I. Introduction

A. Chemical Space. The fundamental challenge of compound design, that is, the reverse engineering of chemical compounds with predefined specific properties, originates in the high-dimensional and combinatorial nature of chemical space.[1] Chemical space is the hyperspace of a given set of molecular observables that is spanned by the grand-canonical variables (particle densities of electrons and nuclei) which define the chemical composition of stable molecules or material. Exploration of this space via screening techniques, such as combinatorial high-throughput techniques or in silico evaluation exploiting quantitative structure−property relationships, consistently faces the problem of having to deal with the factorial scaling of the cardinal number which is due to the combinatorial character of chemical composition, atomic connectivities, and conformational configurations. Consequently, more "rational" endeavors have been undertaken to tackle this problem more deterministically.[2−7]

Unless one possesses knowledge about the relevant chemical space a priori, a first-principles scheme is required for its unbiased exploration since chemical bonding patterns, the number of atoms, and atomic numbers must become variables. Such a scheme can be formulated using the principles of physical chemistry and seeks to address the question of how a compound's property changes upon variation of its chemical composition. Recently, we presented a rigorous physical framework for this purpose by introducing a molecular multicomponent grand-canonical extension[8] of Kohn−Sham density functional theory (KS−DFT).[9,10] How chemical space is distinguished from the well-defined notion of phase space is somewhat arbitrary, as has been extensively discussed and investigated in the context of variations in the electronic chemical potential and the number of electrons in ref 8. In the present work, we focus on the complementary aspect of chemical space, namely, the study of the nuclear chemical potential and the proton density defining the nuclear charge distribution. Specifically, we will investigate how intermolecular energies, important molecular observables which are ubiquitous in all self-assembly processes, depend on the chemical composition of neutral and isoelectronic molecules. The procedure presented herein is prototypical in showing how to rationally "dial-in" certain intermolecular potential energies via compositional and structural changes and without any screening.

Reliable estimates of noncovalent interactions between molecular entities are notoriously difficult to obtain from first principles since they require an accurate description of the relatively weak and long-ranged mutual polarization of

* Corresponding author e-mail: ovt203@nyu.edu.
† Department of Chemistry.
‡ Courant Institute of Mathematical Sciences.

10.1021/ct700002c CCC: $37.00
Published on Web 04/06/2007

molecular entities. Within KS−DFT, the accuracy of current approximations to the exact exchange-correlation potential in describing intermolecular potentials can vary significantly[11−13,31,32] To achieve an accuracy that is state-of-the-art, the generalized gradient approximation to the exchange-correlation potential employed herein has been augmented by an atom-centered London dispersion correction as introduced, assessed, and employed in refs 14−17.

B. Motivation. Before presenting the numerical results, a brief summary of the idea of compound design is given in order to provide the necessary background. It was proposed by one of us[5] that the problem of tuning intermolecular energies through alchemical variations can be mapped onto a minimization problem in a variable space spanned by the electron and proton densities. Within density functional theory, any observable, $\mathcal{O}$, can be expressed in terms of a statistical mechanical average of a functional of the electron density which, due to the Hohenberg−Kohn theorem, is uniquely defined by the total number of electrons, N_e, and the external potential due to the nuclear charge distribution, $Z(\mathbf{r})$. If a target observable, $\mathcal{O}_0$, is known a priori, a variational principle, $\mathscr{P} \geq 0$, can be applied to the general penalty functional

$$\mathscr{P} = (\mathcal{O}_0 - \mathcal{O}[N_e, Z])^2 \tag{1}$$

$\mathscr{P}$ can be minimized by varying the proton distribution, $Z(\mathbf{r})$, and the number of electrons, N_e, defining the observable, $\mathcal{O}$, of the system. Variation with respect to the particle number, however, suggests the use of a grand-canonical ensemble theory. First- and second-order derivatives of $\mathscr{P}$ with respect to Z and N_e are thus "grand-canonical derivatives",[8] reminiscent of chemical potentials and are likely, as is standard in all gradient-based optimization algorithms,[18] to significantly enhance the efficiency of any algorithm which minimizes $\mathscr{P}$ in the elementary particle space of protons and electrons. If the system contains N_n nuclei at positions $\mathbf{R}_1$, ..., $\mathbf{R}_{N_n}$, then the atomic number of the Ith nucleus will be $Z(\mathbf{R}_I)$. Thus, taking into account the inhomogeneity of possible proton distributions, the number of degrees of freedom available in chemical space is $1+N_n$, the additional degree of freedom arising from the fact that the electron number N_e is one additional minimization parameter.

The functional $\mathscr{P}$ is stationary if

$$0 = \frac{\delta \mathscr{P}}{\delta Z(\mathbf{r})} = 2(\mathcal{O}[N_e, Z] - \mathcal{O}_0)\frac{\delta \mathcal{O}[N_e, Z]}{\delta Z(\mathbf{r})} \tag{2}$$

and

$$0 = \frac{\partial \mathscr{P}}{\partial N_e} = 2(\mathcal{O}[N_e, Z] - \mathcal{O}_0)\frac{\partial \mathcal{O}[N_e, Z]}{\partial N_e} \tag{3}$$

Clearly, the local nature of the proton density is reflected in the $\mathbf{r}$ dependence of the derivative $\delta\mathscr{P}/\delta Z(\mathbf{r})$.

When considering the change of molecular observables due to noninteger variations of elementary particles, it must be recognized that, in actuality, isolated charge densities integrate to integers. This can be accounted for by either assuming integer steps in N_e and $Z(\mathbf{r})$ or by requiring $\mathcal{O}$ to be a state function that can be changed along arbitrary paths, specifically via continuous variables, permitting the *virtual* existence of fractional numbers of electrons and protons.

Two further subtleties arise: (1) unless the band gap is zero, potential energy derivative discontinuities emerge for variation in the number of electrons at integer occupation numbers;[19,20] (2) if unphysical, negatively charged "protons" are to be avoided, then for all $Z(\mathbf{r}) = 0$, the derivative of $\mathcal{O}$ with respect to $Z(\mathbf{r})$ is only defined for $dZ(\mathbf{r}) \rightarrow 0^+$. The latter constraint, however, must be relaxed when considering the transmutational, or alchemical, case, $\delta\mathcal{O}/\delta Z(\mathbf{R}_I)$, when a given atomic number, that is, $Z(\mathbf{R}) > 0$, is allowed both to increase and to decrease.

II. Theory—Nuclear Gradients of Intermolecular Energies

A. The Intermolecular Nuclear Chemical Potential. For this study, the observable $\mathcal{O}$ has been chosen to be an intermolecular potential energy. Since this quantity is a state function, its value can be tuned using fractional particle variables. We limit ourselves to the variation in the proton density. Nuclear grand-canonical gradients are derived, computed, and assessed in order to determine their utility for the control of intermolecular energies through compositional and structural tuning. The electronic analog to the nuclear chemical potential, the electronic chemical potential, μ_e, as well as changes in N_e, have already been investigated within applications of the molecular grand-canonical ensemble DFT introduced in ref 8.

The total potential energy of interaction, E^{int}, of an arbitrary substrate (s) with a binding template (t) is defined as

$$E^{int} = E^{t+s} - E^t - E^s \tag{4}$$

In order to maximize the energy of interaction, we set $\mathcal{O} = E^{int}[N_e, Z]$ and $\mathcal{O}_0 = -\infty$ and find a simplified variational principle, namely, $\mathscr{P}[N_e, Z] \geq 0$. $\mathscr{P}$ is minimal when the gradients of $\mathcal{O}$ are zero and the second-order derivatives are positive. In the following, we will denote the nuclear chemical potential of our intermolecular energy as $\mu_n^{int}(\mathbf{r}) = \delta E^{int}/\delta Z(\mathbf{r})$. According to eq 4, these gradients are

$$\begin{aligned}\mu_n^{int}(\mathbf{r}) &= \frac{\delta E^{t+s}}{\delta Z(\mathbf{r})} - \frac{\delta E^s}{\delta Z(\mathbf{r})} - \frac{\delta E^t}{\delta Z(\mathbf{r})} \\ &= \mu_n^{t+s}(\mathbf{r}) - \mu_n^s(\mathbf{r}) - \mu_n^t(\mathbf{r})\end{aligned} \tag{5}$$

Note that $\mu_n^{int}(\mathbf{r})$ is constructed from functional derivatives of the total potential energy with respect to the proton distribution as they are generally formulated in terms of a continuous nuclear charge distribution as discussed in ref 8. It becomes evident that the intermolecular nuclear chemical potential, $\mu_n^{int}(\mathbf{r})$, corresponds to the difference between the nuclear chemical potentials,[8] $\mu_n(\mathbf{r})$, of the complex, t + s, and the isolated systems, t and s. Note that, due to the indistinguishability and the spatial distribution of the proton distribution, $\mu_n^t(\mathbf{r})$ is not zero since a variation of the proton density in the vicinity of an isolated template affects its total potential energy value.

As shown in ref 8, one can use a modified electrostatic potential

$$\bar{V}^{\mathrm{ESP}}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \frac{Z(\mathbf{r}')\ \mathrm{erf}[\sigma|\mathbf{r}-\mathbf{r}'|] - \rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{6}$$

for $\mu_{\mathrm{n}}(\mathbf{r})$, where the parameter σ is chosen to be sufficiently small to exclude the Coulomb repulsion between protons only within the same nucleus. Here, $\rho(\mathbf{r})$ represents the electron, and $Z(\mathbf{r}) = \sum_I N_I \delta(\mathbf{R}_I - \mathbf{r})$ is the proton density, N_I being the atomic number of atom I. These two particle densities must fulfill the constraint that they integrate to their total number, N_{e} and N_{p} for the electrons and protons, respectively.

From eq 5, the intermolecular nuclear chemical potential can be expressed as

$$\mu_{\mathrm{n}}^{\mathrm{int}}(\mathbf{r}) = \bar{V}_{\mathrm{t+s}}^{\mathrm{ESP}}(\mathbf{r}) - \bar{V}_{\mathrm{s}}^{\mathrm{ESP}}(\mathbf{r}) - \bar{V}_{\mathrm{t}}^{\mathrm{ESP}}(\mathbf{r}) \tag{7}$$

Apparently, the change of the total potential energy of interaction with respect to the variation of the number of protons at $\mathbf{r}$ in the system corresponds to the difference in the modified electrostatic potential between the complex and the isolated moieties. If one considers the "vertical" derivative (vert), for which all geometries in the evaluation of eq 7 are kept rigid, that is, $\{\mathbf{R}_I\}$ is constant, the nuclear contribution to the difference in the modified electrostatic potentials cancels, and one finally finds

$$\left.\frac{\delta E^{\mathrm{int}}}{\delta Z(\mathbf{r})}\right|_{\{\mathbf{R}_I\}} = \mu_{\mathrm{n}}^{\mathrm{int}}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \frac{\rho^{\mathrm{s}}(\mathbf{r}') + \rho^{\mathrm{t}}(\mathbf{r}') - \rho^{\mathrm{t+s}}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{8}$$

This expression is exact and does not involve any fractional atomic numbers per se. Consequently, it is not limited to the use of a plane-wave basis set but could also be evaluated within localized basis function calculations. The numerical applications of this study aim at assessing eq 8 for its meaning and utility when navigating through chemical space in order to control the interaction between a fixed target binding template t and the variable substrate s. For comparison, finite difference estimates, $\mu_{\mathrm{n}}^{\mathrm{fd}}$, of the derivative in eq 5 have also been evaluated for rigid geometries according to

$$\mu_{\mathrm{n}}^{\mathrm{fd}}(\mathbf{r}) = \frac{E^{\mathrm{int}}[Z+dZ] - E^{\mathrm{int}}[Z]}{dZ(\mathbf{r})} \tag{9}$$

where $E^{\mathrm{int}}[Z+dZ] = E^{\mathrm{t+s}}[Z+dZ] - E^{\mathrm{t}}[Z+dZ] - E^{\mathrm{s}}[Z+dZ]$ and $dZ(\mathbf{r}) \rightarrow 0^+$ at all $\mathbf{r}$ where $Z(\mathbf{r}) = 0$.

B. Definitions Employed. Formic acid has been selected to represent a rigid binding target t. This is an arbitrary choice, solely motivated by the fact that formic acid can act as hydrogen-bond donor and acceptor and consequently corresponds to a realistic, meaningful, yet simple (though nontrivial) target model. For the substrate s, we have chosen to limit the chemical space of interest to $N_{\mathrm{e}} = N_{\mathrm{p}} = 10$ and all the molecules containing a central atom "interpolated" between the carbon ($Z_{\mathrm{c}} = 6$) and fluorine ($Z_{\mathrm{c}} = 9$) and a number of hydrogen protons, N_{H}, defined by $N_{\mathrm{H}} = 10 - Z_{\mathrm{c}}$.

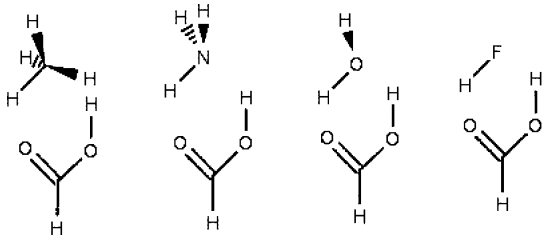

Figure 1. Sketch of the four supramolecular complexes which represent realistic members of all the possible 10-proton systems interacting with formic acid.

This defines an inhomogeneous spatial distribution of 10 protons in the electronic Kohn−Sham Hamiltonian within the Born−Oppenheimer approximation and includes compounds such as CH_4, NH_3, H_2O, and HF. While this restriction of chemical space keeps the number of systems requiring evaluation tractable, there is no loss of generality since the entire scheme is general enough that any other compounds could be used as templates (targets) or substrates.

In order to facilitate the analysis and to ensure smooth alchemical variations, several geometrical constraints have been imposed throughout all calculations. These constraints for determining the interaction energy *vertically* as

$$E^{\mathrm{int}} = E^{\mathrm{fa+s}} - E^{\mathrm{fa}} - E^{\mathrm{s}} \tag{10}$$

that is, without any further geometry optimization of the isolated moieties for the systems displayed in Figure 1, are described in the subsections of section IV. In eq 10, E^{fa} is the energy of the isolated formic acid and $E^{\mathrm{fa+s}}$ is the energy of the complex.

For the alchemical variations of the intermolecular energies, six arbitrary quantum chemical "λ paths" have been considered, as discussed for chemical space in ref 8. The usual parameter λ ($0 \le \lambda \le 1$) is used to measure the progress of the state function of interest, E^{int}, for going reversibly from state A to state B,

$$E^{\mathrm{int}}(\lambda = 0) = E_{\mathrm{A}}^{\mathrm{int}}, \text{ and } E^{\mathrm{int}}(\lambda = 1) = E_{\mathrm{B}}^{\mathrm{int}} \tag{11}$$

where the states are of different chemical composition, see refs 21−23 for discussion. The transmutation of CH_4 into NH_3 is controlled by λ_1; λ_2 controls the $NH_3 \rightarrow H_2O$ transmutation, and λ_3 transforms H_2O into HF. These transformations have been carried out first for geometrically rigid complexes starting with an optimized methane structure and then for a transformation where the structure of the 10-proton system has been relaxed at every point of λ. As the number of protons in the central atom is continuously increased, the atomic number of one of the protons in the surrounding hydrogen atoms is simultaneously decreased from 1 to 0. Which of the hydrogens has been chosen for annihilation can be recognized from inspection of Figure 1.

Alternatively, many other, more efficient or convenient λ paths could have been imagined. For example, three of the

four hydrogen protons in CH_4 could have been simultaneously annihilated while increasing the atomic number of the central atom correspondingly until HF is reached. However, such a path would not enable the system to "pass through" the realistic points in chemical space which correspond to NH_3 and H_2O. Realistic points in chemical space have only integer proton distributions. A systematic way to ensure that a maximal number of realistic points is visited would be to restrict the path to a single certain site in the system until a full proton has been added or removed.

III. Computational Details

All KS−DFT calculations have been carried out using the generalized gradient approximation BLYP[24−26] to the exact exchange-correlation potential as implemented in the plane-wave code CPMD[27] together with the analytical pseudopotentials from refs 28 and 29, a large plane-wave energy cutoff of 100 Ry, periodic boundary conditions, and converged box sizes. London dispersion-corrected atom-centered potentials (DCACPs) as introduced and employed in refs 14−17 and 30 have been used for all atoms. The corresponding reference is the MP2/aug-cc-pVTZ (counterpoise-corrected) level of theory. Noninteger atomic numbers have been obtained through linear scaling of all the parameters $\{\sigma_i\}$ occurring in the analytical pseudopotential of hydrogen, $\{\sigma_i(\lambda)\} = \{\lambda\sigma_i\}$, through linear interpolation between the parameters for carbon, nitrogen, oxygen, and fluorine.

IV. Results and Discussion

A. Intermolecular Energies and Electronic Structure for Integer Atomic Numbers. Total interaction energies, E^{int}, for the four integer atomic number complexes with formic acid and as sketched out in Figure 1 amount to attractive interaction energies of 6.2, 34.1, 33.3, and 48.6 kJ/mol for CH_4, NH_3, H_2O, and HF, respectively. These values have been obtained as follows: first, the geometry of the isolated formic acid was optimized; second, the various substrate (s) geometries were relaxed in the presence of the rigid formic acid and under the constraint that the heavy atoms of the substrate as well as that hydrogen that interacts with the carbonyl oxygen of the formic acid are in the same plane as the formic acid.

At one extreme, CH_4 with its weak octupole moment exhibits only a relatively small attraction to formic acid, while at the other, HF acts with its strong dipole moment as a good hydrogen-bond donor. NH_3 and H_2O both yield comparable interaction energies of ∼34 kJ/mol. This can be rationalized by referring to the relative atomic electronegativities of nitrogen and oxygen. While the NH bond in ammonia is less polar than the OH bond in water, the lone-pair electrons of the nitrogen are more polarizable than the lone pair on the oxygen in H_2O. Consequently, while NH_3 is a weaker hydrogen-bond donor than H_2O, it is also a stronger hydrogen-bond acceptor, resulting in an overall comparable interaction energy.

This explanation is consistent with the effect on the electronic structure due to the binding, $\Delta\rho = \rho^{\text{fa+s}} - \rho^{\text{s}} - \rho^{\text{fa}}$, and along the change from CH_4 to HF as displayed in Figure 2. Specifically, when considering that isosurface

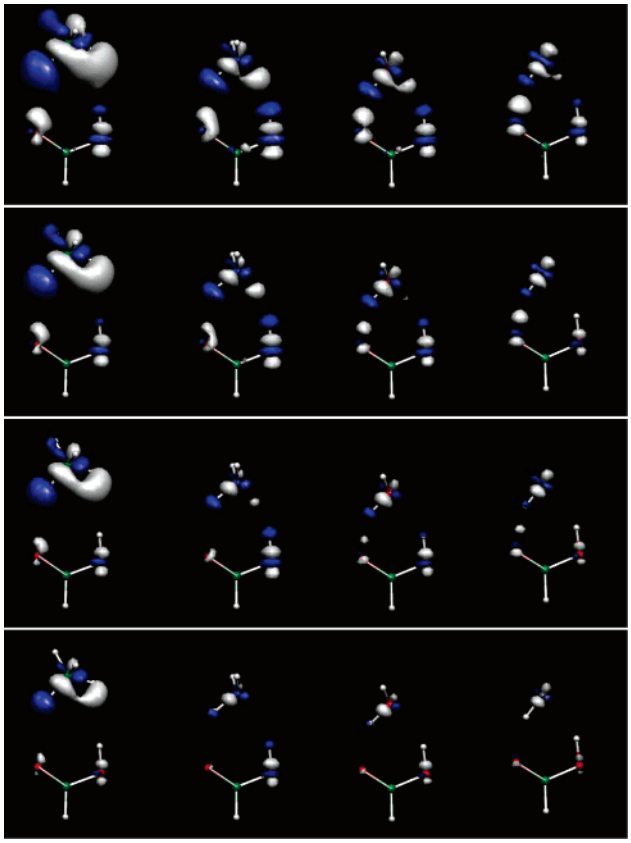

Figure 2. Isosurfaces of the difference in electron density, $\Delta\rho(\mathbf{r})$, due to the interaction of formic acid, $\Delta\rho(\mathbf{r}) = \rho^{fa+s}(\mathbf{r}) - \rho^{fa}(\mathbf{r}) - \rho^{s}(\mathbf{r})$, with the four 10-proton systems CH_4, NH_3, H_2O, and HF (left to right). Negative values are blue and represent depletion of the electron density upon binding; positive values are white and correspond to an increase of electron density due to binding. The isosurfaces are plotted for cutoff values of 5, 7.5, 10, and 15% of the respective absolute maximal value (top to bottom).

which corresponds to a 10% cutoff value (second from the bottom), it becomes evident that, for ammonia, the intermolecular electron density is first increased between the nitrogen lone pair and the acidic hydrogen of formic acid, while for water, the intermolecular electron density is first increased between the water proton and the carbonyl oxygen. Furthermore, an index of the relative strength of the interaction due to the mutual polarization can be estimated from the electronic structure plots with $CH_4 < NH_3 \approx H_2O < HF$: at a 10% cutoff, the lobe of intermolecular density is largest for HF, roughly equal for NH_3 and H_2O, and nonexistent for CH_4.

B. Intermolecular Energies and Electronic Structure for Noninteger particle Distributions. Results for the six continuous distinct alchemical changes (λ paths) of the intermolecular energy are plotted as a function of the atomic number of the heavy atom in Figure 3. In the case of the structurally relaxed λ paths, the first and third λ paths, that is, between CH_4 and NH_3 (λ_1) and between H_2O and HF (λ_3), the interaction energy falls off according to a power law, which is reminiscent of the scaling of atomic energies with atomic numbers. NH_3 and H_2O, however, are connected through a single, very shallow minimum (λ_2), implying that

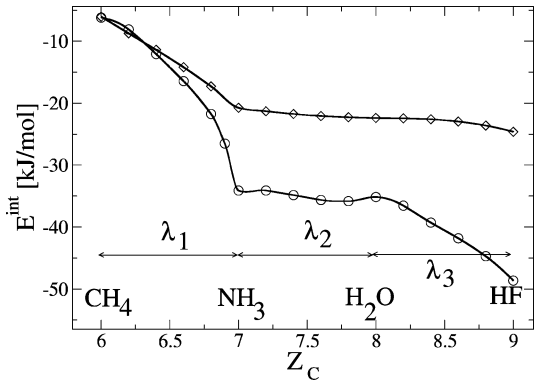

Figure 3. Potential energy of interaction with formic acid for the six λ paths when going from CH_4 ($Z_c = 6$) to HF ($Z_c = 9$) in steps of 0.2 recorded as a function of the atomic number of the central atom, Z_c. The circles correspond to intermolecular energies for substrates which are relaxed in the presence of the rigid formic acid. The diamonds correspond to intermolecular energies obtained for a generally rigid geometry which was optimized for the methane−formic acid complex without any constraint and then kept fixed for all other λ values. The straight lines represent corresponding interpolations.

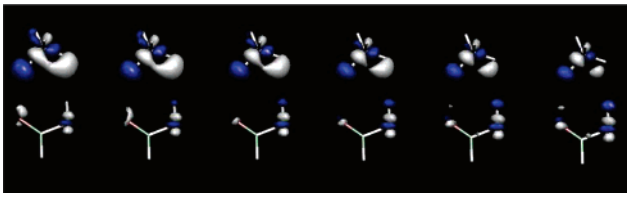

Figure 4. Isosurfaces of the difference in electron density, $\Delta\rho(\mathbf{r})$, due to the interaction with formic acid, $\Delta\rho(\mathbf{r}) = \rho^{fa+s}(\mathbf{r}) - \rho^{fa}(\mathbf{r}) - \rho^{s}(\mathbf{r})$, for the geometrically rigid λ_1 path. Surfaces are displayed for $\lambda_1 = 0.0$ (CH_4), 0.2, 0.4, 0.6, 0.8, and 1.0 (NH_3) (left to right, see also Figure 3). Negative values are blue and represent depletion of the electron density upon binding; positive values are white and correspond to an increase of electron density due to binding. The isosurfaces are plotted for a cutoff value of 10% of the respective absolute maximal value.

the corresponding gradient of the interaction energy, corresponding to the gradient of the plot in Figure 3, must be small between NH_3 and H_2O. In the case of the geometrically rigid λ paths, however, the rapid decrease with λ_1 is much weaker, leading to only ~60% of the interaction energy of ammonia in the relaxed case. λ_2 is a plateau as in the relaxed case. Most interestingly, the effect of the imposed rigidity becomes clearly significant for λ_3, which leads only to very small changes in the interaction energy, implying an underestimation by more than 50% compared to the relaxed formic acid−HF complex.

In Figure 4, isosurfaces of the change in electron density due to the binding, $\Delta\rho = \rho^{fa+s} - \rho^{s} - \rho^{fa}$, are displayed for the geometrically rigid λ_1 path ($CH_4 \rightarrow NH_3$). One can clearly see how the electronic structure continuously adapts to the variation in the proton distribution and how the vanishing hydrogen nucleus of CH_4 loses its influence on the electronic structure.

Overall, the results above indicate the presence of a relatively simple and continuous grand-canonical potential-energy hypersurface. Thus, one can imagine including entropic effects through extended ab initio molecular dynamics schemes. However, derivative discontinuities are expected when switching the λ, that is, when considering the derivatives of the interaction energy in Figure 3 for $E^{int}(\lambda_1 = 1, \lambda_2 = \lambda_3 = 0; NH_3)$ and $E^{int}(\lambda_1 = \lambda_2 = 1, \lambda_3 = 0; H_2O)$. In other words

$$\lim_{d\lambda\to 0} \frac{E^{int}(\lambda_1 = 1 - d\lambda, \lambda_2 = \lambda_3 = 0)}{d\lambda} \neq \lim_{d\lambda\to 0} \frac{E^{int}(\lambda_1 = 1, \lambda_2 = d\lambda, \lambda_3 = 0)}{d\lambda} \quad (12)$$

and

$$\lim_{d\lambda\to 0} \frac{E^{int}(\lambda_1 = 1, \lambda_2 = 1 - d\lambda, \lambda_3 = 0)}{d\lambda} \neq \lim_{d\lambda\to 0} \frac{E^{int}(\lambda_1 = \lambda_2 = 1, \lambda_3 = d\lambda)}{d\lambda} \quad (13)$$

In the case of NH_3, the left side of eq 12 describes the trend in the interaction energy when creating a fourth proton and decreasing the atomic number of the central nitrogen atom toward carbon. This tendency appears to be uncoupled from the tendency of the interaction energy to annihilate the third proton in NH_3 and to increase the atomic number of the central nitrogen atom toward oxygen (right side of that equation). The same is true for eq 13. Consequently, if the force of a given λ transformation is to be measured, care should be taken regarding the correct choice of the corresponding linear combinations of alchemical potentials.

The interaction energy for the geometrically rigid and relaxed alchemical paths in Figure 3 diverges rapidly already between CH_4 and NH_3. This suggests that, for a quantitative integration over the forces of the state function along the transformation, it is crucial to "decouple" frequently the canonical geometrical degrees of freedom from the grand-canonical variations in particle distribution space. Hence, if adiabaticity is desired between the degrees of freedom corresponding to momenta and positions and the degrees of freedom corresponding to the particle number,[23] the step in λ space must be chosen to be sufficiently small. This raises the problem of dealing with noninteger proton densities. That is, while a molecular grand-canonical scheme that is restricted to integer atomic numbers might yield the right trends, it is likely to be quantitatively inaccurate when employed within a molecular dynamics[23] or thermodynamic integration scheme.[21,22] This is an issue that will be dealt with in subsequent studies.

C. The Nuclear Chemical Potential for Integer Proton Distributions. Isosurfaces of the intermolecular nuclear chemical potentials, $\mu_n^{int}(\mathbf{r})$, according to eq 8 are shown in Figure 5. One can recognize how the change in interaction energy due to a positive charge perturbation varies from CH_4 to HF. When the isosurface cutoff is gradually increased, one can also identify the regions and the number of extrema of $\mu_n^{int}(\mathbf{r})$. In order to assess the correctness of eq 8 for

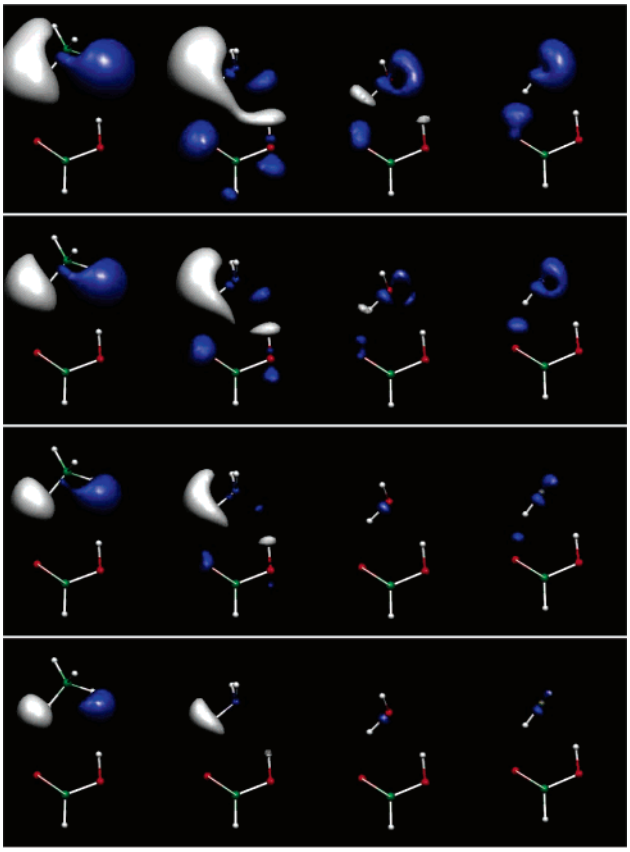

Figure 5. Isosurfaces of the nuclear chemical potential of the interaction energy of formic acid (according to eq 8) with the four 10-proton systems CH_4, NH_3, H_2O, and HF (left to right). The surfaces are plotted for cutoff values of 30, 40, 50, and 60% of the respective absolute maximal value (top to bottom). Negative values are blue and predict an increase; positive values are white and predict a decrease of interaction energy due to proton density variation.

computing this quantity, numerical values of the intermolecular nuclear chemical potential (displayed in Figure 5) at spatial positions where the function is maximal or minimal are presented in Table 1 along with finite difference estimates resulting from the change in interaction energy due to the presence of 0.1% of a proton at the same spatial locations according to eq 9. The finite difference results deviate negligibly by at most 0.2 kJ/mol from the results computed according to eq 8.

Estimates for alchemical potentials, $\{\mu_n^{int}(\mathbf{R}_I)\}$, have likewise been obtained from volumetric data (at grid points closest to corresponding atomic positions) for the geometrically rigid alchemical path. The "left-hand-side derivative" results for the integer atomic number cases CH_4, NH_3, and H_2O are reported in Table 1, together with finite difference results due to a 1% proton perturbation of the atomic number. Clearly, while the overall agreement with the finite difference prediction is still acceptable, it is less accurate than it is for the intermolecular nuclear chemical potential at spatial positions where there is no nucleus. Furthermore, the agreement with finite difference results is significantly better when applied to vanishing hydrogens than when applied to an increase in atomic number by the addition of protons to heavier nuclei.

Table 1. Intermolecular Nuclear Chemical Potentials, $\mu_n^{int}(\mathbf{r})$

	CH_4	NH_3	H_2O	HF
$\mu_n^{int}(\mathbf{r}_{min})$	−32.4	−42.2	−122.8	−156.6
$\mu_n^{fd}(\mathbf{r}_{min})$	−32.3	−42.1	−123.0	−156.5
$\mu_n^{int}(\mathbf{r}_{max})$	34.6	61.8	57.9	32.4
$\mu_n^{fd}(\mathbf{r}_{max})$	34.7	61.9	58.0	32.4
$\mu_n^{int}(\mathbf{R}_H)$	−21.5	8.7	−6.7	
$\mu_n^{fd}(\mathbf{R}_H)$	−18.5	7.2	−7.7	
$\mu_n^{int}(\mathbf{R}_c)$	−2.5	−4.7	−13.6	
$\mu_n^{fd}(\mathbf{R}_c)$	−7.3	−9.5	−8.9	

[a] Values are obtained according to eq 8 and as extracted from the volumetric data displayed in Figure 5 at positions at which μ_n^{int} is maximal,$\mathbf{r}_{max}$, or Minimal, $\mathbf{r}_{min}$. For comparison, μ_n^{fd} is given as a finite difference perturbation of the energy of interaction due to the presence of 0.1% of a proton. For the geometrically fixed alchemical path, intermolecular alchemical potentials, $\mu_n^{int}(\mathbf{R})$, with respect to an increase in proton number are given at the positions of the vanishing hydrogens, $\mu_n^{int}(\mathbf{R}_H)$, and at the position of the central atom, $\mu_n^{int}(\mathbf{R}_c)$, along with corresponding finite difference estimates. Since no λ path involves a Z_c larger than that for fluorine, no values are presented for HF. A linear combination of the alchemical potentials would correspond to the right-hand-side derivatives of the geometrically rigid intermolecular energy curve in Figure 3. All μ_n^{int} values are given in kJ/mol.

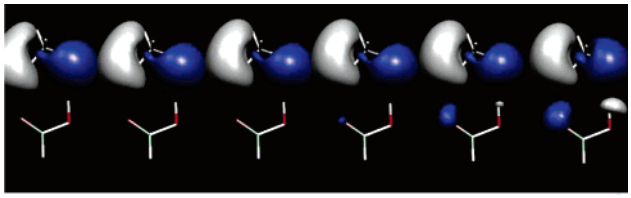

Figure 6. Isosurfaces of the nuclear chemical potential of the interaction energy, $\mu_n^{int}(\mathbf{r})$ (according to eq 8), for the geometrically rigid λ_1 path from CH_4 to NH_3. $\lambda_1 = 0.0$ (CH_4), 0.2, 0.4, 0.6, 0.8, and 0.1 (NH_3) (left to right, see Figure 3). The surfaces are plotted for a cutoff value of 30% of the respective absolute maximal value. Negative values are blue and predict an increase; positive values are white and predict a decrease of interaction due to proton density variation.

D. The Nuclear Chemical Potential for Noninteger Proton Distributions. In order to investigate this effect in more detail, alchemical potentials of the intermolecular energy have been computed according to eq 8, $\mu_n^{int}(\mathbf{R}_I)$, and according to the finite difference in eq 9 with 1% proton perturbation, $\mu_n^{fd}(\mathbf{R}_I)$, along the geometrically rigid alchemical transformations CH_4 ($\lambda_1 = 0$) → NH_3 ($\lambda_1 = 1$), NH_3 ($\lambda_2 = 0$) → H_2O ($\lambda_2 = 1$), and H_2O ($\lambda_3 = 0$) → HF ($\lambda_3 = 1$) in λ steps of 0.0, 0.2, 0.4, 0.6, and 0.8 (see Figure 3). For the purpose of further illustration, isosurfaces of $\mu_n^{int}(\mathbf{r})$ along the λ_1 path are also displayed in Figure 6.

The overall correlation of these results is remarkably good (see Figure 7), and the derivative discontinuities in Figure 3 are reproduced. However, the correlation is not as good as it is for the nonalchemical nuclear chemical potentials, and we find that the aforementioned trend, namely, that the alchemical potential at the vanishing hydrogen correlates better with the finite difference results than the alchemical potential of the central atom, is also confirmed. Specifically, Figure 7 appears to indicate a systematic error in the computation of the alchemical potential of the central atom

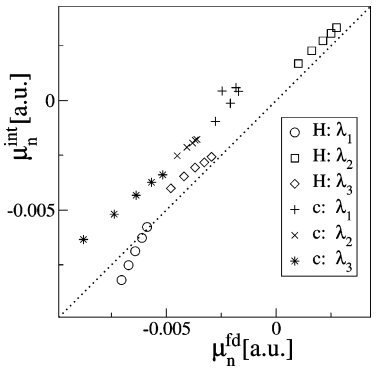

Figure 7. Alchemical potentials according to eq 8, $\{\mu_n^{int}(\mathbf{R}_I)\}$, versus finite difference results, μ_n^{fd}, at the position of the vanishing hydrogens (H) and the central atom (c). Results have been collected for the three, geometrically rigid transformations CH_4 ($\lambda_1 = 0$) → NH_3 ($\lambda_1 = 1$), NH_3 ($\lambda_2 = 0$) → H_2O ($\lambda_2 = 1$), and H_2O ($\lambda_3 = 0$) → HF ($\lambda_3 = 1$) in λ steps of 0.0, 0.2, 0.4, 0.6, and 0.8.

according to eq 8; that is, there is a constant shift of ~0.002 au by which the finite difference results are underestimated. The strongest deviation is found for the λ_1 path.

A possible reason for this behavior might be the use of the electrostatic potential at positions where pseudopotentials are localized, that is, regions in space where the electron density is artificially reduced to zero, which in turn leads to an approximate result for the otherwise exact result from eq 8. Alternatively, the geometrical constraints might be a reason. Methods to remedy this deficiency will be considered in future studies using either eq 8 directly for electron densities obtained within all electron calculations, exploiting an appropriate linear combination of total potential energy derivatives with respect to pseudopotential parameters as proposed in refs 5 and 30, or accounting for geometrical relaxation.

V. Conclusion

Using the molecular grand-canonical ensemble density functional theory framework introduced in refs 5 and 8, we have studied potential energies of interaction and how they are connected for continuously varying a 10-proton system from CH_4 to HF via NH_3 and H_2O in the presence of a formic acid binding target. The effects on interaction energies, electronic structures, and electrostatic potentials due to the alchemical variations for gradually transforming CH_4 into HF (via NH_3 and H_2O) have been illustrated.

Predictions of the nuclear chemical potential of the interaction energy compare well to finite difference results, showing unambiguously the suitability of the molecular grand-canonical DFT ensemble approach for indexing where in space protons should be annihilated or created if an enhancement in the interaction energy is sought. While the *alchemical* potential of second-row elements leads to expected trends, it is estimated only approximately within the presented computational setup. Further investigations will be required in order to achieve satisfactory accuracy in predicting this quantity, which is crucial for the guided exploration of chemical space.

The advantage of such a molecular grand-canonical scheme compared to individual geometry optimizations is that, in order to perform the latter, one must know which compounds to consider a priori. A molecular grand-canonical theory, on the other hand, not only provides the means to perform geometry optimizations but can also help to identify those compounds which are more relevant than others.

We conclude that, in combination with the theoretical framework and the results of the calculations presented in ref 8, numerical evidence has been collected that, within molecular grand-canonical ensemble DFT, the elementary particle number variation of electrons and protons and of their corresponding intensive conjugates, the electronic and the nuclear chemical potential, is perfectly feasible. The procedure presented here should prove useful for devising more efficient and de novo compound design algorithms involving the unbiased control of intermolecular energies via compositional tuning.

Acknowledgment. O.A.v.L. acknowledges the SNF postdoctoral research grant no. PBEL2-110243. M.E.T. acknowledges funding from NSF CHE-0121375 and NSF CHE-0310107.

References

(1) Kirkpatrick, P.; Ellis, C. *Nature* **2004**, *432*, 823.

(2) Wolverton, C.; Zunger, A. *Phys. Rev. Lett.* **1998**, *81*, 606.

(3) Franceschetti, A.; Zunger, A. *Nature* **1999**, *402*, 60.

(4) Jóhannesson, G. H.; Bligaard, T.; Ruban, A. V.; Skriver, H. L.; Jacobsen, K. W.; Nørskov, J. K. *Phys. Rev. Lett.* **2002**, *88*, 255506.

(5) von Lilienfeld, O. A.; Lins, R.; Rothlisberger, U. *Phys. Rev. Lett.* **2005**, *95*, 153002.

(6) Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. *J. Am. Chem. Soc.* **2006**, *128*, 3228.

(7) Keinan, S.; Hu, X.; Beratan, D. N.; Yang, W. *J. Phys. Chem. A* **2007**, *111*, 176.

(8) von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 154104.

(9) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(10) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(11) Koch, W.; Holthausen, M. C. Hydrogen Bonds and Weakly Bound Systems. In *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2001; Vol. B, pp 217−238.

(12) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.

(13) Ortmann, F.; Bechstedt, F.; Schmidt, W. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 205101.

(14) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(15) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 195119.

(16) von Lilienfeld, O. A.; Andrienko, D. *J. Chem. Phys.* **2006**, *124*, 054307.

(17) Tkatchenko, A.; von Lilienfeld, O. A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *73*, 153406.

(18) Press, W. H.; Teukoldky, S. A.; Vetterling, W. T.; Flannery, B. P. Minimization or Maximization of Functions. In *Numerical Recipes in Fortran 77*, 2nd ed.; Cambridge University Press: Cambridge, U. K., 1992; Vol. 1, pp 387−436.

(19) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. *Phys. Rev. Lett*. **1982**, *49*, 1691.

(20) Perdew, J. P.; Levy, M. *Phys. Rev. Lett*. **1983**, *51*, 1884.

(21) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300.

(22) Pearlman, D. A.; Govinda Rao, B. Free Energy Calculations: Methods and Applications; Jorgensen, W. L. Free Energy Changes in Solution. In *Encyclopedia of Computational Chemistry*, 2nd ed.; Ragué von Schleyer, P., Allinger, N., Clark, T., Gasteiger, J., Kollman, P., Schaefer, H. F., III, Eds.; John Wiley & Sons Ltd: West Sussex, U. K., 1998; Vol. 2, pp 1036−1070.

(23) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. *J. Chem. Phys.* **2006**, *125*, 074115.

(24) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(25) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(26) Colle, R.; Salvetti, D. *Theor. Chim. Acta* **1975**, *37*, 329.

(27) Hutter, J. *Computer code CPMD*, version 3.11; IBM Corp./Max-Planck Institute-FKF: Stuttgart, Germany, 1990−2006/1997−2001.http://www.cpmd.org (accessed Jan. 1, 2007).

(28) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev.: Condens. Matter Mater. Phys.* **1996**, *54*, 1703.

(29) Krack, M. *Theor. Chim. Acta* **2005**, *114*, 145.

(30) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.

(31) Langreth, D. C.; Dion, M.; Rydberg, H.; Schroder, E.; Hyldgaard, P.; Lundqvist, B. I. *Int. J. Quantum Chem.* **2005**, *101*, 599.

(32) Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2005**, *122*, 114102.

CT700002C