Computational and Theoretical Chemistry 1043 (2014) 38–46

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

The development and validation of a pseudoatoms approach for combined quantum mechanics and molecular mechanics calculations of polypeptide

Zhiqiang Zhang *, Liancai Xu, Guoqing Wang, Chen Shao, Ling Jiang

Department of Material and Chemical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

ARTICLE INFO

Article history:
Received 20 February 2014
Received in revised form 13 May 2014
Accepted 23 May 2014
Available online 4 June 2014

Keywords:
Boundary problem
Link atom's scheme
Effective core potential
Pseudobond approach
Genetic algorithm
Enzymatic reaction

ABSTRACT

A pseudoatoms approach was proposed for combined quantum mechanics and molecular mechanics (QM/MM) calculations of polypeptide to ensure that each amino acid residue was an indivisible unit in the partitioning of the polypeptide into QM and MM subsystems. The pseudoatoms approach broke the QM/MM boundary peptide bonds as follow: the pseudoatom $O_{ps}(sp^3)$ replacing the QM frontier amine $N(sp^3)$, and the pseudoatoms $N_{ps}(sp^2)$=$O_{ca}(sp^2)$ replacing the QM frontier $C(sp^2)$=$O(sp^2)$ carbonyl group. Conventional *ab initio* effective core potential (ECP) were constructed on the pseudoatoms to duplicate the all-electron molecular structure and charge distribution. The pseudoatoms' ECP parameters were optimized using a genetic algorithm method, and consequently validated by a test set of 19 amino acids. The pseudoatoms approach was further validated by simulating an oligopeptide involved reaction with QM/MM and full-QM methods, respectively. The results from the QM/MM simulations are similar to those from the corresponding full-QM simulations. In summary, the pseudoatoms approach may be useful for QM/MM calculations of polypeptide. Additionally, it can be implemented without program modifications due to the use of conventional ECP.

1. Introduction

Although high speed advances occurred in hardware and computational methodology, it is still a challenge to simulating large biological and chemical systems with *ab initio* or density functional theory (DFT) methods. However, these quantum mechanical (QM) methods, *ab initio*, DFT and so on, provide accurate reaction energetics which is critical for understanding the reactions involving large biological or chemical systems. On the other hand, molecular mechanical methods are efficient and computationally affordable for macromolecule systems, but are incapable of describing electron transfers which are generally involved in a chemical reaction. Thus, the accuracy and the expediency is the constant tradeoff in studies on enzymatic systems. One of the most promising approaches for calculations of enzymatic systems to date is the combined quantum mechanics and molecular mechanics (QM/MM) method which was first introduced by Warshel and Levitt [1]. The entire enzyme molecule is generally divided into the active center and its environment, namely, the QM and MM subsystems. The calculations for the active center are performed using QM methods, whereas the calculations for the environment are performed using molecular force fields. One of pivotal problems occurred in the QM/MM calculations, namely the boundary problem, is how to treat QM boundary atoms which are covalently bonded to the MM atoms. There have developed several techniques to treat the boundary atoms, for example generalized hybrid orbitals [2–4], frozen orbitals [5,6], approximate DFT method (SCC-DFTB [7]), effective capping potentials [8,9], hydrogen-capping [10,11], and fluorine-capping (based on a pseudobond approach [12–15]). The hydrogen-capping and fluorine-capping methods belong to the link atom's scheme in which link atom(s) is(are) introduced to the QM subsystem to reserve the QM/MM boundary bond(s).

The pseudobond approach is one of the most popular QM/MM methods, partially due to its implementation requiring no modifications on the source code of current QM software packages. The pseudobond approach replaced the MM frontier atom by a pseudo-fluorine to saturate the free valence on the QM frontier atom. Effective core potential (ECP) located at the pseudo-fluorine was parameterized to mimic the MM moiety at the link. The form of the ECP was same as CEP [16] and therefore the implementation of the QM calculation of pseudobond approach can be carried out without source code modifications of QM calculation software.

* Corresponding author. Tel.: +86 37186609676.
E-mail address: zhangzq@zzuli.edu.cn (Z. Zhang).

http://dx.doi.org/10.1016/j.comptc.2014.05.013
2210-271X/

This advantage makes the pseudobond approach one of the most popular QM/MM methods, and this approach has been applied for many simulations on biological systems, especially for enzymatic reactions [17–23].

The pseudo-fluorine used in the pseudobond approach has Janus-like character [24], because it participates as an ordinary MM atom in the MM calculation but also carries QM features. The pseudobond approach is in somewhat superior to hydrogen-capping approaches due to the Janus-like fluorine which avoid introducing artificial structural degrees of freedom [24,25]. However, the Janus-like pseudo-fluorine may bring a little trouble, especially for the calculations of QM subsystem's charge. The "real" QM subsystem was generally a neutral group or a group with integer charges. However, the "real" QM subsystem was linked to one or more pseudo-fluorine(s) in QM calculations, and the calculated atomic charges of the pseudo-fluorines are generally not zero. Thus, the atomic charges of the fluorine-linked QM subsystem are inconsistent with those of the "real" QM subsystem. This problem cannot be solved, unless the pseudo-fluorine is moved from the MM to the QM subsystem and abandons its Janus-like character.

Based on the pseudobond approach [14], a pseudoatoms approach was proposed to solve the inconsistency problem of the atomic charges. The pseudoatoms approach replaced the QM frontier atom by a corresponding pseudoatom which contain one extra valence electron to break the QM/MM boundary bond and saturate the free valence on it. Traditional ECP was constructed on the pseudoatom to mimic the structural effects on the QM subsystem by the broken bond at the QM/MM interface. On the other hand, amino acids are the building blocks of proteins. What is more important is that an entire amino acid in a protein is assigned an integer charge in many molecular force fields. In order to make the QM calculations consist with the MM calculations, the pseudoatoms approach considered each amino acid as an indivisible unit in the partitioning of a protein molecule into the QM and MM subsystems. It required the breaking of two peptide bonds at QM/MM interferes. For the amine group of a peptide bond, the nitrogen N(sp^3) was replaced by the pseudo-oxygen O_{ps}(sp^3). For the carbonyl group of the other peptide bond, the carbonyl group C(sp^2)=O(sp^2) was replaced by the N_{ps}(sp^2)=O_{ca}(sp^2) group. The O_{ca} was different from N_{ps} and O_{ps}, because it has the same valence electrons as oxygen. To a certain extent, the O_{ca} was an assistant pseudoatom, and it was always used with N_{ps}. The inner 1*s* electrons of all pseudoatoms including O_{ca} were placed in ECP. The major objective of this work was to optimize the ECP parameters with which the pseudoatoms can break the peptide bonds at the QM/MM interfaces without significant structural effects on QM subsystem.

2. Computational details

2.1. The effective core potentials of pseudoatoms

The ECP form of the pseudoatoms was same as CEP [16] programmed in G09 [26]. Nonrelativistically, the potential was written in the form:

$$V^{eff}(r) = V_L^{eff}(r) + \sum_{l=0}^{L-1}\left[V_l^{eff}(r) - V_L^{eff}\right]\sum_m |lm\rangle\langle lm|,$$
$$V_L^{eff} = \frac{a_L \times \exp(-b_L \times r^2)}{r}, \quad (1)$$
$$V_l^{eff}(r) - V_L^{eff}(r) = a_l \times \exp(-b_l \times r^2), \quad l = 0, 1, \ldots, L-1,$$

where L, l and m are quantum numbers of wave function, a_i and b_i are fitted parameters. The number of parameters is determined by the adopted basis-sets, for an instance, there are four parameters for 3-21G and six parameters for 6-31G* basis-set.

2.2. The training set for the optimizations of pseudoatoms' ECP parameters

N-methylacetamide was used as the all-electron model of the peptide bond (shown in Fig. 1a). The truncated models of the peptide bond were the $O_{ps}HCH_3$ and CH_3N_{ps}=O_{ca} (shown in Fig. 1b and c), corresponding to the partial structure of methylamino group and carbonyl group in the all-electron model, respectively. The carbonyl was treated as a group in which multi-centered ECP [27,28] was added into the two atoms of the carbonyl. The training set for the O_{ps} and N_{ps}=O_{ca} consist of bond length, bond angle, dihedral and Mulliken atomic charges of heavy atoms, namely the Mulliken charge of hydrogens summed into heavy atoms. Geometric optimization for all-electron model and truncated model were performed at two theoretical levels, namely, the B3LYP/3-21G and B3LYP/6-31G*. The Mulliken atomic charges of the partial structures in the all-electron model were not directly used to calculate the rms error, because the sum of the Mulliken atomic charges was not equal to zero in the two partial structures. For the model using 3-21G basis-set, the Mulliken atomic charges of N, C_1, C, O, C_2 were assigned to −0.24, 0.24, 0.50, −0.50 and 0.0, respectively. For the 6-31G* model, these charges were assigned to −0.19, 0.19, 0.50, −0.50 and 0.0, respectively. These partial charges were assigned to produce charges in accordance to Charmm27 [29] force field, and the assignment was mainly based on the following rules: (1) The Mulliken charge of carbonyl oxygen approximates to the value of −0.5 which is the Charmm27 defined partial charge. Additionally, the carbonyl group is a neutral group in Charmm27 force field. Thus, the Mulliken atomic charge of carbonyl oxygen is assigned to −0.5, and the Mulliken atomic charge of carbonyl carbon is assigned to +0.5. (2) The methyl in the acetyl group is a neutral group according to Charmm27 force field. Thus, the Mulliken atomic charge, which is the heavy atomic charge with hydrogen's summed into it, of the carbon of methyl is assigned to 0.0. (3) N-methyl group is a neutral group according to Charmm27 force field. The Mulliken atomic charge is mainly determined by the carbon but not the nitrogen which tends to receive electrons from adjacent group due to its high electronegativity. Although some other methods for computing partial charges (CHELPG, NPA, Merz-Kollman, etc.) are superior to Mulliken and exhibit less basis-set dependence, Mulliken charge was chosen in the training set for developing pseudoatoms approach mainly due to the pseudoatoms' ECP was constructed based on specific basis-set.

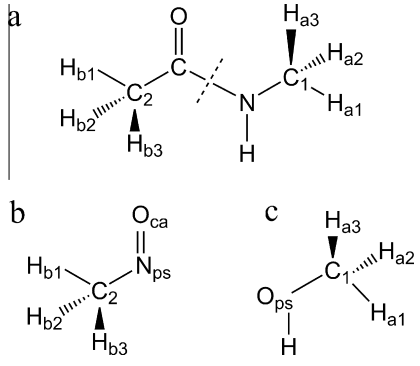

Fig. 1. The Models of all-electron and truncated peptide bond. (a) The all-electron model of peptide bond. (b) The N_{ps} model of truncated peptide bond: the peptide bond was broken by replacing the "real" carbonyl C(sp^2)=O(sp^2) to pseudoatoms group N_{ps}(sp^2)=O_{ca}(sp^2). The 1*s* electrons of N_{ps} and O_{ca} were replaced in ECP. (c) O_{ps} model of truncated peptide bond: the peptide bond was broken by replacing the "real" N(sp^3) to pseudoatom O_{ps}(sp^3) with the 1*s* electrons replaced in ECP.