Computational and Theoretical Chemistry 1034 (2014) 7–16

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

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

Accuracy of the fragment molecular orbital (FMO) calculations for DNA: Total energy, molecular orbital, and inter-fragment interaction energy

Kaori Fukuzawa [a,b,*], Chiduru Watanabe [b], Ikuo Kurisaki [c], Naoki Taguchi [d], Yuji Mochizuki [b,d], Tatsuya Nakano [e], Shigenori Tanaka [f], Yuto Komeiji [g]

[a] Mizuho Information & Research Institute, Inc., 2-3 Kanda Nishiki-cho, Chiyoda, Tokyo 101-8443, Japan
[b] Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan
[c] Graduate School of Information Science, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan
[d] Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima, Tokyo 171-8501, Japan
[e] National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya, Tokyo 158-8501, Japan
[f] Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada, Kobe 657-8501, Japan
[g] Biomedical Research Institute, AIST Tsukuba Central 2, Tsukuba 305-8568, Japan

ARTICLE INFO

Article history:
Received 14 November 2013
Received in revised form 3 February 2014
Accepted 3 February 2014
Available online 17 February 2014

Keywords:
Fragment molecular orbital method
DNA
Polynucleotide
Interaction energy
Molecular orbital
MP2

ABSTRACT

The fragment molecular orbital (FMO) method can calculate the electronic structure of macromolecules such as DNA by dividing them into several fragments and introducing suitable approximations. To establish guiding principles for FMO calculation of DNA, benchmark tests were performed for several small DNA models consisting of one or two bases or two base pairs. The effects of several factors on the accuracy of FMO calculations were investigated, including the methods used to fragment the nucleotide units, approximations for the electrostatic potential, charge neutralization, and electron correlation. It was found that charge neutralization is indispensable for the reliable calculation of energies and spatial distribution of molecular orbitals, but not necessarily so for inter-fragment interaction energy analyses, such as calculation of the base–base interaction. The electrostatic approximations were shown to have only an insignificant effect on the qualitative nature of the calculations. It was also confirmed that the base–base stacking energy can be reproduced semi-quantitatively by the Møller–Plesset second-order perturbation (MP2) method though with some overestimation, and that the overestimation can be alleviated by the spin-component-scaled MP2 method.

1. Introduction

The importance of quantum chemistry for calculating the electronic structure of proteins, sugars, lipids, DNA, RNA and other biological macromolecules cannot be denied. Nevertheless, the large size of these macromolecules still hampers their all electron *ab initio* calculations. Fortunately, recent advances in fragment-based quantum chemical methods are rapidly improving this situation [1]. Among the fragment-based methods, the fragment molecular orbital (FMO) method proposed by Kitaura et al. [2] is a promising one that has already been applied to a vast array of macromolecules, particularly proteins. See Fedorov and Kitaura [3,4] for extensive reviews.

FMO calculations of polynucleotides, *i.e.*, DNA and RNA, are more difficult than those of proteins, however. The difficulty originates mostly from the charged phosphate backbone of the polynucleotides. For this reason, molecular modeling methods such as neutralization and solvation should also be closely considered in addition to FMO calculation methods such as fragmentation and calculation levels. In calculations of polynucleotides, we thus need to investigate both the accuracy of the FMO calculation itself and the molecular modeling protocol.

Presumably reflecting this difficulty, only a limited number of papers have been published on the accuracy of FMO of DNA. The few papers include that by Watanabe et al. [5], in which the total energy was calculated at the Hartree Fock (HF)/STO-3G level under various conditions. Another paper by Sekino et al. [6] reported the total and orbital energies calculated at the HF/STO-3G and PW91/STO-3G levels. These earlier papers examined only small basis sets without dispersion; nonetheless, use of a larger basis-set and inclusion of dispersion via an appropriate electron correlation

* Corresponding author at: Mizuho Information & Research Institute, Inc., 2-3 Kanda Nishiki-cho, Chiyoda, Tokyo 101-8443, Japan. Tel.: +81 3 5281 5271; fax: +81 3 5281 5331.
E-mail address: kaori.fukuzawa@mizuho-ir.co.jp (K. Fukuzawa).

http://dx.doi.org/10.1016/j.comptc.2014.02.002
2210-271X/

method would be necessary to reproduce the stacking interaction between DNA bases [7–10].

The first choice for inclusion of the electron correlation is the Møller-Plesset perturbation theory to the second order (MP2). Hence, various FMO-MP2 calculations have been published so far and have shown that upgrading the calculation level from HF to MP2 drastically improves the accuracy of the interaction between hydrophobic amino acid residues within a protein [11], that of the π–π base–base stacking interaction within a DNA [8], and so on. Nevertheless, MP2 is known to overestimate the stabilization of any interaction energy, and thus a higher-order correlation method such as a coupled-cluster method (CCSD(T)) would be desirable [7,12]. Unfortunately, these higher-order methods are too costly to apply to macromolecules such as DNA and proteins, and therefore we have to seek a less expensive one. The spin-component-scaled MP2 approach (SCS-MP2) is one such method that can improve the overestimation by the original MP2 [13–15]. In this paper, we report FMO calculations of DNA at the HF, MP2, and SCS-MP2 levels with several basis sets and compare the accuracy of the calculations by these different FMO methods.

In FMO, in addition to the electron correlation and basis set, one needs to find an optimal way to fragment the molecule(s) and to approximate the electrostatic potential (ESP) [16] from the environment. The most reasonable way to fragment DNA may be to regard the whole nucleotide unit consisting of a base, sugar, and phosphate as a fragment. However, division of the nucleotide unit into several fragments becomes necessary if the FMO calculation is performed for an inter-fragment interaction energy (IFIE) analysis [17], by which the base–base Watson–Crick hydrogen bonding and stacking interactions can be calculated. The approximation of ESP is another factor that can alter the accuracy of FMO of DNA, which is a highly charged molecular system. Thus we here investigate several options for the optimal division of a nucleotide unit for IFIE analysis, as well as the effect of the ESP approximation on the IFIE.

In structure modeling of DNA, charge neutralization of the molecular system is regarded as necessary for FMO, much as in the classical molecular dynamic simulations of DNA, though FMO calculations are sometimes performed for charged DNA molecules [5,8,18]. There are roughly two ways to neutralize the phosphate group of DNA: one is to place a counterion on the group and the other is to cap the group with hydrogen. Sengoku et al. [19] compared the two ways and found that the former is more effective. Thus, in this paper, we examined the effect of the counterions on FMO. For further comparison, we also examined neutralization by deletion of the phosphate. These test calculations were performed for small DNA models consisting of a base, two bases, and a double strand of two base pairs (bp).

In this study, we investigate the effects of the fragmentation options, the ESP approximations, neutralization methods, basis sets, and incorporation of the electron correlation on the total energy, molecular orbital (MO) energies and spatial distribution, and IFIEs of several DNA models so as to identify guiding principles for FMO calculations. Thus, this study is intended to serve as a benchmark for future FMO calculations of DNA and, possibly, RNA.

2. Methods

2.1. DNA models

We constructed DNA models 1–5 in order to test the FMO calculations (Fig. 1). Model 1 consisted of a single deoxyribonucleoside – the base and sugar without the phosphate (A, T, G, and C; we omit the prefix *d* for deoxy throughout this paper for the sake of simplicity). Model 2 was a single deoxyribonucleotide – the base, sugar, and phosphate. Model 3 was a single-stranded two

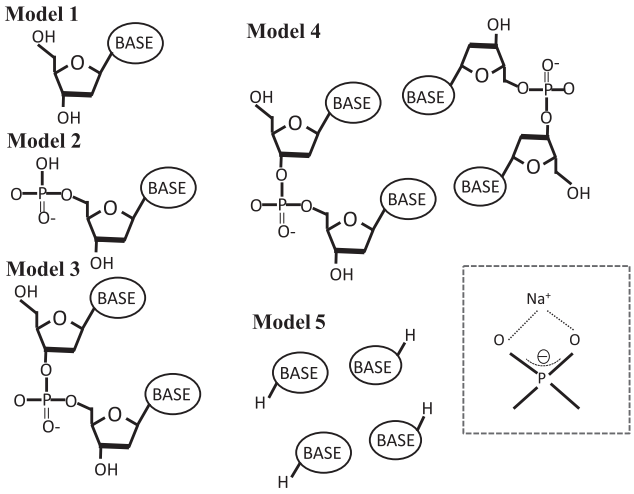

Fig. 1. Chemical definitions of models 1-4. BASE refers to A, T, G, or C. Inset: neutralization of models 2-4.

base DNA (AT, CG, GA, and AA). Model 4 was a double-stranded 2 bp DNA (AT:AT, CG:GC, GA:CT, and AA:TT). Model 5 was an H-capped two-bp DNA (AT:AT, CG:GC, GA:CT, and AA:TT) without sugar and phosphate groups.

The models were generated as follows. Model 1 was optimized at MP2/6-31G** and model 2 at HF/6-31G** by the Gaussian 03 suite program [20]. Models 3 and 4 were constructed from the crystal structure of two strands of 5′-CGCGAATTCGCG-3′ in the B-type conformation (PDB 335D) [21] as follows. The crystal structure was solvated with explicit water solvent and neutralized by a sufficient number of Na^+ ions. The whole system was annealed by the AMBER99 classical force field [22]. Then, the water and ions were stripped off, and from the remaining 12 bp DNA, models 3 and 4 were constructed by excising appropriate nucleotide(s) and by capping them with H atoms at the 5′ and 3′ ends. Model 1 itself is neutral. Models 2–4 shown in Fig. 1 are negatively charged, but a neutral structure was also provided for each model by addition of an Na^+ ion about 3 Å from a phosphate group (Fig. 1 inset). The added H and Na^+ were optimized by the AMBER99 force field.

2.2. Fragmentation of DNA

We compared three options for DNA fragmentation (Fig. 2). In option 1, the one used in our previous study [8], the formal charge on each base becomes $-1e$, which might deteriorate the accuracy in calculating a larger DNA. Hence, we tried options 2 and 3, which set the formal charge of the base to 0*e*. Both options 2 and 3 include some sugar carbons that bring about some drawbacks, however. In option 2, a carbon atom at the 1′ position of sugar ring serves both as the bond detached atom (BDA) and bond attached atom (BAA), and the base fragment containing the carbon therefore has a formal charge of 0*e*. In option 3, the sugar ring should be divided and assigned to two fragments. Carbon atoms at the 1′ and 2′ positions play roles of BDA and BAA, respectively, and the base fragment's formal charge becomes 0*e* also in this option. Each of the three options thus had a possible disadvantage for achieving a good level of accuracy, and hence their effects on the FMO accuracy were investigated in detail. Na^+ ion was included in the phosphate fragment.

2.3. Parameters of FMO calculations

We also investigated several combinations of optional FMO parameters as follows. Calculations without the electron