JCTC Journal of Chemical Theory and Computation

Protein Influence on Electronic Spectra Modeled by Multipoles and Polarizabilities

Pär Söderhjelm,[†] Charlotte Husberg,[†] Angela Strambi,[‡] Massimo Olivucci,[‡] and Ulf Ryde*,[†]

Department of Theoretical Chemistry, Lund University, Chemical Centre, P.O. Box 124, SE-221 00 Lund, Sweden, and Dipartimento di Chimica, Università di Siena, via Aldo Moro 2, I-53100 Siena, Italy

Received October 30, 2008

Abstract: We have developed automatic methods to calculate multipoles and anisotropic polarizabilities for all atoms and bond centers in a protein and to include such a model in the calculation of electronic properties at any level of quantum mechanical theory. This approach is applied for the calculation of the electronic spectra of retinal in rhodopsin at the CASPT2//CASSCF level (second-order multiconfigurational perturbation theory) for the wild-type protein, as well as two mutants and isorhodopsin in QM/MM structures based on two crystal structures. We also perform a detailed investigation of the importance and distance dependence of the multipoles and the polarizabilities for both the absolute and the relative absorption energies. It is shown that the model of the surrounding protein strongly influences the spectrum and that different models give widely different results. For example, the Amber 1994 and 2003 force fields give excitation energies that differ by up to 16 kJ/mol. For accurate excitation energies, multipoles up to quadrupoles and anisotropic polarizabilities are needed. However, interactions with residues more than 10 Å from the chromophore can be treated with a standard polarizable force field without any dipoles or quadrupoles.

Introduction

Electronic spectroscopy is one of the most important sources of information about biological systems. Therefore, one of the major goals in theoretical chemistry has been to obtain methods that accurately predict electronic spectra of chromophores in proteins. This is a formidable task, because it is well-known that the surrounding medium strongly affects the spectrum of a chromophore. Therefore, many attempts have been made to model such effects also in theoretical calculations. For pure solvents, detailed methods exist, which explicitly include the closest solvent molecules in the calculations using a combination of accurate quantum mechanical (QM) methods, detailed electrostatic models, and statistical simulation methods.[1] Thereby, the most important effects are included: electrostatics, polarization, dispersion, exchange repulsion, and dynamic effects.

For proteins, calculations have not reached this level yet, but many attempts have been made to include at least some of these effects, for example, by point-charge models[2−11] and multipole models[12] for the electrostatics, atomic isotropic polarizabilities for polarization,[13−17] extended QM systems (typically at the semiempirical level) to include some dispersion and exchange repulsion,[18] self-consistent reaction-field methods for solvation effects,[19] and sampling of many conformations.[3,14,20] However, few investigations have included several of those effects at the same time.[21] The reason for this is mainly the heterogeneous nature of a protein, which requires a very large number of parameters that vary with geometry. In this Article, we take a first step in this direction, by implementing QM methods to calculate a multicenter-multipole + anisotropic polarizabilities model for a full protein, as well as methods to include such a detailed model of the electrostatics and polarization into accurate ab initio

* Corresponding author phone: +46-46 2224502; fax: +46-46 2228648; e-mail: ulf.ryde@teokem.lu.se.

[†] Lund University.

[‡] Università di Siena.

10.1021/ct800459t CCC: $40.75
Published on Web 01/28/2009

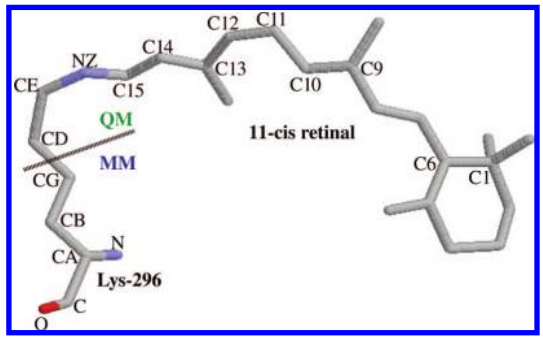

Figure 1. Structure of the protonated Schiff-base 11-cis retinal chromophore in rhodopsin with the quantum (QM) and molecular mechanics (MM) systems, as well as atom names marked out.

QM calculations. Moreover, we study the influence and distance dependence of the various terms in this model on both absolute and relative absorption energies. Such information can point out how much of the model needs to be recalculated if dynamics effects are included by repeating the calculations on several conformations.

As a test case, we use the retinal chromophore in the protein rhodopsin, which already has been the subject for many theoretical studies of the protein influence on the spectrum.[22] This G protein-coupled receptor is responsible for the vision in mammals, and the protein is responsible for the tuning of the same chromophore in the various color pigments in the cones of our eyes.[23] The 11-cis retinal chromophore is covalently bound to the Lys-296 residue (numbering according to the bovine enzyme) via a protonated Schiff-base linkage (Figure 1). After the absorption of a photon, the chromophore is converted to all-trans retinal, which triggers the neuronal signal.

In this Article, we study the wild-type protein, as well as two mutants: In the first, the counterion to the protonated Schiff-base, Glu-113 is converted to Asp, with one CH_2 group less, leading to a movement of a charged group close to the chromophore.[24] In the second, the small Gly-121 group, which forms van der Waals contact with the β-ionone ring of the chromophore, is converted to a much more bulky Leu residue.[25] Finally, we also study isorhodopsin, in which the 11-cis retinal chromophore is replaced with 9-cis retinal.[26] All calculations are based on previous QM/MM structures of the chromophore[10,11] in two different crystal structures[27,28] and a well-tested CASPT2 protocol for the spectra calculations.[29]

Methods

Structures. All calculations are based on previously determined QM/MM structures,[10,11] in which the chromophore, Lys-296, and two nearby water molecules were optimized by the CASSCF (complete active space self-consistent field) method and 6-31G* basis set, keeping the rest of the protein fixed at the crystal structure (and treated by the Amber 1994 force field[30]). Two different crystal structures were used, one 2.8 Å-resolution structure of bovine rhodopsin (1HZX)[27] and a newer structure of the same protein at 2.2 Å resolution (1U19).[28] Following the QM/MM calculations, these residues were charged: Met-1, Ala-348 (amino and carboxy terminals), all seven Arg residues, Lys-245 and 248, Asp-190 and 300, Glu-25, 113, 134, 150, 197, 232, 247, and 249, as well as His-195. All of these residues, except the counterion Glu-113, are solvent-exposed and ensure that the full protein is neutral. All of the other amino acids (including three Asp, nine Glu, and nine Lys residues) were kept neutral, in agreement with experiments for this membrane-bound protein.[31] For isorhodopsin, the chromophore, Lys-296, and the two water molecules were optimized, whereas in the Glu113Asp and Gly121Leu mutants, the mutated groups were also optimized.[10] The full retinal molecule, as well as a $CH_3NH{=}$ model of the Lys-296 protonated Schiff-linkage (from the NZ, HZ, CE, HE, and CD atoms; cf., Figure 1), were treated at the QM level, whereas the rest of Lys-296, the two water molecules, as well as Asp-113 or Leu-121 were treated at the MM level. Calculations with the 1U19 crystal structure were restricted to the wild-type protein and isorhodopsin.

Spectra Calculations. The electronic spectra were calculated at the CASPT2//CASSCF level (complete active space second-order perturbation theory), as has been described before.[10,29] The same QM system as in the QM/MM optimizations was used. The wave function was optimized at the CASSCF level (complete active space self-consistent field)[32] with the 6-31G* basis set and an active space consisting of 12 electrons in 12 active orbitals. An average over the three lowest states was optimized, but the polarization field for the lowest state was used. Next, the energy of the ground state was calculated by the CASPT2 method,[33,34] with the 22 core orbitals frozen and with a level shift of 0.2 to remove intruder states. To be consistent with the old calculations, the new zeroth order Hamiltonian of CASPT2[35] was not employed (i.e., the keyword IPEA was set to 0). Finally, the CASSCF energy was optimized again for the three-level average, but using the polarization field for the second state, and the energy for the second state was optimized with another CASPT2 calculation. All calculations were run with a developmental version of the Molcas 7.1 software.[36]

Multipoles and Polarizabilities. The influence of the surrounding protein was simulated by a detailed multicenter-multipole + polarizabilities model. This classical model consisted typically of charges, dipoles, and quadrupoles, as well as anisotropic polarizabilities for all atoms and bond centers in the protein (in total 10 824 and 11 012 centers for the HZX and U19 structures, respectively).

The quantum-mechanical and classical subproblems were solved in a self-consistent manner at the CASSCF level. In each CASSCF iteration, the electric field from the charge density of either the ground state or an excited state was computed at each center and added to the field from multipoles and induced dipoles. The induced dipoles were updated to this field until self-consistency within the classical system was obtained. The resulting electric field from the classical system (polarization field) was used to perturb the Hamiltonian operator, thus contributing to the change in charge density in the next CASSCF iteration. Similar approaches have been used before.[15,37−39] The final pertur-

bation obtained by this procedure was applied in the CASPT2 calculation. In the computed energy, both the internal polarization energy of the classical system and the coupling between the quantum-mechanical and classical systems were included.

The multipoles and polarizabilities were obtained with the LoProp approach,[40] using the Molcas software. The calculations were performed either at the Hartree−Fock level with the 6-31G* basis set[41] or at the density functional B3LYP[42] level with the aug-cc-pVDZ basis set.[43] Each basis set was turned into the atomic natural orbital form (as required by the LoProp procedure) by a linear transformation that does not affect the orbital optimization. The properties were calculated for the whole protein by dividing it into the individual amino-acid residues, which all were capped with CH_3CO- and $-NHCH_3$ groups (dipeptides). The effect of the capping groups was removed by calculating the properties also of the overlapping $CH_3CONHCH_3$ fragments and subtracting them from the properties of the corresponding dipeptides, the molecular fractionation with conjugate caps approach, which has been shown to give excellent results.[44,45] A separate calculation was performed on every residue in the structure, with the correct geometry from the QM/MM structure. Lys-296 was truncated at the CG atom, which was converted to a hydrogen atom. Test calculations were also performed, in which Lys-296 was instead truncated at the CD atom (cf., Figure 1). However, this led to very unstable results, because the hydrogen atoms from the truncated CD atom on the QM and MM sides are only ∼0.5 Å apart.

To accomplish a well-defined classical treatment of intramolecular polarization, the centers were divided into groups. In the calculation of the electric field at a certain center, multipoles and induced dipoles from a list of groups were excluded, as described in the Molcas manual.[32] For the LoProp model, these exclusion lists were constructed so that polarization between any two centers that belong to the same fragment (dipeptide) in the LoProp calculations is omitted, because this polarization has already been treated quantum-mechanically (a rigorous motivation can be found in ref 45). For the polarizable Amber force field (ff02), we followed the Amber practice to ignore polarization between atoms separated by one, two, or three bonds. When testing the distance-dependent transition between these two models, a "generous" exclusion protocol was used to ensure that no unphysical polarization occurs. Thus, polarization between two centers treated by different models was omitted if it would have been omitted in either of the two models. A simpler method, using the LoProp exclusion rule for the Amber part as well, was also tested and gave similar results.

The splitting and splicing of the protein were performed with local software, as described in the Supporting Information. The calculations for a full protein took ∼400 CPU hours at the HF/6-31G* level and ∼6000 CPU hours at the B3LYP/aug-cc-pVDZ level. For isorhodopsin and the two mutants, only groups with new geometries (Lys-296, the two water molecules, and the mutated residue, if any) were recalculated (in fact, in the original calculations,[10,11] hydrogen atoms were reoptimized in U19-isorhodopsin and the HZX-G121L mutant, but these changes were not considered in this investigation, because the other structures were not changed).

In the 1HZX structure, residues 236−240 and 331−333 are missing. In the LoProp model, residues 235 and 330 were truncated by $-NHCH_3$ groups, and residues 241 and 334 were truncated by CH_3CO- groups (this is necessary to make the QM calculations possible). However, with the Amber force fields, no such truncating groups were used, following the original calculations[10,11] and because these capping groups are not available with all force fields.

Following the previous calculations, no effects of the surrounding solvent were included. Test calculations with a spherical cavity surrounding the protein (i.e., with a radius of 44−50 Å) indicated that the solvent effect is only ∼2 kJ/mol. However, its inclusion is complicated by the fact that the solvent reaction field should not (in variance to the polarization field) relax for the excited state.

Result and Discussion

We have calculated the electronic spectra of rhodopsin, isorhodopsin, and two mutants with the CASPT2//CASSCF approach, performing a detailed study of the influence of the surrounding protein. The use of two different crystal structures gives us a first indication of the reproducibility of the results and of their conformational dependence. Although the two lowest excitation energies are included in the CASSCF state average, we concentrate the discussion on the lowest (and main) excitation, which experimentally is found around 498 nm (240 kJ/mol).[46] The goal of this Article is not to reproduce this value (which would require a better basis set in the CASPT2//CASSCF step and an extensive sampling of conformations of the protein and the chromophore), but to take a first step in this direction by investigating what is needed for an accurate modeling of the surrounding protein for a single structure. This is done by using a very detailed model of the electrostatic and polarizing effect of the surroundings, consisting of a multicenter-multipole model up to quadrupoles and anisotropic polarizabilities for all atoms and bond centers of the protein. We then examine what parts of this model are actually needed, both for absolute and for relative excitation energies, by removing the various terms stepwise. We also study the distance dependence of the various terms, that is, how far from the chromophore they have a significant effect.

Spectra. Calculated excitation energies for the various structures are collected in Table 1. The treatment of the surroundings is described by a vector of four entries in square brackets. The four entries are the charges, which may be LoProp (+), various Amber force fields (ff94, ff03, or ff02[30,47,49]), or integer charges only for the charged residues, Asp, Glu, Lys, and Arg (i); the dipoles, which may be LoProp (+) or none (−); the quadrupoles, which may be LoProp (+) or none (−); and the polarizabilities, which may be LoProp anisotropic polarizabilities (a), scalar isotropic polarizabilities, taken as the average of the diagonal elements of anisotropic polarizabilities (s), polarizabilities from Amber (ff02[49]), or none (−). Thus, our standard and most accurate model is [+,+,+,a], whereas the original investigations[10,11] used a [ff94,−,−,−] model.

Table 1. Calculated Excitation Energies for Rhodopsin, Isorhodopsin, and the Glu113Asp (E113D) and Gly121Leu (G121L) Rhodopsin Mutants, Using Two Different Crystal Structures, U19[28] and HZX[27a]

model				U19				HZX			
				Rh		Iso-Rh		Rh	Iso-Rh	E113D	G121L
q	dipole	q-pole	pol	6-31G*	VDZ	6-31G*	VDZ	VDZ	VDZ	VDZ	VDZ
+	+	+	a	257.5	260.7	269.0	277.0	268.4	283.3	265.5	285.3
+	+	+	s	257.6	259.3	269.0	274.2	268.0	281.6	265.7	282.2
+	+	−	a	253.9	257.8	262.1	266.6	267.8	278.5	266.0	272.4
+	−	−	a	263.2	259.4	278.6	276.0	271.7	291.5	268.6	294.8
i	−	−	a		326.6		332.0	306.0	313.1	291.4	298.2
−	−	−	a	230.6	237.4	234.9	241.4	245.3	253.9	246.0	247.9
+	+	+	−	259.9	263.4	271.0	274.3	270.1	283.2	256.5	279.5
+	+	−	−	248.7	251.9	255.8	259.1	263.5	270.8	252.7	264.1
+	−	−	−	267.4	262.7	279.6	274.2	274.0	289.0	260.3	284.3
i	−	−	−		322.1		326.4	311.3	315.1	294.6	305.3
−	−	−	−		214.9		217.9	219.6	228.7	217.9	224.2
ff94	−	−	−		258.1		269.8	275.9	292.4	264.1	289.7
ff03	−	−	−		272.0		285.4	287.0	303.9	267.0	300.7
ff02	−	−	−		264.1		279.5	277.1	293.1	264.6	290.5
ff02	−	−	ff02		254.4		264.7	269.7	285.1	265.6	284.5
ff94CD[b]	−	−	−		230.2		243.7	253.6	266.8	243.0	261.1
ff03CD[b]	−	−	−		283.2		305.7	283.7	318.4	280.3	319.6
−[c]	−	−	−		215.5			219.7	227.2	218.6	223.8
ff94CD[c]	−	−	−		233.2			250.3	262.4	240.2	269.7
exp[d]					240.2		247.7	240.2	247.7	234.7	251.9

[a] Two different methods were used to calculate the multipoles and polarizabilities: HF/6-31G* and B3LYP/aug-cc-pVDZ (VDZ). The treatment of the surroundings is described by the four entries in model: the charges (q), which may be calculated by the LoProp procedure[40] (+), taken from the Amber force fields (ff94,[30] ff03,[47] or ff02[49]), taken as simple integer charges for the charged residues, Asp, Glu, Lys, and Arg, (i), or be ignored (−); the dipoles and quadrupoles (q-pole), which both may be calculated by LoProp (+) or be ignored (−); and the polarizabilities (pol), which may be fully anisotropic and calculated by LoProp (a), be scalar and isotropic, taken as the trace of the anisotropic ones (s), be taken from the Amber 2002 force field[49] (ff02), or be ignored (−). [b] In these calculations, the CD atom was converted to a hydrogen junction atom and charges were present on this atom, as well as on the CG and HG atoms (in all of the other calculations, instead the CG atom was converted to a hydrogen junction atom, and no charge was present on the CD, HD, and HG atoms; cf., Figure 1). [c] Data from refs 10 and 11. In the ff94CD calculations, charges were present on the CG, HG, and HD atoms, but not on the CD atom. [d] Data from refs 24, 25, 26, and 46.

It can be seen that the isolated chromophore (retinal + CH_3N^+H− from the protonated Lys-296 Schiff base) in vacuum [−,−,−,−] gives an excitation energy of 215 kJ/mol in U19 and 220 kJ/mol in the HZX crystal. This is within 1 kJ/mol of previously published values,[10,11] and ∼25 kJ/mol lower than the experimentally observed excitation energy in rhodopsin. The other structures gave slightly different excitation energies: 3 (U19) or 9 (HZX) kJ/mol higher excitation energy for Iso-Rh, 2 kJ/mol lower energy for the E113D mutant, and 5 kJ/mol higher energy for the G121L mutant. These values reflect the influence of the geometry on the excitation energies.

If the chromophores are inserted into the rhodopsin protein with the full LoProp model [+,+,+,a], the spectra shift significantly: The excitation energy shifts by 46−49 kJ/mol for Rh and by 48−61 kJ/mol for Iso-Rh and the mutants. The spectral shift is in the right direction as compared to the vacuum calculation (i.e., the excitation energy increases). However, the correction is too large, giving rise to too high excitation energies, and errors compared to experiments similar to those of the vacuum calculations (18−36 kJ/mol). This is an effect of the small basis sets used in the calculations and the fact that a single minimized structure is used. However, that is no problem in this investigation, because it is a constant factor, and we will only discuss the relative effect of the various terms in the multipole + polarizability model.

Multipole Model. We will first study the effect of the multipoles on the calculated spectra. From Table 1, it can be seen that if the quadrupoles are ignored, the spectra shift by 4−15 kJ/mol without polarizabilities (i.e., the difference between [+,+,+,−] and [+,+,−,−]) and by 1−13 kJ/mol with the polarizabilities ([+,+,+,a]−[+,+,−,a]). The excitation energies are reduced (by on average 11 and 5 kJ/mol), with one exception. Relative energies (i.e., between isorhodopsin and rhodopsin or between rhodopsin and its mutants) change by up to 12 kJ/mol (Table 2). Thus, the effect of the quadrupoles is quite small, but it cannot be ignored if you aim at accurate results.

If the dipoles are also removed, the spectra shift by 8−24 kJ/mol without and by 2−22 kJ/mol with polarizabilities (relative to the calculations without quadrupoles). However, this time, the excitation energies increase (by 16 and 10 kJ/mol on average), so that the energies go back to close to the original [+,+,+,a/−] energies again (average difference 4−5 kJ/mol). Relative energies change by up to 19 kJ/mol. Thus, the dipoles have a slightly larger effect on the excitation energies than do the quadrupoles.

The charges have an even larger effect on the excitation energies: The spectra shift by 42−62 kJ/mol without and by 14−36 kJ/mol with polarizabilities if the charges are removed (as compared to the [+,−,−,a/−] calculations). In all cases, the excitation energies are reduced (by 54 and 29 kJ/mol on average). Relative energies change by a much smaller amount, up to 14 kJ/mol. Thus, the charges have a large effect, as could be expected.

We have also examined the effect of various sets of point charges on the excitation energies. Calculations with CD

Table 2. Shifts in Excitation Energies (in kJ/mol) for Isorhodopsin Relative to Rhodopsin in the Two Crystal Structures and for the Glu113Asp and Gly121Leu Mutants Relative to the Wild-Type Protein Calculated with the Various Methods[a]

model				isorhodopsin		mutants			
q	dipole	q-pole	pol	U19	Hzx	E113D	G121L	MAD	max
+	+	+	a	16.3	14.9	−2.9	16.9	6.0	8.8
+	+	+	s	14.9	13.6	−2.3	14.2	4.8	7.4
+	+	−	a	8.8	10.7	−1.8	4.6	3.8	7.1
+	−	−	a	16.6	19.8	−3.2	23.1	8.8	12.3
i	−	−	a	5.4	7.1	−14.6	−7.8	7.8	19.5
-	−	−	a	4.0	8.6	0.7	2.6	5.0	9.1
+	+	+	−	10.9	13.1	−13.6	9.4	4.9	8.1
+	+	−	−	7.2	7.3	−10.8	0.6	4.2	11.1
+	−	−	−	11.5	15.0	−13.7	10.3	5.3	8.2
i	−	−	−	4.3	3.8	−16.7	−6.0	9.0	17.7
−	−	−	−	3.0	9.1	−1.7	4.6	4.3	7.1
ff94	−	−	−	11.7	16.5	−11.8	13.8	5.4	9.0
ff03	−	−	−	13.4	16.9	−20.0	13.7	7.9	14.5
ff02	−	−	−	15.4	16.0	−12.5	13.4	6.3	8.5
ff02	−	−	ff02	10.3	15.4	−4.1	14.8	3.8	7.9
ff94CD[b]				13.5	13.2	−10.6	7.5	5.3	6.0
ff03CD[b]				22.5	34.7	−3.4	35.9	17.0	27.2
experimental data				7.5	7.5	−5.5	11.7	0.0	0.0

[a] In addition, the mean average deviation (MAD) and the maximum deviation (max) as compared to the experimental data[24−26,46] are given. [b] In these calculations, the CD atom was converted to a hydrogen junction atom and charges were present on this atom, as well as on the CG and HG atoms (in all of the other calculations, instead the CG atom was converted to a hydrogen junction atom, and no charge was present on the CD, HD, and HG atoms; cf., Figure 1).

junctions in the MM system and employing the Amber 1994 force field[30] (ff94CD in Table 1) give results that differ from those of our full multipole model [+,+,+,−] by 15−33 kJ/mol (always more negative; the original calculations[10,11] used a similar approach, but they included charges on the HD atoms, but not on the CD junction atom; these results differ from the ff94CD results by 3−9 kJ/mol). However, this means that there are two hydrogen capping atoms, one in the QM system and one among the point charges, both representing the CD atom of Lys-296 and only 0.5 Å apart. Such a procedure makes the calculations very unstable. For example, it can be seen from Table 1 that if the charges are shifted to Amber 2003 charges[47] instead (ff03CD), the excitation energies change by 30−62 kJ/mol, showing that the results are completely unreliable. The same happens if we use a LoProp model with two capping atoms from CD; the results become unstable and counterintuitive.

Therefore, all of our calculations are performed with a hydrogen atom representing CD in the QM system and a charge representing the CG atom of Lys-296, whereas no charges are included for the CD, HD, and HG atoms. The results of such calculations with the Amber 1994 and 2003 force fields are also included in Table 1 (ff94 and ff03). It can be seen that the deletion of these three charges changes the excitation energies by 21−29 kJ/mol for ff94 and by 3−20 kJ/mol for ff03. Now, the results with the two force fields are much closer to the LoProp model, with differences of 1−9 kJ/mol (with a varying sign) for ff94 and 2−21 kJ/mol (always smaller excitation energies) for ff03. The two force fields differ by 3−16 kJ/mol, which shows that there still is some ambiguity in the use of a point-charge model of the surrounding protein. However, relative energies are more stable and change by less than 8 kJ/mol.

In both Amber force fields, the charges were obtained from QM calculations, using the RESP approach.[48] However, in the 1994 force field,[30] the electrostatic potential was taken from vacuum HF/6-31G* calculations, which are supposed to overestimate the dipole moments in a way similar to the average polarizing effect in solution. In the 2003 force field,[47] the electrostatic potential is instead calculated using the B3LYP/cc-pVTZ method (which gives almost correct dipole moments), explicitly polarized with a (protein-like) continuum solvent with a dielectric constant of 4. Thus, two different methods are used to obtain charges that are polarized in an average way similar to a protein-like surrounding. However, as we see from Table 1, the methods give quite different spectral shifts of excitation energies in a protein.

We have also used the charges of a third Amber force field, the polarizable 2002 force field (ff02).[49] These charges were also obtained with the B3LYP/cc-pVTZ method, but without the continuum solvent (i.e., in vacuum). Of course, the charges are not intended to be used without the polarizabilities, but this is similar to the comparison with the LoProp multipoles without any polarization. The results of the Amber 2002 charges (without the polarizabilities) differ by 1−10 kJ/mol from our full multipole model, that is, by an amount similar to that of the other Amber charges and are always intermediate between those of the ff94 and the ff03 charges.

An even simpler model of the surroundings is obtained with only integer charges for the residues with a net charge [i,−,−,a/−] models. From Table 1, it can be seen that such a model changes the results as compared to the [+,−,−,a/−] models by 22−67 kJ/mol (but only 3 kJ/mol in one case). However, it can also be seen that the results go in the opposite direction as compared to the calculations without any multipole model of the protein with differences of 45−109 kJ/mol and that the results are far (13−66 kJ/mol) from the full LoProp model. This shows that it is a poor model.

Polarizabilities. We next study the effect of the polarizabilities. If we replace all of the anisotropic polarizabilities (symmetric 3 × 3 tensors) with scalar isotropic polarizabilities (a single number, the average of the diagonal elements in the anisotropic tensors), there is only a minimal change in the excitation energies (less than 3 kJ/mol).

However, if the polarizabilities are completely removed, the spectra change by 0−13 kJ/mol if there is a multipole model of the protein and by 16−28 kJ/mol without any multipoles. Relative energies change by a similar amount. In general, the polarizabilities increase the excitation energies, but with only charges or with the full multipole model, the excitation energies decrease in many cases. This is a significant effect, so for accurate results, polarizabilities are important.

The polarizable Amber 2002 force field[49] performs rather well, with differences of 0−12 kJ/mol as compared to the full LoProp model. Relative energies differ by up to 6 kJ/mol. This is similar to the LoProp charges and polarizabilities [+,−,−,a], which give errors of up to 10 kJ/mol.

Distance Dependence. Next, we studied the distance dependence of the multipoles and polarizabilities; that is, we removed (zeroed) multipoles or polarizabilities for all residues outside a certain distance from the chromophore (minimum distance between any atom in the residue and the chromophore). This gives further information about the influence of the multipoles and polarizabilities on the spectra (it lets us identify accidental coincidences). Moreover, it gives information about the range of each type of interactions, which is important especially if we intend to study several conformations of the protein (it tells us how far out the multipoles or polarizabilities need to be recalculated for each new conformation). We will consider both absolute and relative excitation energies, but we concentrate on the U19 crystal and the Rh and Iso-Rh states. To ensure that the same residues are used for the two states, distances were calculated only for the Rh state and were used also for the Iso-Rh state.

The distance dependence of the conversion of anisotropic polarizabilities to isotropic polarizabilities is shown in Figure 2a. It can be seen that the total effect (i.e., the difference between the energies at the distance 35 Å = only anisotropic polarizabilities and 0 Å = only isotropic polarizabilities) is 1−3 kJ/mol, as was also reported in Table 1. However, the curves show that this small difference is coincidental: The true distance variation is actually up to 11 kJ/mol, showing that the anisotropy needs to be considered for accurate results. However, the results are accurate to within 4 kJ/mol from 6 Å. Likewise, the two curves for Rh and Iso-Rh are almost parallel, meaning that the relative excitation energy is almost constant, 12−17 kJ/mol, and insensitive to whether anisotropic or isotropic polarizabilities are used. Thus, the effect of the anisotropic polarizabilities is quite short-ranged.

Next, we considered the effect of removing the polarizabilities completely. As we saw in Table 1, the polarizabilities had a rather small effect on the spectrum, up to 9 kJ/mol, but only 3 kJ/mol for the two U19 structures. Again, the distance dependence in Figure 2b shows that this is coincidental: The total effect of the polarizabilities is up to 46 kJ/mol, and the curves do not level out (within 4 kJ/mol) until at 10 Å. The curves for Rh and Iso-Rh again run reasonably in parallel, but the variation is larger, ranging from 9 to 17 kJ/mol, but converging at 10 Å to 16 ± 1 kJ/mol. Thus, the polarizabilities are important both for the absolute and for the relative excitation energies.

Next, we looked at the effect of the quadrupoles. Figure 2c shows that the effect is up to 12 kJ/mol for rhodopsin and up to 26 kJ/mol for isorhodopsin. The curves level out around 10 Å. The difference between the two curves is 9−22 kJ/mol, but 16 ± 2 kJ/mol from 6 Å. Thus, we can conclude that the quadrupole interactions are quite short-range, as expected.

Interestingly, the effect of the dipoles is similar, but somewhat smaller: up to 10 kJ/mol for rhodopsin and 17 kJ/mol for isorhodopsin. The curves level out at 6 Å. The effect on the difference in excitation energy between Rh and Iso-Rh is even smaller, and the difference is 9 ± 1 kJ/mol, outside 2 Å.

Finally, we also tested the effect of going from a point-charge model to a model with all charges zeroed, except for residues with a net charge, for which a unit integer charge was used at the center of the charge. From Figure 2e, it can be seen that the effect is extensive, up to 97 kJ/mol, and it does not level out until 15−25 Å. For the relative excitation energy, the convergence is better, and the results are fully converged at 15 Å. However, there are variations of up to 22 kJ/mol at shorter distances.

From this, we can conclude that the polarizabilities are necessary, at least at distances up to 10 Å, although the effect of going to scalar polarizabilities is smaller. The effect of quadrupoles and dipoles is also extensive, but also converges within ∼10 Å. Therefore, a proper approach to use to save time if an extensive conformation sampling is employed (as is most like necessary to obtain accurate results) is to use a polarizable force field with scalar polarizabilities and only point charges for all residues outside a distance of 10 Å from the chromophore. There is no reason to use a nonpolarizable force field, because polarizabilities are needed to get the close interactions right. On the other hand, the use of scalar polarizabilities (and omission of dipoles and quadrupoles) makes the treatment of general parameters for the long-range interaction much easier (there is no need to rotate any vectors and matrices from a standard orientation to the actual orientation in the protein). Thus, we recommend the use of the polarizable Amber 2002 force field, or something similar, for the long-range interactions, outside ∼10 Å.

Such an approach is illustrated in Figure 2f, where we go from the full LoProp model to the Amber 2002 force field. It can be seen that the results of ff02 differ from that of the LoProp model by 6−12 kJ/mol, as was also seen in Table 1. Moreover, the difference increases to ∼20 kJ/mol around 4 Å. However, then it rapidly decreases, and at 10 Å and outward, the difference is less than 2 kJ/mol. The relative excitation energy between Iso and Rh converges even more rapidly and is 16 ± 3 kJ/mol already at 8 Å. This shows that it is an excellent approach to replace the LoProp model with the simpler ff02 outside a radius of 10 Å from the chromophore.

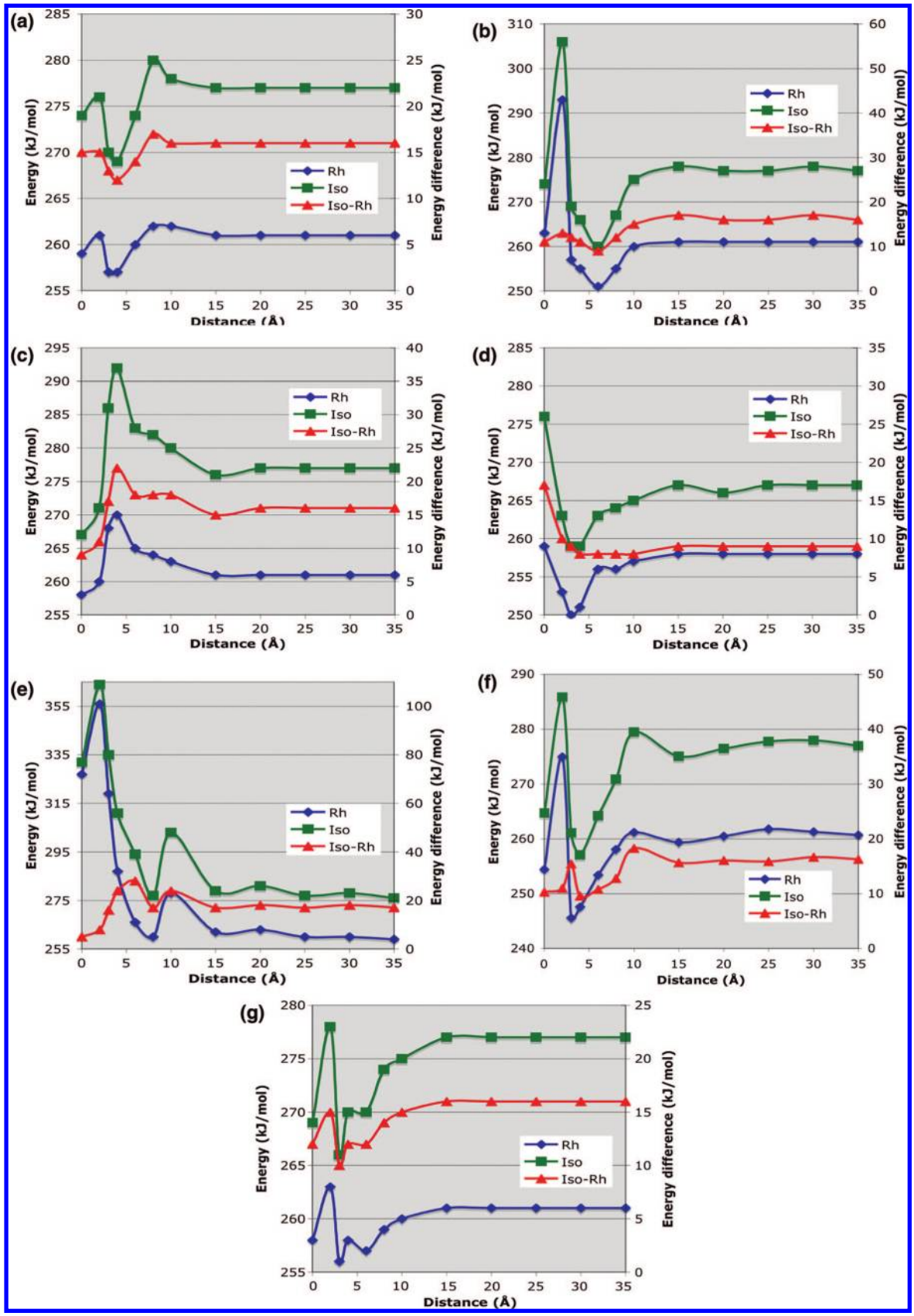

Figure 2. Distance dependence of the effect of the various multipoles and polarizabilities for the first excitation energy of Rh (◆) and Iso-Rh (■) in the U19 crystal. In addition, the difference between the two curves is given (▲, and right axis). All multipoles or polarizabilities for all residues within the given distance from the chromophore are changed. (a) Conversion from anisotropic (35 Å) to isotropic (0 Å) polarizabilities; (b) deletion of the anisotropic polarizabilities (35 Å; no polarizabilities at 0 Å); (c) deletion of the quadrupoles; (d) deletion of the dipoles; (e) conversion of LoProp charges (35 Å) to integer charges (0 Å); (f) conversion from the full LoProp model to the polarizable Amber 2002 force field; and (g) conversion of B3LYP/aug-cc-pVDZ properties to HF/6-31G* properties.

Method and Basis-Set Dependence. Another way to speed up the calculations is to employ a cheaper approach to calculate the LoProp properties. To test this, we have for the U19 structure calculated the multipoles and polarizabilities at two levels of theory, HF/6-31G* and B3LYP/aug-cc-pVDZ. From Table 1, it can be seen that the two methods give quite similar results: With the full multipole and polarizability model [+,+,+,a], the calculated excitation energies differ by only 3 (Rh) or 8 (Iso-Rh) kJ/mol (note that different methods are used only for the calculations of the multipoles and polarizabilities; the excitation energies are always calculated with the same method and basis set, CASPT2/6-31G*). In fact, the cheaper HF/6-31G* is closer to experiments (i.e., lower excitation energies), but this is only fortuitous. Similar differences are obtained for the other calculations, but with only charges [+,−,−,a/−], the HF/6-31G* properties give higher excitation energies (by 3−5 kJ/mol).

This small difference is quite unexpected, because polarizabilities normally depend strongly on the basis sets (and method): diffuse functions are mandatory to obtain converged and accurate polarizabilities, and the double-ζ basis is actually probably still somewhat small to obtain converged results.[50] Likewise, the HF/6-31G* method should overestimate calculated dipole moments, and this is actually employed in the Amber 1994 force field to get results that should simulate the polarizing effect in water solution.[30] In our calculations, we see no such tendency.

The distance dependence of the transition from B3LYP/aug-cc-pVDZ to HF/6-31G* properties is shown in Figure 2g. It can be seen that the actual effect is somewhat larger, up to 7 kJ/mol for Rh and 12 kJ/mol for Iso-Rh. The results are converged at ∼8 Å.

For relative excitation energies (the difference between Rh and Iso-Rh), the difference between the two methods is even smaller, typically 0−1 kJ/mol, but 4−5 kJ/mol for the [+,+,+,a] and [+,+,+,s] calculations. The distance dependence in Figure 2g shows that the difference is converged around 8 Å. Thus, if errors of up to 8 kJ/mol are acceptable, much computer time can be saved by calculating all of the properties at the HF/6-31G* level. If this method is used only for residues outside 8 Å, the error is less than 3 kJ/mol.

Conclusions

In this Article, we have performed a detailed investigation of how the surrounding protein is best modeled in theoretical calculations of excitation energies. As a test case, we have used the retinal chromophore in rhodopsin and calculated the excitation energies at the CASPT2//CASSCF level. We have studied both absolute and relative excitation energies, in the latter case for variations in the chromophore (retinal and 9-cis retinal) and in the protein (two rhodopsin mutants, close to the chromophore). Theoretical excitation energies are typically rather poor (as compared to the experimental uncertainty), so theoretical investigations often concentrate on relative excitation energies. Moreover, we have tested the stability of the results by studying two different crystal structures, giving rise to slightly different QM/MM structures. We have used a very detailed model of the surroundings, a multicenter-multipole model up to quadrupoles and with anisotropic polarizabilities in all atoms and bond centers, as the reference point, and then studied if any parts of this model may be removed without compromising the results. The investigation has led to a number of important and interesting observations and conclusions.

Polarizabilities have a strong influence on the spectrum, up to 46 kJ/mol, and can therefore not be ignored in any detailed study of the effect of the protein on the spectrum of a bound chromophore. The use of anisotropic polarizabilities is important at distances shorter than 6−10 Å. Therefore, we strongly recommend the use of anisotropic polarizabilities in calculations of excitation energies.

The polarizable Amber 2002 force field gives errors of up to 10 kJ/mol, as compared to the full LoProp model (which is calculated for the right conformation of the protein). Thus, a polarizable force field does not solve the problem, but it can be used to speed up the calculations for residues more than 10 Å from the chromophore.

The effect of quadrupoles is up to 26 kJ/mol, that is, quite significant for absolute excitation energies. For relative energies, the effect is up to 13 kJ/mol, but it levels out already around 6 Å.

The effect of dipoles is somewhat smaller, up to 11 kJ/mol for absolute energies and 9 kJ/mol for relative energies.

The effect of point charges is very large, up to 60 kJ/mol for absolute energies and up to 14 kJ/mol for relative energies. The effect is long-ranged.

Different sets of simple point-charge models give quite different results. For example, the Amber 1994 and 2003 charges give an effect that differs by up to 16 kJ/mol for absolute and 8 kJ/mol for relative energies. This shows that calculations with a fixed (standard) point-charge model (as in most previous calculations[2−11]) give quite uncertain results. The results are even more unreliable if there are point charges very close to the chromophore. However, it is possible that the effect is reduced if the geometry is optimized with the respective force field.

Interestingly, multipoles and polarizabilities calculated at the HF/6-31G* level of theory give an effect on the excitation energies quite similar to those calculated at the much more expensive B3LYP/cc-pVDZ level of theory, with errors of less than 8 kJ/mol (6 kJ/mol for relative energies). The results converge within 8 Å from the chromophore. Thus, if errors of this size are acceptable, such an approach may save much computer time.

A natural question is why such seemingly accurate results have been obtained with theoretical calculations if the results are so sensitive to the model of the surroundings. The answer is that only relative excitation energies are normally considered. This is illustrated by Table 2, which shows the difference in excitation energies of rhodopsin and isorhodopsin, as well as the calculated shift in the excitation energies of the E113D and G121L mutants (i.e., the relative energies studied in this Article, for which experimental data are available). It can be seen that all calculations except one (ff03 with CG junctions) give mean absolute deviations of less

than 10 kJ/mol and maximum errors of less than 20 kJ/mol. In fact, even the vacuum model [−,−,−,−] gives excellent results with a maximum error of 7 kJ/mol. Thus, accurate relative energies are easily obtained, whereas correct absolute energies are much harder to calculate.

Our results give us a firm indication of how accurate absolute excitation energies of chromophores in proteins should be calculated. Clearly, both point charges and polarizabilities need to be considered for accurate results, and they must be calculated for the actual conformation of the protein, using an approach similar to the one used in this Article. If the program used allows it, we see no reason not to include a full multipole model for the electrostatics. In fact, previous results indicate that the multipole model is not converged until quadrupoles are included,[21,51] as was done in this Article. There are no computational reasons not to include dipoles and quadrupoles in the calculations; they do not make the calculations significantly slower, neither the calculations of the multipoles, nor the spectra calculations. Instead, it is the polarizabilities that are expensive; they increase the calculation time for the properties by a factor of 6 and that for the spectra calculation by a factor of 2−3, mainly due to slower convergence of the CASSCF iterations. However, all calculations in this Article were run in less than 3 days (two CASSCF and CASPT2 calculations), which is not prohibitively much. In fact, this is much less than the time taken for the calculation of the multipoles and polarizabilities for the surrounding protein at the B3LYP/aug-cc-pVDZ level (∼250 CPU days, but it is trivially parallelizable). However, for accurate results, the basis sets in the CASSCF//CASPT2 calculations should be increased.

Finally, it is most likely important to sample many different conformations of the surrounding protein (and the chromophore). This can be easily done with molecular dynamics simulations. However, the present results indicate that the properties should be recalculated for each new conformation of the protein, at least for residues within 10 Å of the chromophore. However, much time can be saved by using the polarizable Amber 2002 force field for residues outside this distance or by calculating their properties with the HF/6-31G* method (which takes ∼15 CPU days for the whole protein). In future publications, we will test such an approach.

Acknowledgment. This investigation has been supported by grants from the Swedish research council, the Crafoord foundation, and AstraZeneca, as well as by computer resources of Lunarc at Lund University.

Supporting Information Available: Instructions for calculating multipoles for a protein, and a template of the restop file. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Öhrn, A.; Karlström, G. *Mol. Phys.* **2006**, *104*, 3087–3099.

(2) Pierloot, K.; De Kerpel, J. O. A.; Ryde, U.; Roos, B. O. *J. Am. Chem. Soc.* **1997**, *199*, 218–226.

(3) Rajamani, R.; Gao, J. *J. Comput. Chem.* **2002**, *23*, 96–105.

(4) Yamada, A.; Kakitani, T.; Yamamoto, S.; Yamato, T. *Chem. Phys. Lett.* **2002**, *366*, 670–675.

(5) Hayashi, S.; Ohime, I. *J. Phys. Chem. B* **2000**, *104*, 10678–10691.

(6) Hayashi, S.; Tajkhorshid, E.; Pebay-Peyroula, E.; Royant, A.; Landau, E. M.; Navarro, J.; Schulten, K. *J. Phys. Chem. B* **2001**, *105*, 10124–10131.

(7) Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281–1297.

(8) Fujimoto, K.; Jun-ya, H.; Hayashi, S.; Shigeki, K.; Nakatsuji, H. *Chem. Phys. Lett.* **2005**, *414*, 239–242.

(9) Hoffmann, M.; Wanko, M.; Strodel, P.; König, P. H.; Frauenhiem, T.; Schulten, K.; Thiel, W.; Tajkhorshid, E.; Elstner, M. *J. Am. Chem. Soc.* **2006**, *128*, 10808–10818.

(10) Coto, P. B.; Strambi, A.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17154–17159.

(11) Strambi, A.; Coto, P. B.; Ferré, N.; Olivucci, M. *Theor. Chem. Acc.* **2007**, *118*, 185–191.

(12) Bravaya, K.; Bochenkova, A.; Granovsky, A.; Nemukhin, A. *J. Am. Chem. Soc.* **2007**, *129*, 13035–13042.

(13) Warshel, A.; Chu, Z. T.; Hwang, J.-K. *Chem. Phys.* **1991**, *158*, 303–314.

(14) Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857–9871.

(15) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

(16) Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 867–879.

(17) Matsuura, A.; Sato, H.; Houjou, H.; Saito, S.; Hyashi, T.; Sakurai, M. *J. Comput. Chem.* **2006**, *27*, 1623–1630.

(18) Ren, L.; Martin, C. H.; Wise, K. J.; Gillespie, N. B.; Luecke, H.; Lanyi, J.; Spudich, J. L.; Birge, R. R. *Biochemistry* **2001**, *40*, 13906–13914.

(19) Houjou, H.; Inoue, Y.; Sakurai, M. *J. Am. Chem. Soc.* **1998**, *120*, 4459–4470.

(20) Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.

(21) Krauss, M. *Comput. Chem.* **1995**, *19*, 199–204.

(22) Wanko, M.; Hoffmann, M.; Strodel, P.; Kozlowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(23) Kochendoerfer, G. G.; Lin, S. W.; Sakmar, T. T.; Mathies, R. A. *Trends Biochem. Sci.* **1999**, *24*, 300–305.

(24) Sakmar, T. P.; Franke, R. R.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8309–8313.

(25) Han, W.; Lin, S. W.; Smith, S. O.; Sakmar, T. P. *J. Biol. Chem.* **1996**, *271*, 32330–32336.

(26) Hurley, J. B.; Ebrey, T. G.; Honig, B.; Ottolenghi, M. *Nature* **1977**, *270*, 540–542.

(27) Teller, D. C.; Okada, T.; Behnke, C. A.; Palczewski, K.; Stenkamp, R. E. *Biochemistry* **2001**, *40*, 7761–7772.

(28) Okada, T.; Sugihara, M.; Bondar, A. N.; Elstner, M.; Entel, P.; Buss, V. *J. Mol. Biol.* **2004**, *342*, 571–583.

(29) Andruniów, T.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *52*, 17908–17913.

(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631.

(31) Fahmy, K.; Jager, F.; Beck, M.; Zvyaga, T. A.; Sakmar, T. P.; Siebert, F. *Proc. Natnl. Acad. Sci. U.S.A.* **1993**, *90*, 10206–10210.

(32) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157–173.

(33) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(34) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(35) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142–149.

(36) Karlström, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222–239.

(37) Luzhkov, V.; Warshel, A. *J. Am. Chem. Soc.* **1991**, *113*, 4491–4499.

(38) Gao, J.; Byun, K. *Theor. Chem. Acc.* **1997**, *96*, 151–156.

(39) Poulsen, T. D.; Kongsted, J.; Osted, A.; Ogilby, P. R.; Mikkelsen, K. V. *J. Chem. Phys.* **2001**, *115*, 2393–2400.

(40) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494–4450.

(41) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(42) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(43) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(44) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599–3605.

(45) Söderhjelm, P.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 617–627.

(46) Kandori, H.; Schichida, Y.; Yoshisawa, T. *Biochemistry (Moscow)* **2001**, *66*, 1197–1209.

(47) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(48) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(49) Cieplak, P.; Caldwell, J.; Kollman, P. A. *J. Comput. Chem.* **2001**, *22*, 1048–1057.

(50) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2004**, *120*, 9903–9906.

(51) Söderhjelm, P.; Krogh, J. W.; Karlström, G.; Ryde, U.; Lindh, R. *J. Comput. Chem.* **2007**, *28*, 1083–1090.

CT800459T