JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Exploring Chemical Space with the Alchemical Derivatives

Robert Balawender,*[,†] Meressa A. Welearegay,[†] Michał Lesiuk,[‡] Frank De Proft,[§] and Paul Geerlings*[,§]

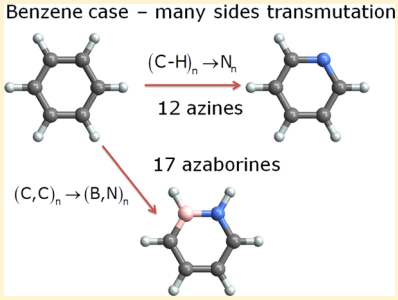
[†]Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka 44/52, PL-01-224 Warsaw, Poland

[‡]Faculty of Chemistry, University of Warsaw, Pasteura 1, PL-02-093 Warsaw, Poland

[§]Eenheid Algemene Chemie, Vrije Universiteit Brussel, Faculteit Wetenschappen, Pleinlaan 2, B-1050 Brussels, Belgium

Supporting Information

**ABSTRACT:** In this paper, we verify the usefulness of the alchemical derivatives in the prediction of chemical properties. We concentrate on the stability of the transmutation products, where the term "transmutation" means the change of the nuclear charge at an atomic site at constant number of electrons. As illustrative transmutations showing the potential of the method in exploring chemical space, we present some examples of increasing complexity starting with the deprotonation, continuing with the transmutation of the nitrogen molecule, and ending with the substitution of isoelectronic B−N units for C−C units and N units for C−H units in carbocyclic systems. The basis set influence on the qualitative and quantitative accuracies of the alchemical predictions was investigated. The alchemical deprotonation energy (from the second order Taylor expansion) correlates well with the vertical deprotonation energy and can be used as a preliminary indicator for the experimental deprotonation energy. The results of calculations for the BN derivatives of benzene and pyrene show that this method has great potential for efficient and accurate scanning of chemical space.




## I. INTRODUCTION

The development of methods for systematic and rational classification of chemical compounds and the search for compounds with desired properties is a fundamental task in chemical, material, and pharmaceutical research.[1−6] Advances in this field led to idea of *chemical space* (CS), which was defined as set of all possible combinations of chemical elements building all geometric isomers,[7,8] but the purposeful design of molecules with optimized properties is daunting because the number of accessible stable molecules is immense.[9,10] From the quantum chemical point of view, exploration of the complete CS is a challenging task as computational timings become prohibitively large (assuming that to predict some property of a molecule at least a single-point calculation has to be performed, often demanding high level ab initio methods). Thus, the straightforward exploration of CS becomes a nearly impossible procedure, and more sophisticated methods such as simulated annealing,[6] genetic algorithms,[11] linear combinations of atomic potentials,[1] or optimization based on alchemical gradients[12−15] have to be used to cope with this problem in an efficient way.

A general description of chemical space involving changes in composition and geometry is possible using density functional reactivity theory (DFRT), often called conceptual density functional theory (DFT).[16−21] Conceptual DFT is based on the study of $E = E[N,v]$ (the energy vs number of electrons ($N$) and external potential $v(\mathbf{r})$ functional); more precisely, it is based on the change the functional undergoes when the number of electrons and/or the external potential is varied, concretized by the corresponding response functions. Traditionally, DFRT has been concerned primarily with the prediction of phenomena associated with electron transfer, that is, responses to changes in the number of electrons. Recently, numerous workers have begun to focus on the response of a system to a change in the external potential,[22−24] which for example results from shifts of the nuclear positions in a molecule, from the approach of reagents or from changing the nuclear charges. Techniques have been developed for the numerical and analytical evaluation of the corresponding response functions including the so-called linear response function, which recently received a lot of attention by some of the present authors.[25−31] Changing the location of the point charges can be useful in geometry optimizations.[32] Addition of a positive point charge on the periphery of the molecule[22,32] is useful for discerning its nucleophilicity and its Brønsted−Lowry acidity. In the present study, the focus is on the modification of the nuclear charge, which is used in the 'alchemical transformations' associated with molecular design.[12,14,15,33,34]

In the alchemical gradients method, developed by von Lilenfeld et al.,[12,33] derivatives of the energy (and other properties) with respect to nuclear charge distribution are employed when exploring CS. When these derivatives are computed for particular molecule from CS, properties of large fraction of the 'surrounding' compounds can be calculated with no additional workload. This approach was used to study reaction energetics[15] energy gradients and to predict orbital energies.[35] Although results obtained by these authors were very promising, no general route was established to treat the higher-order derivatives of the energy or to calculate derivatives of other properties (such as orbital energies) without introduction of some

Received: August 8, 2013

   dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

'empirical' parameters. In a recent paper,[36] a coupled perturbed (CP) scheme for the Hartree−Fock and Kohn−Sham calculation of the second and the third order derivatives with respect to nuclear charges was proposed. The CP equations, which are also known as the response equations, were also successfully used in the past in calculations of the relaxation effect to chemical reactivity indices involving derivatives with respect to the total electron number.[37−41]

In this paper, we verify the usefulness of the alchemical derivatives in the prediction of chemical properties. We concentrate on the stability of the transmutation products, where the term "transmutation" means the change of the nuclear charge at an atomic site. Our interest is restricted only to molecules containing first and second row atoms, but the alchemical calculation scheme is also applicable for systems where the effective core potential approximation is used.[42] As illustrative transmutations showing the potential of the method in exploring chemical space, we present some examples of increasing complexity starting with the deprotonation, continuing with the transmutation of the nitrogen molecule, and ending with the substitution of isoelectronic B−N units for C−C units and N units for C−H units in carbocyclic systems.

The paper is organized as follows. In Section II, we present the theoretical and computational background. In Section III, illustrative examples are discussed. Finally, in Section IV general conclusions and summary is given.

Atomic units are used throughout the paper, except for the deprotonation section, where kJ/mol is used. Naming conventions for the heteromonocycles and for the fused ring systems is in accordance with the IUPAC recommendations.[43,44]

## II. THEORY

Let us at first define precisely the basic concepts used in this paper. The *M-dimensional conformational space* is defined as

$$C_M \equiv \{(\mathbf{R}^M, \mathbf{Z}^M) | \mathbf{R}^M \in \mathcal{R}_M, \mathbf{Z}^M \in \mathcal{Z}_M\} \tag{1}$$

where $\mathcal{R}_M$ is a $3M$-dimensional spatial configuration space

$$\mathcal{R}_M \equiv \{\mathbf{R}^M = (\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_M) | \forall\, A, B \in \{1, 2, .., M\}, A \neq B: \mathbf{R}_A \neq \mathbf{R}_B\} \tag{2}$$

and $\mathcal{Z}_M$ is an $M$-dimensional charge configuration space

$$\mathcal{Z}_M \equiv \{\mathbf{Z}^M = (Z_1, Z_2, ..., Z_M) | Z_A \in \mathbb{R}, Z_A \geq 0\} \tag{3}$$

where $\boldsymbol{R}_A$ and $Z_A$ denote the nuclear locations and the corresponding nuclear charges, respectively. A fractional charge is allowed in the present description. The site with the charge equal zero is called a vacancy and collapsed nuclear coordinates are excluded from $\mathcal{R}_M$.

To each point of the conformational space $(\boldsymbol{R}^M,\boldsymbol{Z}^M)$ corresponds an external potential, defined as

$$v(\mathbf{r}, \mathbf{Z}^M, \mathbf{R}^M) = -\sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \tag{4}$$

The nuclear−nuclear repulsion energy is

$$V_{\mathrm{nn}}[\mathbf{Z}^M, \mathbf{R}^M] = \sum_A \sum_{B>A} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \tag{5}$$

and the total proton number is defined as $N_p[\mathbf{Z}^M] = \sum_A Z_A$.

Every point of the conformational space can be associated with some electronic configuration characterized by the electron number $N$ and some characteristics in the spin space (e.g., singlet, doublet, etc.) This definition can be extended to the systems with noninteger electron number using the ensemble approach.[45−47] The external potential uniquely determines the electronic Hamiltonian, which allows to, at least in principle, solve the molecular Schrödinger equation for a given number of electrons $N$. Thus, the potential function $v(\boldsymbol{r})$ encodes all of the chemical information for a given $N$.

Due to the Hohenberg−Kohn (HK) variational principle,[48] the ground state energy is given by

$$E[N, v] = \min\{\tilde{E}[\rho; v] | \int \rho(\mathbf{r}) \mathrm{d}\mathbf{r} = N\} = \tilde{E}[\rho^{\mathrm{g}}; v] \tag{6}$$

with $\tilde{E}[\rho,v] \equiv F[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) \mathrm{d}\mathbf{r}$, the energy functional for a system moving in an external potential $v(\boldsymbol{r})$. The minimizer, $\rho^{\mathrm{g}}$, represents the GS density of the system. $F[\rho]$ is a universal functional, which contains the sum of the kinetic energy and the electron−electron repulsion energy. The sum of the electronic energy $E$ and the nuclear−nuclear repulsion energy $V_{\mathrm{nn}}$ (without another external filed) is the total energy $W$ of a system,

$$W[N, \mathbf{Z}^M, \mathbf{R}^M] \equiv E[N, v[\mathbf{Z}^M, \mathbf{R}^M]] + V_{\mathrm{nn}}[\mathbf{Z}^M, \mathbf{R}^M] \tag{7}$$

Consequently, the independent variables are the coordinates in the conformational space $(\boldsymbol{R}^M,\boldsymbol{Z}^M)$ and the electron number, $N$.

In this paper, the definition of transmutation extends beyond the historical definition of the transmutation as the conversion of base metals into gold or silver (the origin why these derivative are called alchemical). In our treatment, transmutation means moving from one point to another in the charge space, and it is characterized by a change of the charge vector—*the transmutation vector*

$$\mathrm{d}\mathbf{Z}^M = (\mathrm{d}Z_1, \mathrm{d}Z_2, ..., \mathrm{d}Z_M) \tag{8}$$

where, in general, $\mathrm{d}Z_A$ can be any real number. The change in the total energy in the canonical ensemble, at constant electron number ($\mathrm{d}N = 0$) and frozen geometry ($\mathrm{d}\boldsymbol{R}^M = 0$), can be computed using the functional Taylor series as

$$\begin{aligned} \mathrm{d}W &= W[N, \mathbf{Z}^M + \mathrm{d}\mathbf{Z}^M, \mathbf{R}^M] - W[N, \mathbf{Z}^M, \mathbf{R}^M] \\ &= \sum_A \left(\frac{\partial W}{\partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A + \frac{1}{2} \sum_{AB} \left(\frac{\partial^2 W}{\partial Z_B \partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A \mathrm{d}Z_B \\ &+ (\text{higher order terms}) = \left\{\sum_A \left(\frac{\partial V_{\mathrm{nn}}}{\partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A \right. \\ &\left. + \frac{1}{2} \sum_{AB} \left(\frac{\partial^2 V_{\mathrm{nn}}}{\partial Z_B \partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A \mathrm{d}Z_B \right\} \\ &+ \left[\sum_A \left(\frac{\partial E}{\partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A + \frac{1}{2} \sum_{AB} \left(\frac{\partial^2 E}{\partial Z_B \partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}Z_A \mathrm{d}Z_B \right] \\ &+ ..., \end{aligned} \tag{9}$$

where $\mathbf{Z}_R^M$ means that all other charges are kept fixed.

The terms in braces, {}, are the first- and the second-order terms resulting from $V_{\mathrm{nn}}$:

$$\mu_A^{\mathrm{al,nuc}}[\mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial V_{\mathrm{nn}}[\mathbf{Z}^M, \mathbf{R}^M]}{\partial Z_A}\right)_{\mathbf{Z}_R^M,\mathbf{R}^M} = \sum_{B \neq A} \frac{Z_B}{R_{AB}} \tag{10}$$


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

$$\eta_{AB}^{\mathrm{al,nuc}}[\mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial^2 V_{\mathrm{nn}}[\mathbf{Z}^M, \mathbf{R}^M]}{\partial Z_B \partial Z_A}\right)_{\mathbf{Z}_R^M, \mathbf{R}^M} = (1 - \delta_{AB})\frac{1}{R_{AB}} = \eta_{BA}^{\mathrm{al,nuc}} \tag{11}$$

where $|\boldsymbol{R}_A - \boldsymbol{R}_B| = R_{AB}$. The higher-order derivatives are trivially equal to zero. The terms in brackets, [], are the derivatives of the electronic energy, which depend on $\boldsymbol{Z}^M$ via the external potential, $v(\boldsymbol{r},\boldsymbol{Z}^M,\boldsymbol{R}^M)$. The electronic contribution to the electrostatic potential at site $A$ (the nuclear-electron attraction energy per unit nuclear charge,[49] also involved in the diamagnetic shielding of an atom in a molecule[50]) and the effective 'screened' potential defined by the linear response kernel are

$$\mu_A^{\mathrm{al,el}}[N, \mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial E}{\partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} = \int\left(\frac{\delta E[N, v]}{\delta v(\mathbf{r})}\right)_N \left(\frac{\partial v(\mathbf{r})}{\partial Z_A}\right)_{\mathbf{Z}_R^M,\mathbf{R}^M} \mathrm{d}\mathbf{r} = -\int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_A|}\mathrm{d}\mathbf{r} \tag{12}$$

$$\eta_{AB}^{\mathrm{al,el}}[N, \mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial^2 E}{\partial Z_B \partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} = \int \frac{(\partial\rho(\mathbf{r})/\partial Z_A)_{\mathbf{Z}_R^M,\mathbf{R}^M}}{|\mathbf{r} - \mathbf{R}_B|}\mathrm{d}\mathbf{r} = \int \frac{(\partial\rho(\mathbf{r})/\partial Z_B)_{\mathbf{Z}_R^M,\mathbf{R}^M}}{|\mathbf{r} - \mathbf{R}_A|}\mathrm{d}\mathbf{r} = \eta_{BA}^{\mathrm{al,el}} \tag{13}$$

Here, the relation $(\delta E[N,v]/\delta v(\mathbf{r}))_N = \rho(\mathbf{r};N,v)$ and the linear dependence of $v(\boldsymbol{r})$ on the nuclear charge:

$$\left(\frac{\partial v(\mathbf{r}, \mathbf{Z}^M, \mathbf{R}^M)}{\partial Z_A}\right)_{\mathbf{Z}_R^M,\mathbf{R}^M} = -\frac{1}{|\mathbf{r} - \mathbf{R}_A|} \tag{14}$$

are used.

Finally, the alchemical derivatives in their global version are as follows (the superscript 'al' means alchemical):

The alchemical potential of nucleus at position $\boldsymbol{R}_A$:[12,33]

$$\mu_A^{\mathrm{al}}[N, \mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial W}{\partial Z_A}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} = \mu_A^{\mathrm{al,el}} + \mu_A^{\mathrm{al,nuc}} \tag{15}$$

which measures the transmutational tendency for the atom at position $\boldsymbol{R}_A$ in the molecule. If there is a vacancy at this position, it is related to the proton affinity.

The alchemical hardness[33]

$$\eta_{AB}^{\mathrm{al}}[N, \mathbf{Z}^M, \mathbf{R}^M] \equiv \left(\frac{\partial^2 W}{\partial Z_A \partial Z_B}\right)_{N,\mathbf{Z}_R^M,\mathbf{R}^M} = \eta_{AB}^{\mathrm{al,el}} + \eta_{AB}^{\mathrm{al,nuc}} = \eta_{BA}^{\mathrm{al}} \tag{16}$$

which measures the transmutational resistivity for the atoms at the positions $\boldsymbol{R}_A$ and $\boldsymbol{R}_B$ in the molecule. The alchemical hardness $\eta_{AB}^{\mathrm{al}}$ is simply the effective 'screened' potential.

Higher-order terms are usually assumed to be qualitatively unimportant, although there is a growing realization that the "chemical perturbations" are sometimes too large for this to be true.[21,23]

## III. ILLUSTRATIVE EXAMPLES

To illustrate the usefulness of the alchemical derivatives in exploring chemical space, the deprotonation, the transmutation of the nitrogen molecule, and the substitution of isoelectronic B−N units for C−C units and N units for C−H units were chosen as examples of increasing complexity in the number of sites that are transmuted. To preserve the predictive power of the Taylor expansion and to preserve a realistic perspective from a chemical point of view, only molecules at their equilibrium geometries are used as the basis for an expansion. Moreover, changes in the atomic charges are limited, $\mathrm{d}Z_i \in \{-1,0,1\}$. A $(\boldsymbol{R}^M,\boldsymbol{Z}^M+\mathrm{d}\boldsymbol{Z}^M)$ point in the conformational space will be named a transmutant of the molecule described by the $(\boldsymbol{R}^M,\boldsymbol{Z}^M)$ point in this space. The *alchemical transmutation energy* connected with the displacement from $(\boldsymbol{R}^M,\boldsymbol{Z}^M)$ to $(\boldsymbol{R}^M,\boldsymbol{Z}^M+\mathrm{d}\boldsymbol{Z}^M)$ is defined as

$$D^{\mathrm{al}}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M] \equiv \sum_i \mu_i^{\mathrm{al}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M + \frac{1}{2}\sum_{ij}\eta_{ij}^{\mathrm{al}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M\mathrm{d}Z_j^M = D^{\mathrm{al},\mu}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M] + D^{\mathrm{al},\eta}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M], \tag{17}$$

where $D^{\mathrm{al},\mu}$and $D^{\mathrm{al},\eta}$ are the contributions from the first and second derivatives, respectively.

The energy can be decomposed also into the electronic and the nuclear parts:

$$\begin{aligned} D^{\mathrm{al}}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M] &= D^{\mathrm{al,el}}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M] \\ &+ D^{\mathrm{al,nuc}}[\mathbf{R}^M, \mathbf{Z}^M, \mathrm{d}\mathbf{Z}^M] = \left(\sum_i \mu_i^{\mathrm{al},\mu,\mathrm{el}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M \right. \\ &\left. + \frac{1}{2}\sum_{ij}\eta_{ij}^{\mathrm{al},\eta,\mathrm{el}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M\mathrm{d}Z_j^M\right) \\ &+ \left(\sum_i \mu_i^{\mathrm{al},\mu,\mathrm{nuc}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M \right. \\ &\left. + \frac{1}{2}\sum_{ij}\eta_{ij}^{\mathrm{al},\eta,\mathrm{nuc}}[\mathbf{R}^M, \mathbf{Z}^M]\mathrm{d}Z_i^M\mathrm{d}Z_j^M\right) \end{aligned} \tag{18}$$

The components in the last line will be abbreviated as $D^{\mathrm{al},\mu,\mathrm{el}}$, $D^{\mathrm{al},\eta,\mathrm{el}}$, $D^{\mathrm{al},\mu,\mathrm{nuc}}$, $D^{\mathrm{al},\eta,\mathrm{nuc}}$, respectively. The component resulting from the wave function relaxation, $D^{\mathrm{al},\eta,\mathrm{el}}$, is calculated using the coupled perturbed self-consistent field theory (see ref 36 for computing details and basis set and weights dependencies on the nuclear charge). All calculations were done using locally modified version of GAMESS program package.[51]

**III.A. Deprotonation Energy.** In analogy to the vertical ionization energy, the energy change corresponding to ionization reaction leading to formation of the ion in a configuration which is the same as the equilibrium geometry of the ground state neutral molecule, the *vertical deprotonation energy* is defined as the energy change associated with the deprotonation reaction (@ means at equilibrium geometry of)

$$\mathrm{AH} \rightarrow \mathrm{A}^- + \mathrm{H}^+ \qquad D_{\mathrm{AH}}^{\mathrm{ver,@AH}} = W_{\mathrm{A}^-}^{\mathrm{@AH}} - W_{\mathrm{AH}}^{\mathrm{@AH}} \tag{19}$$

where $W_{\mathrm{AH}}^{\mathrm{@AH}}$, $W_{\mathrm{A}^-}^{\mathrm{@AH}}$ are the total energies of the acid and its conjugate base at the equilibrium geometry of the acid, respectively. $D^{\mathrm{ver}}$ is a quantity defined at 0 K, and is therefore not strictly related to the experimental deprotonation energy,


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

which is the enthalpy change of this reaction. The experimental deprotonation energy can be approximated by ab initio calculations at room temperature as follows

$$D_{AH}^{exp} \approx D_{AH}^{0} \equiv D_{AH}^{ver,@AH} + \Delta W_{A^-}^{rel} + \Delta H_{AH}^{A^-} + H_{corr}^{H^+} \quad (20)$$

where $\Delta W_{A^-}^{rel} = (W_{A^-}^{@relA^-} - W_{A^-}^{@AH})$ is the geometry relaxation correction, $\Delta H_{AH}^{A^-} = (H_{corr}^{A^-} - H_{corr}^{AH})$ is the enthalpy correction and $H_{corr}^{H^+}$ is the calculated gas-phase enthalpy of the proton at room temperature.

The vertical deprotonation energy can be approximated using the Taylor expansion. For the $M$-atomic system with $m$ hydrogens ($1 \le m \le M$) and the nuclear charge vector, $\mathbf{Z}_{AH_m}^{M} = (1,...,1,Z_{m+1},...,Z_M)$, the annihilation of the proton at position $\mathbf{R}_i$ can be described by the transmutation vector

$$d\mathbf{Z}_{AH_m,i}^{M} = (0, ..., dZ_i = -1, 0, ...), \quad 1 \le i \le m \quad (21)$$

The *alchemical deprotonation energy* is defined through the second order Taylor expansion

$$D_{AH_m,i}^{al} = -\mu_i^{al} + \frac{1}{2}\eta_{ii}^{al} = \left(-\mu_i^{al,el} + \frac{1}{2}\eta_{ii}^{al,el}\right) - \mu_i^{al,nuc} \quad (22)$$

and can be used as the approximation to the vertical deprotonation energy. For a system with more than one hydrogen, a set of conjugate bases can be created. The conjugate base with the lowest $\mathbf{Z}_{AH_{m,i}}^{al}$ is assumed to be the most stable one. The nuclear part of the alchemical potential always takes a positive value for the molecule, eq 10, and the hydrogen site with the highest value of $\mu_i^{al,nuc}$, can be put forward as the most probable deprotonation site. A simple example is the methylammonium cation $CH_3 - N^{\AA}H_3$, at the ethane geometry. A calculation shows that $\mu_{H(-N)}^{al,nuc} - \mu_{H(-C)}^{al,nuc} = 1/R_{C-H} > 0$ so that the deprotonation of the hydrogen attached to the nitrogen atom produces the most stable anion. As expected as intuitively $CH_3-NH_2$ is expected to be more stable than the dipolar $C^{\check{z}}H_2 - N^{\AA}H_3$ structure. Another example, the *n*-propane deprotonation which yields two carbanions, shows that the electronic contribution is also important. The $\mu_i^{al,nuc}$ for the hydrogens connected to the central carbon is higher than for the terminal hydrogens. However, the most stable anion is predicted to be the 1-propyl anion rather than the isopropyl anion in agreement with the experimental results[52] and common knowledge.

In general, the one-site deprotonation, eq 19, can be extended to a *n*-tuple deprotonation reaction:

$$AH_m \rightarrow (AH_{m-n})^{n-} + nH^+ \quad (23)$$

This hypothetical gas-phase reaction does not precisely describe at which positions the protons were annihilated, but all conjugate bases of the $AH_m$ acid can be represented by a set of transmutation vectors

$$\mathcal{D}_{AH_m}^{n} \equiv \{d\mathbf{Z}_{AH_m}^{M} = (dZ_1, ..., dZ_m, 0, ...) | \; dZ_i \in \{-1, 0\}, 1 \le i \le m, \sum_{i=1}^{m} dZ_m = n\} \quad (24)$$

For each of these vectors, the alchemical transmutation energy is

$$D_{AH_m}^{al}[d\mathbf{Z}_{AH_m}^{M}] \equiv \sum_{i=1}^{n} \mu_i^{al} dZ_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \eta_{ij}^{al} dZ_i dZ_j \quad (25)$$

The minimizer of the alchemical transmutation energy can be viewed as the transmutation vector which leads to the most stable conjugate base, $(AH_{m-n})^{n-}$.

The usefulness of the alchemical derivatives as chemical reactivity indices in the deprotonation energy was tested on the set of 20 molecules with one and more hydrogen atoms (see Scheme 1). Two exchange-correlation functionals: B3LYP[53,54]

**Scheme 1. Deprotonation of X−H Molecule**

$$X - H \rightarrow X^- + H^+$$
$$X \in \{H, CH_3, CH_2CH_3, CHCH_2, CHO, COOH_3, CN, NH_2, OCH_3, OH, F, CH_2F, CHF_2, CF_3, SiH_3, PH_2, SH, Cl, CH_2Cl, CCl_3\}$$

and PBE[55] were used in a combination with the correlation consistent double- and triple-$\zeta$ basis sets proposed by Dunning and co-workers (along with their augmented counterparts).[56−61] The correlation coefficients and the mean absolute errors (MAEs) are shown in Table 1 (detailed data for these correlations are presented in Supporting Information). We conclude that functional dependencies are insignificant since the results for both functionals used in this work are very similar (the B3LYP results are slightly better than the PBE results if we consider the correlation coefficients and conversely if the MAEs are considered). Increasing the $\zeta$ splitting slightly decreases the correlation coefficient (and the increasing the MAE) for the $D_{XH}^{ver}$ vs $D_{XH}^{al}$ correlation. This can be explained by the presence of the artificial binding of the electron by the basis set. The same effect of the basis set is observed in the case of the electron affinity

**Table 1. Correlation Coefficients and Mean Absolute Error (in Parentheses, in kJ/mol) for Deprotonation Energy for 20 X−H Molecules[a]**

| | | $D_{XH}^{exp}$ vs | | | |
|---|---|---|---|---|---|
| functional | basis | $D_{XH}^{\circ}$ | $D_{XH}^{ver}$ | $D_{XH}^{al}$ | $D_{XH}^{ver}$ vs $D_{XH}^{al}$ |
| B3LYP | cc-pVDZ | 0.859 (77) | 0.851 (138) | 0.804 (181) | 0.973 (319) |
| | cc-pVTZ | 0.941 (41) | 0.893 (94) | 0.849 (260) | 0.969 (354) |
| | aug-cc-pVDZ | 0.964 (19) | 0898 (40) | 0.895 (250) | 0.978 (290) |
| | aug-cc-pVTZ | 0.961 (13) | 0.889 (44) | 0.878 (279) | 0.954 (324) |
| PBE | cc-pVDZ | 0.825 (74) | 0.829 (134) | 0.783 (181) | 0.972 (315) |
| | cc-pVTZ | 0.919 (37) | 0.884 (88) | 0.832 (249) | 0.965 (352) |
| | aug-cc-pVDZ[b] | 0.954 (24) | 0.901 (33) | 0.891 (253) | 0.975 (284) |
| | aug-cc-pVTZ | 0.950 (20) | 0.891 (35) | 0.869 (284) | 0.946 (319) |

[a]Comparison between experimental, vertical and alchemical values is made. Experimental data from http://webbook.nist.gov. [b]$H_2$ is not included in correlation due to numerical instability.


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

calculation.[62] However, contrary to the latter, the addition of diffuse functions improves the correlations with decreasing MAE (see Figure 1). Based on the correlations between $D_{XH}^{exp}$ and $D_{XH}^{0}$ (Table 1), we can explain that this is connected with the fact

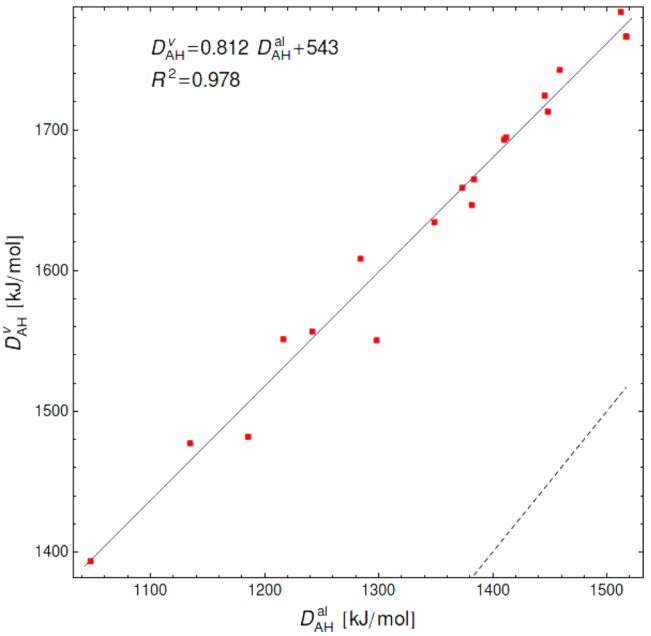
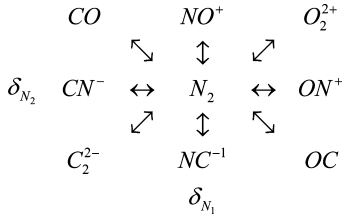



**Figure 1.** Correlation of the alchemical deprotonation energy with the vertical deprotonation energy at B3LYP/aug-cc-pVDZ level. All values in kJ/mol, the dashed line is an exact result line. Data and plots for other levels are accessible from Supporting Information.

that the augmented basis sets most closely reproduce the experimental geometry. A similar effect was observed in the case of silicaceous materials,[63] for example. It is pertinent to comment on the possible influence of the fact that the basis set centered on the proton is still included in the conjugated base basis set after the annihilation of the proton. The calculations of the vertical deprotonation energy with the ghost atom at the deprotonation site barely change the results compared to the no-ghost atom calculations (the average difference is less than 10 kJ/mol). The most important conclusion is the very good correlation between $D_{XH}^{ver}$ and $D_{XH}^{al}$ (e.g., B3LYP/aug-cc-pVDZ, $R^2 = 0.978$) indicating that with a simple alchemical type calculation (requiring only one single point calculation) trends in the deprotonation energies can be retrieved. Note that when comparing $D_{XH}^{exp}$ and $D_{XH}^{al}$ contributions from geometry relaxation and enthalpy correction are to be included to fully reproduce the experimental data. Nevertheless, the alchemical derivative value is a trustworthy indicator of the experimental deprotonation energies in comparative studies as usually performed when exploring CS.

**III.B. Nitrogen Molecule Case.** As the next example, we consider the transmutation of the nitrogen molecule. Five chemically relevant transmutations of the nitrogen molecule are considered in this work (see Scheme 2). The nitrogen molecule and its "transmutants" are examples of the isosteres[64] and isoelectronic molecules. All these transmutants can be expected to exist in the atmosphere of the terrestrial planets which have a dense atmosphere of $N_2$.[65] $C_2^{2-}$ is one of the few doubly negative ions to have bound valence states[66] and has been detected in comets.[67] $O_2^{2+}$ is kinetically stable, and it is considered as the molecule with the shortest bond between any two heavy atoms.[68] The nitrosonium cation is the key species in the process of

**Scheme 2. Transmutation of the Nitrogen Molecule with Possible Changes of Charge at the Nuclear Sites $\delta_{N1},\delta_{N2} \in \{-1,0,1\}$**




nitrosation, an important process in the cell biochemistry.[69] The cyanide ion and carbon monoxide are well-known in organic and inorganic chemistry, both as reagent and ligand.

The set of the transmutation vectors for the nitrogen transmutation is simply

$$\mathcal{D}_{N_2} \equiv \{d\mathbf{Z}^{M}_{N_2} = (dZ_1, dZ_2) | dZ_i \in \{-1, 0, 1\}\} \tag{26}$$

and the alchemical transmutation energies are (the fact that $\eta_{1,1}^{al,N_2} = \eta_{2,2}^{al,N_2}$ is used)

$$D^{al,N_2}_{N_2 \to AB} = \left(\mu_1^{al,N_2}(dZ_1 + dZ_2) + \frac{1}{2}\eta_{1,1}^{al,N_2}\right)(dZ_1 + dZ_2) + (\eta_{1,2}^{al,N_2} - \eta_{1,1}^{al,N_2})dZ_1 dZ_2 \tag{27}$$

In the case of the $N_2 \rightarrow CO$ transmutation, the equation becomes simply

$$D^{al,N_2}_{N_2 \to CO} = \eta_{1,1}^{al,N_2} - \eta_{1,2}^{al,N_2} \tag{28}$$

because $dZ_1 + dZ_2 = 0$. For the reverse transmutation, $CO \rightarrow N_2$ at the nitrogen molecule geometry, the alchemical transmutation energy is

$$D^{al,N_2}_{CO \to N_2} = (\mu_O^{al,N_2} - \mu_C^{al,N_2}) + \frac{1}{2}(\eta_{O,O}^{al,N_2} + \eta_{C,C}^{al,N_2} - 2\eta_{C,O}^{al,N_2}) \tag{29}$$

The difference between the vertical and alchemical transmutation energies, $\Delta D_{N_2 \to CO}{}^{al,N_2} = D_{N_2 \to CO}{}^{ver,N_2} - D_{N_2 \to CO}{}^{al,N_2}$ and $\Delta D_{CO \to N_2}{}^{al,N_2} = D_{N_2 \to CO}{}^{ver,N_2} - D_{CO \to N_2}{}^{al,N_2}$ includes an error connected with the omission of higher-order terms in the Taylor expansion. In case of the $N_2 \rightarrow CO$ transmutation, only the even order terms survive due to the cancellation of the odd terms.

For testing the usefulness of the alchemical derivatives in the case of such transmutations, the same pair of the exchange-correlation functionals as in Section III.A was used. The basis set was expanded by inclusion of the additional tight functions (large exponents) in order to recover the core−core and the core−valence correlation,[70] which is expected to be significant in such transmutations. In Table 2, the mean absolute errors for the $D^{ver,N_2}_{N_2 \to AB}$ vs $D^{al,N_2}_{N_2 \to AB}$, and $D^{ver,N_2}_{AB \to N_2}$ vs $D^{al,N_2}_{AB \to N_2}$ are presented. The performance of the alchemical derivatives can be significantly improved by increasing the $\zeta$- splitting and adding the diffuse functions. However, the most crucial issue for the quantitative description of this type of transmutations turns out to be the presence of the tight functions in the basis set. The change in the nuclear screening and the core electron interaction is better predicted by the alchemical derivative if this augmentation is used. In the case of the $N_2 \rightarrow CO$ transmutation, the error is around 1 mHartree (see Table 3), illustrating that in this more complicated alchemical transmutation as compared to the deprotonation


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

**Table 2. Mean Absolute Error between the Vertical and the Alchemical Transmutation Energy for All Nitrogen Molecule Transmutations (Scheme 2)[a]**

| | B3LYP | | PBE | |
|---|---|---|---|---|
| | n | | n | |
| | D | T | D | T |
| cc-pVnZ | 1.148 (1.125) | 0.275 (0.266) | 1.142 (1.119) | 0.279 (0.266) |
| aug-cc-pVnZ | 0.755 (0.715) | 0.158 (0.122) | 0.751 (0.711) | 0.162 (0.121) |
| cc-pCVnZ | 0.115 (0.121) | **0.034** (**0.025**) | 0.113 (0.119) | **0.034** (**0.023**) |
| aug-cc-pCVnZ | 0.096 (0.079) | 0.041 (0.047) | 0.095 (0.078) | 0.042 (0.045) |

[a]The MAE for $D^{ver,N_2}_{N_2\rightarrow AB}$ vs $D^{al,N_2}_{N_2\rightarrow AB}$ is tabulated, and the values for $D^{ver,N_2}_{AB\rightarrow N_2}$ vs $D^{al,N_2}_{AB\rightarrow N_2}$ are given in parentheses. All data are in a.u.

**Table 3. Vertical and Alchemical Transmutation Energies and Errors (Δ) between the Alchemical Prediction and Vertical Calculation for the $N_2 \rightleftarrows CO$ Transmutation Type[a]**

| | | transmutation type: $N_2 \rightleftarrows CO$ | | | | |
|---|---|---|---|---|---|---|
| | | $D^{ver,N_2}_{N_2\rightarrow CO}$ | $D^{al,N_2}_{N_2\rightarrow CO}$ | $D^{al,N_2}_{CO\rightarrow N_2}$ | $\Delta D^{al,N_2}_{N_2\rightarrow CO}$ | $\Delta D^{al,N_2}_{CO\rightarrow N_2}$ |
| B3LYP | cc-pVDZ | −3.785 | −2.389 | 5.179 | −1.396 | 1.394 |
| | cc-pVTZ | −3.786 | −3.462 | 4.151 | −0.324 | 0.366 |
| | aug-cc-pVDZ | −3.786 | −2.861 | 4.672 | −0.925 | 0.886 |
| | aug-cc-pVTZ | −3.785 | −3.605 | 3.956 | −0.180 | 0.170 |
| | cc-pCVDZ | −3.785 | −3.666 | 3.919 | −0.119 | 0.134 |
| | cc-pCVTZ | −3.786 | −3.782 | 3.802 | **−0.004** | **0.017** |
| | aug-cc-pCVDZ | −3.786 | −3.683 | 3.899 | −0.102 | 0.114 |
| | aug-cc-pCVTZ | −3.785 | −3.784 | 3.801 | **−0.001** | **0.016** |
| PBE | cc-pVDZ | −3.778 | −2.386 | 5.161 | −1.392 | 1.383 |
| | cc-pVTZ | −3.779 | −3.450 | 4.143 | −0.329 | 0.364 |
| | aug-cc-pVDZ | −3.778 | −2.858 | 4.659 | −0.921 | 0.881 |
| | aug-cc-pVTZ | −3.779 | −3.593 | 3.949 | −0.185 | 0.170 |
| | cc-pCVDZ | −3.778 | −3.658 | 3.908 | −0.119 | 0.130 |
| | cc-pCVTZ | −3.779 | −3.775 | 3.793 | **−0.004** | **0.014** |
| | aug-cc-pCVDZ | −3.778 | −3.676 | 3.889 | −0.102 | 0.111 |
| | aug-cc-pCVTZ | −3.779 | −3.777 | 3.792 | **−0.002** | **0.013** |

[a]All data in a.u. Data for the other transmutation depicted in Scheme 2 are accessible from Supporting Information.

in Section III.A, now involving two non-H-atom of importance when exploring, for example, benzene rings (Section III.C), the transmutation energy, and certainly its trend, can very well be obtained using alchemical derivatives, thus offering additional possibilities when exploring CS. In Supporting Information, additional results obtained with the polarization consistent basis sets[71] are presented but the results are worse than in case of the correlation consistent basis.

**III.C. Benzene's Transmutation.** Using benzene as the reference molecule, two types of transmutations are studied. The first one is the substitution of the N units for C−H units, which leads to azines at the benzene geometry,

$$(C{-}H)_n \rightarrow N_n \tag{30}$$

The second one is the substitution of C−C units by the isoelectronic B−N units, which leads to the azaborines:

$$(C{-}C)_n \rightarrow (B{-}N)_n \tag{31}$$

Azines are important because of their use as powerful reagents in synthetic organic chemistry, where their aromaticity is at stake.[72] Azaborines are also important in synthetic organic chemistry, but recently, the incorporation of boron as a part of a functional target structure has emerged as a useful way to generate diversity in organic compounds. The replacement of a C−C unit in an aromatic molecule with an isoelectronic B−N unit has been shown to impart interesting electronic, photophysical, luminescent, and chemical properties, which are often very distinct from those of the C−C containing aromatic analogues.[73−75] The syntheses of the azaborine compounds are, in general, very challenging and often involve multistep reactions or the use of the transition-metal catalysts.[76,77] Thus, prediction of the most stable B−N-containing compounds can be very useful for the development of efficient and simple synthetic methods of this important class of compounds.

In this part, the same basis sets as those described above were used. As the exchange-correlation functional only B3LYP functional was used, which is known as the most widely used functional in trouble-free situations (covalent interaction, excluding $\pi-\pi$ stacking and no transition metal bonds) and which proved to yield highly satisfactory results in Sections III.A and III.B. In addition, the calculations at HF/STO-3 level were included. As can be expected, the HF method predictions are quantitatively wrong, but the main point of these calculations is their qualitatively verification as preliminary results for complex CS scanning.

*III.C.1. Azines.* In general, the energy change for the transmutation, which is a combination of the proton annihilation and the replacement of a carbon by a nitrogen atom in benzene, is

$$D^{al,i}_{C_6H_6} = \mu^{al}_H - \mu^{al}_C + \frac{1}{2}(\eta^{al}_{C,C} + \eta^{al}_{H,H}) - \eta^{al}_{C_i,H_i} \tag{32}$$

Due to the benzene symmetry, the first derivatives are equal for the given type of atom. The same is true for the diagonal terms of the second derivative matrix. Assuming that the charge changes


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

on the carbon atom are numbered as one, the most stable product of the transmutation is indicated by the highest value of the carbon−hydrogen alchemical hardness, $\eta^{al}_{C,H}$. The highest value of $\eta^{al}_{C,H}$ occurs for the hydrogen connected with the replaced carbon (in Table 4, the alchemical derivatives for benzene at B3LYP/cc-pVTZ level are presented) Therefore, the most stable product of this transmutation is pyridine with transmutation energy:

$$D^{al}_{(C-H)\rightarrow N} = \mu^{al}_{H} - \mu^{al}_{C} + \frac{1}{2}(\eta^{al}_{C,C} + \eta^{al}_{H,H}) - \eta^{al}_{C1,H1} \quad (33)$$

The set of the transmutation vectors for the $C_6H_6 \rightarrow C_{6-n}H_{6-n}N_n$ is

$$\mathcal{D}^{C_{6-n}H_{6-n}N_n}_{C_6H_6} = \{d\mathbf{Z}^{12}_{C_6H_6} = (dZ_{C1}, ..., dZ_{C6}, dZ_{H1}, ..., dZ_{H6}) |$$
$$dZ_{C_i} \in \{0, 1\}, dZ_{H_i} \in \{0, -1\}, \sum_{i=1}^{6} dZ_{C_i} = -\sum_{i=1}^{6} dZ_{H_i} = n\} \quad (34)$$

and the transmutation energy is

$$D^{C_{6-n}H_{6-n}N_n}_{C_6H_6}[d\mathbf{Z}] = n\left(\mu^{al}_{H} - \mu^{al}_{C} + \frac{1}{2}(\eta^{al}_{C,C} + \eta^{al}_{H,H})\right) + \sum_{i=1}^{6}\sum_{j>i}^{6}(\eta^{al}_{C_i,C_j} dZ_{C_i} dZ_{C_j} + \eta^{al}_{H_i,H_j} dZ_{H_i} dZ_{H_j}) + \sum_{i=1}^{6}\sum_{j=1}^{6}\eta^{al}_{H_i,C_j} dZ_{C_i} dZ_{H_j} \quad (35)$$

The first term in eq 35 is constant for a given $n$, the second term brings a positive contribution to the sum, while the last term is negative. From simple examination of the carbon−hydrogen hardness values ($\eta^{al}_{Hi,Cj}$ elements in Table 4) for the given carbon replacements, $d\mathbf{Z}_C = (dZ_{C1},...,dZ_{C6})$, the annihilation of the proton connected with these carbons minimizes the last term in eq 35, $\eta^{al}_{H_i,C_i} = 0.338$. This means that $(dZ_{H1},...,dZ_{H6}) = -(dZ_{C1},...,dZ_{C6})$, and eq 35 can be rewritten as

$$D^{al}_{(C-H)_n\rightarrow N_n}[d\mathbf{Z}_C] \equiv D^{C_{6-n}H_{6-n}N_n}_{C_6H_6}[d\mathbf{Z} = (d\mathbf{Z}_C, -d\mathbf{Z}_C)] = nD^{al}_{(C-H)\rightarrow N} + \sum_{i=1}^{6}\sum_{j>i}^{6}(\eta^{al}_{C_i,C_j} + \eta^{al}_{H_i,H_j} - 2\eta^{al}_{C_i,H_j}) \times dZ_{C_i} dZ_{C_j} \quad (36)$$

The change in the interaction energy between two C−H units after their transmutation is $(\eta^{al}_{C_i,C_j} + \eta^{al}_{H_i,H_j} - 2\eta^{al}_{C_i,H_j})$. Using data from Table 4, this change is 0.035, −0.002, 0.004 a.u., when $C_j$ is at the ortho, meta, and para position with respect to $C_i$, respectively. This result means that to obtain the most stable azines, carbons should be replaced in such a way that maximizes the number of carbon pairs at meta position with respect to themselves and minimizes the number of carbon pairs at ortho position with respect to themselves. In Table 5, we present the transmutation energies for the $(C-H)_n \rightarrow N_n$ transmutations (and their components), which confirm the above statement. In the case of the diazines, triazines, and tetrazines, there are three isomers. In all cases, the structure with the maximum number of carbon pairs at meta position with respect to themselves is the most stable for the given $n$: the 1,3-diazine (one pair at (1,3) position), 1,3,5-triazine (three pairs at (1,3),(1,5),(3,5) positions),

**Table 4. Alchemical Derivatives and Their Electronic and Nuclear Components for Benzene at B3LYP/cc-pVTZ Level[a]**

| | first derivatives | | |
|---|---|---|---|
| atom | $\mu^{al,el}$ | $\mu^{al,nuc}$ | $\mu^{al}$ |
| C | −24.453 | 9.713 | −14.740 |
| H | −10.388 | 9.291 | −1.097 |
| | second derivatives | | |
| bond | $\eta^{al,el}$ | $\eta^{al,nuc}$ | $\eta^{al}$ |
| C1−C1 | −1.900 | 0.000 | −1.900 |
| C1−C2 | −0.102 | 0.378 | 0.276 |
| C1−C3 | −0.006 | 0.218 | 0.213 |
| C1−C4 | 0.029 | 0.189 | 0.219 |
| C1−H1 | −0.146 | 0.484 | 0.338 |
| C1−H2 | −0.013 | 0.245 | 0.231 |
| C1−H3 | 0.048 | 0.155 | 0.203 |
| C1−H4 | 0.064 | 0.136 | 0.200 |
| H1−H1 | −1.087 | 0.000 | −1.087 |
| H1−H3 | 0.008 | 0.212 | 0.221 |
| H1−H3 | 0.068 | 0.123 | 0.191 |
| H1−H4 | 0.079 | 0.106 | 0.185 |

[a]All values in a.u. Data for other levels are accessible from Supporting Information.

**Table 5. Transmutation Energy for $(C-H)_n \rightarrow N_n$ Transmutation Type at B3LYP/cc-pVTZ Level[a]**

| n | name | $D^{al}_{C_6H_6}$ | $D^{al,\mu}_{C_6H_6}$ | $D^{al,\eta}_{C_6H_6}$ | $D^{al,el}_{C_6H_6}$ | $D^{al,\mu,el}_{C_6H_6}$ | $D^{al,\eta,el}_{C_6H_6}$ | $D^{al,nuc}_{C_6H_6}$ | $D^{al,\mu,nuc}_{C_6H_6}$ | $D^{al,\eta,nuc}_{C_6H_6}$ | $\Delta D^{ver}_{C_6H_6}$ | $\Delta D^{ver,BSSE}_{C_6H_6}$ | $\Delta D^{rel}_{C_6H_6}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | pyridine | −16.043 | −13.652 | −2.390 | −15.959 | −14.058 | −1.902 | −0.084 | 0.405 | −0.489 | 0.017 | 0.016 | 0.008 |
| 2 | 1,3-diazine | −32.089 | −27.305 | −4.784 | −31.952 | −28.115 | −3.837 | −0.136 | 0.811 | −0.947 | 0.031 | 0.029 | 0.011 |
| | 1,4-diazine | −32.082 | −27.305 | −4.778 | −31.939 | −28.115 | −3.823 | −0.144 | 0.811 | −0.954 | 0.029 | 0.027 | 0.011 |
| | 1,2-diazine | −32.054 | −27.305 | −4.750 | −31.989 | −28.115 | −3.873 | −0.066 | 0.811 | −0.876 | 0.025 | 0.024 | 0.012 |
| 3 | 1,3,5-triazine | −48.137 | −40.957 | −7.180 | −47.979 | −42.173 | −5.806 | −0.158 | 1.216 | −1.374 | 0.047 | 0.045 | 0.015 |
| | 1,2,4-triazine | −48.097 | −40.957 | −7.140 | −48.002 | −42.173 | −5.829 | −0.095 | 1.216 | −1.311 | 0.040 | 0.038 | 0.016 |
| | 1,2,3-triazine | −48.069 | −40.957 | −7.112 | −48.052 | −42.173 | −5.879 | −0.017 | 1.216 | −1.233 | 0.031 | 0.028 | 0.014 |
| 4 | 1,2,3,5-tetrazine | −64.114 | −54.609 | −9.505 | −64.099 | −56.231 | −7.868 | −0.015 | 1.622 | −1.637 | 0.049 | 0.046 | 0.019 |
| | 1,2,4,5-tetrazine | −64.108 | −54.609 | −9.498 | −64.085 | −56.231 | −7.854 | −0.022 | 1.622 | −1.644 | 0.054 | 0.051 | 0.023 |
| | 1,2,3,4-tetrazine | −64.080 | −54.609 | −9.471 | −64.135 | −56.231 | −7.905 | 0.056 | 1.622 | −1.566 | 0.038 | 0.036 | 0.017 |
| 5 | pentazine | −80.094 | −68.262 | −11.832 | −80.253 | −70.289 | −9.964 | 0.159 | 2.027 | −1.868 | 0.050 | 0.047 | 0.022 |
| 6 | hexazine | −96.077 | −81.914 | −14.163 | −96.440 | −84.346 | −12.094 | 0.364 | 2.432 | −2.069 | 0.052 | 0.049 | 0.024 |

[a]$D^{al}_{C_6H_6}$: alchemical prediction for transmutation energy; and its components (see eq 22). $\Delta D^{ver}_{C_6H_6}$, $\Delta D^{ver,BSSE}_{C_6H_6}$, $\Delta D^{rel}_{C_6H_6}$: difference between the alchemical transmutation energy and the vertical transmutation energy, the vertical transmutation energy with the ghost function at deprotonation site, the transmutation energy (including geometry relaxation), respectively. All data in a.u.


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 6. Mean Absolute Error between $D^{al}_{C_6H_6}$ and the Vertical Transmutation Energy, $D^{ver}_{C_6H_6}$, the Vertical Transmutation Energy with the Ghost Function at Deprotonation Site, $D^{ver,BSSE}_{C_6H_6}$, and the Transmutation Energy (Including Geometry Relaxation), $D^{rel}_{C_6H_6}$ [a]

| | | transmutation type | | | | | |
|---|---|---|---|---|---|---|---|
| | | $(C-H)_n \rightarrow N_n$ | | | $C_n \rightarrow N_n$ | $(C,C)_n \rightarrow (B,N)_n$ | |
| | | $D^{ver}_{C_6H_6}$ | $D^{ver,BSSE}_{C_6H_6}$ | $D^{rel}_{C_6H_6}$ | $D^{ver}_{C_6H_6}$ | $D^{ver}_{C_6H_6}$ | $D^{rel}_{C_6H_6}$ |
| B3LYP | cc-pVDZ | 1.802 | 1.809 | 1.825 | 2.221 | 2.466 | 2.532 |
| | cc-pVTZ | **0.039** | **0.036** | **0.016** | 0.414 | 0.355 | 0.419 |
| | aug-cc-pVDZ | 0.920 | 0.921 | 0.944 | 1.329 | 1.431 | 1.495 |
| | aug-cc-pVTZ | 0.309 | 0.309 | 0.287 | 0.128 | 0.052 | 0.116 |
| | cc-pCVDZ | 0.089 | 0.082 | 0.065 | 0.325 | 0.203 | 0.268 |
| | cc-pCVTZ | 0.346 | 0.344 | 0.323 | **0.108** | **0.017** | **0.075** |
| | aug-cc-pCVDZ | 0.128 | 0.127 | 0.104 | 0.280 | 0.184 | 0.247 |
| | aug-cc-pCVTZ | 0.349 | 0.349 | 0.327 | **0.101** | **0.016** | **0.074** |
| HF | STO-3 | 2.817 | 2.895 | 2.858 | 3.714 | 3.701 | 3.768 |

[a] All data in a.u.

1,2,3,5-tetrazine. The 1,2,4-triazine is more stable than 1,2,3-triazine; both have one pair of non C-atoms at meta-position with respect to themselves, but the former has one pair at ortho position with respect to themselves while the latter has two. The nuclear components of $D^{al}$, which depend only on the geometry and the nuclear charges ($D^{al,nuc}$,$D^{al,\mu,nuc}$,$D^{al,\eta,nuc}$ from eq 18), can be used as preliminary indicators. However, their predictions are not always consistent with the $D^{al}$ ordering, that is, the $D^{al,nuc}$ ordering is consistent with the $D^{al}$ ordering for the triazines, but inconsistent in the case of the diazines and tetrazines.

The MAE between $D^{al}$ and $D^{ver}$, $D^{ver,BSSE}$, $D^{rel}$ are collected in Table 6. Inclusion of the ghost atom at the deprotonation site has very little effect on the MAE. In some cases, we observe smaller MAE values between $D^{al}$ and $D^{rel}$ than between $D^{al}$ and $D^{ver}$. This can be explained as an effect of the cancellation of errors connected with the higher-order terms and the geometry relaxation. In general, however, the basis set dependence is complex. The lowest MAE is observed for the cc-pVTZ basis and the introduction of the tight functions significantly increases the MAE. Increase of the $\zeta$- splitting without the tight functions gives the opposite result. The effect of the diffuse functions seems to be inexplicable at present.

The analysis of the following transmutation can illuminate these results. The $(C-H)_n \rightarrow N_n$ transmutation can be split into two steps: the carbon atom to nitrogen cation transmutation and the deprotonation as follows

$$\begin{array}{ccc} (C-H)_n & \rightarrow & (N-H)_n^{n+} \\ \downarrow & \searrow & \downarrow \\ C_n^{n-} & \rightarrow & N_n \end{array} \qquad (37)$$

Since problems connected with the highly negative ions calculations can be expected, only for the $(C-H)_n \rightarrow (N-H)_n^{n+}$ transmutation, the vertical transmutation energies were calculated. The MAEs are presented in Table 6. These results are in line with the results obtained for the $(C-C)_n \rightarrow (B-N)_n$ transmutation (see Table 6) and the nitrogen molecule transmutation, which suggests that the basis set dependency for the complex transmutation is not simply predictable from the basis set dependencies of its components. Alchemical derivatives should be used for qualitative rather than quantitative predictions in these types of transmutations which are essential in comparative studies exploring CS. The quantitative prediction requires to pay special attention to the basis set selection.

*III.C.2. Azaborines.* In case of the $(C-C)_n \rightarrow (B-N)_n$ transmutation, the odd terms in the Taylor expansion cancel as in case of the $N_2 \rightarrow CO$ transmutation and the replacement energy of two carbons by one nitrogen and one boron is simply given by

$$D^{al,i}_{C_6H_6} = \eta^{al}_{C,C} - \eta^{al}_{C_i,C_i} \qquad (38)$$

and the most preferable substitution is to replace two adjacent carbon (see Table 4).

The set of the transmutation vectors for the $C_6H_6 \rightarrow C_{6-2n}H_6(BN)_n$ transmutation is

$$\mathcal{D}^{C_{6-2n}H_6(BN)_n}_{C_6H_6} = \{d\mathbf{Z}^{12}_{C_6H_6} = (dZ_{C1}, ..., dZ_{C6}, 0, ..., 0) | dZ_{C_i} \in \{-1, 0, 1\}, \sum_{i=1}^{6} dZ_{C_i} = 0\} \qquad (39)$$

and the transmutation energy is

$$D^{al}_{(C-H)_n \rightarrow (B-N)_n}[d\mathbf{Z}_C] \equiv D^{C_{6-2n}H_6N_n}_{C_6H_6}[d\mathbf{Z} = (d\mathbf{Z}_C, \mathbf{0})] = n\eta^{al}_{C,C} + \sum_{i=1}^{6}\sum_{j>i}^{6} \eta^{al}_{C_i,C_j} dZ_{C_i} dZ_{C_j} \qquad (40)$$

To simplify the analysis of the above equation, the $\mathbf{Z}_C$ will be replaced by two vectors

$$\mathbf{N}^n = \{(N1, ..., Nn) | Ni < Nj, Z_{Ni} = 1\}$$
$$\mathbf{B}^n = \{(B1, ..., Bn) | Bi < Bj, Z_{Bi} = 1\} \qquad (41)$$

and eq 40 can be rewritten as

$$D^{al}[\mathbf{N}^n, \mathbf{B}^n] = n\eta^{al}_{C,C} + \sum_{i=1}^{n}\sum_{j>i}^{n} (\eta^{al}_{Ni,Nj} + \eta^{al}_{Bi,Bj}) - \sum_{i=1}^{n}\sum_{j=1}^{n} \eta^{al}_{Ni,Bj} \qquad (42)$$

Let us relabel the carbon−carbon alchemical hardness as follows: $\eta^{al}_{ortho} = \eta^{al}_{C1,C2}$, $\eta^{al}_{meta} = \eta^{al}_{C1,C3}$, $\eta^{al}_{para} = \eta^{al}_{C1,C4}$, and note that $\eta^{al}_{ortho} > \eta^{al}_{para} > \eta^{al}_{meta}$. The preferable position for the pair of heteroatoms is the ortho-position with respect to themselves followed by para and meta. Based on this conclusion, the results presented in Table 7 for the azaborines become easy to understand. In case $n = 1$, the 1,2-azaborine is the most stable, the 1,4-azaborine with the substitution at the para position is more stable than the 1,3-azaborine (meta position). In the case $n = 2$, the last term in eq 42 is a combination of six components and the second is a sum of the


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 7. Transmutation Energy for $(C-C)_n \rightarrow (B-N)_n$ Transmutation Type at B3LYP/cc-pCVTZ level $D^{al}_{C_6H_6}$, Alchemical Prediction for Transmutation Energy, and Its Components (see eq 22)[a]

| n | | name | $D^{al}_{C_6H_6}$ | $D^{al,el}_{C_6H_6}$ | $D^{al,nuc}_{C_6H_6}$ | $D^{al,\eta,nuc}_{N}$ | $D^{al,\eta,nuc}_{B}$ | $\Delta D^{ver}_{C_6H_6}$ | $\Delta D^{rel}_{C_6H_6}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1,2-azaborine | −3.406 | −3.025 | −0.380 | 0.000 | 0.000 | −0.001 | −0.027 |
| | 2 | 1,4-azaborine | −3.354 | −3.164 | −0.190 | 0.000 | 0.000 | −0.002 | −0.044 |
| | 3 | 1,3-azaborine | −3.347 | −3.128 | −0.220 | 0.000 | 0.000 | −0.002 | −0.038 |
| 2 | 4 | 1,3,2,4-diazadiborine | −6.877 | −5.984 | −0.892 | 0.220 | 0.220 | 0.003 | −0.038 |
| | 5 | 1,5,2,4- diazadiborine | −6.825 | −6.123 | −0.702 | 0.220 | 0.220 | 0.024 | −0.036 |
| | 6 | 1,3,2,5- diazadiborine | −6.805 | −6.015 | −0.790 | 0.220 | 0.190 | −0.005 | −0.049 |
| | 7 | 1,4,2,6- diazadiborine | −6.805 | −6.015 | −0.790 | 0.190 | 0.220 | −0.001 | −0.052 |
| | 8 | 1,4,2,5- diazadiborine | −6.799 | −5.979 | −0.820 | 0.190 | 0.190 | −0.005 | −0.051 |
| | 9 | 1,4,2,3- diazadiborine | −6.747 | −6.117 | −0.629 | 0.380 | 0.190 | −0.034 | −0.087 |
| | 10 | 1,2,3,6- diazadiborine | −6.747 | −6.117 | −0.629 | 0.190 | 0.380 | −0.025 | −0.099 |
| | 11 | 1,2,3,5- diazadiborine | −6.701 | −6.291 | −0.410 | 0.220 | 0.380 | −0.005 | −0.092 |
| | 12 | 1,3,4,5- diazadiborine | −6.701 | −6.291 | −0.410 | 0.380 | 0.220 | −0.012 | −0.078 |
| | 13 | 1,2,3,4- diazadiborine | −6.643 | −6.394 | −0.249 | 0.380 | 0.380 | −0.037 | −0.126 |
| | 14 | 1,2,4,5- diazadiborine | −6.591 | −6.532 | −0.059 | 0.380 | 0.380 | −0.021 | −0.119 |
| 3 | 15 | 1,3,5,2,4,6-triazatriborinane | −10.412 | −8.877 | −1.535 | 0.659 | 0.659 | 0.019 | −0.023 |
| | 16 | 1,2,5,3,4,6-triazatriborinane | −10.152 | −9.143 | −1.010 | 0.790 | 0.790 | −0.050 | −0.139 |
| | 17 | 1,2,3,4,5,6-triazatriborinane | −9.944 | −9.695 | −0.249 | 0.980 | 0.980 | −0.040 | −0.181 |

[a] $D^{al,\eta,nuc}_{N}$ and $D^{al,\eta,nuc}_{B}$: the homoatom parts of the $D^{al,\eta,nuc}$ (see eq. 47). $\Delta D^{ver}_{C_6H_6}$, $\Delta D^{rel}_{C_6H_6}$: difference between the alchemical transmutation energy and the vertical transmutation energy, the transmutation energy (including geometry relaxation), respectively. All data are in a.u.

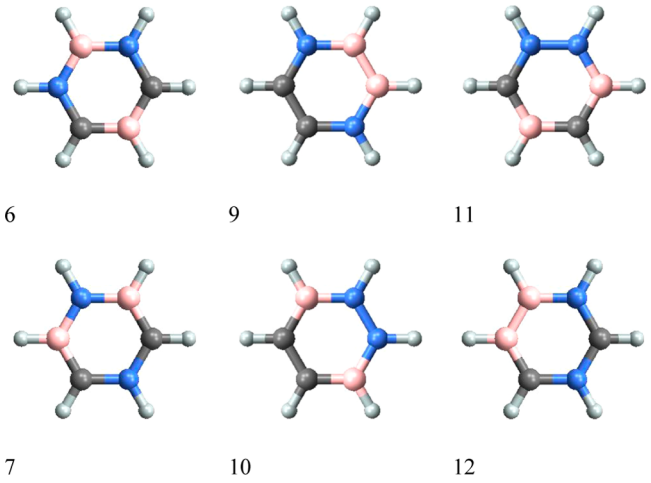



**Figure 2.** Azaborine pairs (6,7), (9,10), (11,12) for which the alchemical transmutation energy is equal (see Table 7).

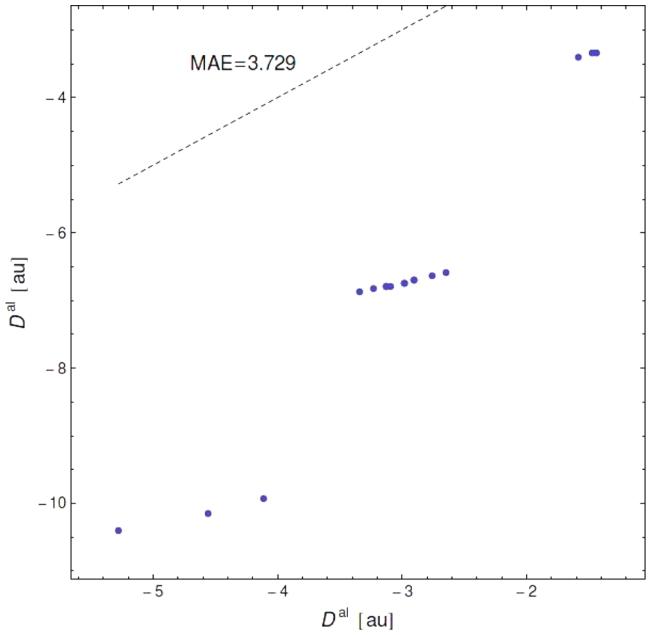



**Figure 3.** Relation between the alchemical transmutation energy HF/STO-3 level (horizontal axis) vs the alchemical transmutation energy at B3LYP/cc-pVTZ level (vertical axis) for $(C-C)_n \rightarrow (B-N)_n$ transmutation type. These two levels are qualitatively related.

nitrogen−nitrogen alchemical hardness and the boron−boron alchemical hardness. The substitution of the next four carbons causes the appearance of a combination of three $\eta^{al}_{ortho}$, two $\eta^{al}_{para}$, and one $\eta^{al}_{meta}$ in eq 42. The $\mathbf{N}^2 = (1,3)$ and $\mathbf{B}^2 = (2,4)$vectors are a minimizer of eq 42, $D^{al}[(1,3),(2,4)] = 2\eta^{al}_{C,C} + 2\eta^{al}_{meta} - (3\eta^{al}_{ortho} + \eta^{al}_{para})$. The 1,4,2,3-diazadiborine and the 1,2,3,6-diazadiborine are less stable than the 1,3,2,4-diazadiborine. The alchemical transmutation energy for the former two diazaborines is $D^{al}[(1,4),(2,3)] = D^{al}[(1,2),(3,6)] = 2\eta^{al}_{C,C} + \eta^{al}_{para} + \eta^{al}_{ortho} - (2\eta^{al}_{ortho} + 2\eta^{al}_{meta})$. This difference in the transmutation energy is an effect of the change from the $N-B-N-B$ to the $B-N-N-B$ or $N-B-B-N$ sequence. Both of these diazaborines have the same value of $D^{al}$ since the second-order Taylor expansion cannot distinguish between these two molecules. The same situation occurs in the case of the 1,3,2,5-diazaborine and 1,4,2,6-diazaborine and in the case of the 1,2,3,5-diazaborine and 1,3,4,5-diazaborine. The splitting into the nitrogen−nitrogen interaction and the boron−boron interaction terms:

$$D^{al,\eta,nuc}_{B}[\mathbf{N}^n, \mathbf{B}^n] = \sum_{i=1}^{n}\sum_{j>i}^{n} \eta^{al}_{Bi,Bj}$$

$$D^{al,\eta,nuc}_{N}[\mathbf{N}^n, \mathbf{B}^n] = \sum_{i=1}^{n}\sum_{j>i}^{n} \eta^{al}_{Ni,Nj} \quad (43)$$

can be used to distinguish these cases. In Table 7, the splitting of the nuclear components, which involve only geometrical data, was used to distinguish isomers presented in Figure 2.

In case of azaborines, the basis set dependence is very easy to figure out (see Table 6). Increasing the $\zeta$- splitting results in a


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

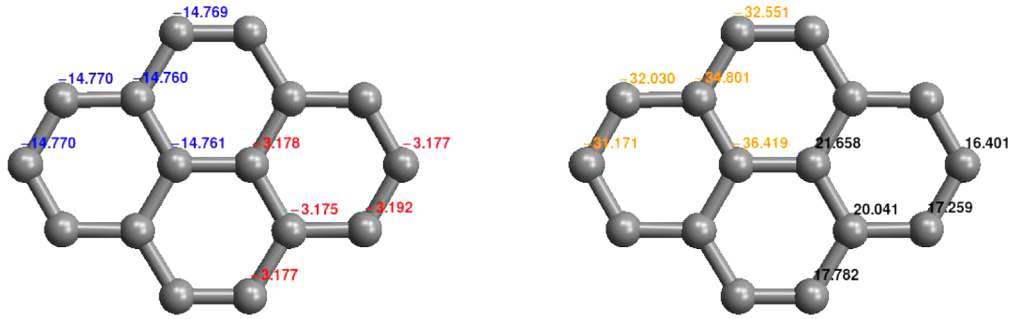



**Figure 4.** Pyrene alchemical derivatives at B3LYP/cc-pVTZ level. The alchemical potential (blue), the diagonal element of the alchemical hardness (red), and carbon label are on the left panel. The electronic (yellow) and nuclear (black) parts of the alchemical potential are on the right panel. Data for other levels are accessible from Supporting Information.

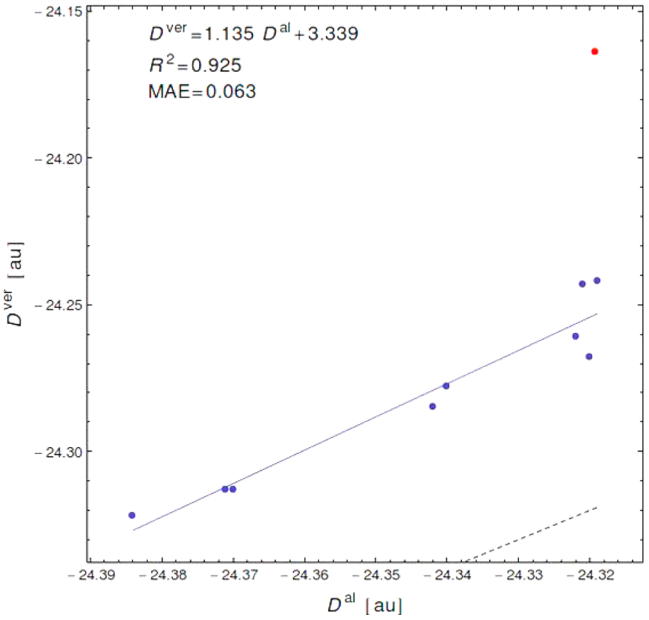



**Figure 5.** Correlation between the alchemical transmutation energy vs the vertical transmutation energy at B3LYP/cc-pcVTZ level for $(C-C)_7 \rightarrow (B-N)_7$ transmutation type. The red point is not included. Data and plots for other levels are accessible from Supporting Information.

decrease of the MAE. As in the case of the nitrogen molecule, the introduction of the tight functions in the basis set is crucial for the quantitative description. The MAEs for the cc-pCVTZ basis and its augmented counterpart are very similar, so that cc-pCVTZ basis is advised, especially when one takes into account the average calculation timings. Of course, the HF/STO-3 results should not be used for a quantitative prediction, but as we can see from Figure 3, they are in qualitative agreement with the B3LYP/cc-pCVTZ results. This is caused by the fact that the difference between the benzene geometry calculated at these two levels is very small. The correlation coefficient between the $D^{al,nuc}$ at these two levels is 1, the intercept equals to zero, and their MEA is 0.006 in a.u. In addition, the effect of the difference in the methods used for the geometry optimization and the alchemical derivatives calculation was investigated for all levels listed in Table 6. The MAE between $D^{al}$ calculated at the same level but at different geometry level is less than 0.004 a.u. (one exception is the MAE between $D^{al}$ from cc-pVTZ//cc-pVTZ and $D^{al}$ cc-pVDZ//cc-pVTZ calculation, which is equal to 0.012 a.u.). The general conclusion from these calculations is that geometry

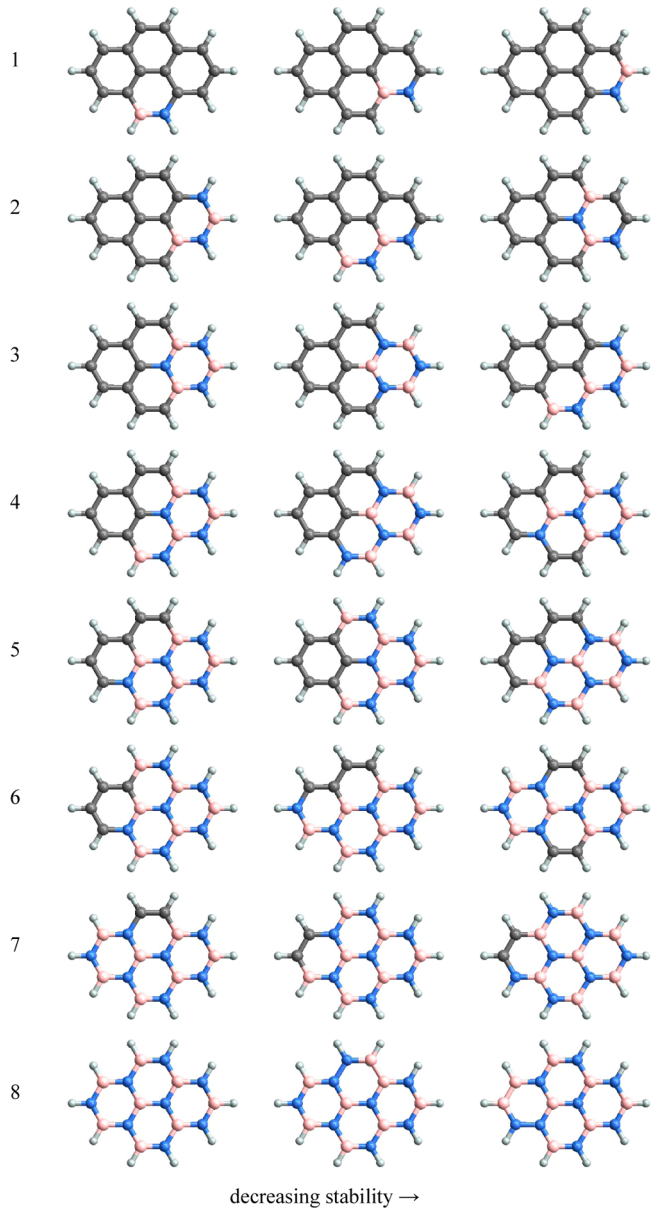



**Figure 6.** The first three most stable isomers of the $(C-C)_n \rightarrow (B-N)_n$ transmutation type of pyrene at B3LYP/cc-pCVTZ level. The bold numbers used in text means the row and the item in this row, for example, **12**, means the second isomer in the first row.


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 8. Qualitative and Quantitative Accuracies of the Alchemical Prediction for the $(C-C)_n \rightarrow (B-N)_n$ Transmutation Type of Pyrene[a]

| | | B3LYP | | | | | | | | HF | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | cc-pVDZ | | cc-pCVDZ | | cc-pVTZ | | cc-pCVTZ | | STO-3 | |
| n | 1 | 0.107 | 1.194 | 0.989 | 0.098 | 0.879 | 0.153 | **0.999** | **0.002** | 0.988 | 1.806 |
| | | 0−1−0 | | 10−10−0 | | 2−5−0 | | **10−10** | | 10−10−1 | |
| | 2 | 0.265 | 2.366 | 0.973 | 0.189 | 0.887 | 0.293 | **0.982** | **0.002** | 0.884 | 3.523 |
| | | 0−0−4 | | 3−6−0 | | 0−2−3 | | **10−10** | | 3−4−1 | |
| | 3 | 0.765 | 3.542 | 0.987 | 0.275 | 0.861 | 0.437 | **0.974** | **0.009** | 0.887 | 5.203 |
| | | 1−2−5 | | 2−8−1 | | 2−4−2 | | **2−8** | | 3−5−2 | |
| | 4 | 0.384 | 4.732 | 0.931 | 0.363 | 0.858 | 0.581 | **0.868** | **0.015** | 0.817 | 6.871 |
| | | 1−3−3 | | 3−4−1 | | 3−4−3 | | **3−6** | | 4−4−6 | |
| | 5 | 0.851 | 5.913 | 0.918 | 0.441 | 0.845 | 0.717 | **0.892** | **0.030** | 0.807 | 8.515 |
| | | 4−5−6 | | 1−4−1 | | 4−7−2 | | **4−5** | | 0−0−1 | |
| | 6 | 0.207 | 7.094 | 0.943 | 0.526 | 0.373 | 0.863 | **0.937** | **0.038** | 0.561 | 10.150 |
| | | 1−1−3 | | 1−4−0 | | 3−5−2 | | **5−5** | | 5−7−3 | |
| | 7 | 0.483 | 8.265 | 0.597 | 0.589 | 0.490 | 0.978 | **0.639** | **0.072** | 0.426 | 11.749 |
| | | 4−4−2 | | 1−5−1 | | 5−6−0 | | **5−5** | | 3−6−1 | |
| | 8 | 0.988 | 9.548 | 0.989 | 0.754 | 0.982 | 1.197 | **0.988** | **0.022** | 0.905 | 13.855 |
| | | 2−8−0 | | 2−4−0 | | 2−7−2 | | **2−8** | | 2−4−1 | |
| total | | 1.000 | 5.332 | 1.000 | 0.404 | 1.000 | 0.652 | **1.000** | **0.024** | 0.998 | 7.709 |

[a]Item for given $n$ and method contains: the correlation coefficient and the mean absolute error between $D^{al}$ and the vertical transmutation energy, $D^{ver}$ (upper part of item), and compatibility, accuracy and deficiency numbers (lower part of item). The deficiency number is with respect to B3LYP/cc-pCVTZ results.

optimizations can be done at a less computationally demanding level.

**III.D. Pyrene Case.** Pyrene is a ubiquitous fluorophore that has found application in various sensors and as a probe in biochemical labeling studies. B−N substitution provides new opportunities for organic electronics.[77] Examples of synthesized BN-pyrenes are very scarce; only 4,10,5,9-diazadiborapyrene,[78] 4,5-azaborapyrene,[79] and 10b,10c-azaborapyrene[80] are known.

In this section, the usefulness of the alchemical derivatives will be tested on the pyrene transmutation to the azaborapyrenes:$C_{16}H_{10} \rightarrow C_{16-2n}H_{10}(BN)_n$. Pyrene is a tetracyclic aromatic hydrocarbon, the smallest polycyclic aromatic hydrocarbon (PAH) on which the effect of a various boron and nitrogen substituents and their positions (internal/periphery) in the PAH can be studied. Pyrene with a molecular formula of $C_{16}H_{10}$ and $D_{2h}$ symmetry has five no-symmetry equivalent carbons (see Figure 4). The alchemical energy of the transmutation of two carbons into one nitrogen and one boron is

$$D^{al,i,j}_{C_{16}H_{10}} = (\mu_{C_i} - \mu_{C_j}) + \frac{1}{2}(\eta^{al}_{C_i,C_i} + \eta^{al}_{C_j,C_j} - 2\eta^{al}_{C_i,C_j}) \quad (44)$$

with an assumption that the nitrogen is placed at the position $i$ and the boron at the position $j$. If substituted carbons are symmetry equivalent, then the contribution from the first derivative vanishes and $D^{al,i,j} = D^{al,j,i}$. Otherwise, the preferable position for boron is the site with the highest value of the alchemical potential, $(\mu_{C_i} < \mu_C)$, such as $\mu_3 < \mu_{3a}$, and the 3,3a-azaborapyrene (see Figures 5 and 6, **12**) is more stable than 3a,3-azaborapyrene. In the case of 4,10,5,9-diazdiborapyrene, all substituted carbons are symmetry equivalent, so that the transmutation energy is

$$D^{al}[(4, 10), (5, 9)] = 2(\eta^{al}_{4,4} - \eta^{al}_{4,5}) + 2(\eta^{al}_{4,10} - \eta^{al}_{4,9}) \quad (45)$$

where $(\eta^{al}_{4,4} - \eta^{al}_{4,5})$ is the (C−C) → (B−N) transmutation energy at the positions (4,5) and (10,9), and $2(\eta^{al}_{4,10} - \eta^{al}_{4,9})$ is the interaction part between these two moieties. This interaction part is equal to 0.0002 a.u. for the 4,10,5,9-diazadiborapyrene and −0.0002 a.u. for the 4,9,5,10-diazadiborapyrene at B3LYP/cc-pCVTZ level (the $\eta^{al}_{4,10}$ and $\eta^{al}_{4,9}$ are 0.1457 and 0.1456 in a.u., respectively). This means that the latter is more stable than the former. Interestingly, the less stable isomer was synthesized from 2,6-diaminobiphenyl, while the more stable isomer could not be obtained in this manner.[73]

Based on the azaborine results, where the cc-pCVTZ results and the aug-cc-pCVTZ results are similar, the basis sets with augmented functions were not included in the test (see Table 6). As in the benzene case, the HF/STO-3 calculation was included to test its qualitative usefulness. The alchemical transmutation energy was calculated for all possible isomers (e.g., for $n = 1$, there are 63 isomers). Then, for the 10 energetically most stable isomers, the calculations of the vertical transmutation energies were performed. As a measure of quantitative accuracy, the correlation coefficient and the MAE were used. The qualitative accuracy is characterized by three numbers: the compatibility number, the accuracy number and the deficiency number (see Appendix for definitions). All these numbers are collected in Table 8. The smallest disagreement between the vertical and the alchemical predictions is observed at B3LYP/cc-pCVTZ level. The worst result (for $n = 7$), after rejecting the most extreme case (the red point in Figure 5), is in line with the other results. These results are consistent with those obtained for the transmutation of nitrogen and benzene: the double $\zeta$- splitted basis sets are simply too small for this type of transmutation. Moreover, the presence of the tight functions in the basis set is crucial for the qualitative accuracy. For example, in case of the azaborapyrenes and diazadiborapyrenes, the MAE is only 2 mHartree. The HF results confirm their usefulness as a preliminary result (e.g., compare the deficiency numbers for the B3LYP/cc-pVDZ level and for the HF/STO-3 level).

The alchemical transmutation energies and their components are presented in Table 9 and the first three most stable isomers are presented in Figure 6. In the case of the fullerenes, two major rules for BN-substitution were formulated: the "hexagon filling"


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 9. Vertical Transmutation Energy, $D^{ver}$, the Alchemical Transmutation Energy, and Its Components for the Most Stable Isomer of the $(C-C)_n \rightarrow (B-N)_n$ Transmutation Type of Pyrene at B3LYP/cc-pCVTZ Level[a]

| | n | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $D^{ver}$ | −3.424 | −6.886 | −10.388 | −13.858 | −17.344 | −20.814 | −24.322 | −27.834 |
| $D^{al}$ | −3.426 | −6.891 | −10.403 | −13.885 | −17.371 | −20.856 | −24.384 | −27.922 |
| $D^{al,\mu}$ | 0.000 | −0.011 | −0.012 | −0.012 | −0.011 | −0.011 | 0.000 | 0.000 |
| $D^{al,\eta}$ | −3.426 | −6.880 | −10.391 | −13.873 | −17.360 | −20.845 | −24.384 | −27.922 |
| $D^{al,el}$ | −3.036 | −4.076 | −8.574 | −11.585 | −12.876 | −15.908 | −20.715 | −23.708 |
| $D^{al,\mu,el}$ | 0.000 | 1.913 | 0.295 | 0.295 | 1.913 | 1.913 | 0.000 | 0.000 |
| $D^{al,\eta,el}$ | −3.036 | −5.989 | −8.869 | −11.880 | −14.788 | −17.821 | −20.715 | −23.708 |
| $D^{al,nuc}$ | −0.390 | −2.815 | −1.828 | −2.300 | −4.496 | −4.947 | −3.670 | −4.214 |
| $D^{al,\mu,nuc}$ | 0.000 | −1.924 | −0.307 | −0.307 | −1.924 | −1.924 | 0.000 | 0.000 |
| $D^{al,\eta,nuc}$ | −0.390 | −0.891 | −1.522 | −1.993 | −2.572 | −3.024 | −3.670 | −4.214 |
| $D_N^{al,\eta,nuc}$ | 0.000 | 0.219 | 0.655 | 1.208 | 1.873 | 2.658 | 3.494 | 4.607 |
| $D_B^{al,\eta,nuc}$ | 0.000 | 0.218 | 0.652 | 1.103 | 1.873 | 2.647 | 3.494 | 4.607 |
| $\Delta D_2^{al}$ | 0.009 | 0.002 | 0.024 | 0.024 | 0.005 | 0.006 | 0.013 | 0.223 |
| $\Delta D_3^{al}$ | 0.018 | 0.017 | 0.032 | 0.040 | 0.022 | 0.009 | 0.014 | 0.286 |

[a] $\Delta D_2^{al}$ and $\Delta D_3^{al}$ are the difference between the most stable isomer and the second and third isomers in the stability order (see Figure 6).

and "continuity" rules. According to the former rule, BN units tend to replace all three-carbon pairs of a hexagon one by one and then spread to adjacent hexagons. The "continuity" rule overshadows this rule where the incoming BN unit connects existing BN units to maintain the continuity of BN units. These two rules lead to an atomic arrangement where BN and C form separate regions in the network, consistent with several experimental investigations. This pattern of grouping heteroatoms and carbons is independent of the number of BN units in the fullerenes case. Beyond that, formation of unfavorable B−B and/or N−N bonds, or separation of BN units might destabilize the hybrid fullerene.[81] Inspection of the structures presented in Figure 6, shows a similar but not identical pattern that in most cases individual hexagons are filled first followed by adjacent ones obeying a continuity rule, and that the avoidance of formation of unfavorable B−B and/or N−N bonds, decreasing the pyrene derivatives' stability, is confirmed. However, in the pyrene case, the most stable isomer is not always simply a combination of the incoming BN units with the existing BN groups. The most stable diazadiborapyrene **21** is a combination of **12** and **13** azaborapyrenes. For $n = 3$, the replacement of the interior carbon by the nitrogen is more preferable than by the boron (see $\Delta D_2^{al}$ for $n = 3$ in Table 5 and **31**, **32** in Figure 6). In the pyrene molecule, we can distinguish two types of rings: two rings with three fusion carbons (e.g., carbons at the positions 3a,10b and 10a) and two rings with four fusion carbons (e.g., carbons at the positions 3a,10b, 10c and 5a). These rings will be labeled *R3* and *R4*, respectively. In pyrene, at first the *R3* ring is filled and then the *R4* (**31** and **51**). The more stable isomer for $n = 6$ suggests than in the next step the second *R4* ring will be filled but the more stable form for $n = 7$ is the structure with two *R3* rings and one *R4* ring completely filled (compare **71** and **72**). The maximization of the nitrogen−boron connection rule is confirmed by isomer **81**. We should mention that these predictions are valid at the pyrene geometry and the geometry relaxation can change the observed ordering.

## IV. CONCLUSION

In this article, the usefulness of the alchemical derivatives in the prediction of the chemical properties was tested. Additionally, the basis set influence on the qualitative and quantitative accuracies of the alchemical predictions was investigated. As "test" transmutations, the deprotonation, the transmutation of the nitrogen molecule, and the substitution of the isoelectronic B−N units for C−C units and N units for C−H units were used. In all cases, the quantitative accuracy was more than satisfactory. The worst performance is observed for the deprotonation. This is connected with the slow convergence of the Taylor expansion in the case of hydrogen (compare the values of the derivatives for hydrogen and carbon in Table 4) and the fact that the product of the deprotonation is an anion (some vertical calculations would not lead to the carbanion energy, but rather to the energy of the complex of the radical plus an electron). The alchemical deprotonation energy (from the second order Taylor expansion) correlates well with the vertical deprotonation energy and can be used as a preliminary indicator for the experimental deprotonation energy.

The transmutation of the nitrogen molecule shows that the quantitative accuracy is more sensible to the basis set choice than to the choice of the functional. The introduction of the tight functions, to recover the core−core, and the core−valence correlation, is crucial to achieve high quality results.

The transmutations of the benzene molecule were extensively studied. The obtained results are very good from the quantitative point of view (the MAE is around a few tenths of a mHartree) and the quantitative point of view (the received stability order for azines and azaborines is consistent with the literature data). In the azines case, the basis set dependence was striking. This can be explained by the occurrence of two different transmutations for which the basis set dependencies are incompatible. Therefore, this type of transmutation requires special attention and careful choice of the basis set. The results for the BN derivatives of benzene and pyrene show that the alchemical derivatives and the alchemical transmutation energies are very efficient and effective tools for the stability prediction. The splitting of the alchemical transmutation energy into electronic and nuclear parts (or into the first and second derivatives contribution) are very useful in the analysis of the dominant effect in the $(C-C) \rightarrow (B-N)$ transmutation. It must be noted that the ratio between the derivatives and SCF calculation times for a single azaborapyrene molecule is 8:1 (the azaborapyrene number is 63). As a result, the presented method can be effectively applied to the problems such as the BN substitution in fullerenes, graphenes and similar systems.


dx.doi.org/10.1021/ct400706g | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

In summary, this method has great potential for efficient and accurate scanning of chemical space.

## APPENDIX

To characterize the accuracy qualitatively let us introduce the compatibility number, the accuracy number and the deficiency number for a given series of isomers. Let us explain their meaning through the following example. For a given level of calculations, two lists are defined: $\mathbf{p}^{al} = \{p_1^{al},...,p_{10}^{al}\}$, $p_i^{al} = i$, which represents new labels of the isomers numbering the 10 energetically lowest isomers, $D_i^{al} < D_{i+1}^{al}$, and $\mathbf{p}^{ver} = \{p_1^{ver},...,p_{10}^{ver}\}$, $p_i^{ver} \in \{1,...,10\}$, which gives the positions of the 10 lowest vertical transmutation energies for these isomers. If $p_i^{ver} \neq i$ then the alchemical prediction for the isomers position is inconsistent with the vertical prediction (using the above labels). The *compatibility number* is defined as a number of isomers for which the alchemical position is equal the vertical position ($p_i^{ver} = i$). The *accuracy number* is defined as the number of the vertical positions of the lowest isomers, which are consistent with the alchemical predictions.

As an example, for all calculations $\boldsymbol{p}^{al}$ is simply {1,2,3,4,5,6,7,8,9,10} and for the isomers with three BN moieties $\boldsymbol{p}^{ver}$={1,2,4,3,5,6,7,8,9,10}(at B3LYP/cc-pCVTZ level). This means that for the isomers with the labels 4 and 3, the alchemical prediction for the position was wrong. The compatibility and accuracy numbers are 8 and 2, respectively. The deficiency number for the given method is defined as the number of the 10 energetically lowest isomers which are not in the reference set; for example, at HF/STO-3 level, only one isomer is not in the B3LYP/cc-pCVTZ set (see Table 9).

## ASSOCIATED CONTENT

### Supporting Information

Alchemical derivatives, transmutation energies, data used for basis set dependency studies. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Authors

*E-mail: rbalawender@ichf.edu.pl.

*E-mail: pgeerlin@vub.ac.be.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was partially supported by the Ministry of Science and Higher Education of Poland through Grant No. N204275939, the International PhD Projects Programme of the Foundation for Polish Science, cofinanced from European Regional Development Fund within Innovative Economy Operational Programme "Grants for Innovation" and the Interdisciplinary Center for Mathematical and Computational Modeling computational grant. P.G. and F.D.P. thanks the VUB and the FWO (Fund for Scientific Research-Flanders) for continuous support for their group. RB is grateful to A. Holas for helpful discussions.

## REFERENCES


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

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

(3) d'Avezac, M.; Zunger, A. *J. Phys.: Condens. Matter* **2007**, *19*, 402201.

(4) Xiao, D.; Yang, W.; Beratan, D. N. *J. Chem. Phys.* **2008**, *129*, 044106.

(5) Rinderspacher, B. C.; Andzelm, J.; Rawlett, A.; Dougherty, J.; Beratan, D. N.; Yang, W. *J. Chem. Theory Comput.* **2009**, *5*, 3321.

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

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

(8) Dobson, C. M. *Nature* **2004**, *432*, 824.

(9) Lipinski, C.; Hopkins, A. *Nature* **2004**, *432*, 855.

(10) Ertl, P. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 374.

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

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

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

(14) von Lilienfeld, O. A.; Tuckerman, M. E. *J. Chem. Theory Comput.* **2007**, *3*, 1083.

(15) Sheppard, D.; Henkelman, G.; von Lilienfeld, O. A. *J. Chem. Phys.* **2010**, *133*, 084104.

(16) Ayers, P. W.; Anderson, J. S. M.; Bartolotti, L. J. *Int. J. Quantum Chem.* **2005**, *101*, 520.

(17) Parr, R. G. *Annu. Rev. Phys. Chem.* **1983**, *34*, 631.

(18) Geerlings, P.; De Proft, F.; Langenaeker, W. *Chem. Rev.* **2003**, *103*, 1793.

(19) Chermette, H. *J. Comput. Chem.* **1999**, *20*, 129.

(20) Gazquez, J. L. *J. Mex. Chem. Soc.* **2008**, *52*, 3.

(21) Geerlings, P.; De Proft, F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3028.

(22) Cedillo, A.; Contreras, R.; Galvan, M.; Aizman, A.; Andres, J.; Safont, V. S. *J. Phys. Chem. A* **2007**, *111*, 2442.

(23) Liu, S.; Li, T.; Ayers, P. W. *J. Chem. Phys.* **2009**, *131*, 114106.

(24) Ayers, P. W.; Liu, S. B.; Li, T. L. *Chem. Phys. Lett.* **2009**, *480*, 318.

(25) Boisdenghien, Z.; Van Alsenoy, C.; De Proft, F.; Geerlings, P. *J. Chem. Theory Comput.* **2013**, *9*, 1007.

(26) Yang, W.; Cohen, A. J.; De Proft, F.; Geerlings, P. *J. Chem. Phys.* **2012**, *136*, 144110.

(27) Sablon, N.; De Proft, F.; Sola, M.; Geerlings, P. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3960.

(28) Sablon, N.; De Proft, F.; Geerlings, P. *J. Phys. Chem. Lett.* **2010**, *1*, 1228.

(29) Sablon, N.; De Proft, F.; Geerlings, P. *Chem. Phys. Lett.* **2010**, *498*, 192.

(30) Sablon, N.; Proft, F. D.; Ayers, P. W.; Geerlings, P. *J. Chem. Theory Comput.* **2010**, *6*, 3671.

(31) Ayers, P. W.; De Proft, F.; Borgoo, A.; Geerlings, P. *J. Chem. Phys.* **2007**, *126*, 224107.

(32) Ayers, P. W.; Parr, R. G. *J. Am. Chem. Soc.* **2001**, *123*, 2007.

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

(34) von Lilienfeld, O. A. *Int. J. Quantum Chem.* **2013**, *113*, 1676.

(35) Anatole von Lilienfeld, O. *J. Chem. Phys.* **2009**, *131*, 164102.

(36) Lesiuk, M.; Balawender, R.; Zachara, J. *J. Chem. Phys.* **2012**, *136*, 034104.

(37) Balawender, R.; Komorowski, L. *J. Chem. Phys.* **1998**, *109*, 5203.

(38) Balawender, R.; Komorowski, L.; De Proft, F.; Geerlings, P. *J. Phys. Chem. A* **1998**, *102*, 9912.

(39) Balawender, R.; Geerlings, P. *J. Chem. Phys.* **2001**, *114*, 682.

(40) Balawender, R.; De Proft, F.; Geerlings, P. *J. Chem. Phys.* **2001**, *114*, 4441.

(41) Balawender, R.; Geerlings, P. *J. Chem. Phys.* **2005**, *123*, 124103.

(42) Lesiuk, M.; Zachara, J. *J. Chem. Phys.* **2013**, *138*, 074107.

(43) Moss, G. P. *Pure Appl. Chem.* **1998**, *70*, 143.

(44) Powell, W. H. *Pure Appl. Chem.* **1983**, *55*, 409.

(45) Mermin, N. D. *Phys. Rev.* **1965**, *137*, A1441.

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

(47) Balawender, R. arXiv:1212.1367, 2012.

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

(49) Nalewajski, R. F. *J. Chem. Phys.* **1982**, *77*, 399.

(50) Ray, N. K.; Parr, R. G. *J. Chem. Phys.* **1980**, *73*, 1334.

(51) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
(52) Depuy, C. H.; Gronert, S.; Barlow, S. E.; Bierbaum, V. M.; Damrauer, R. *J. Am. Chem. Soc.* **1989**, *111*, 1968.
(53) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(54) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(55) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(56) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
(57) Peterson, K. A.; Dunning, T. H. *J. Chem. Phys.* **2002**, *117*, 10548.
(58) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1995**, *103*, 4572.
(59) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. *Vib. Spectrosc.* **1992**, *96*, 6796.
(60) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98*, 1358.
(61) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100*, 2975.
(62) De Proft, F.; Sablon, N.; Tozer, D. J.; Geerlings, P. *Faraday Discuss.* **2007**, *135*, 151.
(63) Zhang, Y.; Li, Z. H.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 593.
(64) Langmuir, I. *J. Am. Chem. Soc.* **1919**, *41*, 1543.
(65) Thissen, R.; Witasse, O.; Dutuit, O.; Wedlund, C. S.; Gronoff, G.; Lilensten, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18264.
(66) Weltner, W.; Vanzee, R. J. *Chem. Rev.* **1989**, *89*, 1713.
(67) Palma, A.; Sandoval, L.; Churyumov, K.; Chavushyan, V.; Berezhnoy, A. *Int. J. Quantum Chem.* **2007**, *107*, 2650.
(68) Wong, M. W.; Nobes, R. H.; Bouma, W. J.; Radom, L. *J. Chem. Phys.* **1989**, *91*, 2971.
(69) Hughes, M. N. *Biochim. Biophys. Acta* **1999**, *1411*, 263.
(70) Jensen, F. *Wires. Comput. Mol. Sci.* **2013**, *3*, 273.
(71) Jensen, F. *Vib. Spectrosc.* **2001**, *115*, 9113.
(72) Wang, Y.; Wu, J. I.; Li, Q.; Schleyer, P. *Org. Lett.* **2010**, *12*, 4824.
(73) Bosdet, M. J. D.; Piers, W. E. *Can. J. Chem. Rev. Can. Chim.* **2009**, *87*, 8.
(74) Campbell, P. G.; Marwitz, A. J.; Liu, S. Y. *Angew. Chem., Int. Ed.* **2012**, *51*, 6074.
(75) Islas, R.; Chamorro, E.; Robles, J.; Heine, T.; Santos, J. C.; Merino, G. *Struct. Chem.* **2007**, *18*, 833.
(76) Lu, J. S.; Ko, S. B.; Walters, N. R.; Kang, Y.; Sauriol, F.; Wang, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 4544.
(77) Wang, X. Y.; Lin, H. R.; Lei, T.; Yang, D. C.; Zhuang, F. D.; Wang, J. Y.; Yuan, S. C.; Pei, J. *Angew. Chem., Int. Ed.* **2013**, *52*, 3117.
(78) Dewar, M. J. S.; Dietz, R. *J. Chem. Soc.* **1959**, 2728.
(79) Dewar, M. J. S.; Poesche, W. H. *J. Org. Chem.* **1964**, *29*, 1757.
(80) Bosdet, M. J.; Piers, W. E.; Sorensen, T. S.; Parvez, M. *Angew. Chem., Int. Ed.* **2007**, *46*, 4940.
(81) Kar, T.; Pattanayak, J.; Scheiner, S. *J. Phys. Chem. A* **2003**, *107*, 8630.