

# Comparison of Basis Set Effects and the Performance of *Ab Initio* and DFT Methods for Probing Equilibrium Fluctuations

ROSS C. WALKER,<sup>1\*</sup> IAN P. MERCER,<sup>2†</sup> IAN R. GOULD,<sup>1</sup> DAVID R. KLUG<sup>2</sup>

<sup>1</sup>*Biophysics and Biological Chemistry Group, Department of Chemistry, Imperial College London, London, United Kingdom*

<sup>2</sup>*Molecular Dynamics Group, Department of Chemistry, Imperial College London, London, United Kingdom*

Received 4 October 2005; Revised 19 April 2006; Accepted 12 May 2006

DOI 10.1002/jcc.20559

Published online 21 December 2006 in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** The electronic absorption and emission spectra of large molecules reflect the extent and timescale of electron-vibration coupling and therefore the extent and timescale of relaxation/reorganization in response to a perturbation. In this paper, we present a comparison of the calculated absorption and emission spectra of NADH in liver alcohol dehydrogenase (LADH), using quantum mechanical/molecular mechanical methods, in which we vary the QM component. Specifically, we have looked at the influence of basis set (STO-3G, 3-21G\*, 6-31G\*, CC-pVDZ, and 6-311G\*\*), as well as the influence of applying the DFT TD-B3LYP and *ab initio* TD-HF and CIS methods to the calculation of absorption/emission spectra and the reorganization energy (Stokes shift). The *ab initio* TD-HF and CIS methods reproduce the experimentally determined Stokes shift and spectral profiles to a high level of agreement, while the TD-B3LYP method significantly underestimates the Stokes shift, by 45%. We comment on the origin of this problem and suggest that *ab initio* methods may be naturally more suited to predicting molecular behavior away from equilibrium geometries.

© 2006 Wiley Periodicals, Inc. J Comput Chem 28: 478–490, 2007

**Key words:** LADH; TD-HF; TD-DFT; CIS; basis set; fluctuation dissipation theory

## Introduction

The difference in free energy between the reactants and products of a chemical reaction is important in determining the extent of the reaction. The free energy of activation greatly affects the rate constant of a catalytic step. To be able to predict the extent, origin, and rate of development of this free energy difference in condensed-phase chemical reactions would thus be very useful. When subjected to a rapid change in charge distribution, a liquid or protein will respond by undergoing structural rearrangements that seek to minimize the energy. The energy associated with these rearrangements is often termed the reorganization energy and has well established roles in, for example, nonadiabatic electron-transfer theory and the response of a chemical system to optical excitation.<sup>1</sup> In the latter, the absorption of a photon leads to a rearrangement of the electron density in the molecule creating an electronically excited state. This causes a rearrangement of the surrounding medium, lowering the excited-state energy and simultaneously raising the ground-state energy. This reorganization is related to the fluorescence Stokes shift.

In one sense the creation of an excited state and the subsequent fluorescence Stokes shift can be viewed as an extremely simple type of chemical reaction. The reactant in this case is the

molecule's excited state created by the applied optical field, while the product state is the state created by the relaxation of the system. The shift in the peak of the steady-state absorption band from that of the steady-state emission band indicates the free energy gap between the reactants and products. The steady-state emission spectrum is a good representation of the product energy because most of the relaxation is complete well before the excited state emits a photon. Moreover, the spectral width of the absorption and emission spectra reflect the time scale of the fluctuations, as is apparent from optical response theory, which is discussed later.

A natural first step prior to calculating free energy gaps—free energies of activation—and relaxation rates during complex chemical processes is attempting to calculate the free energy dif-

\*Present address: Department of Molecular Biology TPC15, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA, 92037, USA

†Present address: Centre for Synthesis and Chemical Biology, School of Physics, UCD Dublin, Ireland

**Correspondence to:** R.C. Walker; e-mail: ross@rosswalker.co.uk

Contract/grant sponsor: Engineering and Physical Sciences Research Council; contract/grant number: GR/M46730

ference between initial and relaxed states in optically excited systems. Previously, we have used a quantum mechanical/molecular mechanical (QM/MM) method to calculate the relaxations in solvated chlorophylls<sup>1,2</sup> and liver alcohol dehydrogenase (LADH),<sup>3</sup> having found that it reproduces both the reorganization energies and spectral widths to within 20% of their experimental values.

In our LADH work,<sup>3</sup> we showed that calculating the fluctuation in the ground to excited state energy gap of the NADH cofactor as a function of time, using classical molecular dynamics (MD) trajectories of LADH inhibited with *n*-cyclohexyl formamide postprocessed using an *ab initio* HF-CIS<sup>4</sup> method, we could accurately evaluate the time dependence of the ground to excited state energy gap fluctuations. This allowed us to recreate the absorption and emission spectra for NADH in LADH with accuracy in the reorganization energy of 13%. Initially we concentrated on the effect played by the classical MD parameters in accurately describing the equilibrium dynamics of the NADH coenzyme and LADH protein matrix. This work showed that the results obtained were very sensitive to the parameters used in the classical MD simulations. The carefully parameterized NADH parameter set of Pavelites et al.<sup>5</sup> was found to yield results closest to experiment.

Our original choice of QM method (CIS) and basis set (3-21G\*<sup>6</sup>) was greatly influenced by the computational cost of these calculations. At the time the original work was done, the computational cost of using a basis set larger than 3-21G\* meant that such calculations were not tractable to the available computing resources. For similar reasons, we could not readily vary the QM component in order to optimize the approach. Advances in computer technology have now made the use of larger basis sets more tractable and have also shortened the computation time sufficiently to allow us to do a more detailed investigation of the effect played by the choice of method for the evaluation of the ground to excited state as a function of time.

The CIS approach, although a computationally efficient method for calculating excitation energies, is not the only method available that is computationally tractable to the QM/MM approach to relaxation dynamics employed in this work. Alternative post SCF approaches to calculating electronic transition energies are the time-dependent DFT (TD-DFT) and HF (TD-HF) methods.<sup>7</sup> In the work described in this paper, the TD-HF, also known as the random phase approximation (RPA), and TD-DFT implementations within Gaussian 98<sup>8-10</sup> have been used in an identical fashion to the CIS method discussed in our previous work.

The TD-DFT method has been reformulated within the last ten years to compute discrete transition energies and oscillator strengths and has been applied to a number of different atoms and molecules.<sup>8,11-16</sup> TD-DFT is now a standard algorithm in many QM computational packages. The method has recently been extended to include second derivatives allowing the optimization of excited state geometries.<sup>17,18</sup> There are doubts, however, about the validity of the results obtained from TD-DFT calculations, particularly with reference to excited state energy surfaces—an issue discussed by Tozer et al.,<sup>15</sup> who state that “in order for DFT to become a useful tool for studies of excited states it should be able to . . . give a correct description of excited state energy surfaces. . . necessary in order to be able to describe not only vertical excitation spectra, but also adiabatic transitions and emission

spectra.” It is this application of TD-DFT that we have investigated in this work.

Increases in available computing power since our original LADH work was published have made the choice of basis set for postprocessing MD simulations less critical. Our original choice of 3-21G\* was a factor of the need to keep the computation time acceptable, and the fact that a series of single point calculations on test structures using different basis sets showed that the 3-21G\* basis set gave accurate peak separations but with a constant offset in energy to experiment. Doubts were voiced, however, about potential problems with the accuracy of using a small basis set with our QM/MM approach, and so in this work we have addressed these doubts by repeating the CIS calculations with the STO-3G,<sup>19,20</sup> 3-21G\*,<sup>6</sup> 6-31G\*,<sup>21</sup> CC-pVDZ<sup>22</sup> and 6-311G\*\*<sup>23</sup> basis sets.

### System of Study

The system of study chosen for this work is the reduced form of LADH. The Stokes shift in dihydronicotinamide adenine dinucleotide (NADH), the chemically active cofactor in alcohol dehydrogenase and many other enzymes, is 1.04 eV (100.36 kJ/mol) in water. This is reduced to 0.87 eV when bound to the protein matrix of horse LADH with the inhibitor *n*-cyclohexyl formamide in the substrate binding site. LADH was initially chosen as a test case for our methodology because the large relaxation of almost 1 eV from an absorption in the range of 4 eV provides a demanding test of our approach, as the assumption of linear response is pushed to the limit by such a large deviation from equilibrium. Moreover any problems associated with accurately calculating variations in electronic energy as a function of nuclear position are likely to reveal themselves in cases where the electron-vibration coupling is strong, as is the case in LADH. LADH<sup>24</sup> is a very well characterized enzyme which, with the use of NAD<sup>+</sup> as a cofactor, catalyses the reversible oxidation of a wide range of alcohols to their corresponding aldehydes. It has a molecular weight of 80,000 and is composed of a dimer of two identical parts as reported in the X-ray structure.<sup>25</sup> Each subunit of the dimer binds one molecule of NADH and two Zn(II) ions. One zinc is in the active site and is associated with the aldehyde oxygen of the substrate or substrate analogue, while the other zinc performs a structural role. At pH 7, the equilibrium position of the chemistry favors NAD<sup>+</sup> and alcohol,<sup>26</sup> and so to isolate the LADH–NADH ternary complex, in which we are interested, an inhibitor must be used. We therefore use the LADH–NADH–*n*-cyclohexyl formamide ternary complex as the system of study.

### Methodology

#### Obtaining Absorption and Emission Spectra from a QM/MM Simulation

The theory used for calculating optical spectra from a QM/MM simulation has been discussed in our previous work<sup>1,3</sup> but will be covered briefly here for the benefit of the reader.

The excitation due to an incident photon is taken to occur between the ground and lowest excited singlet state of the mole-

cule. The energy of these levels fluctuates due to interactions with the surrounding bath leading to a temporal evolution of the transition frequency (Bohr frequency). The time evolution of the system is described by the Liouville–von Neumann equation,<sup>27</sup>

$$\frac{d\rho(t)}{dt} = -\frac{i}{\hbar}[(H(t), \rho(t))] \quad (1)$$

where  $\rho(t)$  is the density operator for a two-level system with levels  $g$  and  $e$  and  $H(t)$  is the Hamiltonian operator. Following an initial optical interaction at time  $t = 0$ , at a subsequent time  $t$ , eq. (1) yields for an off-diagonal element of the density matrix  $\rho_{eg}(t)$

$$\rho_{eg}(t) = \rho_{eg}(0)e^{-i\omega_{eg}t} \left\langle e^{-i\int_0^t \delta\omega_{eg}(\tau)d\tau} \right\rangle \quad (2)$$

where  $\omega_{eg}$  is the mean Bohr frequency for the electronic excitation,  $\delta\omega_{eg}(\tau)$  is the time varying deviation from the mean Bohr frequency, and the angled brackets denote averaging over the ensemble of system trajectories.

Finding the polarization of the ensemble requires taking the trace of the product of the transition dipole operator with the density matrix. Within the rotating wave approximation, this yields the linear optical response function,  $R(t)$ ,

$$R(t) = \left\langle \mu(t)\mu(0)e^{i\int_0^t \delta\omega(\tau)d\tau} \right\rangle \quad (3)$$

where for the linear process of optical absorption and emission,  $\mu(0)$  and  $\mu(t)$  are the transition dipole operators at the times of the first and second interactions with the light field, respectively. Two optical field-matter interactions are required to complete the absorption or emission of a photon, with probability of absorption being linearly proportional to light intensity or equivalently to the product of two electric field interactions.

Taking the distribution of fluctuations of the energy gap to be Gaussian and assuming that sufficient phase space has been sampled, the cumulant expansion<sup>28</sup> can be applied to eq. (3). Taking the transition dipole moment to be constant in time (Condon approximation) and setting it to unity, the response function can be expressed as

$$R(t) = e^{-g(t)} \quad (4)$$

where  $g(t)$  is the line broadening function as defined in the following section. The optical spectra are generated via

$$\begin{aligned} \sigma_{\text{abs}}(\omega) &\propto \text{Re} \left[ \int_0^t dt R(t)e^{i(\omega-\omega_{eg})t} \right] \\ \sigma_{\text{ems}}(\omega) &\propto \text{Re} \left[ \int_0^t dt R^*(t)e^{i(\omega-\omega_{eg})t} \right] \end{aligned} \quad (5)$$

#### Generation of the Line Broadening Function, $g(t)$ , and Assertion of Detailed Balance

In principle, an exact quantum calculation captures all properties. Such an approach is, however, intractable, with simulation requir-

ing matrix diagonalization, the effort for which scales exponentially with the number of degrees of freedom. The computational effort required for classical MD, however, scales as the number of force evaluations necessary to run a trajectory, which at its most expensive, remains proportional to the square of the number of degrees of freedom. In adopting the mixed quantum-classical approach of this work, we are in particular not directly sensitive to information pertaining to detailed balance; the relative probability of phonon absorption to emission at a given phonon frequency. Subsequent inclusion of detailed balance is required to account for reorganization in system energy after optical perturbation, generating an associated optical Stokes shift. The relationship between Stokes shift and detailed balance is an example of linear response and conforms to the fluctuation dissipation theorem.<sup>29</sup>

The time autocorrelation function of the fluctuating transition frequency,  $M(t)$ , forms the basis for subsequent calculation as is given by

$$M(t) = \frac{\int_{-\infty}^{\infty} \delta\omega(t+\tau)\delta\omega(\tau)d\tau}{\int_{-\infty}^{\infty} \delta\omega(\tau)\delta\omega(\tau)d\tau} \quad (6)$$

The inclusion of detailed balance is performed by operating on the power spectrum,  $J(\omega)$ , given by the Fourier transform of  $M(t)$ . With  $M(t)$  being real and symmetric,  $J(\omega)$  is also real and symmetric. However, to satisfy detailed balance, it is required that positive frequencies be related to their negative counterparts by a Boltzmann coefficient<sup>30</sup> such that

$$J(-\omega) = e^{-\frac{\hbar\omega}{kT}}J(\omega) \quad (7)$$

To satisfy this relationship, a modified semiclassical form of the spectral density is generated in the following transformation<sup>31</sup>

$$J_{\text{sc}}(\omega) = \frac{2J}{\left[1 + e^{-\frac{\hbar\omega}{kT}}\right]} \equiv \left[1 + \tanh\left(\frac{-\hbar\omega}{kT}\right)\right]J(\omega) \quad (8)$$

Back Fourier transforming [eq. (8)] yields a modified semiclassical form of the time correlation function  $M_{\text{sc}}(t)$ . In general, this function is a complex Hermitian. The real component of the time correlation function remains unaltered, it being determined by the symmetric component of its Fourier complement  $J(\omega)$ , which also remains unaltered. The imaginary component being introduced to the time correlation function results from the real asymmetric component introduced in eq. (8) as  $t$ 's Fourier complement. This semiclassical time correlation function is then used to generate a complex line-broadening function with the following

$$g(t) = \Delta^2 \int_0^t d\tau_1 \int_0^{\tau_1} M_{\text{sc}}(\tau_2)d\tau_2 \quad (9)$$

Taking this line broadening function, eqs. (4) and (5) are applied to produce the optical response function and associated absorption and emission spectra.

The general approach adopted in this work asserts the condition of detailed balance without recourse to a high temperature approximation. In the high temperature limit, however, this method is readily demonstrated in keeping with that of a multi-mode Brownian oscillator (see Appendix).

### Molecular Dynamics Protocol

The parameterization and molecular dynamics (MD) protocol used in this work is identical to that used for the simulation discussed in our previous paper on LADH,<sup>3</sup> and the reader should refer to this paper for details of the parameters used in simulating LADH. We feel that it is important to state here that the force field parameters used are those developed by Pavelites et al.<sup>5</sup> These were developed independently of this work and were optimized for classical dynamics and not with regard to the QM/MM method.

The protocol used for the MD simulation is reproduced here for the benefit of the reader and was as follows. The system was first subjected to 200 steps of steepest descent followed by 300 steps of conjugate-gradient minimization on just the solvent molecules followed by 200 steps of steepest descent and 800 steps of conjugate-gradient minimization on the whole system to alleviate incorrect van der Waals contacts created by hydrogenation and solvation of the system. An atom based nonbonded cut off of 12 Å was used for this and all subsequent simulations. The system was then subjected to 20 ps of slow heating from 0 to 300 K using the method of Berendsen et al.<sup>32</sup> to control the temperature. Constant pressure periodic boundary conditions using the particle mesh Ewald method<sup>33</sup> were employed, the integration time step was set at 1 fs, and all interactions were calculated at every step. No atoms had their positions fixed or their motions damped.

After slow heating, the system was equilibrated for 100 ps at 300 K. Equilibration was deemed to have been successfully obtained when the root mean square deviation (RMSD) of the protein  $\alpha$  carbons was reasonably small ( $<0.9$  Å) and both the RMS and classical energies of the system fluctuated by less than 10% over a time scale of approximately 20 ps. An energy gap production run was then performed for the subsequent 10 ps with the complete system coordinates being recorded every 2 fs, resulting in a trajectory of 5000 coordinate sets.

### Quantum Mechanical Calculation of Energy Gaps

Our QM/MM method for calculating optical spectra requires calculation of an energy gap correlation function with an energy gap calculated typically every 2 fs for 5 ps.

The trajectory obtained from the 10-ps production run of the MD simulation was used to create time-ordered input structures for the QM calculations. The QM methods employed in this study are implemented within the software package Gaussian 98.<sup>34</sup> The method consists of incorporating the point charges of the water and protein surrounding the selected NADH residue into the one-electron Hamiltonian. This has the effect of polarizing the wave function with respect to the protein structure and solvent distribution.

Taking structures from the classical MD trajectory, single point calculations were run using the CIS, TD-HF, and TD-DFT methods in order to calculate the singlet energy gap between the ground and first excited state for the system of interest at 2-fs intervals. Trial CIS calculations showed that the orbitals involved in the excitation of interest were centered purely on the nicotinamide moiety of the NADH, and thus it was decided to treat the chosen NADH residue quantum mechanically, while treating the

rest of the protein and surrounding waters classically, giving a system consisting of 71 atoms corresponding to 470 basis functions (BF) (3-21G\*). In this way, the need for link atoms was negated, because the NADH residue is not formally bound to the protein matrix. The point charges of the classical system were included in the one-electron Hamiltonian of the quantum element of the calculations.

A total of 2500 points representing the first 5 ps of the 10-ps production run were postprocessed using the three different QM methods described earlier, allowing the effect of QM method on the predicted spectra to be evaluated. To allow direct comparison, all 2500 points were evaluated using each QM method, but in the interests of lowering computational expense, only the energy gaps for NADH residue 753 were evaluated. The reason for choosing to evaluate the energy gaps for one NADH residue over 5 ps rather than both residues for 2.5 ps each was a function of the way in which the Stokes shift prediction converges with time. Our experience has shown that 5 ps is sufficient, with LADH, to obtain a converged correlation function while 2.5 ps is not. The structures and environment about both active sites are broadly equivalent and thus a 5-ps well-converged run is more attractive for the same computational cost, than two runs for which the convergence is unclear.

To test the effect of basis set on the predicted spectra and Stokes shift, we reran the single point CIS calculations with STO-3G (255 BF), 3-21G\* (470 BF), 6-31G\* (722 BF), CC-pVDZ (759 BF) and 6-311G\*\* (970 BF) basis sets. From the results, discussed later, we found that the 3-21G\* basis set we used in our earlier work gives acceptable results within our QM/MM framework. For this reason, we used the 3-21G\* basis set for the TD-HF and TD-DFT calculations. For the TD-DFT calculations, the hybrid B3LYP<sup>35,36</sup> functional was used. Although other functionals, including pure exchange functionals, could have been tried, B3LYP was chosen because it is a commonly used functional and has been reported to give the best results when used with the time dependent approach.<sup>8</sup>

For the TD-HF case, the calculations were restricted to the first four excited states, while for the TD-DFT case it was found to be necessary to solve for the first eight states in order to ensure that the results included a state with nonzero oscillator strength. This significantly increased the time required for the TD-DFT calculations, making the method less efficient than the CIS method.

The calculated absorption and emission spectra shown are shifted in energy by the same amount, such that the peaks of the calculated absorption spectrum coincide with those of the experimental spectrum. This shift is applied after the spectra are calculated by finding the difference in energy between the maximum of the experimental absorption spectrum and the maximum of the calculated absorption spectrum. All data points making up the calculated absorption and emission spectrum were then shifted in energy by this difference. The spectral profile and Stokes shift remain unaltered. By these means, each method's ability to predict the extent of relaxation becomes apparent by comparing the experimental and calculated Stokes shifts. The CIS method together with the 3-21G\* basis set gives an average transition energy of approximately 5.12 eV, while the TD-HF method gives an average of approximately 4.92 eV, and the TD-DFT method

**Table 1.** Summary of the Effect of Basis Set on the Results Obtained from CIS Evaluation of the Energy gaps.

Basis set	Average time per point (1 × P4 3.4 GHz)	Mean energy gap (eV) (Exp = 3.64)	Calculated peak shift (eV) (Exp = 0.872)	Emission width [FWHM] (Exp = 0.595)
STO-3G	1227s (0.34 h)	6.27	0.995	0.526
3-21G*	5043s (1.40 h)	5.12	0.850	0.491
6-31G*	27874s (7.74 h)	5.00	0.869	0.496
CC-pVDZ	60482s (16.8 h)	4.92	0.865	0.493
6-311G**	72400s (20.1 h)	4.90	0.864	0.492

an average of approximately 3.85 eV. This compares to an experimental absorption maximum of 3.64 eV.

The absorption and emission spectra were generated from the time-correlated energy gap fluctuations by using the methodology of Mercer et al.<sup>1,2</sup> discussed above. Our previous calculations and arguments based on the nature of electron vibrational coupling in proteins had shown that we could expect a reasonable convergence of the Stokes shift within 5 ps. We chose our convergence criteria as previously with this methodology, namely that convergence was deemed to have occurred once the change in predicted Stokes shift was less than 10% for a doubling of the run length. The measurement of the experimental spectra is discussed in our previous work.<sup>3</sup>

## Results and Discussion

### Basis Set Effects

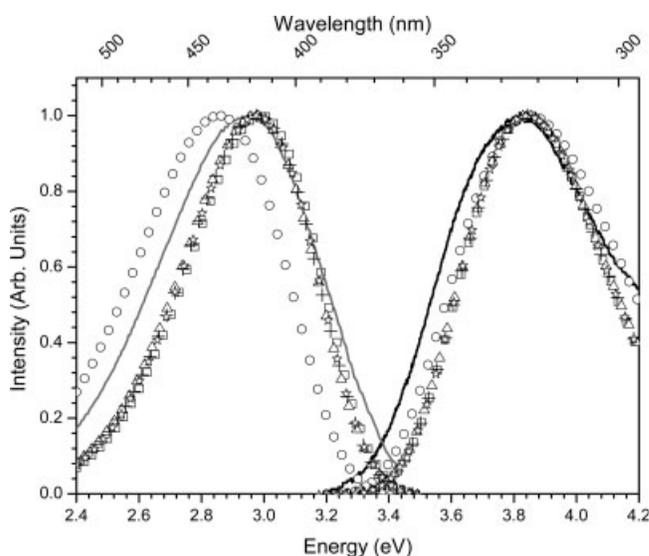
The results obtained from varying the basis set used in the CIS evaluation of the energy gaps are summarized in Table 1 and Figure 1. From the table, it can be seen that all basis sets with the exception of STO-3G get within 2.3% of the experimental peak shift. The larger 6-31G\*, CC-pVDZ, and 6-311G\*\* basis sets all get within 1% of the experimental value but with vastly increased computational expense over the 3-21G\* basis set. From the figure, it can be seen that the spectral widths for basis sets beyond STO-3G are also similar. The larger basis sets do show a trend in the mean absolute value of energy gap fluctuations towards the experimental mean, but since our interest is in accurately reproducing the relative energy gap fluctuations, the 3-21G\* basis set is sufficient. Indeed these results, coupled with our earlier work that looked at the effect played by the classical MD parameters, suggest that the structural fluctuations sampled during the MD simulation have significantly more influence on the predicted spectra than the choice of basis set.

### QM Method Effects

Table 2 summarizes the predicted energy gaps for the first structure from the MD production run. Here it can be seen that the TD-HF and CIS methods give similar excitation profiles. The TD-B3LYP method appears to perform quite well, giving a single major excitation at 3.7081 eV, which is in much closer agreement with the experimental absorption maximum of 3.64 eV than that of the TD-HF or CIS methods.

However, while the TD-B3LYP method predicts energy gaps that are closer to the experimental value, it suffers from convergence problems. Over the 2500 structures, the CIS and TD-HF methods always took between 14 and 17 SCF cycles to converge, while the TD-B3LYP method ranged from a minimum of 14 cycles to a maximum of 211 cycles, with 1 point failing to converge within 256 cycles.

The TD-B3LYP method also suffers from another problem in that a large number of the predicted states from the calculations have zero oscillator strength. In the vast majority of cases, this is not a problem as the excitation corresponding to the nicotinamide chromophore of the NADH cofactor is the first state reported. However, some calculations gave a large number of states with zero oscillator strength below the nicotinamide excitation. Increasing the number of states that were evaluated to eight ensures that in the majority of cases the nicotinamide excitation is present in the reported states. However, for seven cases out of the 2500 points all of the reported states have zero oscillator strengths. Two cases also reported the only nonzero oscillator



**Figure 1.** Calculated absorption and emission spectra for LADH (NADH + CXF) generated from energy gap fluctuations using the CIS method and the STO-3G (circles), 3-21G\* (squares), 6-31G\* (triangles), ccPVDZ (crosses) and 6-311G\*\* (stars). The experimental absorption and emission spectra are shown as solid lines for comparison.

**Table 2.** Excited State Data for QM/MM Single Point Calculations (NADH 753) on the First Structure of the MD Production Run.

Method	Excited state	Energy (eV)	$\lambda/\text{nm}$	Oscillator strength, $f$
CIS	<b>1</b>	<b>4.7848</b>	<b>259.12</b>	<b>0.1756</b>
	2	5.9388	208.77	0.0123
	<b>3</b>	<b>6.4270</b>	<b>192.91</b>	<b>0.4459</b>
	4	6.6436	186.62	0.0135
TD-HF	<b>1</b>	<b>4.5626</b>	<b>271.74</b>	<b>0.1412</b>
	2	5.7836	214.37	0.0115
	<b>3</b>	<b>6.1250</b>	<b>202.42</b>	<b>0.4060</b>
	4	6.4773	191.41	0.0098
TD-B3LYP	<b>1</b>	<b>3.7081</b>	<b>334.36</b>	<b>0.1413</b>
	8	4.4583	278.10	0.0118

All states with oscillator strengths  $> 0.009$  are shown and states with oscillator strengths  $> 0.1000$  are highlighted.

strength states to be as low as 2.65 eV. The large number of zero oscillator strength states also meant that the excitation corresponding to the adenine portion of NADH was not predicted. The origin of these zero oscillator strength states and low energy excitations is unknown but has been noted by other researchers.<sup>16</sup> Analysis of the structures that showed either poor convergence and/or a large number of zero oscillator strength states yields no obvious correlations between structural fluctuations and convergence issues. A more detailed investigation might yield more insight, although such a study would be hampered by the fact, as discussed later, that the convergence behavior of the SCF with the TD-B3LYP method is highly basis set dependent.

Unfortunately the need to evaluate a larger number of excitation states and the higher average number of SCF cycles required per step, 17.2 for TD-B3LYP compared with 15.7 for TD-HF and 16.0 for CIS, results in the TD-B3LYP method being significantly less efficient than the TD-HF and CIS methods for probing the excited state fluctuations in LADH.

One could foresee that the poor performance, both in spectral prediction and CPU time, of the TD-B3LYP method could be a function of the small size of the 3-21G\* basis set used for the comparison with CIS. To address this, we reran the TD-B3LYP calculations discussed above with the 6-31G\* basis set. The results of these calculations indicate that this is not the case. In fact the convergence issues got considerably worse in going from 3-21G\* to 6-31G\*. The mean number of SCF cycles required for convergence with 6-31G\* was 48.0 compared with 17.2 for 3-21G\*. The maximum number of cycles required was similar at 208 compared with 211, while the minimum number of cycles required was 15 as opposed to 14 with 3-21G\*. Overall the increase in the average number of SCF cycles required and the larger number of basis functions resulted in the TD-B3LYP 6-31G\* calculations requiring more than 18 times the CPU time as the TD-B3LYP 3-21G\* calculations and more than 25 times the time required for the CIS 3-21G\* calculations. The increased basis set size also did nothing to reduce the large number of zero oscillator strength states, and the predicted Stokes shift (0.459 eV) actually moved further from the experimental value. Although not thoroughly tested these results suggest that the problems

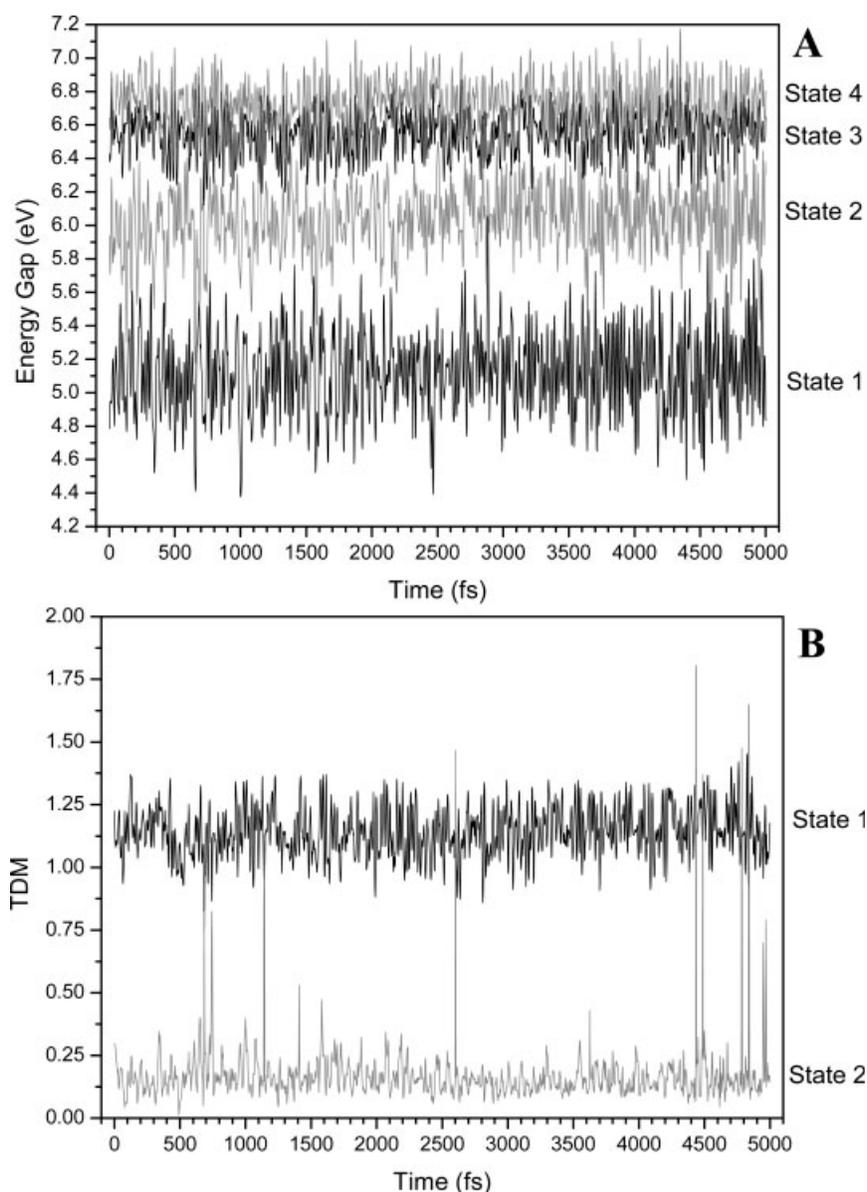
encountered in using the TD-B3LYP methodology with our QM/MM method are not related to the basis set size.

### Energy Gap Extraction

One of the biggest problems encountered when applying our QM/MM methodology is in the selection of which excitation state from each energy calculation should be used to construct the autocorrelation function. At present the implementation is such that only a single excitation can be considered for each structural snapshot. This limitation, which is currently the focus of ongoing work, limits the method to systems that do not show large degrees of mixing in the excited states. As such the consistency of state ordering is an important factor that must be addressed when calculating the autocorrelation function of the energy gap trace. Fortunately the system chosen for this study shows a high degree of stability in the state ordering as illustrated by Figure 2. Figure 2A shows the four states, ordered by energy, calculated from the CIS/3-21G\* calculations. Here it can be seen that for the lowest energy state the one corresponding to the nicotinamide absorption rarely mixes with the other states. Figure 2B shows the transition dipole moment (TDM) of the two lowest energy states. State 1 typically has the highest TDM with State 2 only occasionally showing a large TDM. For the LADH system studied in this work the energy gaps used for the spectra calculation were constructed by taking the lowest energy state, State 1, except where State 2 showed a high TDM. In these cases the orbitals involved in the two states were checked manually and the one corresponding to the nicotinamide absorption was selected. In almost all cases, where State 2 showed a high TDM, it corresponded to a mixing with State 3, which originated from the adenine absorption. As such the energy gap traces used in the calculation of the spectra were almost exclusively from the lowest energy state that showed appreciable oscillator strength (see Fig. 3).

### Calculated Trajectories, Energy Gaps, and Optical Spectra

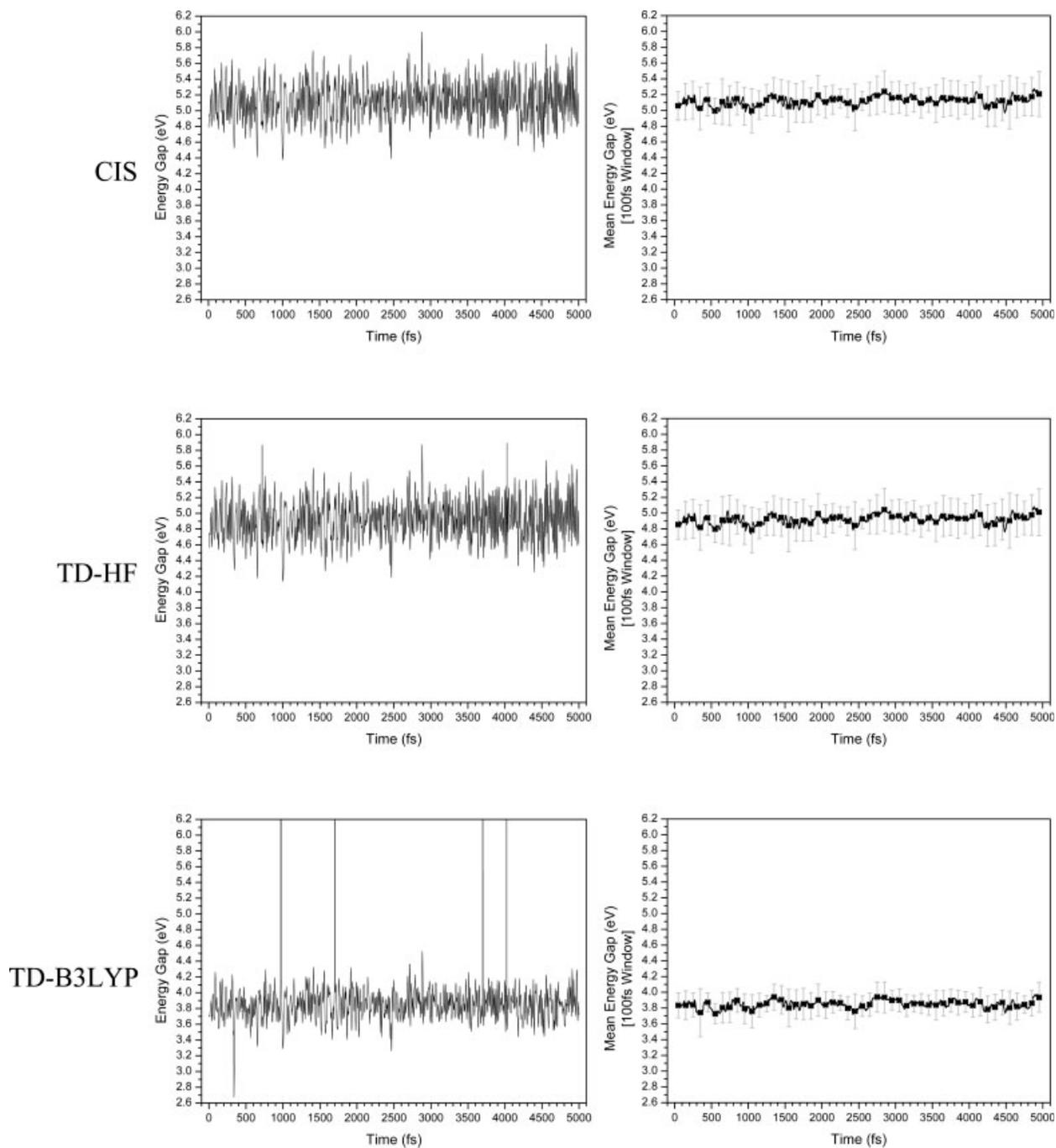
Figure 3 shows the predicted energy gaps for the CIS, TD-HF, and TD-B3LYP methods. The missing points in the TD-B3LYP data, indicated by the vertical lines in the energy gap trace, were replaced with data points found by performing spline interpolation of the surrounding points. In this way, the small number of missing points could be accounted for in a way that was not expected to have a major effect on the predicted spectra. The energy gap averages, also shown in Figure 3, and all further TD-B3LYP theoretical predictions were made using the interpolated energy gap data. It can be seen from the energy gap averages that the three methods gave stable energy gap traces with an overall mean of 5.12 eV for CIS, 4.92 eV for TD-HF, and 3.85 eV for TD-B3LYP. The TD-B3LYP method gives an energy gap mean that is much closer to the experimental value of 3.64 eV. However, although the TD-B3LYP method predicts an energy gap that is closer to experiment, it does not calculate the energy gap fluctuations adequately for agreement of the predicted spectra to experiment. It may be that such a method suited to the accurate evaluation of equilibrium geometries may however be inappropriate for studying dynamical fluctuations about equilibrium.



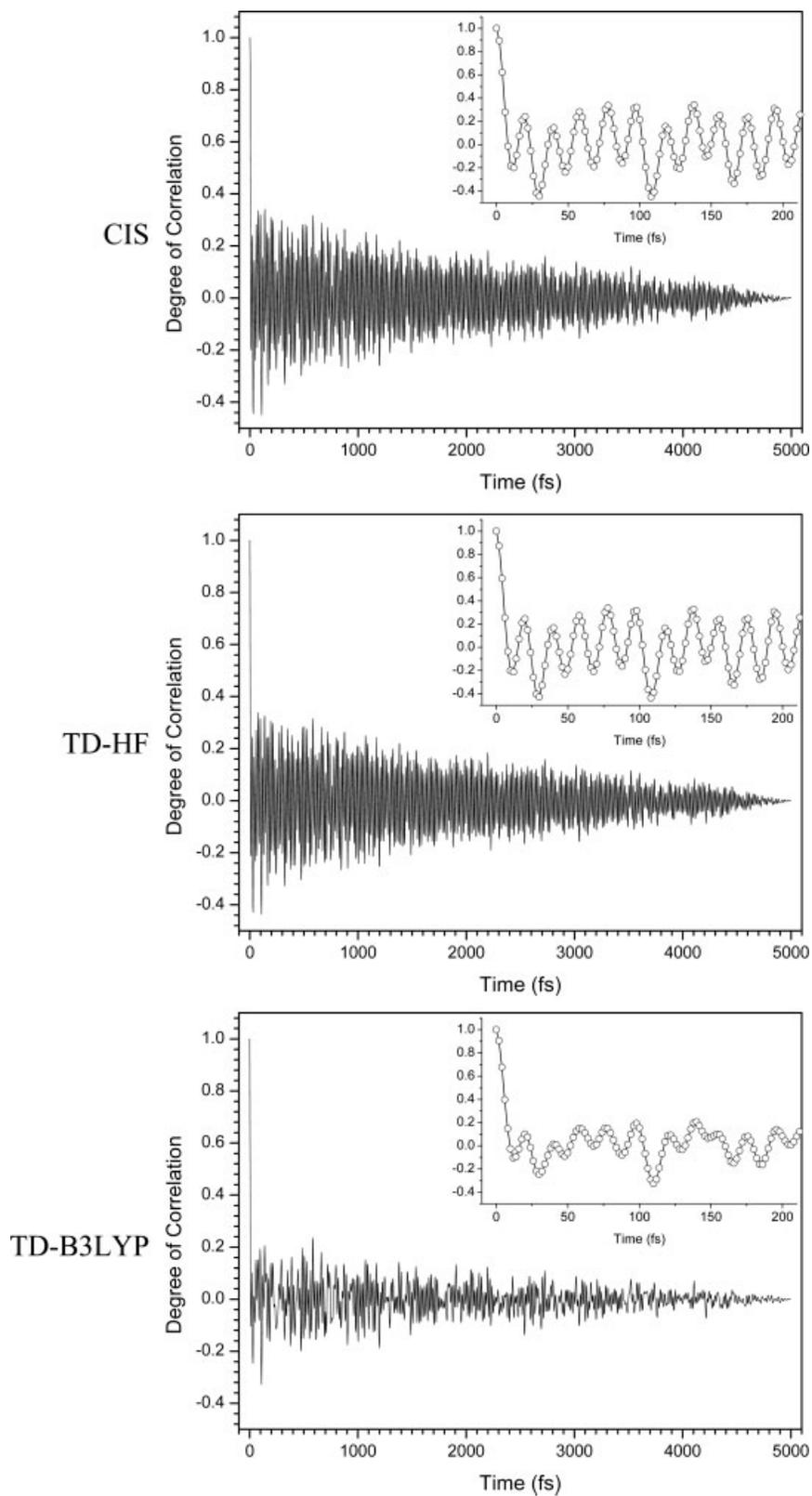
**Figure 2.** Excited state energy gaps (A) and transition dipole moments (TDM) (B) vs. time for the CIS/3-21G\* calculations. The four states calculated are shown ordered by energy in (A). For clarity the TDM data is only shown for States 1 and 2. The excited state energy profile used in the spectra calculations was extracted based on a combination of energy and TDM.

The energy gap autocorrelation functions, shown in Figure 4, for CIS and TD-HF are very similar, showing the same initial ultrafast decay of 8 fs followed by a slow component that decays gradually over the remaining 5 ps. The TD-B3LYP correlation function is different, however. The initial ultrafast decay is slightly slower, taking approximately 10 fs to reach zero, and the long term correlation is much less pronounced. Figure 5 shows the power spectrum of these autocorrelations functions. From these it can be seen that the low frequency and low intensity modes are broadly similar, although there are cases where the TD-B3LYP method gives slightly higher intensities for frequen-

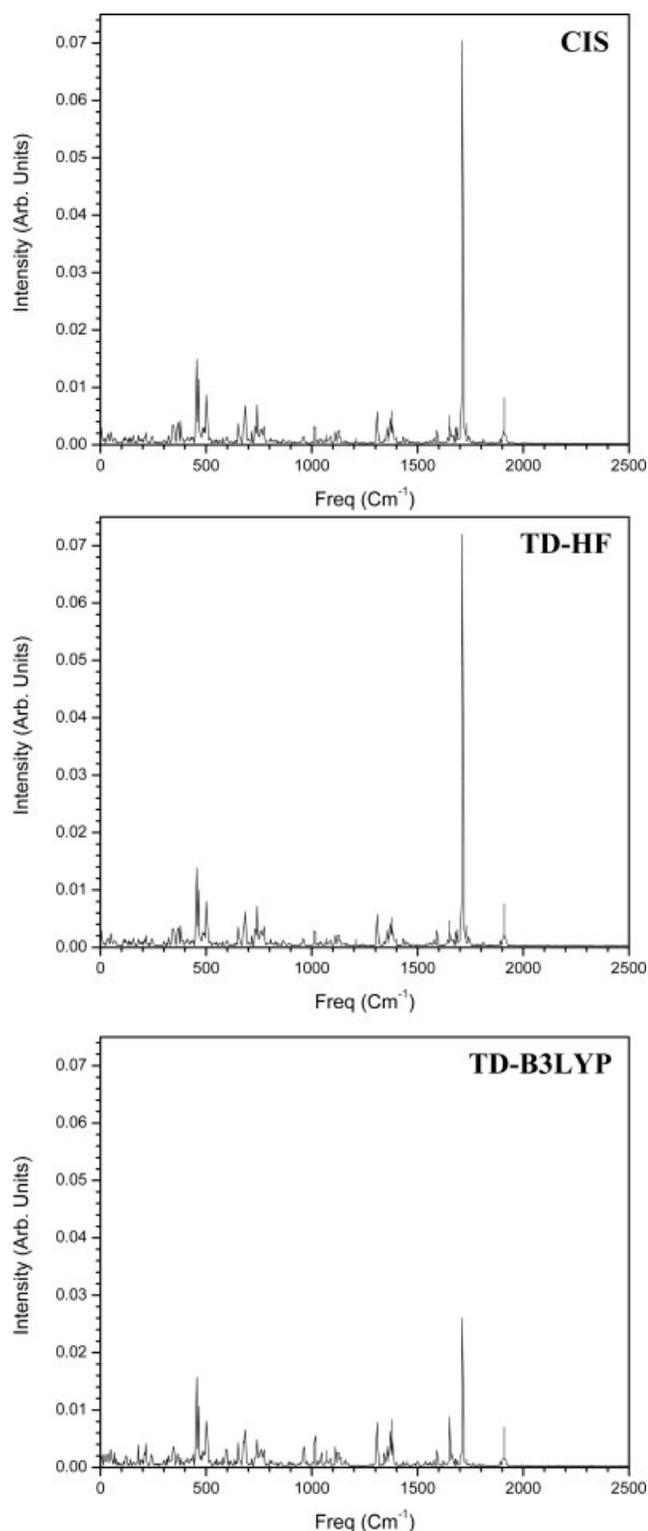
cies in the range of 1000–1500  $\text{cm}^{-1}$ . The major difference, however, between the TD-B3LYP results and the TD-HF and CIS results is in the modes around 1700  $\text{cm}^{-1}$ . This suggests that the majority of the difference in the predicted spectra potentially stems from correlated motion of the nicotinamide carbonyl group. For the high intensity mode at 1710  $\text{cm}^{-1}$ , both the TD-HF and CIS methods give intensities of around 0.07, while with TD-B3LYP this mode has an intensity of only 0.026. Conversely the mode at 1655  $\text{cm}^{-1}$  has an intensity of 0.009 with TD-B3LYP but only 0.005 with TD-HF and CIS. These small differences in the results manifest themselves as significant differences in the



**Figure 3.** Excited state energy gaps (left hand column) for NADH residue 753 in QM region for the first 5 ps of the MD production run and the mean values of the energy gaps (right hand column black line) and standard deviations (right hand column grey lines), computed using CIS, TD-HF, and TD-B3LYP. The energy gap averages and standard deviations were calculated using a  $\pm 50$ -fs moving box car average. \*The 7 TD-B3LYP points in which all states had zero oscillator strength and the 1 point that failed to converge are shown by the vertical lines in the energy gap trace. These points were interpolated prior to calculating the energy gap averages.



**Figure 4.** Excited state energy gap autocorrelation functions calculated from CIS, TD-HF, and TD-B3LYP data. Insets show magnified views. Missing energy gap points for TD-B3LYP data were interpolated prior to calculating the autocorrelation function.



**Figure 5.** Power spectrum of the excited state energy gap autocorrelation functions (see Fig. 4). The scales are directly comparable. The CIS and TD-HF methods give very similar spectra, while the TD-B3LYP method shows similar behavior for some modes; however, others are underestimated by as much as 60%.

calculated optical spectra, and in particular the calculated Stokes shift/reorganization energy.

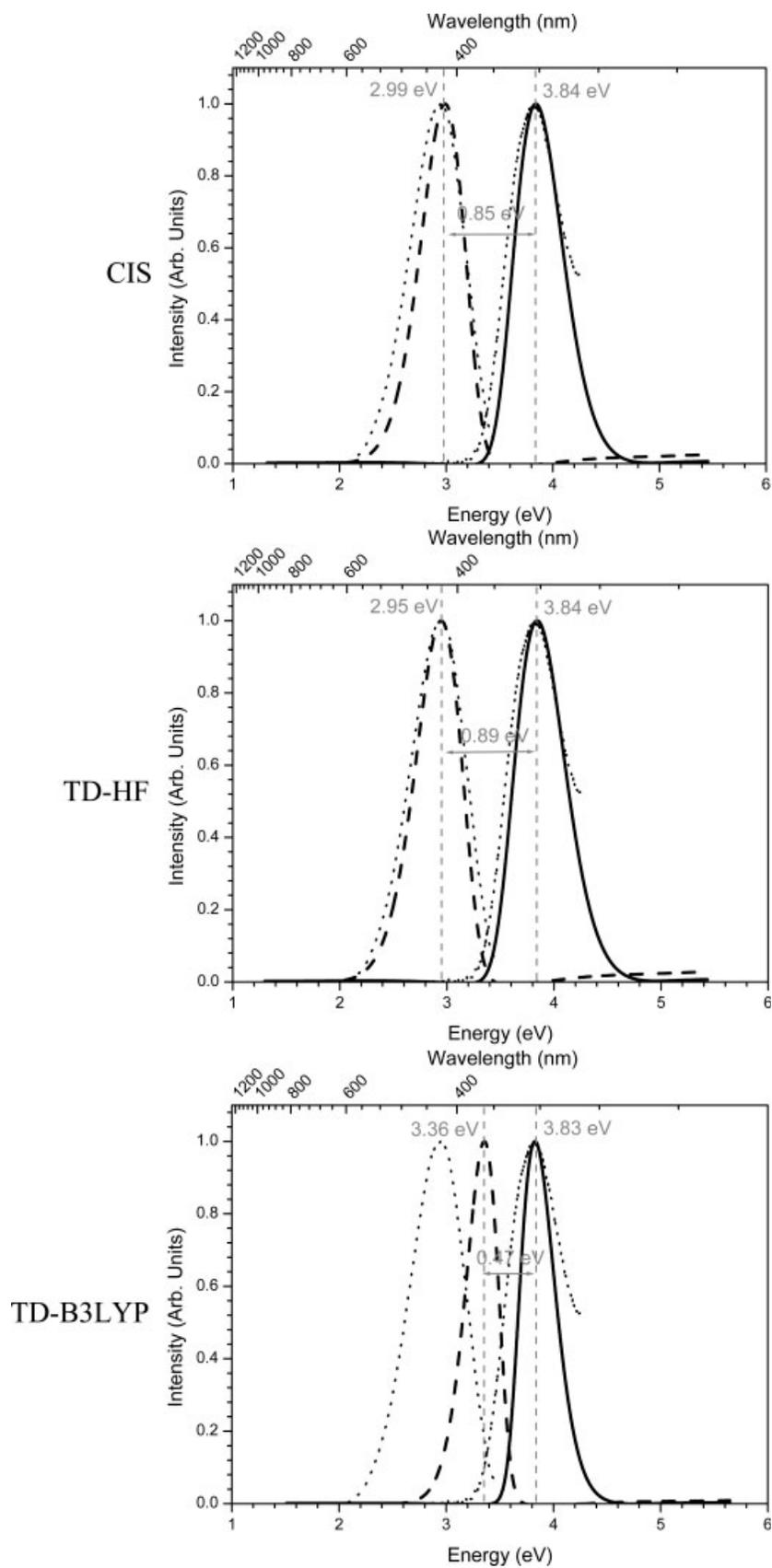
The predicted spectra for each method are shown, overlaid with experiment, in Figure 6. Overall the TD-HF method performs the best, doing slightly better than the CIS method, giving a predicted peak shift of 0.897 eV, which is within 2.9% of the experimental value of 0.872 eV. The peak shift for the TD-B3LYP data shows that it underestimates the shift by more than 45%, yielding a prediction of only 0.473 eV. The convergence of the results with run length was also found to be much poorer for the TD-B3LYP method than for the TD-HF or CIS methods.

These results imply that the TD-HF method is as good, if not better, than the CIS method for evaluating energy gap fluctuations of equilibrium structures. The similarity between the TD-HF and CIS results suggests that our QM/MM approach is not overly sensitive to the method used to evaluate the energy gap from the converged SCF. However, the technique is sensitive to the SCF method employed. The TD-B3LYP method performed poorly with convergence problems as well as poor spectral prediction. While it is possible that the TD-B3LYP method might be more sensitive to the basis set than the TD-HF or CIS methods, and may give better results with a larger basis set, such changes would require significantly more CPU time, making the TD-B3LYP method an inefficient choice.

The poor predictive performance of the TD-B3LYP method may be due to a number of reasons. Other researchers<sup>15</sup> have noted that TD-DFT approaches can show difficulty in describing charge transfer states in dipeptide models. The orbitals involved in the excitation of interest in LADH are centered on the nicotinamide region of the NADH cofactor, the amide group of which is very similar to a peptide group and may show significant charge transfer. The nuclear mode at 1710  $\text{cm}^{-1}$  shows greatly reduced coupling to the electronic transition for TD-B3LYP, which may imply that the TD-B3LYP method does not correctly describe the charge transfer that occurs as the nicotinamide carbonyl group bond length fluctuates. Since we are concerned with evaluating single point energies of dynamic structures, it is essential that any method used for calculating the ground to excited state energy gap be sufficiently flexible that it can handle structures that may be significantly displaced from the optimum geometry. Unlike the CIS and TD-HF approaches, the TD-B3LYP approach required both a basis set approximation and a functional approximation. The poor performance of the TD-B3LYP approach may be due to insufficient flexibility in the functional. Indeed the absence of a system similar to nicotinamide in the functional training set may be a contributing factor. More exhaustive investigations are needed before firm conclusions can be drawn, but from this work it appears that TD-DFT approaches may be limited in their usefulness for the studying of excited state relaxation with our QM/MM approach.

## Conclusions

As stated previously, the QM/MM approach that we have developed can be used to calculate activation energies and reorganization energies based on reproducing fluctuations about equilibrium.



**Figure 6.** Calculated absorption (solid line) and emission (dashed line) spectra for LADH (NADH + CXF) generated from energy gap fluctuations calculated using CIS, TD-HF, or TD-B3LYP. Peak locations and shifts are shown in grey. Dotted lines show experimental data.

Since accurate absolute energies of the equilibrium structures can be obtained experimentally from the absorption spectrum, it is thus more important that the method used to calculate the ground to excited state energy as a function of time be highly automatable and able to reproduce the magnitude of the fluctuations than it is for the method to give accurate absolute energies for optimized structures. Notably, computational approaches that may be suitable for accurate evaluation of equilibrium geometries may not be appropriate for studying dynamical fluctuations about equilibrium.

The results discussed above have shown that the *ab initio* CIS and TD-HF methods of evaluating the equilibrium energy gap fluctuations between ground and excited state of LADH in water yield very similar results that agree well with the experimental data. Basis set effects, for the CIS method that was tested, were found to be minimal for this system with the 3-21G\* basis set performing almost as well as the much larger 6-311G\*\* basis set but with 14 times less CPU time.

The TD-B3LYP method did not perform as well for these calculations. Several regions of phase space were required, because of poor SCF convergence, significantly more computing time was required to evaluate the energy gap with TD-B3LYP than it did for the CIS or TD-HF methods. It is possible that the structures for which poor convergence was encountered are a long way from the average ground state structure, and as such the form of the ground state electron density functional was not appropriate; however, no simple correlation between the SCF energy or structural properties and the number of SCF cycles required for convergence could be found. Overall the TD-B3LYP method is found to be less suitable than TD-HF or CIS methods for studying relaxation dynamics with the QM/MM approach developed in our earlier work.

## Acknowledgment

The authors would like to thank The Scripps Research Institute for providing the computational facilities.

## Appendix

In the high temperature limit, the line broadening function can in principle also be generated from a fluctuating transition energy via the multimode Brownian oscillator model<sup>27</sup>

$$g(t) = \Delta^2 \int_0^t d\tau_1 \int_0^{\tau_1} M(\tau_2) d\tau_2 + i\lambda \int_0^t M(\tau) d\tau \quad (A1)$$

where in the high temperature limit, coupled oscillators conform to classical behavior and

$$\lambda = \frac{\hbar\Delta^2}{2kT} \quad (A2)$$

where  $\lambda$  is the reorganization energy,  $k$  is Boltzmann's constant, and  $T$  is temperature. The complex component of the line broadening function generates a Stokes shift.

Equivalence of methods can be readily demonstrated in the high temperature limit, where the semiclassical transfor-

mation of the spectrum of oscillators equation [eq. (8)] approximates to

$$J_{sc}(\omega) = \frac{2J(\omega)}{[2 - \hbar\omega/kT]} \equiv [1 + \hbar\omega/kT]J(\omega) \quad (A3)$$

if follows

$$M_{sc}(t) = \mathfrak{S}^{-1}J(\omega) + \frac{\hbar}{2kT}\mathfrak{S}^{-1}[\omega J(\omega)]$$

$$M_{sc}(t) = M(t) + \frac{i\hbar}{2kT} \frac{dM(t)}{dt} \quad (A4)$$

Substituting the semiclassical correlation function of eq. (A4) into eq. (9) is equivalent to that generated via the high temperature multimode Brownian oscillator picture, eq. (A1).

## References

1. Mercer, I. P.; Gould, I. R.; Klug, D. R. *J Phys Chem B* 1999, 103, 7720.
2. Mercer, I. P.; Gould, I. R.; Klug, D. R. *Faraday Discuss* 1997, 108, 51.
3. Walker, R. C.; de Souza, M. M.; Mercer, I. P.; Gould, I. R.; Klug, D. R. *J Phys Chem B* 2002, 106, 11658.
4. Foresman, J. B.; Headgordon, M.; Pople, J. A.; Frisch, M. J. *J Phys Chem* 1992, 96, 135.
5. Pavelites, J. J.; Gao, J. L.; Bash, P. A.; Mackerell, A. D. *J Comput Chem* 1997, 18, 221.
6. Binkley, J. S.; Pople, J. A.; Herhe, W. J. *J Am Chem Soc* 1980, 102, 939.
7. Casida, M. E. *Recent Advances in Density Functional Methods*, Vol. 1; World Scientific: Singapore, 1995.
8. Bauernschmitt, R.; Ahlrichs, R. *Chem Phys Lett* 1996, 256, 454.
9. Casida, M. E.; Jamorski, C.; Casida, K.; Salahub, D. R. *J Chem Phys* 1998, 108, 4439.
10. Stratmann, R. E.; Scuseria, G. E. *J Chem Phys* 1998, 109, 8218.
11. Burcl, R.; Amos, R. D.; Handy, N. C. *Chem Phys Lett* 2002, 355, 8.
12. Hirata, S.; Head-Gordon, M.; Szczepanski, J.; Vala, M. *J Phys Chem A* 2003, 107, 4940.
13. Jamorski, C.; Casida, M. E.; Salahub, D. R. *J Chem Phys* 1996, 104, 5134.
14. Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. *Phys Rev Lett* 1996, 76, 1212.
15. Tozer, D. J.; Amos, R. D.; Handy, N. C.; Roos, B. O.; Serrano-Andres, L. *Mol Phys* 1999, 97, 859.
16. Tozer, D. J.; Handy, N. C. *Phys Chem Chem Phys* 2000, 2, 2117.
17. Van Caillie, C.; Amos, R. D. *Chem Phys Lett* 1999, 308, 249.
18. Van Caillie, C.; Amos, R. D. *Chem Phys Lett* 2000, 317, 159.
19. Collins, J. B.; Schleyer, P. V. R.; Binkley, J. S.; Pople, J. A. *J Chem Phys* 1976, 64, 5142.
20. Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J Chem Phys* 1969, 51, 2657.
21. Hariharan, P. C.; Pople, J. A. *Theor Chim Acta* 1973, 28, 213.
22. Dunning, T. H. *J Chem Phys* 1989, 90, 1007.
23. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J Chem Phys* 1980, 72, 650.
24. Brändén, C.-I.; Jörnvall, H.; Eklund, H.; Furugren, B. *The Enzymes*, Vol. 11; Academic Press: New York, 3rd ed., 1975.
25. Ramaswamy, S.; Scholze, M.; Plapp, B. V. *Biochemistry* 1997, 36, 3522.
26. Shearer, G.; Kim, K.; Lee, K.; Wang, C.; Plapp, B. *Biochemistry* 1993, 32, 11186.

27. Mukamel, S. Principles of Nonlinear Optical Spectroscopy; Oxford University Press: New York, 1995.
28. Kubo, R. J Phys Soc Jpn 1962, 17, 1100.
29. Chandler, D. Introduction to Modern Statistical Mechanics; Oxford University Press: New York, 1987.
30. Kubo, R. Rep Prog Phys 1966, 29, 255.
31. Mukamel, S. J Phys Chem 1985, 89, 1077.
32. Berendsen, H. J.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
33. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. J Chem Phys 1995, 103, 8577.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98 (Revision A.11); Gaussian Inc.: Pittsburgh, PA, 2001.
35. Lee, C. T.; Yang, W. T.; Parr, R. G. Phys Rev B 1988, 37, 785.
36. Becke, A. D. J Chem Phys 1993, 98, 1372.