Published in final edited form as:

J Chem Phys. 2007 September 28; 127(12): 124102. doi:10.1063/1.2774980.

Design-atom approach for the QM/MM covalent boundary: A design-carbon atom with five valence electrons

Chuanyun Xiao and Yingkai Zhang^{*}

Department of Chemistry, New York University, New York, NY, 10003

Abstract

A critical issue underlying the accuracy and applicability of the combined quantum mechanical/molecular mechanical (QM/MM) methods is how to describe the QM/MM boundary across covalent bonds. Inspired by the *ab initio* pseudo-potential theory, here we introduce a novel design-atom approach for a more fundamental and transparent treatment of this QM/MM covalent boundary problem. The main idea is to replace the boundary atom of the active part with a design-atom, which has a different number of valence electrons but very similar atomic properties. By modifying the Troullier-Martins scheme, which has been widely employed to construct norm-conserving pseudo-potentials for density functional calculations, we have successfully developed a design-carbon atom with five valence electrons. Tests on a series of molecules yield very good structural and energetic results, and indicate its transferability in describing a variety of chemical bonds, including double and triple bonds.

I. INTRODUCTION

With the increase in the speed of computers and the development of more efficient algorithms for electronic structure calculations, the field of the development and application of *ab initio* quantum mechanical/molecular mechanical (QM/MM) methods is expanding rapidly [1-13]. The methods have become increasingly powerful in complementing experimental methods to elucidate the chemistry of complex systems, from the solid state and surface catalysis to chemical reactions in solution and enzymes. In the QM/MM framework [14-17], a small chemically active region is treated by a quantum mechanical method, while the remainder of the system containing a large number of atoms is described by a molecular mechanical force field. Such a combined approach can take advantage of the applicability and accuracy of high level QM methods for chemical reactions and of the computational efficiency of the MM calculations, and thus extend the realm of state-of-the-art electronic structure calculations to large systems.

A critical issue underlying the accuracy and applicability of the combined QM/MM method is how to describe the QM/MM boundary across the covalent bonds [14,16-20,13,5,1]. For example, Fig.1 illustrates the side chain of a glutamine residue participating in the chemical reaction, which needs to be treated by quantum mechanical methods. However, the active part resulting from the cleavage of the C_α-C_β bond is a radical with one free valence. The behavior of the radical is clearly very different from the original closed-shell system. Thus, it is not acceptable to simply treat the active part quantum mechanically, while the rest molecular mechanically.

^{*} yz22@nyu.edu

Over the years, a number of methods have been developed to tackle this covalent-bond-cutting boundary problem, which can be generally classified into three categories. The first is the link-atom approach [14-16,18,21-26], where additional link-atoms, generally hydrogen atoms, are introduced to cap the free valence of the active part. The second category involves the use of bonding hybrid orbitals, including the hybrid orbital method [14,27], local self-consistent field (LSCF) method [28-31], generalized hybrid orbital (GHO) method [20,32,33] and frozen orbital method [34,35]. One main drawback of the link-atom approach is the introduction of additional degrees of freedom into the system. For the methods in the second category, due to the use of hybrid orbitals, extensive theoretical formulation and substantial code development are required. Meanwhile, it has been realized that the use of hybrid orbitals alone can not lead to a satisfactory description of the QM/MM interface, and some specific parameterizations are needed.

The third category of methods neither introduces additional atoms into the system nor employs bonding hybrid orbitals. Instead, the parameterized effective core potentials (pseudo-potentials) have been employed for QM/MM boundary atoms to mimic certain desired molecular properties. These methods include the pseudobond approach [36,37], the connection-atom scheme [38], the quantum capping potential method [39,40], the effective group potential technique [41], effective Hamiltonians from a minimum principle [42], variational optimization of effective core potentials for molecular properties [43] and multicentered valence-electron effective potentials [44]. Among this category of methods, the pseudobond method developed by Zhang *et al.* [36] was the first approach developed for *ab initio* QM/MM methods and has been most widely employed in studying enzyme reactions [45-60]. It uses a seven-valence-electron atom with an effective core potential to replace the boundary carbon atom of the environmental part (C_α in Fig.1) and to form a pseudobond with the boundary atom of the active part. The pseudobond was first developed for the $C(sp^3)$ - $C(sp^3)$ bond only, and the Gaussian basis set of fluorine was used for the pseudobond atom [36]. Recently, Zhang [37] developed a new formulation to construct this seven-valence-electron boundary atom, which has its own basis set instead of that of fluorine. By parameterizing both the basis set and the effective core potential to mimic certain desired molecular properties, it has been able to not only significantly improve the $C_{ps}(sp^3)$ - $C(sp^3)$ pseudobond, but also develop accurate $C_{ps}(sp^3)$ - $C(sp^2, \text{carbonyl})$ and $C_{ps}(sp^3)$ - $N(sp^3)$ pseudobonds for the cutting of protein backbones and nucleic acid bases.

In spite of the success of the pseudobond approach [36,37], it has some limitations. First, the parameters in the effective core potentials and basis sets are parameterized against molecular properties with a limited set of molecules. The pseudobond parameters thus developed are semi-empirical in nature. Second, different pseudobonds need to be developed for different types of bonds at the boundary. Up to now, only three types of pseudobonds in connection with the Gaussian basis set 6-31G* have been developed, which limits its applicability. It should be noted that other available link-atom-free methods in treating the QM/MM covalent boundary problems, including the hybrid orbital methods and effective core potential approaches, nevertheless, share very similar limitations. Thus, a more fundamental and transparent treatment is clearly needed.

In this paper, we introduce a novel design-atom approach for this QM/MM covalent boundary problem. The main idea is to replace the boundary atom of the active part with a design-atom, which has a different number of valence electrons but very similar atomic properties to the original atom. For example, as shown in Fig.1, the C_β atom will be replaced by a design-carbon atom with 5 valence electrons (C_{design}^5) thus the resulting QM subsystem does not have a dangling valence. By constructing its pseudo-potential, the C_{design}^5 atom will be designed to have very similar **atomic** properties to the real carbon atom so that it interacts with other atoms

in the QM region like the real carbon. Inspired by the *ab initio* pseudo-potential theory [61-63], we have developed and tested a modified Troullier-Martins scheme [63] to construct the pseudo-potential of C_{design}^5 . With the constructed numerical pseudo-potential, the basis set of the design-atom is then generated with the same procedure as that for the real carbon atom.

In a series of test calculations, the resulting C_{design}^5 has been found to have good transferability in describing a variety of chemical bonds, including double and triple bonds.

II. THEORY AND METHOD

In this section, first a brief review of the norm-conserving approach to construct pseudo-potentials [61-63] is given, and a generalized norm-conserving (GNC) condition is presented, which is our initial idea to construct design-atoms with difference number of valence electrons. Then, we presented a modified Troullier-Martins scheme to construct the pseudo-potential of the design-atom. In the final subsection, the computational details to design and test the C_{design}^5 atom are described.

A. Brief review of the norm-conserving approach to construct pseudopotentials and a generalized norm-conserving condition

Pseudopotentials or effective core potentials, which mimic the effect of atomic core states on the valence electrons, have been widely used in simplifying *ab initio* electronic structure calculations, especially when dealing with heavy atoms, employing plane-wave or numerical basis sets. Over the years, many methods have been developed to construct pseudo-potentials, among which the norm-conserving approach [61-63] has been most widely employed to construct *ab initio* pseudo-potentials. The norm-conserving pseudo-potentials are mainly constructed to satisfy two key requirements: first, for the chosen atomic reference configuration, the pseudo-atomic valence eigenvalues are the same as those from all-electron calculations; second, beyond a chosen core cutoff radius, the pseudo-atomic valence wavefunction matches the corresponding valence wave-function from all-electron calculations. Since the pseudo-atom and the all-electron atom generally have the same number of valence electrons, the second requirement leads to a desired property that the pseudo-atom and the all-electron atom have the same valence electron density beyond a cutoff radius.

For our proposed design-atom, the key difference is that it has a different number of valence electrons, as well as nuclear charge, in comparison with the real atom. Thus our initial idea is to employ the following three conditions to construct the effective core potential of the design-atom:

- 1. The valence eigenvalues of the design-atom agree with those of the real atom, i.e., for a valence orbital with an angular quantum number l ,

$$\varepsilon_l^{\text{design}} = \varepsilon_l. \quad (1)$$

- 2. Beyond a chosen core cutoff radius r_c , the valence orbital of the design-atom has the same shape as that of the real atom, i.e., for a valence orbital with an angular quantum number l ,

$$u_l^{\text{design}}(r) = \eta_l u_l(r) \quad r \geq r_c \quad (2)$$

where $u_l(r) = r\phi_l(r)$ is the product of the radial orbital $\phi_l(r)$ and r , and the scaling factor η_l is a constant which does not depend on r .

- 3. Beyond a chosen core cutoff radius r_c , the valence electron density of the design-atom is equal to that of the real atom, i.e.,

$$\rho_{\text{valence}}^{\text{design}}(r) = \rho_{\text{valence}}(r) \quad r \geq r_c \quad (3)$$

In comparison with the two requirements in the *ab initio* pseudo-potential theory [61-63], the first requirement is exactly the same as Eq.1, while the second requirement, that the pseudo-atomic valence wave function agrees with that from all-electron calculations beyond a chosen core cutoff radius r_c , has been replaced by two less strict constraints. It should be noted that with Eq.3, it can be easily derived that the total charge enclosed within the core cutoff radius r_c for the design-atom, which is the sum of the electron charge and the nuclear charge, is the same as that for the real atom. Thus this can be considered as a generalized norm-conserving (GNC) condition in analogy to the norm-conserving condition in *ab initio* pseudo-potential theory [61]. Through Gauss' law in electromagnetism, this GNC condition guarantees that the electrostatic potential of the design-atom is the same as that of the real atom beyond a chosen core cutoff radius r_c . It should be noted that Eq.3 has also been directly employed by Carter and her co-workers in the development of transferable local pseudo-potentials [64], and it is consistent with the basic tenet of the density functional theory that only the electron density is of fundamental importance [65].

One way to enforce Eq. 3 is by choosing the scaling factors to satisfy the following condition:

$$\eta_l = \sqrt{\frac{N_l}{N_l^{\text{design}}}} \quad (4)$$

where N_l and N_l^{design} are the total number of valence electrons with an angular quantum number l for the real atom and the design-atom, respectively. In the case of the $\text{C}_{\text{design}}^5$ atom, the electronic configurations for the real atom and the design-atom are $2s^2 2p^2$ ($l = s, p$ and $n_s = n_p = 2$) and $2s^2 2p^3$ ($l = s, p$ and $n_s^{\text{design}} = 2, n_p^{\text{design}} = 3$), respectively. Thus, the scaling factors $\eta_s = 1, \eta_p = \sqrt{\frac{2}{3}} = 0.8165$, which will be referred to as the default GNC factors throughout the paper. Thus, the wave-functions of the real and the design-carbon atoms are related by

$$u_s^{\text{design}}(r) = \eta_s u_s(r) = u_s(r) \quad \text{for } r \geq r_{cs} \quad (5)$$

and

$$u_p^{\text{design}}(r) = \eta_p u_p(r) = 0.8165 u_p(r) \quad \text{for } r \geq r_{cp} \quad (6)$$

It should be noted that in the case of $N_l^{\text{design}} = N_l$ for all valence shells, which is the case in the conventional construction of pseudo-atoms, the scaling factors $\eta_l = 1$. Thus the generalized norm-conserving condition presented above becomes the norm-conserving condition for pseudo-atoms.

In our initial test, however, it is found that the design-atom constructed by satisfying above GNC condition does not have the proper covalent radius. So in our current construction of the

C_{design}^5 atom, we have allowed η_l to deviate from the default GNC values so that the resulting design-atom has a similar covalent radius to the real atom. It should be noted that the Eq.3 is then no longer satisfied, and the valence charge density of the design-atom will be slightly different from that of the real atom beyond the cutoff radius.

B. Modified Troullier-Martins scheme to construct pseudopotentials for the design-atom

Our overall construction scheme for the design-atom is very similar to the Troullier-Martins scheme (tm2) [63], which has been widely employed to generate norm-conserving pseudo-potentials for density functional calculations. It mainly involves the following steps:

- 1. Numerical solution of the radial part of the Kohn-Sham equation for the real atom to obtain its valence orbitals $u_l(r)$, and orbital energies ε_l .

$$\left[-\frac{1}{2} \frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V(r) \right] u_l(r) = \varepsilon_l u_l(r) \quad (7)$$

where $u_l(r) = r\phi_l(r)$. $\phi_l(r)$ is the radial wavefunction, and $V(r)$ is the effective potential.

- 2. Construction of valence orbitals $u_l^{\text{design}}(r)$ to satisfy the following equation:

$$u_l^{\text{design}}(r) = \begin{cases} \eta_l u_l(r), & r \geq r_{cl} \\ r^{l+1} e^{p(r)}, & r \leq r_{cl} \end{cases} \quad (8)$$

We can see that for $r \geq r_{cl}$, the formula is the same as in Eq.2. While in the formula for $r \leq r_{cl}$, $p(r)$ is a polynomial of order 6 in r^2 :

$$p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12} \quad (9)$$

This is the same as in the Troullier-Martins scheme (tm2) [63], and the seven coefficients of the polynomial are determined by enforcing the following seven conditions:

- (I). The normalization condition to ensure $u_l^{\text{design}}(r)$ to be normalized. This leads to:

$$\int_0^{r_{cl}} |u_l^{\text{design}}(r)|^2 dr = 1 - \eta_l^2 + \eta_l^2 \int_0^{r_{cl}} |u_l(r)|^2 dr \quad (10)$$

From the above equation, we can get

$$2c_0 + \ln \left[\int_0^{r_{cl}} r^{2(l+1)} e^{[2p(r)-2c_0]} dr \right] = \ln \left[1 - \eta_l^2 + \eta_l^2 \int_0^{r_{cl}} |u_l(r)|^2 dr \right] \quad (11)$$

It should be noted that for the regular pseudo-atom, which has $\eta_l = 1$, the above equations become the same as those in the Troullier-Martins scheme (tm2) [63]:

$$\int_0^{r_{cl}} |u_l^{\text{ps}}(r)|^2 dr = \int_0^{r_{cl}} |u_l(r)|^2 dr \quad (12)$$

and

$$2c_0 + \ln \left[\int_0^{r_{cl}} r^{2(l+1)} e^{[2p(r)-2c_0]} dr \right] = \ln \left[\int_0^{r_{cl}} |u_1(r)|^2 dr \right] \quad (13)$$

- II. Continuity of the wavefunction at r_{cl} , which yields:

$$p(r_{cl}) = \ln \left[\frac{\eta_l u_l(r_{cl})}{r_{cl}^{l+1}} \right]. \quad (14)$$

- III. Continuity of the first derivative of the wavefunction at r_{cl} , which leads to:

$$p'(r_{cl}) = \frac{u'_l(r_{cl})}{u_l(r_{cl})} - \frac{l+1}{r_{cl}}. \quad (15)$$

This equation is the same as that for the regular pseudo-atom. The equations for the second, third and fourth derivatives of the wavefunction, resulting from further differentiation of this equation, therefore, will also be the same as those for the regular pseudo-atom.

- IV. Continuity of the second derivative at r_{cl} :

$$p''(r_{cl}) = \frac{2}{a} V_{AE}(r_{cl}) - \frac{2}{a} \varepsilon_l^{AE} - \frac{2(l+1)}{r_{cl}} p'(r_{cl}) - p'(r_{cl})^2. \quad (16)$$

- V. Continuity of the third derivative at r_{cl} :

$$p'''(r_{cl}) = \frac{2}{a} V'_{AE}(r_{cl}) + \frac{2(l+1)}{r_{cl}^2} p'(r_{cl}) - \frac{2(l+1)}{r_{cl}} p''(r_{cl}) - 2p'(r_{cl}) p''(r_{cl}). \quad (17)$$

- VI. Continuity of the fourth derivative at r_{cl} :

$$p''''(r_{cl}) = \frac{2}{a} V''_{AE}(r_{cl}) - \frac{4(l+1)}{r_{cl}^3} p'(r_{cl}) + \frac{4(l+1)}{r_{cl}^2} p''(r_{cl}) - \frac{2(l+1)}{r_{cl}} p'''(r_{cl}) - 2[p''(r_{cl})]^2 - 2p'(r_{cl}) p'''(r_{cl}). \quad (18)$$

- VII. Zero curvature of the screened pseudo-potential at origin:

$$\left[\frac{d^2 V_{l,scr}^{design}(r)}{dr^2} \right]_{r=0} = 0 \quad (19)$$

This leads to

$$c_2^2 + c_4(2l+5) = 0 \quad (20)$$

- 3. Once the valence orbital for the design-atom is constructed, the screened pseudo-potential for the design atom, $V_{l,scr}^{design}(r)$, which produces the valence orbital $u_l^{design}(r)$ with the valence eigen-value ϵ_l , can be determined by inverting the radial Kohn-Sham equation.

$$V_{l,scr}^{design}(r) = \epsilon_l - \frac{l(l+1)}{2r^2} + \frac{1}{2u_l^{design}(r)} \frac{d^2 u_l^{design}(r)}{dr^2} \quad (21)$$

Here the first condition (Eq. 1) has been employed.

- 4. Finally, the ionic pseudo-potential $V_l^{design}(r)$ for the design-atom can be obtained by subtracting the Coulomb and exchange-correlation potentials from the screened pseudo-potential $V_{l,scr}^{design}(r)$.

$$V_l^{design}(r) = V_{l,scr}^{design}(r) - V_c[\rho^{design}(r)] - V_{xc}[\rho^{design}(r)] \quad (22)$$

In comparison with the Troullier-Martins scheme (tm2) [63], our modified scheme for the construction of design-atom pseudo-potential only differs in Eq. 10,11, and 14. With $\eta_l = 1$, this modified scheme is the same as the original tm2 scheme [63].

C. Computational details

We have implemented the modified Troullier-Martins (tm2) scheme in the ATOM code [66] of the SIESTA package [67-69] and employed it to generate numerical pseudo-potential for the design-atom C_{design}^5 . The cutoff radius of $r_c = 1.25$ au is used for all shells of the design-atom, which is the same as the value for the pseudo-carbon atom in the database at SIESTA website [70]. The cut-off values for generating pseudo-potentials for all other atoms are also taken from the same database [70]. All atomic calculations are spin-unpolarized, and have used the BLYP exchange-correlation functional [71,72]. The constructed numerical pseudo-potentials are employed for molecular calculations using the SIESTA package with numerical basis sets.

In SIESTA, the pseudoatomic orbitals (PAO's) associated with the pseudo-potential are used as the basis sets [73-75]. Since the free-atom orbitals are generally too diffuse as the basis sets, a confinement potential [74] is added so that the numerical atomic orbitals are zero beyond a cutoff radius. We have chosen the DZP (double zeta plus polarization) basis sets [68,74,75] for all atoms in our test calculations. In generating the numerical basis sets, the default parameters provided by SIESTA are used for the confinement potentials of the C_{design}^5 , C, H, and N atoms. The basis set for the C_{design}^5 atom is generated with the same procedure as that for the real carbon atom. For the O and S atoms, the basis sets generated with the default parameters can result in reasonable structures but yield poor deprotonation energies in comparison with the corresponding all-electron BLYP calculations using 6-31G* Gaussian basis set. Therefore, for the O atom, we have chosen the DZP basis set that was optimized for the H₂O molecule [74,70]. For the S atom, no optimized basis set is available. Considering the fact that S and Si have similar covalent bond lengths, we have chosen the parameters optimized for bulk Si (DZP 7 bohr) [68,70] to generate the numerical basis set for the S atom.

A series of test calculations are made on small molecules shown in Fig.2 which involve various types of bonds between C and other atoms, including $C_{design}\text{-H}$, $C_{design}\text{-C}$, $C_{design}\text{-N}$, $C_{design}\text{-O}$, and $C_{design}\text{-S}$ single bonds; $C_{design}=\text{C}$, and $C_{design}=\text{O}$ double bonds; and $C_{design}\equiv\text{C}$ triple bond. BLYP exchange-correlation functional [71,72] has been used in all calculations. For each molecule, we first optimized the structure with all carbon atoms being the real atoms.

Then we replaced the boundary C atom by the design-atom C_{design}^5 and removed one H atom attached to it (labeled by H in Fig.2) since C_{design}^5 will form one less bond than the regular C atom. By comparing the minimized design-atom system with the corresponding optimized real-atom system in their structural and energetic properties, we examine whether the C_{design}^5 atom acts like the original carbon atom in different chemical environments. It should be noted that in QM/MM calculations, this H* should be described as an MM atom. In our current test calculations, no QM/MM optimizations has been carried out. The main rationale is that we do not want the test results to be dependent on a specific force field employed or a specific way to treat QM/MM electrostatic interactions. However, the removal of H* also brings one potential problem: the bond angles centered at C_{design}^5 can be significantly different from those of the corresponding real-atom systems. Thus in our tests, we have performed two kinds of optimizations for design-atom systems: for the one is to perform full optimization all atoms using Cartesian coordinates, and the other is to do partial optimization by freezing the angles of the atoms attached to C_{design}^5 to values of the corresponding real-atom system using Z-Matrix. Both results are presented and being compared to the standard real-atom calculations to examine the performance of the C_{design}^5 atom.

III. TESTS AND RESULTS

A. Determination of proper scaling factors η_l

At first, we attempted to construct the pseudo-potential for the C_{design}^5 atom with the default GNC scaling factors ($\eta_s = 1.0$ and $\eta_p = 0.8165$ for the s and p shells). The resulting design-atom reproduces the valence charge density of the real carbon atom beyond the cutoff radius, and can yield very good atomic energetic properties, including ionization energies, electron affinity and valence orbital energies. However, it does not have the proper covalent radius since the resulting $C_{design}\text{-M}$ bonds are significantly longer than corresponding $C_{ps}\text{-M}$ bonds, where C_{ps} is the regular C pseudo-atom. Thus in our current construction of the C_{design}^5 atom, we have allowed the scaling factors η_l to deviate from the default GNC values so that the resulting design-atom would have a similar covalent radius to the real atom.

To examine how the scaling factors affect the covalent radius of the design atom, we have tested a series of scaling factors to calculate the $C_{design}\text{-M}$ bond lengths for the first seven molecules listed in Table II. It is found that η_s has a significant effect on the resulting $C_{design}\text{-M}$ bond lengths, while the effect of η_p is smaller. By fixing the value of η_p to the default GNC value of 0.8165, we have scanned the value of η_s from 0.8 to 1.0. Fig.3 is the dependence of the standard deviations in the $C_{design}\text{-M}$ bond lengths on the s scaling factor for the first seven molecules listed in Table II. For calculations with both the full and partial optimizations, the standard deviation is quite high for the scaling factor to be around the default GNC value of 1.0, and it becomes significantly smaller when η_s becomes less than 1.0. From Fig. 3, we can see that the curve with the partial optimization is quite flat in the range between 0.80 and 0.91 while the curve with the full optimization has the lowest value at $\eta_s = 0.91$. Meanwhile, it is found that the dependence of the standard deviation in the same set of $C_{design}\text{-M}$ bond lengths on the p scaling factor is quite insignificant, and the default GNC value $\eta_p = 0.8165$ performs

quite well. Therefore in this work, $\eta_s = 0.91$ and $\eta_p = 0.8165$ have been chosen to construct the C_{design}^5 atom, which has been used for further tests and discussions in the rest of the paper.

B. Comparison of the real atom and design-atom in their valence atomic orbitals, pseudo-potentials and valence charge densities

To see how different is the C_{design}^5 atom in comparison with the normal carbon atom, we have calculated the radial atomic orbitals, radial charge densities and pseudo-potentials for the real carbon (all-electron and pseudo-atom) and design-carbon atom, respectively, as shown in Figs. 4-8. For the better comparison, the results for the nitrogen atom which has the same number of valence electrons as the C_{design}^5 atom are also presented in the figures.

As seen from Fig.4 and Fig.5, the radial orbitals for the regular pseudo-carbon atom in the s and p shells well reproduce the all-electron orbitals beyond the cutoff radius $r_c = 1.25$ au. For the design-carbon atom, the radial orbitals for both the s and p shells are scaled down beyond the cutoff radius and they are magnified in the region with $r_c \leq 1.25$ au so that the entire orbital is normalized. The radial charge density is presented in Fig.6. For the regular pseudo-carbon atom, the norm-conserving condition is satisfied so the pseudo and all-electron charge densities agree beyond the cutoff radius $r_c = 1.25$ au. For the design-carbon atom, if the generalized norm-conserving condition were satisfied, the charge density would also reproduce the all-electron one beyond the cutoff radius. In our case, we have used a scaling factor of $\eta_s = 0.91$ for the s shell beyond the cutoff radius, so that the total charge density is different from that of the real carbon atom beyond the cutoff radius $r_c = 1.25$ au. Since the charge density beyond the cutoff radius is reduced, and the design-carbon atom has one more valence electron than the real carbon atom, the charge distribution of the design-carbon atom is largely shifted to the core region with $r \leq 1.25$, forming a high peak.

The pseudo-potentials for the regular pseudo-atom and design-atom are presented in Figs.7 and 8. Also included in these figures are the asymptotic forms of the pseudo-potentials for the

regular pseudo-atom and the design-atom at large distance, namely, $-\frac{4}{r}$ and $-\frac{5}{r}$, respectively. The striking feature is that the pseudo-potentials for both s and p shells of the design-atom are considerably deeper than those of the real atom and there is a bump near the cutoff radius. This result comes from the fact that the scaling of the wave-functions for the s and p shells of the design-atom makes them less smooth than those of the regular pseudo-atom, and the pseudo-potentials are associated with the second derivatives of the wave-functions.

Since the design-carbon atom has the same number of valence electrons as a nitrogen atom. It is interesting to make a comparison between them. Significant difference can be observed between the design-carbon and the nitrogen atoms in their orbitals, charge density and pseudo-potentials from Figs.4-8. Orbitals and charge density of the C_{design}^5 atom is more similar to the carbon atom than the nitrogen atom. The pseudo-potentials for the nitrogen atom look more similar to the real C atom than to the design-carbon atom. Thus it is clear that the C_{design}^5 atom is very different from the nitrogen atom, although both of them have five valence electrons.

C. Comparison of atomic properties between real and design carbon atoms

A comparison of the atomic properties between real carbon and design-carbon atoms is given in Table I. It indicates that the design-atom has been constructed to reproduce the orbital energies of the real carbon atom for the s and p valence shells very well. To see how the design-atom behaves in different charged states, we have calculated the ionization potential (IP) with partial charge of 0.1, 0.5 and 1.0e and the electron affinity (EA) with charge of -0.1 and $-0.5e$.

Calculation with charge $-1e$ does not converge due to the well-known problem of the long-range behavior of the exchange-correlation potentials [76]. For the smaller charge of $\pm 0.1e$, the design-atom reproduces the IP and EA very well, with an error less than 0.1 kcal/mol. For the charge of $\pm 0.5e$, the performance of the design-atom is still quite decent. The IP with 0.5e is 92.6 kcal/mol in comparison with the 90.9 kcal/mol for the real atom, while the EA with $-0.5e$ has an error of 0.5 kcal/mol. With even larger charge of 1.0e, the performance becomes not as good, which indicates a limitation of the current design-atom. Overall, the C_{design}^5 atom behaves quite well in comparison with the real carbon atom given that the charge state of the carbon atom is no more than 0.5e from the neutral state.

D. Test on small molecules

We have tested the performance of the design-carbon atom on a series of small molecules as shown in Fig.2. For each molecule, the boundary C atom is replaced by C_{design}^5 and the H* atom attached to it is removed in accordance with its modified number of valence electrons. Each structure has been optimized in two ways: fully relaxed in Cartesian coordinates (Full Opt) and partially relaxed in Z-matrix (Partial Opt) where the angles of the atoms attached to C_{design}^5 are frozen to the values of the corresponding real-atom system. The test results of the bond lengths, bond angles, and the energy differences between pairs of molecules are tabulated in Tables II-IV. These results are compared with corresponding real-atom systems.

It should be noted that the molecules in Fig.2 represent very diverse types of bonds formed between C and other atoms. It not only includes single bonds C-H, C-C, C-O, C-N and C-S, but also double bonds C=O, C=C and triple bond C $\equiv$ C. Tables II and III indicate that the C_{design}^5 atom can describe a variety of bonds quite well. The standard deviations (SD) between the design-atom systems and the corresponding real-atom systems are 0.027 Å in bond length and 2.3° in bond angle. Table IV presents the energy differences for four pairs of molecules, which are the deprotonation energies for molecules CH_3NH_3^+ , CH_3OH , CH_3SH and CH_3COOH , respectively. The mean absolute deviation (MAD) between the design-atom systems and the corresponding real-atom systems are 2.6 and 4.3 kcal/mol respectively for full and partial geometry optimizations. It should be noted that this is a very stringent test, since the design-atom is directly bonded to the deprotonation atom in these test molecules. All of these indicate that the developed C_{design}^5 atom performs quite well in describing a variety of chemical bonds.

IV. CONCLUSIONS

In this paper, inspired by the *ab initio* norm-conserving pseudopotential theory [61-63], we have introduced a novel design-atom approach to treat the QM/MM covalent boundary problem. The main idea is to replace the boundary atom of the active part with a design-atom, which has a different number of valence electrons but very similar atomic properties to the original atom. By modifying the Troullier-Martins scheme [63], which has been widely employed to construct norm-conserving pseudo-potentials for density functional calculations, we have successfully developed a design-carbon atom with five valence electrons. The pseudo-potential of C_{design}^5 is constructed to satisfy the following three conditions: first, the valence eigenvalues agree with those of the real carbon atom; second, beyond a chosen cutoff radius r_c , the valence orbital of the design-atom has the same shape as that of the real atom; and third, it has very similar covalent radius to the real carbon atom. With the constructed numerical pseudo-potential, the numerical basis set of the design-atom is then generated with the same procedure as that for the real carbon atom. In a series of test calculations, the resulting C_{design}^5

has been found to have good transferability in describing a variety of chemical bonds, including double and triple bonds. The proposed scheme to construct design-atom presented here is general. It can be employed to develop other design-atoms, and be further extended and tested for calculations in the codes using plane wave or gaussian basis sets. Work along these directions is currently in progress.

Acknowledgments

Y.Z. is grateful for the support from National Institute of Health (R01-GM079223), National Science Foundation (CHE-CAREER-0448156, CHE-MRI-0420870), NYSTAR (James D. Watson Young Investigator Award) and NYU (Whitehead Fellowship for Junior Faculty in Biomedical and Biological Sciences).

REFERENCES

- [1]. Senn HM, Thiel W. *Top. Curr. Chem* 2007;268:173.
- [2]. Lin H, Truhlar DG. *Theor. Chem. Acc* 2007;117:185.
- [3]. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM. *Chem. Rev* 2006;106:3210. [PubMed: 16895325]
- [4]. Gao J, Ma S, Major DT, Nam K, Pu J, Truhlar DG. *Chem. Rev* 2006;106:3188. [PubMed: 16895324]
- [5]. Zhang Y. *Theor. Chem. Acc* 2006;116:43.
- [6]. Mulholland AJ. *Drug Discov. Today* 2005;10:1393. [PubMed: 16253878]
- [7]. Friesner RA, Guallar V. *Annu. Rev. Phys. Chem* 2005;56:389. [PubMed: 15796706]
- [8]. Moret ME, Tapavicza E, Guidoni L, Rohrig UF, Sulpizi M, Tavernelli I, Rothlisberger U. *Chimia* 2005;59:493.
- [9]. Cui Q, Karplus M. *Adv. Protein Chem* 2003;66:315. [PubMed: 14631822]
- [10]. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, Billeter S, Terstegen F, Thiel S, Kendrick J, Rogers SC, Casci J, Watson M, King F, Karlsen E, Sjøvoll M, Fahmi A, Schafer A, Lennartz C. *Theochem-J. Mol. Struct* 2003;632:1.
- [11]. Shurki A, Warshel A. *Adv. Protein Chem* 2003;66:249. [PubMed: 14631821]
- [12]. Gao J, Truhlar DG. *Annu. Rev. Phys. Chem* 2002;53:467. [PubMed: 11972016]
- [13]. Field MJ. *J. Comput. Chem* 2002;23:48. [PubMed: 11913389]
- [14]. Warshel A, Levitt M. *J. Mol. Bio* 1976;103:227. [PubMed: 985660]
- [15]. Singh UC, Kollman P. *J. Comp. Chem* 1986;7:718.
- [16]. Field MJ, Bash PA, Karplus M. *J. Comp. Chem* 1990;11:700.
- [17]. Gao, J. *Review in Computational Chemistry*. Vol. 7. VCH; New York: 1995. p. 119-185.
- [18]. Eurenium KP, Chatfield DC, Brooks BR, Hodoscek M. *Int. J. Quantum. Chem* 1996;60:1189.
- [19]. Bakowies D, Thiel W. *J. Phys. Chem* 1996;100:10580.
- [20]. Gao J, Amara P, Alhambra C, Field MJ. *J. Phys. Chem. A* 1998;102:4714.
- [21]. Maseras F, Morokuma K. *J. Comp. Chem* 1995;16:1170.
- [22]. Lyne PD, Hodoscek M, Karplus M. *J. Phys. Chem. A* 1999;103:3462.
- [23]. Das D, Eurenium KP, Billings EM, Sherwood P, Chatfield DC, Hodoscek M, Brooks BR. *J. Chem. Phys* 2002;117:10534.
- [24]. Amara P, Field MJ. *Theor. Chem. Acc* 2003;109:43.
- [25]. König PH, Hoffmann M, Frauenheim T, Cui Q. *J. Phys. Chem. B* 2005;109:9082. [PubMed: 16852081]
- [26]. Shao YH, Kong J. *J. Phys. Chem. A* 2007;111:3661. [PubMed: 17429951]
- [27]. Kairys V, Jensen JH. *J. Phys. Chem. A* 2000;104:6656.
- [28]. Thery V, Rinaldi D, Rivail J-L. *J. Comp. Chem* 1994;15:269.
- [29]. Monard G, Loos M, Thery V, Baka K, Rivail J. *Int. J. Quantum. Chem* 1996;58:153.
- [30]. Assfeld X, Rivail J-L. *Chem. Phys. Lett* 1996;263:100.

- [31]. Ferre N, Assfeld X, Rivail JL. J. Comput. Chem 2002;23:610. [PubMed: 11939595]
- [32]. Pu JZ, Gao J, Truhlar DG. J. Phys. Chem. A 2004;108:632.
- [33]. Pu J, Gao J, Truhlar DG. ChemPhysChem 2005;6:1853. [PubMed: 16086343]
- [34]. Philipp DM, Friesner RA. J. Comput. Chem 1999;20:1468.
- [35]. Murphy RB, Philipp DM, Friesner RA. Chem. Phys. Lett 2000;321:113.
- [36]. Zhang Y, Lee TS, Yang W. J. Chem. Phys 1999;110:46.
- [37]. Zhang Y. J. Chem. Phys 2005;122:024114. [PubMed: 15638579]
- [38]. Antes I, Thiel W. J. Phys. Chem. A 1999;103:9290.
- [39]. Dilabio GA, Hurley MM, Christiansen PA. J. Chem. Phys 2002;116:9578.
- [40]. Dilabio GA, Wolkow RA, Johnson ER. J. Chem. Phys 2005;122:044708.
- [41]. Poteau R, Ortega I, Alary F, Solis AR, Barthelat JC, Daudey JP. J. Phys. Chem. A 2001;105:198.
- [42]. Yasuda K, Yamaki D. J. Chem. Phys 2004;121:3964. [PubMed: 15332942]
- [43]. von Lilienfeld OA, Tavernelli I, Rothlisberger U, Sebastiani D. J. Chem. Phys 2005;122:014113.
- [44]. Slavicek P, Martinez TJ. J. Chem. Phys 2006;124:084107. [PubMed: 16512708]
- [45]. Zhang Y, Liu H, Yang W. J. Chem. Phys 2000;112:3483.
- [46]. Liu H, Zhang Y, Yang W. J. Am. Chem. Soc 2000;122:6560.
- [47]. Zhang Y, Kua J, McCammon JA. J. Am. Chem. Soc 2002;124:10572. [PubMed: 12197759]
- [48]. Zhang Y, Kua J, McCammon JA. J. Phys. Chem. B 2003;107:4459.
- [49]. Cisneros GA, Liu H, Zhang Y, Yang W. J. Am. Chem. Soc 2003;125:10384. [PubMed: 12926963]
- [50]. Cisneros GA, Wang M, Silinski P, Fitzgerald MC, Yang W. Biochemistry 2004;43:6885. [PubMed: 15170325]
- [51]. Cheng Y, Zhang Y, McCammon JA. J. Am. Chem. Soc 2005;127:1553. [PubMed: 15686389]
- [52]. Cheng Y, Zhang Y, McCammon JA. Protein Sci 2006;15:672. [PubMed: 16522793]
- [53]. Hu P, Zhang Y. J. Am. Chem. Soc 2006;128:1272. [PubMed: 16433545]
- [54]. Corminboeuf C, Hu P, Tuckerman ME, Zhang Y. J. Am. Chem. Soc 2006;128:4530. [PubMed: 16594663]
- [55]. Lin P, Yang W, Pedersen LC, Negishi M, Pedersen LG. Int. J. Quantum Chem 2006;106:2981.
- [56]. Cisneros GA, Wang M, Silinski P, Fitzgerald MC, Yang W. J. Phys. Chem. A 2006;110:700. [PubMed: 16405343]
- [57]. Wang S, Hu P, Zhang Y. J. Phys. Chem. B 2007;111:3758. [PubMed: 17388541]
- [58]. Wang L, Yu X, Hu P, Broyde S, Zhang Y. J. Am. Chem. Soc 2007;129:4731. [PubMed: 17375926]
- [59]. Cheng Y, Cheng X, Radic Z, McCammon JA. J. Am. Chem. Soc 2007;129:6562. [PubMed: 17461584]
- [60]. Xiao C, Zhang Y. J. Phys. Chem. B 2007;111:6229. [PubMed: 17503802]
- [61]. Hamann DR, Schluter M, Chiang C. Phys. Rev. Lett 1979;43:1494.
- [62]. Bachelet GB, Hamann DR, Schluter M. Phys. Rev. B 1982;26:4199.
- [63]. Troullier N, Martins JL. Phys. Rev. B 1991;43:1993.
- [64]. Zhou BJ, Wang YA, Carter EA. Phys. Rev. B 2004;69:125109.
- [65]. Hohenberg P, Kohn W. Phys. Rev. B 1964;136:864.
- [66]. Garcia, A. ATOM User Manual, version 3.2. 2002.
- [67]. Artacho, E.; Gale, J.; Garca, A.; Junquera, J.; Martin, RM.; Ordejn, P.; Snchez-Portal, D.; Soler, JM. SIESTA 2.0, User's Guide. 2006.
- [68]. Soler JM, Artacho E, Gale JD, Garcia A, Junquera J, Ordejon P, Sanchez-portal D. J. Phys.-Condes. Matter 2002;14:2745.
- [69]. Ordejon P, Artacho E, Soler JM. Phys. Rev. B 1996;53:10441.
- [70]. <http://www.uam.es/departamentos/ciencias/fismateriac/siesta/>
- [71]. Becke AD. Phys. Rev. A 1988;38:3098. [PubMed: 9900728]
- [72]. Lee CT, Yang W, Parr RG. Phys. Rev. B 1988;37:785.
- [73]. Anglada E, Soler JM, Junquera J, Artacho E. Phys. Rev. B 2002;66:205101.
- [74]. Junquera J, Paz O, Sanchez-portal D, Artacho E. Phys. Rev. B 2001;64:235111.

- [75]. Artacho E, Sanchez-portal D, Ordejon P, Garcia A, Soler JM. Phys. Status Solidi B-Basic Res 1999;215:809.
- [76]. Rosch N, Trickey SB. J. Chem. Phys 1997;106:8940.

FIG. 1. Illustration of the QM/MM boundary across a covalent bond and the proposed design-atom approach for this problem.

FIG. 2.

Molecules used for testing the performance of the C_{design}^5 atom in a variety of chemical bonding environments. In the molecular calculations with the design-atom, the boundary carbon atom bonded to the H* atom is replaced by the design-atom C_{design}^5 and the H* atom is removed. All numerical subscripts in the figure are for labeling only.

FIG. 3. Dependence of the standard deviation in C_{design} -M bond lengths on the scaling factor in s shell (η_s) with $\eta_p = 0.8165$ for the first seven molecules listed in Table II.

FIG. 4. The radial orbital $u_{2s}(r)$ for the 2s shell. The cutoff radius $r_c = 1.25$ au.

FIG. 5.
The radial orbital $u_{2p}(r)$ for the 2p shell. The cutoff radius $r_c = 1.25$ au.

FIG. 6.
The radial charge density for the valence electrons. The cutoff radius $r_c = 1.25$ au.

FIG. 7.
The ionic pseudo-potential for the 2s shell. The cutoff radius $r_c = 1.25$ au.

FIG. 8.
The ionic pseudo-potential for the 2p shell. The cutoff radius $r_c = 1.25$ au.

TABLE I

Comparison of atomic properties between real and design carbon atoms. Orbital energies are in Rydberg, ionization potentials (charge 0.1, 0.5 and 1.0e) and electron affinity (charge -0.1 and -0.5e) are in kcal/mol. Scaling factors: $\eta_{\text{I}} = 0.91$, $\eta_{\text{p}} = 0.8165$

	Orbital			IP			EA		
	2s	2p	0.1e	0.5e	1.0e	-0.1e	-0.5e		
C_{AE}	-0.9844	-0.3704	13.2	90.9	249.7	10.9	33.8		
C_{TS}	-0.9844	-0.3704	13.2	90.9	249.4	10.9	33.8		
C_{design}	-0.9843	-0.3703	13.3	92.6	260.3	10.9	33.3		

TABLE II

Test results of the bond lengths ($\text{\AA}$) for the molecules in Fig.2. SD refers to the standard deviation between design-atom systems and corresponding real-atom systems. Scaling factors: $\eta_s = 0.91$, $\eta_p = 0.8165$

Molecule	Bond	Real-atom	Design-atom	
			Full Opt	Partial Opt
CH ₃ CH ₃	C-C1	1.537	1.534	1.574
	C1-H1	1.109	1.111	1.110
	C1-H2	1.109	1.106	1.105
CH ₃ NH ₂	C-N	1.464	1.458	1.496
	N-H1	1.023	1.024	1.031
	N-H2	1.022	1.024	1.024
CH ₃ NH ₃ ⁺	C-N	1.522	1.521	1.493
	N-H1	1.036	1.040	1.039
	N-H2	1.036	1.035	1.035
CH ₃ O ⁻	C-O	1.364	1.421	1.422
CH ₃ OH	C-O	1.448	1.449	1.480
	O-H1	0.977	0.980	0.983
CH ₃ S ⁻	C-S	1.853	1.862	1.894
CH ₃ SH	C-S	1.844	1.795	1.804
	S-H1	1.377	1.389	1.386
CH ₃ COO ⁻	C-C1	1.563	1.526	1.569
	C1-O1	1.283	1.272	1.263
	C1=O2	1.282	1.275	1.264
CH ₃ COOH	C-C1	1.513	1.458	1.533
	C1-O1	1.396	1.388	1.382
	C1=O2	1.227	1.226	1.216
	O1-H1	0.984	0.983	0.987
CH ₃ CHO	C-C1	1.512	1.512	1.541
	C1=O	1.230	1.215	1.212
	C1-H1	1.126	1.122	1.124
CHOCH ₃	C-C1	1.512	1.565	1.549
	C=O	1.230	1.296	1.290
	C1-H1	1.105	1.104	1.107
C ₂ H ₄	C1-H2	1.111	1.105	1.104
	C=C1	1.344	1.376	1.367
	C1-H1	1.101	1.104	1.103
C ₂ H ₂	C1-H2	1.101	1.101	1.103
	C≡C1	1.220	1.268	1.268
SD			0.027	0.027

TABLE III

Test results of the bond angles (degree) for the molecules in Fig.2. SD refers to the standard deviation between design-atom systems and corresponding real-atom systems. Scaling factors: $\eta_s = 0.91$, $\eta_p = 0.8165$

Molecule	Angle	Real-atom	Design-atom	
			Full Opt	Partial Opt
CH ₃ CH ₃	C-C1-H1	111.4	112.0	111.3
	C-C1-H2	111.4	108.7	108.2
CH ₃ NH ₂	C-N-H1	112.1	111.2	109.3
	C-N-H2	112.1	111.1	107.7
CH ₃ NH ₃ ⁺	C-N-H1	111.3	112.8	109.1
	C-N-H2	111.3	109.4	110.3
CH ₃ OH	C-O-H1	107.3	105.1	107.1
CH ₃ SH	C-S-H1	96.4	99.5	96.5
CH ₃ COO ⁻	C-C1-O1	115.3	112.5	113.1
	C-C1=O2	116.1	115.4	113.5
	C-C1-O1	111.0	109.0	108.4
CH ₃ COOH	C-C1=O2	126.5	125.9	128.0
	C1-O1-H1	105.7	105.5	106.7
	C-C1=O	124.8	121.1	121.7
CH ₃ CHO	C-C1-H1	115.2	116.4	115.4
	C-C1-H1	111.0	109.4	111.5
CHOCH ₃	C-C1-H2	109.4	104.8	104.0
	C-C1-H1	121.8	123.1	120.3
C ₂ H ₄	C-C1-H2	121.8	117.6	120.3
SD			2.3	2.4

TABLE IV

Energy difference (kcal/mol) for four pair of molecules in Fig.2. MAD refers to the mean absolute deviation between design-atom systems and the corresponding real-atom systems. Scaling factors: $\eta_s = 0.91$, $\eta_p = 0.8165$

process	Real-atom	Design-atom	
		Full Opt	Partial Opt
$\text{CH}_3\text{NH}_3^+ \rightarrow \text{CH}_3\text{NH}_2$	-225.0	-220.7	-222.7
$\text{CH}_3\text{OH} \rightarrow \text{CH}_3\text{O}^-$	-392.9	-394.6	-390.1
$\text{CH}_3\text{SH} \rightarrow \text{CH}_3\text{S}^-$	-370.6	-371.5	-367.7
$\text{CH}_3\text{COOH} \rightarrow \text{CH}_3\text{COO}^-$	-354.2	-350.6	-344.9
MAD		2.6	4.3