Optimization of General Molecular Properties in the Equilibrium Geometry Using Quantum Alchemy: An Inverse Molecular Design Approach

Takafumi Shiraogawa,*[a] Jun-ya Hasegawa[a,b]

[a]Institute for Catalysis, Hokkaido University, N21, W10, Kita-ku, Sapporo, Hokkaido 001-0021, Japan

[b]Interdisciplinary Research Center for Catalytic Chemistry, National Institute of Advanced Industrial Science and Technology, Central 5, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8565, Japan

*Corresponding author: T.S. (E-mail: shiraogawa@cat.hokudai.ac.jp)

Abstract

Inverse molecular design allows optimization of molecules in chemical space and is promising for accelerating the development of functional molecules and materials. To design realistic molecules, it is necessary to consider geometric stability during optimization. In this work, we introduce an inverse design method that optimizes molecular properties by changing the chemical composition in the equilibrium geometry. The optimization algorithm of our recently developed molecular design method has been modified to allow molecular design for general properties at a small computational cost. The proposed method is applicable to large chemical space based on quantum alchemy without empirical data. We demonstrate the applicability of the present method in the optimization of the electric dipole moment and atomization energy in chemical spaces for (BF, CO), (N_2, CO), and BN-doped benzene derivatives. Moreover, we also investigate and discuss the applicability of quantum alchemy to the electric dipole moment.

Introduction

Inverse molecular design proposes molecules with predefined properties and is a reverse process of traditional direct design processes.[1-3] In the reverse design process, molecules are optimized for the properties. One approach is to explore chemical space, a set of candidate molecules, and functionality space mapped onto the chemical space. In contrast to early efforts toward inverse design,[1, 4-5] exploration leads directly to chemical structures. Only the subspace of the chemical space is explored by using the information of the functionality space, and the exhaustive enumeration of the properties of all the candidate molecules is avoided. This efficiency is appealing because the chemical space combinatorially expands with molecular size.[6] Thus, inverse design is a promising approach for accelerating the development of functional molecules and materials. Importantly, the choice of representation of molecules and their properties dramatically impacts the performance of molecular design. Electronic structure theories can map molecular structures onto various physical properties without experimental records and are a natural choice for inverse design. The inverse molecular design approaches based on electronic structure theories involve the variational particle approach,[7] linear combination of atomic potentials (LCAP),[8] and quantum algorithm-based alchemical optimization.[9]

Considering the geometric stability of molecules in the design process is essential for designing functional molecules available in experiments under ambient conditions.[3] Recently, we proposed an inverse molecular design method based on quantum alchemy to efficiently explore chemical space composed of molecules with equilibrium geometries.[10] Hereafter, this method is referred to as MDM. Quantum alchemy models the change in composition ("alchemical" changes) from a reference molecule to a target molecule at the level of quantum mechanics.[7, 11-15] MDM simultaneously predicts the molecular species, target properties, and equilibrium geometry of the

functional molecule without empirical data. In the design process, MDM gradually varies the molecular species and equilibrium geometry using the gradient of the target property with respect to the molecular species that accounts for the change in equilibrium structure. MDM was applied to various chemical spaces, one of which contains 3.1×10^5 BN-doped phenanthrene derivatives,[16] and its capabilities in the design of energy-related functionalities were demonstrated. MDM can target various molecular properties calculated by the quantum alchemy method. However, evaluating the property gradient other than for energy-related properties (e.g., energy, atomization energy, and reaction energy) is computationally expensive. This greatly limits the scope of applications. To overcome this problem, MDM applicable to the general properties with a small computational cost is needed.

In this work, we have extended MDM for general functionalities by improving the optimization algorithm. The present MDM separates two coupled optimization problems for the molecular species and equilibrium geometry; MDM gradually and iteratively changes the molecular species and equilibrium geometry during the exploration of chemical space but avoids evaluating the contribution of the geometry change in the gradient. This allows for a dramatic reduction in the computational cost. This type of approach has been demonstrated in discrete optimization using semiempirical AM1-based LCAP.[17] Moreover, we achieved rapid design by improving the constraint optimization method for updating molecular species. The proposed MDM was successfully applied to molecular design for electric dipole moment and atomization energy with small chemical spaces. The electric dipole moment is important for optical properties and intermolecular interactions. The atomization energy is related to the stability of molecules. We adopted alchemical perturbation density functional theory (APDFT)[18] as the quantum alchemy

method. The electric dipole moment is less investigated in APDFT, and it is worth examining its accuracy. We have assessed the applicability of APDFT for estimating the electric dipole moment.

Methods

We describe the MDM algorithm introduced for general molecular properties. A conceptual illustration is shown in Figure 1(a). The present MDM searches for a functional molecule in the equilibrium geometry by optimizing the molecular species, as introduced in our previous MDM paper.[10] However, the previous search algorithm is designed for energy-associated properties and is not necessarily appropriate for general properties. In this section, a common strategy of our MDM is first shown. Then, the algorithm of the present MDM algorithm is compared with that of Ref. 10.

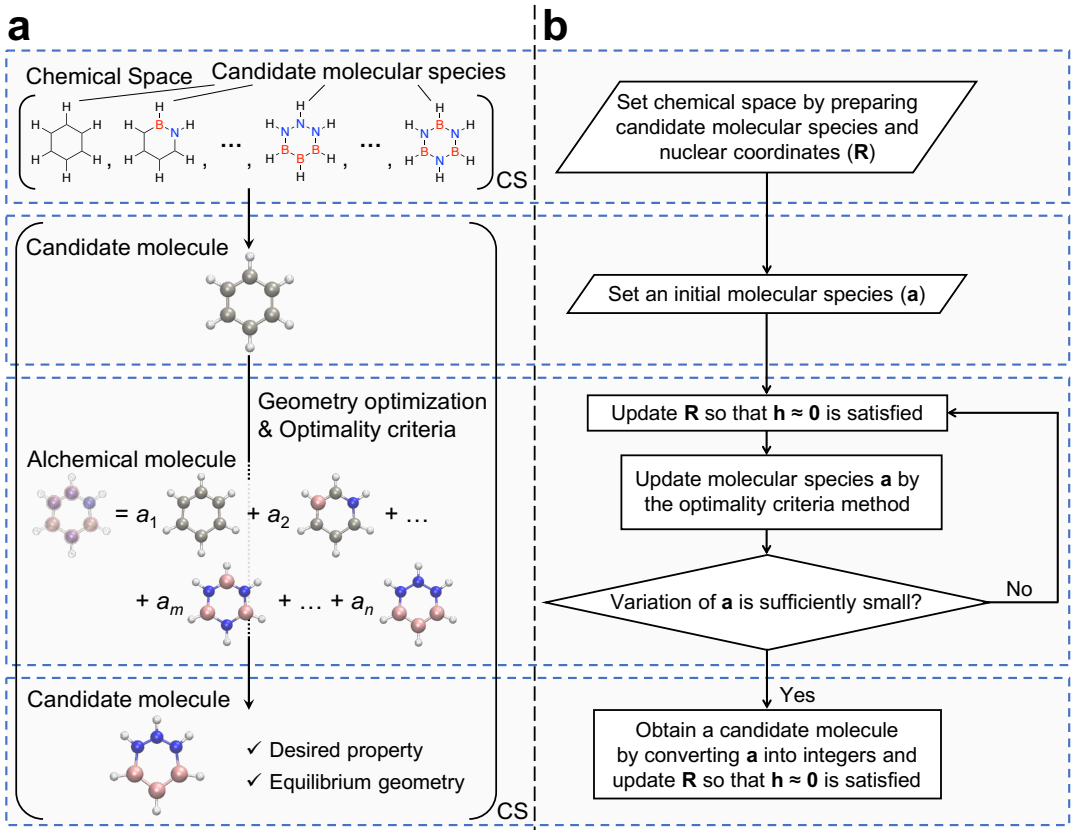

Figure 1. (a) Conceptual illustration and (b) simplified flowchart of the presently introduced MDM.

MDM begins by preparing candidate molecular species. Then, starting from a candidate molecule with initial nuclear coordinates, it searches for the candidate molecule with the desired functionality in the equilibrium geometry. To efficiently explore the chemical space using the gradient, we use continuous chemical space interpolated by introducing virtual alchemical molecules. The alchemical molecules have already been utilized in inverse design methods to facilitate exploration: LCAP[8] and quantum algorithm-based alchemical optimization[9] approaches represent alchemical molecules at the atomic level; in the design of molecular aggregates and plasmonic systems, alchemical molecules are described at the molecular level.[19-20] MDM adopts the latter approach and represents the alchemical molecular species as a weighted average of candidate molecular species.[10] The weights $\mathbf{a}$ ($\{a_i\}_{i=1}^{n}$) are called participation coefficients. i is the index of the candidate molecular species. n is the total number of candidate molecular species. The candidate molecules are defined with $a_i = 1,\ a_{j\neq i} = 0,\ \forall i$. The alchemical molecules are defined with $0 < a_i < 1,\ \forall i$. Corresponding to the molecular species, the physical properties of the alchemical molecule specified by $\mathbf{a}$ and the nuclear coordinates $\mathbf{R}$ ($\{\mathbf{R}_I\}_{I=1}^{N}$) are represented as $f(\mathbf{a},\mathbf{R}) = \sum_i^n a_i f_i(\mathbf{R})$, where $\{f_i(\mathbf{R})\}_{i=1}^{n}$ are the corresponding properties of the candidate molecules with the geometry $\mathbf{R}$. I is the index of the atomic nucleus. N is the total number of nuclei. MDM designs the molecule with the desired property by optimizing $\mathbf{a}$ using the property gradient while suitably changing the equilibrium geometry. The calculations of $\{f_i(\mathbf{R})\}_{i=1}^{n}$ and geometry optimization are performed using quantum alchemy during the design. The MDM algorithm introduced for general target properties is described in detail in the following.

We consider a computational procedure for molecular design as an optimization problem. The functional molecule is obtained by maximizing the target property f with respect to the molecular species represented by **a** in the chemical space. This is expressed by minimizing objective function p:

$$\min_{\mathbf{a}} \; p\left(\mathbf{a},\mathbf{R}\right) = \min_{\mathbf{a}} \; -f\left(\mathbf{a},\mathbf{R}\right). \tag{1}$$

Since **a** are the weights for the weighted average, the constraints on **a** are expressed as

$$\text{subject to } g\left(\mathbf{a}\right) = \sum_{i}^{n} a_i - 1 = 0, \tag{2}$$

$$\begin{aligned} &\text{subject to } 0 \le a_i \le 1, \\ &i = 1,\, 2,\, \cdots,\, n \end{aligned} \tag{3}$$

These constraints are necessary to limit the value of p. In addition, the equilibrium geometry is imposed on the molecule during the exploration of the chemical space. This criterion for geometric stability is employed in the gradient-driven molecular construction approach.[21-23] In the equilibrium geometry, the derivatives of energy with respect to the nuclear coordinates are zero:

$$\begin{aligned} &\text{subject to } h_I\left(\mathbf{a},\mathbf{R}\right) = \frac{\partial E\left(\mathbf{a},\mathbf{R}\right)}{\partial \mathbf{R}_I} = \sum_{i}^{n} a_i \frac{\partial E_i\left(\mathbf{R}\right)}{\partial \mathbf{R}_I} = 0, \\ &I = 1,\, 2,\, \cdots,\, N \end{aligned} \tag{4}$$

This can be solved by using the (quasi-)Newton method with respect to **R**. For the full optimization problem consisting of Eqs. (1)–(4), we have two coupled optimization problems with respect to participation coefficients **a** and nuclear coordinates **R**. The former seeks **a** that maximizes the functionality p for a given fixed **R**. The latter searches for the equilibrium geometry with local energy minimization for a given **a**.

Here, MDMs of this work and Ref. 10 are compared. It is possible to efficiently solve the full optimization problem using the gradient of p with respect to **a**, $dp/d\mathbf{a}$, taking into account the

change in the equilibrium geometry.[10] Assuming the equilibrium geometry, $dp/d\mathbf{a}$ can be expressed as

$$\frac{dp(\mathbf{a},\mathbf{R})}{d\mathbf{a}} = \frac{\partial p(\mathbf{a},\mathbf{R})}{\partial \mathbf{a}} - \frac{\partial p(\mathbf{a},\mathbf{R})}{\partial \mathbf{R}}\left[\frac{\partial \mathbf{h}(\mathbf{a},\mathbf{R})}{\partial \mathbf{R}}\right]^{-1}\frac{\partial \mathbf{h}(\mathbf{a},\mathbf{R})}{\partial \mathbf{a}} \tag{5}$$

The first term on the right-hand side represents the gradient in the iso-geometric chemical space and contains the elements of $-f_i$. The second term $-(\partial p/\partial \mathbf{R})[\partial \mathbf{h}/\partial \mathbf{R}]^{-1}(\partial \mathbf{h}/\partial \mathbf{a})$ takes into account the displacement of the equilibrium geometry. $\partial p/\partial \mathbf{R}$ in the second term has the elements of $-f_i/\mathbf{R}_I$. $\partial \mathbf{h}/\partial \mathbf{R}$ is the Hessian matrix of energy with respect to the nuclear coordinates. $\partial \mathbf{h}/\partial \mathbf{a}$ contains the elements of $\partial E_i/\partial \mathbf{R}_I$. While quantum alchemy can efficiently compute these derivatives as described later, the computational costs of $\partial p/\partial \mathbf{R}$ and $\partial \mathbf{h}/\partial \mathbf{a}$ are still high. When p is the energy-related functionality, Eq. (5) becomes $dp/d\mathbf{a} = \partial p/\partial \mathbf{a}$, and their evaluations are avoided.[10] For the other functionalities, $-(\partial p/\partial \mathbf{R})[\partial \mathbf{h}/\partial \mathbf{R}]^{-1}(\partial \mathbf{h}/\partial \mathbf{a})$ must be evaluated. That is, the computational cost of MDM becomes significantly higher when targeting general properties. Therefore, in our previous study, MDM was only applied for optimizing energetic functionalities.[10] For the design of general properties, the present MDM avoids evaluating $-(\partial p/\partial \mathbf{R})[\partial \mathbf{h}/\partial \mathbf{R}]^{-1}(\partial \mathbf{h}/\partial \mathbf{a})$ by separating the full optimization problem into two subproblems: Eqs. (1)–(3) and Eq. (4). MDM iterates updating **a** and optimizing **R**. First, geometry optimization is performed to solve constraint Eq. (4). Next, **a** are updated based on Eqs. (1)–(3). This procedure is repeated until the variation in **a** becomes sufficiently small (Figure 1(b)). The full optimization problem is solved under that condition since the equilibrium geometry change is

caused by the update of **a**. When the full optimization problem is separated, the substantial update of **a** results in a large error due to the large equilibrium geometry change. Therefore, MDM gradually updates **a** in the continuous chemical space. It is expected that this also leads to the small displacement of the equilibrium geometry at each design step. Therefore, in practice, the geometry optimization that suitably changes **R** and approximately solves Eq. (4) is not performed at all design steps. As the update of **a** depends on the current equilibrium geometry **R**, their coupling persists through the iterations. The LCAP discrete optimization performs geometry optimization of nearby candidate molecules in discrete chemical space and subsequently computes the gradient by finite difference.[24-25] This strategy is costly in our MDM since all the candidate molecules are located at equal distances in the chemical space. Moreover, the error in the discrete update of **a** should be large.

a are updated using the optimality criteria method[26] used in topology optimization for macroscopic material design. The similarities between this type of molecular design and topology optimization have already been pointed out in the context of LCAP.[27] The use of the optimality criteria method in MDM is expected to lead to rapid optimization convergence. To obtain the update formula, we introduce the Lagrangian function with respect to constraint Eq. (2):

$$L(\mathbf{a},\mathbf{R},\lambda) = p(\mathbf{a},\mathbf{R}) + \lambda g(\mathbf{a}) \tag{6}$$

where λ is the Lagrange multiplier. The optimality condition with respect to **a** is expressed as

$$\frac{\partial L(\mathbf{a},\mathbf{R},\lambda)}{\partial a_i} = \frac{\partial p(\mathbf{a},\mathbf{R})}{\partial a_i} + \lambda \frac{\partial g(\mathbf{a})}{\partial a_i} = 0, \quad i = 1, 2, \cdots, n \tag{7}$$

Based on this equation, we obtain the scale factors **B** for updating **a**:

$$B_i = -\frac{\partial p(\mathbf{a},\mathbf{R})/\partial a_i}{\lambda \partial g(\mathbf{a})/\partial a_i} = \frac{f_i(\mathbf{R})}{\lambda}, \quad i = 1, 2, \cdots, n \tag{8}$$

Here, $\{f_i\}$ are positive. The optimality condition is satisfied when $B_i = 1$. Increasing a_i is effective in making p smaller when $B_i > 1$. When $B_i < 1$, the opposite is true. Multiplying **a** by **B**, **a** are gradually updated while satisfying Eq. (3). A heuristic updating scheme is as follows:

$$a_i := \begin{cases} \max(0,\ a_i - \Delta a) \\ \quad \text{if } \sqrt{B_i} a_i \le \max(0,\ a_i - \Delta a), \\ \sqrt{B_i} a_i \\ \quad \text{if } \max(0,\ a_i - \Delta a) < \sqrt{B_i} a_i < \min(1,\ a_i + \Delta a), \\ \min(1,\ a_i + \Delta a) \\ \quad \text{if } \min(1,\ a_i + \Delta a) \le \sqrt{B_i} a_i \end{cases} \quad i = 1, 2, \cdots, n \tag{9}$$

Δa is a parameter for limiting each update. To gradually update the molecular species, Δa was set to 0.01 at each design step. λ is obtained by a bisection search to satisfy constraint Eq. (2). The effectiveness of the algorithm comes from the fact that each a_i is updated independently of the others, except for the scaling that must take place to satisfy Eq. (2). To obtain the candidate molecular species ($a_i = 1,\ a_{j \ne i} = 0,\ \forall i$) as the numerical procedure, optimized **a** are rounded off and converted into integers. From Eq. (9), it is found that the candidate molecules are stationary points. Therefore, the candidate molecule can be obtained by MDM. When starting the exploration from the candidate molecule, **a** must be varied slightly.

MDM requires the target properties $\{f_i\}$ and energy derivatives $\{\partial E_i/\partial \mathbf{R}_I\}$ in the design process. Quantum alchemy can efficiently compute them. In particular, alchemical perturbation density functional theory (APDFT)[18] can predict the properties of a combinatorially large number

of molecules from a combination of derivatives of the electron density of a reference molecule with respect to nuclear charges based on perturbation expansions. While the change in the number of electrons is allowed,[28-29] this work considers isoelectronic chemical space. In the MDM procedure, $\{f_i\}$ and $\{\partial E_i/\partial \mathbf{R}_I\}$ for all the candidate molecules with equivalent nuclear coordinates are calculated using APDFT. The accuracy of APDFT energy has been intensively investigated for the perturbation order and basis sets.[30-31] In the iso-geometric chemical space, APDFT gives accurate predictions.[18, 30, 32] An efficient analytical energy derivative has been proposed within the restricted Hartree–Fock theory.[33] Because the accuracy of APDFT depends on the reference electronic structure theory,[18] the direct derivative of the APDFT energy with respect to the nuclear coordinates[33] was employed for MDM.[10] The convergence of the perturbation expansion was confirmed for several atoms and molecules.[18, 31, 33] Quantum alchemy has been widely applied to investigate catalysis,[34-36] covalent bond energies,[18, 37] non-covalent interactions,[18] mixtures,[12-13, 38] protonation and deprotonation energies,[33, 39-40] and chemical reactions.[41-42] Although APDFT can calculate various response properties, to the best of our knowledge, few studies[18] have applied APDFT to the prediction of non-energetic properties. In this work, we investigate the accuracy of APDFT for the electric dipole moment for molecular design.

The complexity of the functionality space explored by MDM depends on the chemical space. This means that the efficiency and results of MDM depend on the chemical space. To examine this issue, the chemical space containing two candidate molecules was linearly interpolated, and we investigated the corresponding atomization energy and electric dipole strength (EDS) (Supporting Information). EDS is the norm of the electric dipole moment. The adopted chemical spaces are $(N_2, CO)_{CS}$, $(BF, CO)_{CS}$, and $(N_2, BF)_{CS}$ for the atomization energy and $(N_2, CO)_{CS}$ for EDS, where

"$(X, Y, \ldots)_{CS}$" denotes a chemical space including molecular species X, Y, and so on. The atomization energy and EDS smoothly and monotonically vary with the interpolation parameter in all cases (Figure S1). This result indicates that molecular design can be easily achieved by MDM.

Computational details

The energies of the candidate molecules and their derivatives with respect to the nuclear coordinates were calculated using APDFT with the second-order perturbation expansion (APDFT2). The reference electronic structure theory is coupled cluster singles and doubles (CCSD) or Kohn–Sham DFT with the PBE0 functional.[43] We employed def2-TZVP[44] as the basis set. All the molecules are assumed to be in the spin singlet state. APDFT/def2-TZVP accurately predicts energies and energy derivatives.[18, 30, 33] The energy and equilibrium geometries at the adopted calculation level are reasonably accurate, as shown in our previous study.[10] The electric dipole moment was calculated from the electron density derivatives for the APDFT2 energy according to the Hellmann–Feynman theorem (Supporting Information). Details of the atomization energy calculations are given in Ref. 10. MDM was performed using our original code. The implementation of the optimality criteria method follows the code for topology optimization.[45] For the APDFT calculations, we used a modified APDFT code[46] combined with PySCF.[47] Geometry optimization was performed with ASE.[48] Molecular structures were drawn with VMD.[49] Further details can be found in Supporting Information.

Accuracy of the electric dipole moment in APDFT

The MDM results depend on the adopted quantum alchemy method. We adopted APDFT in this work. Here, we have investigated the accuracy of APDFT EDS with respect to $(N_2, CO)_{CS}$ and

(BN-doped benzene derivatives)$_{CS}$. MDM for EDS will be demonstrated with these chemical spaces in the next section. APDFT is an approximation of the reference electronic structure theory. Therefore, we compared the results of APDFT with those of the reference theory.

The APDFT calculations of electronic EDS of CO were performed using reference N_2 (Supporting Information). Although CCSD-based APDFT EDM for CO has already been calculated,[18] we focus on the investigation of the dependence of EDS on the reference electronic structure theory. We adopted CCSD, Hartree–Fock, and PBE0 theories as the reference electronic structure theory. Regardless of the reference theory and perturbation expansion order, APDFT EDS is reasonably accurate (Table S1). As the perturbation order of APDFT increases, the error from the reference theory decreases. In particular, the first-order perturbation expansion considerably improves EDS.

The calculation results for EDSs of BN-doped benzene derivatives **1**–**18** (Figure 2) at the APDFT and PBE0 levels are shown in Figure 3. PBE0-based APDFT is used. The reference molecule is benzene **1**. The lower-order APDFT up to the third-order perturbation expansion fails to reproduce the PBE0 trend (see also Table S2). At the APDFT level, the derivative with the highest EDS is **16** (7.66 Debye), and **5** has the second-highest EDS (7.34 Debye). However, at the PBE0 level, the derivative with the highest EDS is **5** (7.80 Debye), and **16** has the second-highest EDS (6.74 Debye). APDFT underestimates EDS of **5** by 0.47 Debye but overestimates EDS of **16** by 0.92 Debye. As a result, APDFT gives an incorrect trend. This result is independent of the reference theory, perturbation expansion order, basis sets, and geometry relaxation (Figures S5, S6, S9, and S11). Moreover, the change in the reference molecule from **1** to **17** or **18** worsens the overall trend (Figure S11). At the PBE0 level, the difference in EDSs between **5** and **16** is 1.07 Debye. Accurate quantum alchemy predictions are required to search for more desirable molecules.

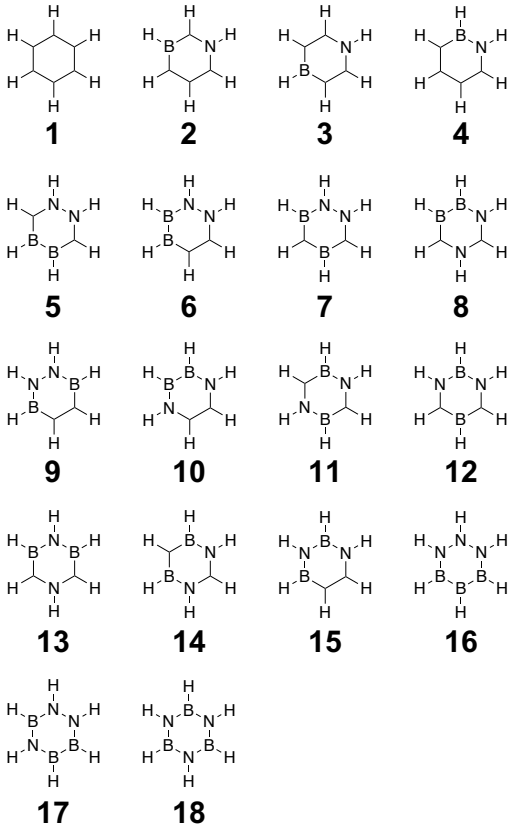

Figure 2. Candidate BN-doped benzene derivatives **1**–**18** in descending energy order.

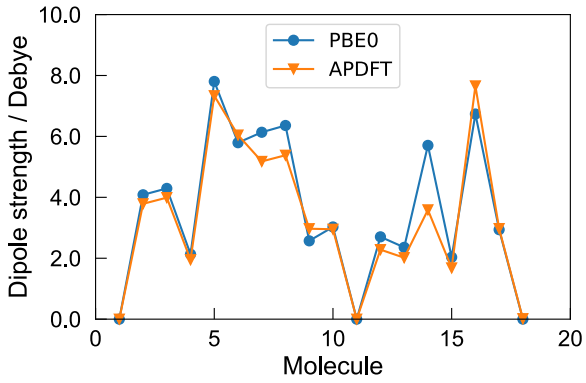

Figure 3. EDSs of BN-doped benzene derivatives **1**–**18** at the APDFT and PBE0 levels of theory. EDS was calculated in the equilibrium geometries of the derivatives. EDSs in ascending EDS order can be found in Figure S11.

We discuss possible improvements for precise molecular design. To improve the accuracy of APDFT, assuming convergence, the perturbation expansion order must be increased. Analytic derivatives of energy and electron density[33, 50-52] allow numerically stable yet computationally efficient evaluations of higher-order APDFT. Finite differences are also useful for systematically increasing the expansion order.[31, 50, 53] A hybrid approach of high- and low-level reference theories for the derivatives of the electron density is effective for reducing the computational cost of higher-order APDFT calculations.[40] A very recently developed alchemical integral transformation (AIT)[54] transforms the integral in APDFT, allowing the calculation of the target system from only the electron density of the reference system. We believe that advances in quantum alchemy will allow more precise molecular design by MDM.

Results and discussion

We apply MDM to search for functional molecules with equilibrium geometries in various chemical spaces. First, the search for a molecule with high atomization energy in simple (BF, $CO)_{CS}$ is performed to verify the performance of the optimality criteria method. The results are compared with those in Ref. 10. Second, we search for a molecule with high electric dipole strength (EDS) in simple (N_2, $CO)_{CS}$ and realistic chemical space (BN-doped benzene derivatives $B_nN_nC_{6-2n}H_6$, n = 1–3)$_{CS}$. This work focuses on the proof of concept and is limited to applications to small chemical spaces. Future work will address larger chemical space and candidate molecules with considerably different equilibrium geometries. For the APDFT reference molecules, we used N_2, CO, and benzene for (N_2, $CO)_{CS}$, (BF, $CO)_{CS}$, and (BN-doped benzene derivatives)$_{CS}$, respectively. The reference electronic structure theory is CCSD for (N_2, $CO)_{CS}$ and (BF, $CO)_{CS}$ and Kohn–Sham DFT with the PBE0 functional for (BN-doped benzene derivatives)$_{CS}$.

1. Atomization energy design for (BF, CO)$_{CS}$

We used MDM to search for a molecule with high atomization energy in (BF, CO)$_{CS}$. The results are compared with our previous work[10] (Figure 4). MDM started with BF and designed CO with higher atomization energy than BF. The atomization energy, participation coefficients $\{a_i\}$ representing the molecular species, and equilibrium bond length vary monotonically in the MDM procedure. The update of the molecular species accounts for the change in the equilibrium geometry in the optimization of the atomization energy (see Methods section). This design was also implemented in our previous study Ref. 10; the difference between this work and Ref. 10 is the algorithm for updating the molecular species. In Ref. 10, the constraints on the domain of definition for **a** (Eqs. (2) and (3)) are removed by the variable transformation, and **a** are updated according to the steepest descent formula. However, the present MDM uses the optimality criteria method (Eq. (9)). Both the steepest descent and optimality criteria methods require a few parameters. In addition, the convergence conditions for $\{a_i\}$ are different. Nonetheless, we compared the design efficiencies. We obtained identical design results regardless of the optimization algorithm. On the other hand, there is an obvious difference in design efficiency. The linear changes in atomization energy, $\{a_i\}$, and equilibrium bond length contrast with Ref. 10, where Sigmond-like changes are observed. As a result, the present MDM requires significantly fewer design steps (only 7% of steps in the previous algorithm). The bottleneck of MDM is geometry optimization, which requires APDFT calculations. The geometry optimization was performed only 33 times, which is 10 times less than in Ref. 10. Thus, the use of the optimality criteria method results in rapid convergence and small computational cost.

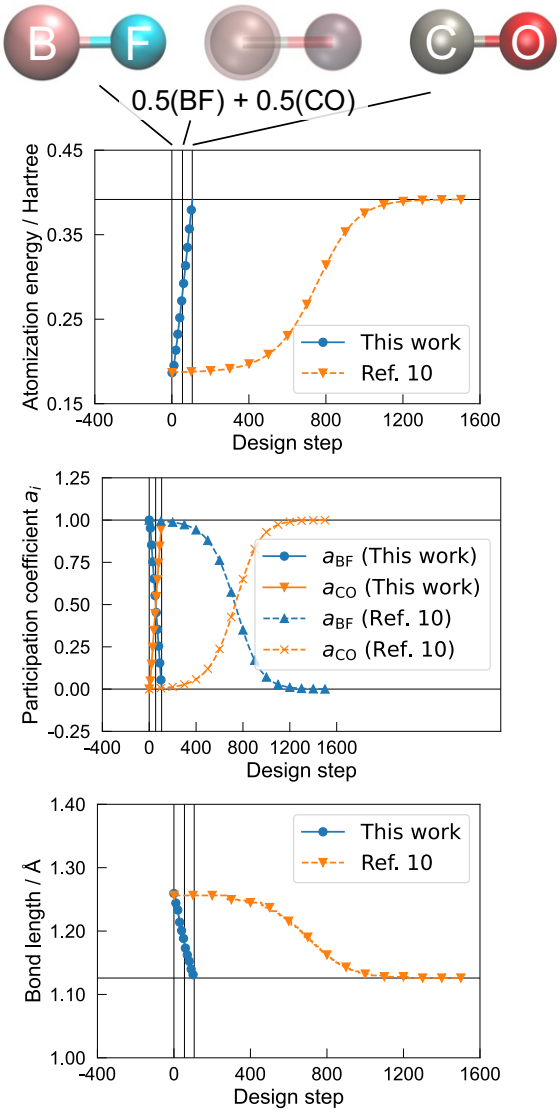

Figure 4. Results of the search for a molecule with high atomization energy in $(BF, CO)_{CS}$. Changes in atomization energy, participation coefficients, and equilibrium bond length during the search process are shown in the upper, middle, and lower panels, respectively. The results of this work and Ref. 10 are shown.

2. Electric dipole strength (EDS) design

2.1. $(N_2, CO)_{CS}$

MDM was applied to search for a molecule with high EDS in $(N_2, CO)_{CS}$. EDS of N_2 is zero because of its symmetry. CO has nonzero EDS. Therefore, the answer to the design problem was set to CO. Initiating with N_2, MDM successfully obtained CO. Figure 5 shows the variation of the EDS, $\{a_i\}$, and equilibrium bond length during the optimization process. EDS becomes higher with the progress of the design, but the change is not monotonic. When geometry relaxation occurs, EDS can slightly decrease. As mentioned in the Methods section, geometry optimization does not occur at every design step due to the finite convergence criterion. $\{a_i\}$ monotonically and linearly vary. The equilibrium bond distance monotonically and gradually varies from N_2 to CO.

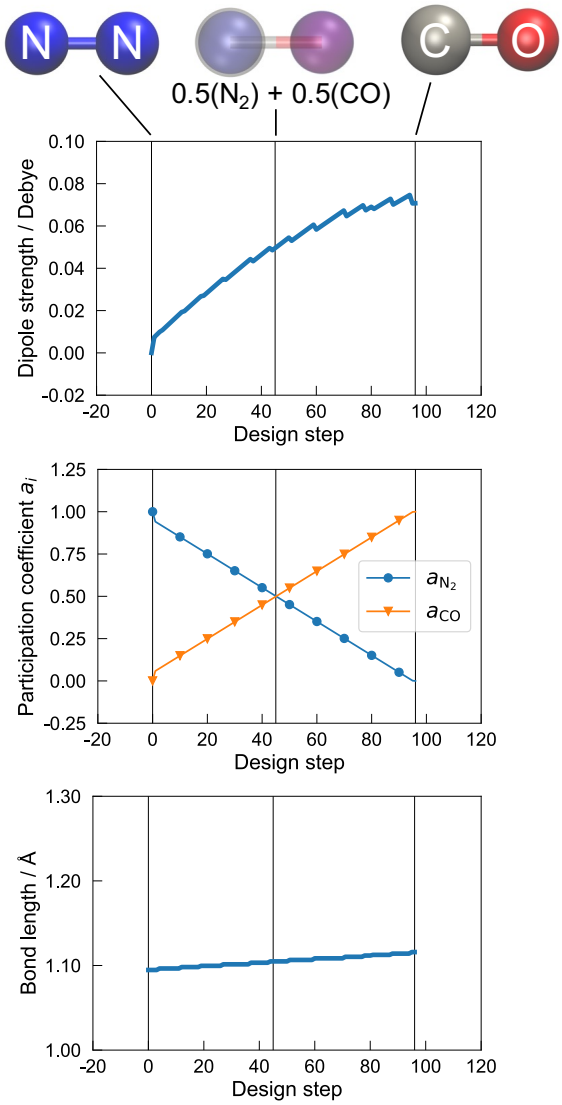

Figure 5. Results of the search for a molecule with high EDS in $(N_2, CO)_{CS}$. Changes in EDS, participation coefficients, and equilibrium bond length during the search process are shown in the upper, middle, and lower panels, respectively.

2.2. (BN-doped benzene derivatives)$_{CS}$

MDM was used to find a derivative with high EDS among 18 BN-doped benzene derivatives **1**–**18** (Figure 2). The search started with benzene **1**. EDS of **1** is zero due to its symmetry. Figure

6 shows the changes in EDS and $\{a_i\}$ during the search process. As the design proceeds, a_i of **1** rapidly decreases, and a_i of **16** rapidly increases. a_i of **5** slowly increases. The other $\{a_i\}$ remain almost unchanged. The optimized $\{a_i\}$ values are 0.05 and 0.94 for **5** and **16**, respectively, and are not fully localized, unlike the previous designs. This alchemical molecule shows EDS slightly higher than that of **16**. MDM obtains the nearest neighbor **16** in the chemical space by rounding off $\{a_i\}$ at the last design step. The results of the APDFT screening (Figure 3) show that MDM designs the derivative with the highest EDS. **16** has a six-membered ring with separated B and N moieties, and its EDS is 7.6 Debye higher than that of **1**.

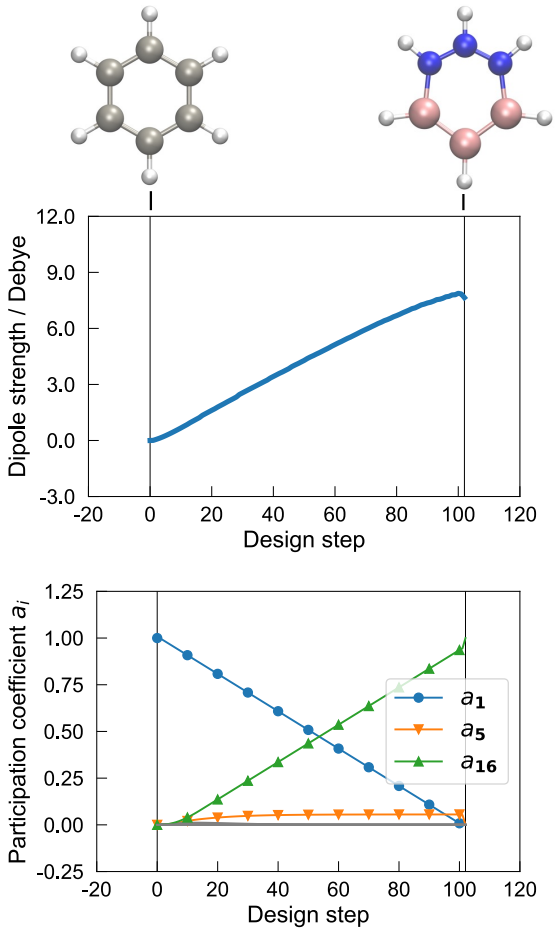

Figure 6. Results of the search for a molecule with high EDS in (BN-doped benzene derivatives)$_{CS}$. Changes in EDS and participation coefficients during the search process are shown in the upper and lower panels. The gray lines represent the changes in the participation coefficients of the derivatives other than **1**, **5**, and **16**. The initial and final molecular structures (pink, boron; gray, carbon; blue, nitrogen; white, hydrogen) are displayed.

Conclusions

A molecular design method (MDM) that explores chemical space accounting for geometric stability is demonstrated for general functionalities. We improved the optimization algorithm of MDM introduced in our previous work, which enables computationally efficient design targeting

general properties. MDM simultaneously predicts the molecular species, desired property, and equilibrium geometry of functional molecules without experimental records. MDM is based on quantum alchemy and is effective for large chemical space. APDFT was adopted as a quantum alchemy method. The electric dipole strength (EDS) and atomization energy are set to target properties for molecular design as examples. Few studies have applied APDFT to the electric dipole moment. Therefore, we investigated the accuracy with $(N_2, CO)_{CS}$ and (BN-doped benzene derivatives)$_{CS}$ prior to molecular design. The accurate APDFT EDS of CO was obtained with reference N_2. In (BN-doped benzene derivatives)$_{CS}$, we showed that APDFT up to the third-order perturbation expansion cannot correctly reproduce the EDS trend of derivatives at the reference PBE0 level of theory, regardless of the reference electronic structure theory, perturbation order, and basis set. Higher-order APDFT calculations based on analytical and numerical differentiation or integral transformation are expected to enable more accurate molecular design by MDM. We applied MDM to search for molecules with high EDS and atomization energy in the chemical spaces $(BF, CO)_{CS}$, $(N_2, CO)_{CS}$, and (BN-doped benzene derivatives)$_{CS}$. In $(BF, CO)_{CS}$, BF with high atomization energy was obtained. This result was compared with Ref. 10. It was found that the update of the molecular species by the optimality criteria method leads to rapid convergence of the optimization with small computational cost. In $(N_2, CO)_{CS}$, CO with high EDS was designed. In (BN-doped benzene derivatives)$_{CS}$, MDM designed the derivative with the highest EDS at the lower-order APDFT level of theory.

Associated content

Supporting Information

Interpolated chemical space and corresponding properties; design details; calculation methods and results of the electric dipole moment in APDFT; full citations of the references of PySCF, ASE, and Gaussian 16.

Author Information

Corresponding Author

*E-mail: shiraogawa@cat.hokudai.ac.jp

ORCID

Takafumi Shiraogawa: 0000-0002-2191-1263

Jun-ya Hasegawa: 0000-0002-9700-3309

Notes

The authors declare no competing financial interest.

Acknowledgments

T.S. acknowledges support from the Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 21J00210. J.H. thanks JSPS KAKENHI grant number JP20H02685. The present study is also supported by the Photoexcitonix Project at Hokkaido University and the Integrated Research Consortium on Chemical Science. The computation was partially performed using the Research Center for Computational Science, Okazaki, Japan (Project: 22-IMS-C002).

References

(1) Kuhn, C.; Beratan, D. N., Inverse Strategies for Molecular Design. *J. Phys. Chem.* **1996**, *100*, 10595–10599.

(2) Sanchez-Lengeling, B.; Aspuru-Guzik, A., Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360–365.

(3) Zunger, A., Inverse Design in Search of Materials with Target Functionalities. *Nat. Rev. Chem.* **2018**, *2*, 0121.

(4) Marder, S. R.; Beratan, D. N.; Cheng, L. T., Approaches for Optimizing the First Electronic Hyperpolarizability of Conjugated Organic Molecules. *Science* **1991**, *252*, 103–106.

(5) Risser, S. M.; Beratan, D. N.; Marder, S. R., Structure-Function Relationships for *B*, the First Molecular Hyperpolarizability. *J. Am. Chem. Soc.* **1993**, *115*, 7719–7728.

(6) Kirkpatrick, P.; Ellis, C., Chemical Space. *Nature* **2004**, *432*, 823.

(7) von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U., Variational Particle Number Approach for Rational Compound Design. *Phys. Rev. Lett.* **2005**, *95*, 153002.

(8) Wang, M.; Hu, X.; Beratan, D. N.; Yang, W., Designing Molecules by Optimizing Potentials. *J. Am. Chem. Soc.* **2006**, *128*, 3228–3232.

(9) Barkoutsos, P. K.; Gkritsis, F.; Ollitrault, P. J.; Sokolov, I. O.; Woerner, S.; Tavernelli, I., Quantum Algorithm for Alchemical Optimization in Material Design. *Chem. Sci.* **2021**, *12*, 4345–4352.

(10) Shiraogawa, T.; Hasegawa, J., Exploration of Chemical Space for Designing Functional Molecules Accounting for Geometric Stability. *J. Phys. Chem. Lett.* **2022**, *13*, 8620–8627.

(11) Wilson, E. B., Jr., Four-Dimensional Electron Density Function. *J. Chem. Phys.* **1961**, *36*.

(12) de Gironcoli, S.; Giannozzi, P.; Baroni, S., Structure and Thermodynamics of Si_xGe_{1-x} Alloys from Ab Initio Monte Carlo Simulations. *Phys. Rev. Lett.* **1991**, *66*, 2116–2119.

(13) Marzari, N.; de Gironcoli, S.; Baroni, S., Structure and Phase Stability of $Ga_xIn_{1-x}P$ Solid Solutions from Computational Alchemy. *Phys. Rev. Lett.* **1994**, *72*, 4001–4004.

(14) von Lilienfeld, O. A.; Tuckerman, M. E., Molecular Grand-Canonical Ensemble Density Functional Theory and Exploration of Chemical Space. *J. Chem. Phys.* **2006**, *125*, 154104.

(15) Muñoz, M.; Cárdenas, C., How Predictive Could Alchemical Derivatives Be? *Phys. Chem. Chem. Phys.* **2017**, *19*, 16003–16012.

(16) Chakraborty, S.; Kayastha, P.; Ramakrishnan, R., The Chemical Space of B, N-Substituted Polycyclic Aromatic Hydrocarbons: Combinatorial Enumeration and High-Throughput First-Principles Modeling. *J. Chem. Phys.* **2019**, *150*, 114106.

(17) Keinan, S.; Hu, X.; Beratan, D. N.; Yang, W., Designing Molecules with Optimal Properties Using the Linear Combination of Atomic Potentials Approach in an AM1 Semiempirical Framework. *J. Phys. Chem. A* **2007**, *111*, 176–181.

(18) von Rudorff, G. F.; von Lilienfeld, O. A., Alchemical Perturbation Density Functional Theory. *Phys. Rev. Res.* **2020**, *2*, 023220.

(19) Shiraogawa, T.; Ehara, M., Theoretical Design of Photofunctional Molecular Aggregates for Optical Properties: An Inverse Design Approach. *J. Phys. Chem. C* **2020**, *124*, 13329–13337.

(20) Shiraogawa, T.; Dall'Osto, G.; Cammi, R.; Ehara, M.; Corni, S., Inverse Design of Molecule-Metal Nanoparticle Systems Interacting with Light for Desired Photophysical Properties. *Phys. Chem. Chem. Phys.* **2022**, *24*, 22768–22777.

(21) Weymuth, T.; Reiher, M., Toward an Inverse Approach for the Design of Small-Molecule Fixating Catalysts. *MRS Proc.* **2013**, *1524*.

(22) Weymuth, T.; Reiher, M., Gradient-Driven Molecule Construction: An Inverse Approach Applied to the Design of Small-Molecule Fixating Catalysts. *Int. J. Quantum Chem.* **2014**, *114*, 838–850.
(23) Krausbeck, F.; Sobez, J. G.; Reiher, M., Stabilization of Activated Fragments by Shell-Wise Construction of an Embedding Environment. *J. Comput. Chem.* **2017**, *38*, 1023–1038.
(24) Keinan, S.; Paquette, W. D.; Skoko, J. J.; Beratan, D. N.; Yang, W.; Shinde, S.; Johnston, P. A.; Lazo, J. S.; Wipf, P., Computational Design, Synthesis and Biological Evaluation of *para*-Quinone-Based Inhibitors for Redox Regulation of the Dual-Specificity Phosphatase Cdc25B. *Org. Biomol. Chem.* **2008**, *6*, 3256–3263.
(25) Keinan, S.; Therien, M. J.; Beratan, D. N.; Yang, W., Molecular Design of Porphyrin-Based Nonlinear Optical Materials. *J. Phys. Chem. A* **2008**, *112*, 12203–12207.
(26) Bendsøe, M. P.; Sigmund, O., *Topology Optimization: Theory, Methods and Applications*; Springer-Verlag: Berlin, 2004.
(27) d'Avezac, M.; Zunger, A., Finding the Atomic Configuration with a Required Physical Property in Multi-Atom Structures. *J. Phys.: Condens. Matter* **2007**, *19*, 402201.
(28) Balawender, R.; Lesiuk, M.; De Proft, F.; Van Alsenoy, C.; Geerlings, P., Exploring Chemical Space with Alchemical Derivatives: Alchemical Transformations of H through Ar and Their Ions as a Proof of Concept. *Phys. Chem. Chem. Phys.* **2019**, *21*, 23865–23879.
(29) Fias, S.; Chang, K. Y. S.; von Lilienfeld, O. A., Alchemical Normal Modes Unify Chemical Space. *J. Phys. Chem. Lett.* **2019**, *10*, 30–39.
(30) Domenichini, G.; von Rudorff, G. F.; von Lilienfeld, O. A., Effects of Perturbation Order and Basis Set on Alchemical Predictions. *J. Chem. Phys.* **2020**, *153*, 144118.
(31) von Rudorff, G. F., Arbitrarily Accurate Quantum Alchemy. *J. Chem. Phys.* **2021**, *155*, 224103.
(32) Chang, K. Y.; Fias, S.; Ramakrishnan, R.; von Lilienfeld, O. A., Fast and Accurate Predictions of Covalent Bonds in Chemical Space. *J. Chem. Phys.* **2016**, *144*, 174110.
(33) Domenichini, G.; von Lilienfeld, O. A., Alchemical Geometry Relaxation. *J. Chem. Phys.* **2022**, *156*, 184801.
(34) Saravanan, K.; Kitchin, J. R.; von Lilienfeld, O. A.; Keith, J. A., Alchemical Predictions for Computational Catalysis: Potential and Limitations. *J. Phys. Chem. Lett.* **2017**, *8*, 5002–5007.
(35) Griego, C. D.; Saravanan, K.; Keith, J. A., Benchmarking Computational Alchemy for Carbide, Nitride, and Oxide Catalysts. *Adv. Theory Simul.* **2018**, *2*, 1800142.
(36) Griego, C. D.; Kitchin, J. R.; Keith, J. A., Acceleration of Catalyst Discovery with Easy, Fast, and Reproducible Computational Alchemy. *Int. J. Quantum Chem.* **2020**, *121*, e26380.
(37) Balawender, R.; Lesiuk, M.; De Proft, F.; Geerlings, P., Exploring Chemical Space with Alchemical Derivatives: Bn-Simultaneous Substitution Patterns in C_{60}. *J. Chem. Theory Comput.* **2018**, *14*, 1154–1168.
(38) Alfé, D.; Gillan, M. J.; Price, G. D., Constraints on the Composition of the Earth's Core from Ab Initio Calculations. *Nature* **2000**, *405*, 172–175.
(39) Muñoz, M.; Robles-Navarro, A.; Fuentealba, P.; Cárdenas, C., Predicting Deprotonation Sites Using Alchemical Derivatives. *J. Phys. Chem. A* **2020**, *124*, 3754–3760.
(40) von Rudorff, G. F.; von Lilienfeld, O. A., Rapid and Accurate Molecular Deprotonation Energies from Quantum Alchemy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10519–10525.
(41) Sheppard, D.; Henkelman, G.; von Lilienfeld, O. A., Alchemical Derivatives of Reaction Energetics. *J. Chem. Phys.* **2010**, *133*, 084104.

(42) Al-Hamdani, Y. S.; Michaelides, A.; von Lilienfeld, O. A., Exploring Dissociative Water Adsorption on Isoelectronically Bn Doped Graphene Using Alchemical Derivatives. *J. Chem. Phys.* **2017**, *147*, 164113.
(43) Adamo, C.; Barone, V., A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **1999**, *110*, 6158.
(44) Weigend, F.; Ahlrichs, R., Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
(45) Sigmund, O., A 99 Line Topology Optimization Code Written in Matlab. *Struct. Multidisc. Optim.* **2001**, *21*, 120–127.
(46) https://github.com/ferchault/APDFT
(47) Sun, Q.; Zhang, X.; Banerjee, S.; Bao, P.; Barbry, M.; Blunt, N. S.; Bogdanov, N. A.; Booth, G. H.; Chen, J.; Cui, Z. H., et al., Recent Developments in the PySCF Program Package. *J. Chem. Phys.* **2020**, *153*, 024109.
(48) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C., et al., The Atomic Simulation Environment-a Python Library for Working with Atoms. *J. Phys. Condens. Matter* **2017**, *29*, 273002.
(49) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33–38.
(50) von Lilienfeld, O. A., Accurate Ab Initio Energy Gradients in Chemical Compound Space. *J. Chem. Phys.* **2009**, *131*, 164102.
(51) Lesiuk, M.; Balawender, R.; Zachara, J., Higher Order Alchemical Derivatives from Coupled Perturbed Self-Consistent Field Theory. *J. Chem. Phys.* **2012**, *136*, 034104.
(52) Flores-Moreno, R.; Cortes-Llamas, S. A.; Pineda-Urbina, K.; Medel, V. M.; Jayaprakash, G. K., Analytic Alchemical Derivatives for the Analysis of Differential Acidity Assisted by the *h* Function. *J. Phys. Chem. A* **2021**, *125*, 10463–10474.
(53) Eikey, E. A.; Maldonado, A. M.; Griego, C. D.; von Rudorff, G. F.; Keith, J. A., Evaluating Quantum Alchemy of Atoms with Thermodynamic Cycles: Beyond Ground Electronic States. *J. Chem. Phys.* **2022**, *156*, 064106.
(54) Krug, S. L.; von Rudorff, G. F.; von Lilienfeld, O. A., Relative Energies without Electronic Perturbations via Alchemical Integral Transform. *J. Chem. Phys.* **2022**, *157*, 164109.

Table of Contents (TOC) Graphics