Asymptotic Extrapolation Scheme for Large-Scale Calculations with Hybrid Coupled Cluster and Molecular Dynamics Simulations

Karol Kowalski* and Marat Valiev*

William R. Wiley Environmental Molecular Sciences Laboratory, Battelle, Pacific Northwest National Laboratory, K8-91, P. O. Box 999, Richland, Washington 99352

Received: July 6, 2006; In Final Form: September 7, 2006

In this paper we discuss a simple extrapolation scheme based on the asymptotic behavior of the electronic energies considered as functions of cutoff factor for orbital energies corresponding to virtual orbitals. The performance of this approach is illustrated in the context of large-scale dynamic simulations for excitation energies of the cytosine molecule in its native DNA environment. We demonstrate that the extrapolation errors are significantly smaller than the excitation-energy fluctuations, due to the fluctuating environment.

1. Introduction

The low-scaling variants of the coupled cluster (CC) method[1−5] have been a focus of quantum chemists since the inclusion of triply or triply and quadruply connected clusters proved to provide a chemical level of accuracy for equilibrium ground-state properties. For obvious reasons, mainly related to relatively low-scaling and inherent parallelism, the noniterative CC approaches are the best candidates to be applicable in large-scale calculations performed on massively parallel computers. Many ground- and excited-state methods that account for the effect of triply excited clusters such as the standard CCSD[T][6] and CCSD(T)[7] approaches, the excited-state EOMCCSD(T) corrections,[8,9] the methods based on similarity transformed Hamiltonians,[10−18] and the ground- and/or excited-state renormalized[19−23] and locally renormalized approaches[24]—despite their N^7 numerical scaling with the system size N—have became or will soon become standard tools in routine highly accurate *ab initio* calculations. Several attempts were undertaken to alleviate the numerical overhead of these methods. The techniques based on Laplace or Cholesky decomposition[25−28] of perturbative denominators, methods striving at the reduction of the virtual orbital space,[29−32] or localized approaches[33−38] are very efficient in reducing the overall costs of noniterative approaches.

Some of the aforementioned methods, such as the local CC approaches, reparametrize the genuine CC methodology and lead to energies that become functions of additional parameters. Similar ideas, based on the reparametrization and subsequent extrapolation of the correlation energy, have been recently pursued by Bytautas and Ruedenberg in the context of correlation energy extrapolation by intrinsic scaling (CEEIS),[39−42] which relies on the extrapolation of the correlation energy obtained for a sequence of truncated basis sets to the complete basis set limit. Preliminary results clearly show that the CEEIS procedure is capable of providing results which are fairly close to the full configuration-interaction (FCI) energies. Other attempts to harness the various extrapolation schemes were discussed by Ayala, Scuseria, and Savin[43] in order to extrapolate the exact MBPT2 (second order of Møller−Plesset perturbation theory) results. A different approach, discussed by Iyengar, Scuseria, and Savin,[44] led to mathematically rigorous bounds for extrapolated correlation energies.

Since extrapolation approaches can merge both the accuracy of underlying methodology and relatively low computational demands, they can be considered as fit methods for treatment of large-size systems currently beyond the reach of standard CC approaches, especially in excited-state calculations. Recently, we have developed, using NWChem[45] capabilities, a suite of programs that combines classical molecular dynamics with high-level *ab initio* methods for excited states.[46] The main goal of this effort was to create a framework for realistic, temperature-dependent, excited-state calculations for biochemical systems with an approximate description of the effects of the native environment, including its dynamic fluctuations. Since thermal averaging involves multiple calls to rather expensive ab initio procedures, the low-scaling extrapolation schemes may play a pivotal role in further advancing this area. However, this is likely to happen only if the errors due to the extrapolation procedure are negligible compared to standard fluctuations in the excitation-energy values due to a fluctuating environment.

Usually, the size of systems of biological importance prohibits the use of extensive basis sets that include diffuse functions or functions of triple- or quadruple-ζ quality. On the other hand, large basis sets are required to obtain a quantitative consensus between experimental and theoretical predictions.

The main purpose of this paper is to address, on a very basic level, these important issues. We use a very simple τ-parameter-dependent cutoff scheme for the virtual orbitals with orbital energies lying above the cutoff factor. On the basis of that, we will derive simple heuristic formulas, subsequently used in the extrapolation procedures, that describe an asymptotic dependence of τ-dependent energies. The performance of the asymptotic extrapolation scheme for completely renormalized equation-of-motion approach with the singles, doubles, and noniterative triples (CR-EOMCCSD(T)) method is illustrated on the excited-state calculation of cytosine base in its native DNA environment.

2. Theory

In the combined coupled-cluster and molecular mechanics (CC/MM) approach, the system is described by the Hamiltonian

$$H = H_{\mathrm{QM}} + H_{\mathrm{QM/MM}} + H_{\mathrm{MM}} \quad (1)$$

---

* To whom correspondence should be addressed. E-mail: karol.kowalski@pnl.gov.; marat.valiev@pnl.gov.

10.1021/jp064266p CCC: $33.50

where H_{QM} is the standard many-electron Hamiltonian describing the internal energy of the quantum-mechanical (QM) region

$$H_{\mathrm{QM}} = E_{\mathrm{QM}}^{(0)} + \sum_{\mu,\nu} f^{\mu}_{\nu} a^{\dagger}_{\nu} a_{\mu} + \frac{1}{4} \sum_{\mu,\nu,\lambda,\kappa} v^{\mu\nu}_{\lambda\kappa} a^{\dagger}_{\lambda} a^{\dagger}_{\kappa} a_{\nu} a_{\mu} \qquad (2)$$

where the indices μ, ν, λ, and κ designate single-particle states and the elements f^{μ}_{ν} and $v^{\mu\nu}_{\lambda\kappa}$ represent one- and two-electron integrals, respectively, whereas $a^{\dagger}_{\lambda}$ (a_{λ}) operators are the usual creation (annihilation) operators. The interaction between the QM region and its surroundings (MM region) is contained in the second term $H_{\mathrm{QM/MM}}$

$$H_{\mathrm{QM/MM}} = \sum_{i,\mu,\nu} \left\langle \mu \left| \frac{Q_i^m}{|\mathbf{R}_i^m - \mathbf{r}|} \right| \nu \right\rangle a^{\dagger}_{\mu} a_{\nu} + V(\{\mathbf{R}^m\},\{\mathbf{R}\}) \qquad (3)$$

where Q_i^m and $\mathbf{R}_i^m$ denote charges and coordinates of the MM region. The $V(\{\mathbf{R}^m\},\{\mathbf{R}\})$ term represents the interaction between nuclei in MM and QM regions ($\{\mathbf{R}\}$ represents symbolically the set of nuclear coordinates in the QM region). The third term in eq 1, H_{MM}, describes the internal energy of the MM region represented by an Amber-type force field.[47] The Hamiltonian $\tilde{H}$, defined as

$$\tilde{H} = H_{\mathrm{QM}} + H_{\mathrm{QM/MM}} \qquad (4)$$

effectively includes, through the $H_{\mathrm{QM/MM}}$ term, the interaction of the environment with the QM region.

In our CC/MM simulations of excited states we decided to use the EOMCC formalism as one of the most established methodologies to calculate excited-state properties. Its basic variant with singles and doubles (EOMCCSD) (EOMCCSD-related methods were used in the context of advanced QM/MM simulations described in refs 48−52) has proven to provide a satisfactory description of vertical excitation energies for a class of excited states dominated by single excitations with respect to the reference function $|\Phi\rangle$—customarily chosen to be a Hartree−Fock (HF) determinant. In the EOMCCSD approach the wave function corresponding to the K state takes the form

$$|\Psi_K\rangle = (R_{K,0} + R_{K,1} + R_{K,2}) e^{T_1+T_2} |\Phi\rangle \qquad (5)$$

where T_1 and T_2 are singly and doubly excited cluster operators and $R_{K,i}$ ($i = 0-2$) represent i-tuply excited components of the excitation operators for a given state. For obtaining the excitation energies, a two-step procedure is invoked. First we solve the CCSD equations for cluster amplitudes

$$\langle \Phi^{a_1 \ldots a_n}_{i_1 \ldots i_n} | \bar{H}^{\mathrm{CCSD}} | \Phi \rangle = 0 \qquad (i = 1, 2) \qquad (6)$$

where $\bar{H}^{\mathrm{CCSD}}$ is the CCSD similarity transformed Hamiltonian, $\bar{H}^{\mathrm{CCSD}} = e^{-(T_1+T_2)} \tilde{H} e^{T_1+T_2}$. Second, excited-state energies (E_K^{EOMCCSD}) are obtained by diagonalizing the matrix representation of the $\bar{H}^{\mathrm{CCSD}}$ operator, $\bar{\mathbf{H}}^{\mathrm{CCSD}}$, in the space spanned by the reference function and all singly and doubly excited configurations

$$\bar{\mathbf{H}}^{\mathrm{CCSD}} \mathbf{R}_K^{\mathrm{EOMCCSD}} = E_K^{\mathrm{EOMCCSD}} \mathbf{R}_K^{\mathrm{EOMCCSD}} \qquad (7)$$

In the above equation, the $\mathbf{R}_K^{\mathrm{EOMCCSD}}$ vector components correspond to $R_{K,0}$ and all singly and doubly excited amplitudes defining the $R_{K,1}$ and $R_{K,2}$ operators, respectively. As a rule of thumb, the EOMCCSD method works well for the singly excited states. To cope with more complicated states that have non-negligible doubly excited components, one should include the effect of triply excited configurations in either an iterative or a noniterative manner. In this paper we will use the noniterative CR-EOMCCSD(T) δ(IA) approximation in which the due-to-triples corrections expressed in terms of triply excited moments are directly added to the EOMCCSD excitation energies,[22] $\omega_K^{\mathrm{EOMCCSD}}$:

$$\omega_K^{\mathrm{CR-EOMCCSD(T)}} = \omega_K^{\mathrm{EOMCCSD}} + \delta_K(\mathrm{IA}) \qquad (8)$$

$$\delta_K(\mathrm{IA}) = \sum_{i<j<k;a<b<c} Z^{abc}_{K,ijk} M^{ijk}_{K,abc} / D_K \qquad (9)$$

where the tensor $Z^{abc}_{K,ijk}$ is defined as

$$Z^{abc}_{K,ijk} = \left\langle \Phi \left| \left[R_{K,0}\left(T_1 T_2 + \frac{1}{6} T_1^3\right) + R_{K,1}\left(T_2 + \frac{1}{2} T_1^2\right) + R_{K,2} T_1 + \tilde{R}_{K,3} \right]^{\dagger} \right| \Phi^{abc}_{ijk} \right\rangle \qquad (10)$$

with $\tilde{R}_{K,3}$ representing an approximation of the exact, triply excited $R_{K,3}$ operator, where the amplitudes, $\tilde{R}^{ijk}_{K,abc}$, are set equal to $M^{ijk}_{K,abc} / (\omega_K^{\mathrm{EOMCCSD}} + \epsilon_i + \epsilon_j + \epsilon_k - \epsilon_a - \epsilon_b - \epsilon_c)$ (ϵ's correspond to the HF orbital energies), and $M^{ijk}_{K,abc}$ represent the three-body moments of the EOMCCSD equations. Even for single-point calculations the CCSD/EOMCCSD and CR-EOMCCSD(T) approaches are characterized by quite large computational demands of $n_{\mathrm{o}}^2 n_{\mathrm{u}}^4$ and $n_{\mathrm{o}}^3 n_{\mathrm{u}}^4$, respectively, where n_{o} (n_{u}) designates the number of correlated occupied (unoccupied) orbitals. For QM/MM simulations this situation is even worse, due to the need for multiple calls of expensive QM-related procedures. To reduce this prohibitively large numerical cost, in the last few decades we have witnessed an enormous effort striving at reducing the scaling of the approximate CC/MBPT/EOMCC/CI approaches in order to make them applicable in situations characterized by considerable system size or large dimension of the basis set. Also, in all approaches that attempt to achieve the infinite basis set limit, the correlation energies (as well as the underlying Hartree−Fock) were parametrized with respect to the quality of the basis set.[53−57] In this paper we want to discuss a simple extrapolation model for calculating excitation energies, which is parametrized with respect to a cutoff factor for orbital energies. We will show that the errors of the extrapolation model developed here are an order of magnitude smaller than the excitation energy discrepancies characteristic for a typical QM/MM simulation. Therefore, in the context of high-level QM/MM calculations these simplified schemes may assume more practical dimension by making, at least at an approximate level, the use of more extensive basis sets in realistic simulations feasible.

Let us assume that we decided to use in the calculations of energies corresponding to a given basis set only a limited subset of all virtual orbitals with corresponding orbital energies less than a given threshold τ. The set of these virtual orbitals along with all correlated occupied orbitals will be denoted as Ω_τ while all remaining virtual orbitals form the set $\bar{\Omega}_\tau$, so we have

$$\Omega = \Omega_\tau + \bar{\Omega}_\tau \qquad (11)$$

where the set Ω is composed of all correlated orbitals. In the next step we will try to relate the results of a given correlated method that uses only orbitals from the Ω_τ set with the results of the same method which employs the full set of orbitals, Ω, for sufficiently large values of τ parameter. This analysis can be conveniently carried out using standard perturbative reasoning. Without loss of generality, let us focus our attention on any truncated configuration interaction method that uses a subspace spanned by up to n-tuply excited configurations with respect to a given Hartree−Fock determinant. This space can be defined by the projection operator P. The orbital set decomposition (11) naturally entails related decomposition of the P space:

$$P = P_\tau + Q_\tau \tag{12}$$

where P_τ represents the subspace of P space which contains the Slater determinants constructed from the orbitals defining the Ω_τ set. Its orthogonal complement in the P space will be referred to as the Q_τ space. To relate the low-lying energies $E_{K,\tau}$ and corresponding states $|\Phi_{K,\tau}\rangle$ ($K = 0, 1, 2, ..., M$, $M \ll N_\tau$, $N_\tau = dim\ P_\tau$) obtained by diagonalizing the matrix representation of the Hamiltonian operator in the space P_τ, $H_\tau = P_\tau H P_\tau$, with the target eigenvalues E_K and eigenvectors $|\Psi_K\rangle$ ($K = 0, 1, 2, ..., M$) of the PHP operator, let us consider the perturbed problem

$$(H_{0,\tau} + \lambda V_\tau)|\Psi_{K,\tau}(\lambda)\rangle = E_{K,\tau}(\lambda)|\Psi_{K,\tau}(\lambda)\rangle \tag{13}$$

where τ-dependent $H_{0,\tau}$ and V_τ operators are defined as

$$H_{0,\tau} = P_\tau H P_\tau + Q_\tau F Q_\tau \tag{14}$$

$$V_\tau = Q_\tau H P_\tau + P_\tau H Q_\tau + Q_\tau(H - F)Q_\tau \tag{15}$$

In the above equations the F operator represents the Fock operator. The solutions of eq 13 for $\lambda = 0$ correspond to eigenvalues of the H_τ operator, i.e., $E_{K,\tau}(0) = E_{K,\tau}$ and $|\Psi_{K,\tau}(0)\rangle = |\Phi_{K,\tau}\rangle$ ($K = 0-2, ..., M$), while for $\lambda = 1$ we obtain $E_{K,\tau}(1) = E_K$ and $|\Psi_{K,\tau}(1)\rangle = |\Psi_K\rangle$ ($K = 0, 1, 2, ..., M$). Once the values of cutoff factor τ are chosen to be sufficiently large, we can expect that each $|\Phi_{K,\tau}\rangle$ constitutes a very good approximation for the target state $|\Psi_K\rangle$ and, therefore, the V_τ operator for these states can be considered as a small perturbation. In such a situation it is justified to expect that the second-order corrections, $E^{(2)}_{K,\tau}$, to $E_{K,\tau}$ recover E_K almost entirely: i.e.

$$E_K \simeq E^{(2)}_{K,\tau} + E_{K,\tau} \tag{16}$$

Using the standard MBPT methodology and assuming intermediate normalization with respect to a given reference $|\Phi_{K,\tau}\rangle$ one can show that $E^{(2)}_{K,\tau}$ takes the form

$$E^{(2)}_{K,\tau} = \langle\Phi_{K,\tau}|\ V_\tau R^{(0)}_{K,\tau} V_\tau|\Phi_{K,\tau}\rangle \tag{17}$$

where

$$R_{K,\tau} = \sum_\Delta \frac{|\Phi_{\Delta,\tau}\rangle\langle\Phi_{\Delta,\tau}|}{E_{K,\tau} - E_{\Delta,\tau}} \tag{18}$$

where the index Δ is used to label the configurations from the Q_τ space and $E_{\Delta,\tau} = \langle\Phi_{\Delta,\tau}|\ F|\Phi_{\Delta,\tau}\rangle$. Again, once τ is sufficiently large all $E_{\Delta,\tau} - E_{K,\tau}$ values are simply on the order of τ. If so, in this asymptotic limit, the E_K energies can be expressed as

$$E_K \simeq E_{K,\tau} + \frac{A_{K,\tau}}{\tau} \tag{19}$$

where $A_{K,\tau} = -\langle\Phi_{K,\tau}|V_\tau Q_\tau V_\tau|\Phi_{K,\tau}\rangle$. The addition of the higher order corrections to $E_{K,\tau}$ energies results in a more detailed description of the asymptotic behavior

$$E_K \simeq E_{K,\tau} + \frac{A_{K,\tau}}{\tau} + \frac{B_{K,\tau}}{\tau^2} + \frac{C_{K,\tau}}{\tau^3} + ... \tag{20}$$

Since all basis sets used in calculations are of finite dimension, we should formally require that

$$A_{K,\tau} = B_{K,\tau} = C_{K,\tau} = ... = 0\ \forall_{\tau \geq \epsilon_U} \tag{21}$$

where ϵ_U is the orbital energy of the highest unoccupied orbital. These constraints simply reflect the fact that all one-particle basis sets used in calculations are of finite dimension. In terms of excitation energies, which are the focus of our interest, these formulas translate to

$$E_K - E_0 = \omega_K = \omega_{K,\tau} + \frac{\alpha_{K,\tau}}{\tau} + \frac{\beta_{K,\tau}}{\tau^2} + \frac{\gamma_{K,\tau}}{\tau^3} + ... \tag{22}$$

where $\alpha_{K,\tau} = A_{K,\tau} - A_{0,\tau}$, $\beta_{K,\tau} = B_{K,\tau} - B_{0,\tau}$, $\gamma_{K,\tau} = C_{K,\tau} - C_{0,\tau}$, etc. Our attention will be on the simplest case when only the first two terms on the right-hand side of eq 22 are taken into account, i.e.

$$\omega_K \simeq \omega_{K,\tau} + \frac{\alpha_{K,\tau}}{\tau} \tag{23}$$

This formula can be used in defining a simple extrapolation scheme. Let us assume that we performed a number of calculations $\{\omega_{K,\tau_i}\}_{i=1}^m$ corresponding to increasing values of the τ parameter, i.e., $\tau_i > \tau_j$ for $i > j$. For sufficiently large values of τ parameters, which ensure fast convergence of the MBPT expansion, (13), one can expect functions of the type

$$f(\tau,\mathbf{x}) = x_1 + \frac{\Gamma(\tau,\mathbf{x})}{\tau} \tag{24}$$

where components of the vector $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ are the parameters whose values are variationally optimized to provide the best fit to the set of $\{\omega_{K,\tau_i}\}_{i=1}^m$ values. We will assume that for the optimum $\mathbf{x}$ vector we have

$$\bar{\omega}_K = f(\epsilon_U,\mathbf{x}) \simeq \omega_K \tag{25}$$

Obviously, we do not know the functional dependence of the $\Gamma(\tau,\mathbf{x})$ function but it is justified to ponder the $\Gamma(\tau,\mathbf{x})$ function also as a decreasing function of the τ parameter. The most apparent choice of $\Gamma(\tau,\mathbf{x})$ can be defined as $\sum_{i=1}^p x_{i+1}/\tau^i$, which leads to the class of f functions

$$f_p(\tau,\mathbf{x}) = x_1 + \sum_{i=1}^p \frac{x_{i+1}}{\tau^{i+1}} \tag{26}$$

In the next section we will employ the f_p functions for $p = 1-3$. We believe that the results derived for the truncated CI method are also valid for other parametrizations of the wave function. In fact, using the second quantized formalism, similar results can be derived for the CC and EOMCC methodology. We will apply the results of this section to the CR-EOMCCSD(T) vertical excitation energies corresponding to different values of τ parameters.

On the basis of the decomposition of the second quantized form of the Hamiltonian, cluster operator T, and excitation operator R

$$X = X_\tau + \bar{X}_\tau \qquad (X = \tilde{H}, T, R) \tag{27}$$

TABLE 1: Results of Several Asymptotic Extrapolation Models for the $\pi\pi^*$ Excitation Energy of the Cytosine Molecule in the cc-pVDZ, aug-cc-PVDZ, cc-pVTZ, and cc-pVQZ[68,69] Basis Sets[a]

basis	$\omega^{\mathrm{CR(T)}}_{\tau=1.0}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.2}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.4}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.6}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.8}$	$\omega^{\mathrm{CR(T)}}$	$f_1(\epsilon_{\mathrm{U}},\mathbf{x})$	$f_2(\epsilon_{\mathrm{U}},\mathbf{x})$	$f_3(\epsilon_{\mathrm{U}},\mathbf{x})$	ρ
Five-Point Extrapolation										
cc-pVDZ	5.22	5.15	5.13	5.13	5.10	5.02	5.06	5.11	4.92	0.18
aug-cc-pVDZ	5.00	4.98	4.96	4.93	4.91	4.85	4.90	4.83	4.77	0.13
cc-pVTZ	5.13	5.08	5.06	5.03	5.02		4.98	4.97	4.92	0.06
cc-pVQZ	5.08	5.04	5.01	4.99	4.98		4.94	4.93	4.90	0.04
Three-Point Extrapolation										
cc-pVDZ	5.22	5.15	5.13			5.02	5.04	5.17	5.13	0.13
aug-cc-pVDZ	5.00	4.98	4.96			4.85	4.93	4.88	4.87	0.06
cc-pVTZ	5.13	5.08	5.06				4.98	5.01	4.98	0.03
cc-pVQZ	5.08	5.04	5.01				4.94	4.94	4.91	0.03

[a] Five and three excitation energies corresponding to listed values of τ parameters were used in the extrapolation. The exact CR-EOMCCSD(T) excitation energy is denoted as $\omega^{\mathrm{CR(T)}}$. The values of ϵ_U are discussed in the text. The ρ quantities are defined as $\max_{i,j}|f_i - f_j|$.

where X_τ represents part of the X operator expressed in terms of spin−orbital indices from the set Ω_τ while each term in $\bar{X}_\tau$ contain at least one index from $\bar{\Omega}_\tau$, we can easily derive similar results for CC-based approaches not only for excitation energies but also for properties. Using this decomposition for the T and $\tilde{H}$ operators, the corresponding CC equations can be written in the form

$$P_\tau[(\tilde{H}_\tau e^{T_\tau})_C + (\bar{\tilde{H}}_\tau e^{T_\tau+\bar{T}_\tau})_C]|\Phi = 0 \tag{28}$$

$$Q_\tau[(\tilde{H}_\tau e^{\bar{T}_\tau})_C + (\bar{\tilde{H}}_\tau e^{T_\tau+\bar{T}_\tau})_C]|\Phi\rangle = 0 \tag{29}$$

where P_τ and Q_τ are projection operators (defined in the same way as the operators in eq 12) on the manifold of excited configurations used to define the cluster operator $T = T_\tau + \bar{T}_\tau$. Simple analysis shows that the $\bar{T}_\tau$ operator should reveal $1/\tau$ bahavior for sufficiently large values of τ (for details see ref 58). Similar conclusions can be inferred from the analysis of equations for Λ operator used in the context of gradient and property calculations. Using the bivariational expression for CC expectation value for operator ρ

$$\langle\rho\rangle = \langle\Phi|(1 + \Lambda_\tau + \bar{\Lambda}_\tau)(e^{-(T_\tau+\bar{T}_\tau)}(\rho_\tau + \bar{\rho}_\tau)e^{T_\tau+\bar{T}_\tau})|\Phi\rangle \tag{30}$$

we arrive at the formula

$$\langle\rho\rangle = \langle\rho_\tau\rangle + \bar{\nu}_\tau \tag{31}$$

where $\bar{\nu}_\tau$ decays at least as $1/\tau$ in the asymptotic limit. Obviously the minimum value of the τ parameter that guarantees the $1/\tau$ bahavior can be contingent upon the system and basis set employed.

3. Results and Discussion

As a representative application to test our extrapolation models, we have focused on the calculation of the lowest $\pi\pi^*$ excited state of cytosine base in the DNA environment. The main motivation behind this choice of system is the experimentally observed efficient excited-state deactivation mechanism that protects DNA bases against photochemical damage. To understand the mechanism of radiationless internal conversion, several excited-state models have been intensively studied in the past decade (see refs 59−66 and references therein). All of these calculations were performed either for cytosine or the cytosine−guanine pair in the gas phase or for the hydrated form of the cytosine. Since for the DNA bases the effect of fluctuating environment can be quite sizable, we have recently addressed this problem in our hybrid CC/MM approach, which allows for sampling the excitation energies in the course of molecular dynamics (MD) simulation. To calculate the excitation energies of the quantum region, we chose to use the variant of the CR-EOMCCSD(T) method defined by eqs 8 and 9. In all calculations the core orbitals were kept frozen.

Let us briefly describe only the basic tenets of this calculation (for details see ref 46). The system considered in this work consisted of the 12-mer fragment of B-DNA (3′-TCGCGT-TGCGCT-5′) solvated in a rectangular box (51 × 51 × 69 Å) of SPC/E[67] water. To neutralize the charge, 22 sodium ions were also added to the system, resulting in a total of 18 060 atoms. After initial optimization the system was brought to equilibrium by warming in stages (50 K increments) over the course of 60 ps of classical molecular dynamics simulation. Dynamic trajectories were generated with constant temperature and pressure (298.15 K, 1.025 × 10^5 Pa) molecular dynamics simulations using a 15 Å cutoff. The excited-state calculations were based on a quantum representation of cytosine base capped with a hydrogen link atom in the field of the entire DNA−water complex (18 048 point charges).

The first test consisted of excited-state calculations on a single reference snapshot taken from the classical MD trajectory. The snapshot was chosen to provide the smallest discrepancy between resulting vertical excitation energy for the $\pi\pi^*$ state and its time-averaged value within full cc-pVDZ calculations. All snapshot calculations were performed in the field of surrounding environment. We used four basis sets, cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ,[68,69] in order to illustrate the performance and dependence of our extrapolation models on the dimension of basis set employed.

Table 1 collects the CR-EOMCCSD(T) results obtained using various extrapolation schemes, which employ $f_1(\tau,\mathbf{x})$, $f_2(\tau,\mathbf{x})$, and $f_3(\tau,\mathbf{x})$ trial functions described in the previous section. The components of $\mathbf{x}$ vectors are subject to variational optimization in order to provide the best fit to the calculated points. Two schemes have been studied: the first one uses five points corresponding to $\tau_1 = 1.0$, $\tau_2 = 1.2$, $\tau_3 = 1.4$, $\tau_4 = 1.6$, and $\tau_4 = 1.8$ to extrapolate the results for ϵ_{U}, while the other is based on the three-point extrapolation for $\tau_1 = 1.0$, $\tau_2 = 1.2$, and $\tau_3 = 1.4$. For each value of the τ parameter used in CR-EOMCCSD(T) calculations, a significant number of virtual orbitals were excluded from the calculations. For example, for the cc-pVQZ basis the total number of molecular orbitals amounts to 590. Of those, for $\tau = 1.4$, 476 virtual orbitals are not correlated in the CC/EOMCC calculations. The values of the highest orbital energy ϵ_{U} for cc-pVDZ, aug-cc-pVDZ, cc-pVTZ, and cc-pVQZ basis sets are equal to 3.84632, 4.10223, 15.90914, and 43.06471 hartree, respectively. Let us start our analysis from Table 1 with the results for the five-point extrapolation. For two basis sets, cc-pVDZ and aug-cc-pVDZ, we were able to perform full CR-EOMCCSD(T) calculations for the $\pi\pi^*$ state of cytosine. For the aug-cc-pVDZ basis set, extrapolation results, except for the f_3 model, nicely compare with full CR-EOMCCSD(T) excitation energies ($\omega^{\mathrm{CR(T)}}$), with errors within 0.05 eV. This picture is also valid for the smaller cc-pVDZ basis set, where, except for the f_3 scheme, the errors with respect to the exact values of excitation energies are less than 0.1 eV. In both cases the extrapolation based on the simplest f_1 function seems to provide the most reliable results. To better characterize the discrepancies between results obtained with f_1, f_2, and f_3 trial functions, we define the ρ parameter, which is defined as the largest difference between any two of these extrapolation approaches, i.e., $\rho = \max_{ij}|f_i - f_j|$. Notably, the values of ρ are monotonically decreasing with the basis set size. Indeed, the value of ρ varies from 0.13 eV for cc-pVDZ to the small value of 0.03 eV for the cc-pVQZ basis set. This observation means that the specific choice of the trial function is not so important. The observed large ρ discrepancies characteristic for smaller cc-pVDZ and aug-cc-pVDZ basis sets can be attributed to the relatively small size of basis set and lack of homogeneous distribution of the orbital energies. Consequently, the use of more extensive basis sets leads to a more stable behavior of our extrapolation models.

These general observations are supported by the cheaper extrapolation model based only on three calculated points (see Table 1). For the cc-pVDZ and aug-cc-pVDZ basis set we observe that f_1-based extrapolation gives reasonable results, within 0.08 eV of $\omega^{\mathrm{CR(T)}}$. Also, the discrepancies between different extrapolation schemes gradually vanish (see the ρ parameter values) and become as small as 0.03 eV for cc-pVQZ. At this point, a word about the numerical savings is in order. If n_{u} and $n_{\mathrm{u}}(\tau)$ designate the number of virtual orbitals in full calculations and in truncated calculations defined by the τ coefficient, then, per a single point CR-EOMCCSD(T) due to triples correction calculation that scales such as $n_{\mathrm{o}}{}^3 n_{\mathrm{u}}{}^4$ (n_{o} is the number of occupied correlated orbitals), the speed-up factor η should be given by the formula

$$\eta = \left(\frac{n_u}{n_u(\tau)}\right)^4 \tag{32}$$

For example, for the cc-pVQZ basis set with $n_{\mathrm{u}} = 561$ and $n_{\mathrm{u}}(\tau = 1.4) = 126$, the speedup is as large as 393. Assuming that for the large basis sets the extrapolated energies lead to errors not exceeding 0.1 eV, the extrapolation model may be considered as a viable alternative to other low-scaling approaches.

In the second test we have calculated the dynamic average of the energy of the first excited state ($\pi\pi^*$) by resampling 20 ps classical MD trajectory at a rate of 0.5 ps. The dynamic calculations were based on the triply threaded extrapolation model that consisted of three independent CR-EOMCCSD(T) calculations defined by three different values of cutoff factor τ_1, τ_2, and τ_3 ($\tau_1 < \tau_2 < \tau_3$) for each sampling event. Subsequently, the extrapolated value of $\tilde{\omega}_K^{\mathrm{CR-EOMCCSD(T)}}$ (in all cases $x_1 + x_2/\tau^2$ was used to find the best fit) has been used in thermal averaging of the vertical excitation energy. The results of our studies performed with the cc-pVDZ and cc-pVTZ basis sets are collected in Table 2. For comparison, the gas-phase CR-EOMCCSD(T)/cc-pVDZ calculations for the isolated cytosine fragment using B3LYP optimized ground-state geometry yield vertical excitation energy for $\pi\pi^*$ equal 4.76 eV.[46] As in the snapshot calculations, we found that the $f_1(\tau,\mathbf{x})$ scheme provides us with the most reliable results. It is remarkable to notice that for the cc-pVDZ basis set this simple extrapolation scheme brings us as close as 0.02 eV to the thermally averaged results obtained with a nontruncated set of virtual orbitals. This fact clearly shows that the errors made in the asymptotic extrapolation approach are negligible compared to typical fluctuation due to environment, which for CR-EOMCCSD(T) excitation energies translates into several tenths of an electronvolt.

TABLE 2: Averaged Excitation Energies of $\pi\pi^*$ State Obtained with the Triply Threaded Scheme on the Basis of the CR-EOMCCSD(T) Approach[a]

basis	$\omega^{\mathrm{CR(T)}}_{\tau=1.0}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.2}$	$\omega^{\mathrm{CR(T)}}_{\tau=1.4}$	$f_1(\epsilon_{\mathrm{U}}, \mathbf{x})$	$\omega^{\mathrm{CR(T)}}$
cc-pVDZ	5.25	5.17	5.16	5.06	5.04
cc-pVTZ	5.16	5.12	5.10	5.03	

[a] All averaged excitation energies are reported in eV. In this simulation the classical MD trajectory was sampled at a rate of 0.5 ps.

In conclusion, we hope that our asymptotic extrapolation scheme will help to reduce effectively a very large numerical overhead associated with the use of the high-level CC/EOMCC ab initio methods that account for the effect of triply excited configurations in the context of large-scale QM/MM simulations. Our results also indicate that the approximate QM/MM calculations exploiting more extensive basis sets of cc-pVXZ or aug-cc-PVXZ (X = D, T, Q) quality are feasible. In the snapshot calculations we have also demonstrated that for the larger basis sets the asymptotic extrapolation approach reveals much more stable behavior compared to analogous calculations exploiting smaller basis sets. The development of more efficient extrapolation methods will be the focus of our future studies.

Acknowledgment. This work has been performed using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory. We also wish to acknowledge support to M.V. from the Advanced Scientific Computing Research program of the U.S. Department of Energy, Office of Science (DE-AC06-76RLO 1830). The William R. Wiley Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory is funded by the Office of Biological and Environmental Research in the U.S. Department of Energy. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by the Battelle Memorial Institute under Contract DE-AC05-76RL01830.

References and Notes

(1) (a) Coester, F. *Nucl. Phys.* **1958**, *7*, 421. (b) Coester, F.; Kümmel, H. *Nucl. Phys.* **1960**, *17*, 477.

(2) (a) Čížek, J. *J. Chem. Phys.* **1966**, *45*, 4256. (b) Čížek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.

(3) Čížek, J.; Paldus, J. *Int. J. Quantum Chem.* **1971** *5*, 359.

(4) Bartlett, R. J. In *Modern Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: Singapore, 1995; Part I, p 1047.

(5) Paldus, J.; Li, X. *Adv. Chem. Phys.* **1999**, *110*, 1.

(6) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041.

(7) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.

(8) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1995**, *233*, 81.

(9) Watts, J. D.; Bartlett, R. J. *Chem. Phys. Lett.* **1996**, *258*, 581.

(10) Stanton, J. F.; Gauss, J. *J. Chem. Phys.* **1995**, *103*, 1064.

(11) Stanton, J. F. *Chem. Phys. Lett.* **1997**, *281*, 130.

(12) Crawford, T. D.; Stanton, J. F. *Int. J. Quantum Chem.* **1998**, *70*, 601.

(13) Gwaltney, S. R.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *323*, 21.

(14) Gwaltney, S. R.; Sherrill, C. D.; Head-Gordon, M.; Krylov, A. I. *J. Chem. Phys.* **2000**, *113*, 3548.

(15) Gwaltney, S. R.; Head-Gordon, M. *J. Chem. Phys.* **2001**, *115*, 2014.

(16) Gwaltney, S. R.; Byrd, E. F. C.; Van Voorhis, T.; Head-Gordon, M. *Chem. Phys. Lett.* **2002**, *353*, 359.

(17) (a) Hirata, S.; Nooijen, M.; Grabowski, I.; Bartlett, R. J. *J. Chem. Phys.* **2001**, *114*, 3919. (b) Hirata, S.; Nooijen, M.; Grabowski, I.; Bartlett, R. J. *J. Chem. Phys.* **2001**, *115*, 3967.

(18) Hirata, S.; Fan, P.-D.; Auer, A. A.; Nooijen, M.; Piecuch, P. *J. Chem. Phys.* **2004**, *121*, 12197.

(19) (a) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2000**, *113*, 18. (b) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2000**, *113*, 5644.

(20) Piecuch, P.; Włoch, M. *J. Chem. Phys.* **2005**, *123*, 224105. Piecuch, P.; Włoch, M.; Gour, J. R.; Kinal, A. *Chem. Phys. Lett.* **2006**, *418*, 467.

(21) Piecuch, P.; Włoch, M. *J. Chem. Phys.* **2005**, *123*, 224105.

(22) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2004**, *120*, 1715.

(23) Kowalski, K. *Chem. Phys. Lett.* **2005**, *411*, 306.

(24) (a) Kowalski, K.; Piecuch, P. *J. Chem. Phys.* **2005**, *122*, 074107. (b) Kowalski, K. *J. Chem. Phys.* **2005**, *123*, 014102.

(25) Almlöf, J. *Chem. Phys. Lett.* **1991**, *181*, 319.

(26) Constans, P.; Ayala, P. Y.; Scuseria, G. E. *J. Chem. Phys.* **2000**, *113*, 10451.

(27) Koch, H.; Sánches de Merás, A. *J. Chem. Phys.* **2000**, *113*, 508.

(28) Koch, H.; Sánches de Merás, A.; Pedersen, T. B. *J. Chem. Phys.* **2003**, *118*, 9481.

(29) Adamowicz, L.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 6314.

(30) Adamowicz, L.; Bartlett, R. J.; Sadlej, A. J. *J. Chem. Phys.* **1988**, *88*, 5749.

(31) Sosa, C.; Geertsen, J.; Trucks, G. W.; Bartlett, R. J. *Chem. Phys. Lett.* **1989**, *159*, 148.

(32) Neogrady, P.; Pitonak, M.; Urban, M. *Mol. Phys.* **2005**, *103*, 2141.

(33) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370.

(34) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986.

(35) Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772.

(36) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.

(37) Crawford, T. D.; King, R. A. *Chem. Phys. Lett.* **2002**, *366*, 611.

(38) Korona, T.; Werner, H.-J. *J. Chem. Phys.* **2003**, *118*, 3006.

(39) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *121*, 10905.

(40) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *121*, 10919.

(41) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2004**, *121*, 10852.

(42) Bytautas, L.; Ruedenberg, K. *J. Chem. Phys.* **2005**, *122*, 154110.

(43) Ayala, P. Y.; Scuseria, G. E.; Savin, A. *Int. J. Quantum Chem.* **1999**, *307*, 227.

(44) Iyengar, S.; Scuseria, G. E.; Savin, A. *Int. J. Quantum Chem.* **2000**, *79*, 222.

(45) Aprà, E.; Windus, T. L.; Straatsma, T. P.; Bylaska, E. J.; de Jong, W.; et al. NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7; Pacific Northwest National Laboratory, Richland, WA 99352-0999, 2005.

(46) Valiev, M.; Kowalski, K. Accepted for publication in *J. Chem. Phys.*

(47) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(48) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *Mol. Phys.* **2002**, *100*, 1813.

(49) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Chem. Phys.* **2003**, *118*, 1620.

(50) Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2003**, *107*, 2578.

(51) Osted, A.; Kongsted, J.; Mikkelsen, K. V.; Christiansen, O. *J. Phys. Chem. A* **2004**, *108*, 8646.

(52) Osted, A.; Kongsted, J.; Mikkelsen, K. V.; Astrand, P. O.; Christiansen, O. *J. Chem. Phys.* **2006**, *124*, 124503.

(53) Kutzelnigg, W.; Morgan, J. D., III. *J. Chem. Phys.* **1992**, *96*, 4484.

(54) Halkier, A.; Helgaker, T.; Jrgensen, P.; Klopper, W.; Koch, H.; Olson, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

(55) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.

(56) Peterson, K. A.; Dunning, T. H., Jr. *J. Phys. Chem.* **1995**, *99*, 3898.

(57) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.

(58) Hammond, J. R.; Valiev, M.; de Jong, W. A.; Kowalski, K. Manuscript in preparation.

(59) Fülscher, M. P.; Roos, B.-O. *J. Am. Chem. Soc.* **1995**, *117*, 2089.

(60) Ismail, N.; Blancafort, L.; Olivucci, M.; Kohler, B.; Robb, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 6818.

(61) Shukla, M. K.; Leszczyński, J. *J. Phys. Chem. A* **2002**, *106*, 11338.

(62) Merchan, M.; Serrano-Andres, L. *J. Am. Chem. Soc.* **2003**, *125*, 8108.

(63) Sobolewski, A. L.; Domcke, W. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2763.

(64) Sobolewski, A. L.; Domcke, W.; Hättig, C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17903.

(65) Zgierski, M. Z.; Patchkovskii, S.; Lim, E. C. *J. Chem. Phys.* **2005**, *123*, 081101.

(66) Zgierski, M. Z.; Patchkovskii, S.; Fujiwara, T.; Lim, E. C. *J. Phys. Chem. A* **2005**, *109*, 9384.

(67) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(68) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(69) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6769.