THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Analysis of Electronic Structure, Binding, and Vibrations in Biotin−Streptavidin Complexes Based on Density Functional Theory and Molecular Mechanics

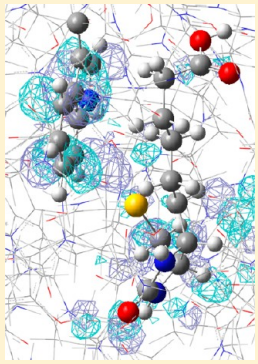
Alexei Bykhovski,*[,†] Weidong Zhang,[†] James Jensen,[‡] and Dwight Woolard[†,§]

[†]Department of Electrical & Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695, United States

[‡]U.S. Army Edgewood Chemical Biological Center, Aberdeen Proving Ground, Maryland 21010, United States

[§]U.S Army Research Office, Research Triangle Park, North Carolina 27709, United States

ABSTRACT: In this work, the biotin−streptavidin complex was studied with density functional theory (DFT), molecular mechanical methods (MM), and a hybrid DFT/MM approach in order to obtain the theoretical predictions for electronic structures, binding, optical transitions, harmonic vibrations, and absorption spectra. It was demonstrated that biotin solvation in water can reduce the binding strength to streptavidin by more than half. All studied properties, including the biotin binding and the UV absorption of the biotin−streptavidin complex, are predicted to be protonation state dependent. The absorption edge of the complex calculated with TDDFT/MM was found to be virtually insensitive to the choice of the MM force field and strongly dependent on the type of embedding of the DFT partition. Both UV and terahertz light absorption spectra are predicted to be sensitive to the presence of biotin in the streptavidin tetramer.



INTRODUCTION

Biotin−streptavidin and the related biotin−avidin are examples of exceptionally stable noncovalent complexes that find many uses in biochemical sensing applications. In particular, they are used for identifying possible new drug targets.[1] The complex (biological assembly) is formed by a tetrameric protein and up to four molecules of biotin, a 32-atom vitamin H, which is a flexible and terahertz (THz) light-absorbing ligand.[2−5] Biotinylated molecules (e.g., proteins, DNA, RNA, etc.) can easily be detected with streptavidin derivatives or efficiently captured on streptavidin-coated solid supports.[6] The fact that streptavidin has four binding sites for biotin means that it can be used in the nanoassembly process, for example, as a bridge between a surface-bound biotin and a biotin-terminated DNA molecule to be used as a nanosensor component. This was exploited in an approach to binding a single-stranded DNA molecule onto a silicon substrate.[7] In this approach, thiolated biotin molecules are attached to gold nanoclusters bound to the silicon substrate. Also, carbon nanotubes were positioned on top of the DNA origami using interactions between streptavidin molecules and biotinylated DNA strands.[8] Use of biotin is one of the most preferred DNA modifications because biotin can react with the surface immobilized protein molecule of streptavidin with a significant binding strength. Streptavidin can be used interchangeably with a related protein NeutrAvidin, which is a deglycosylated charge neutral form of avidin, a protein of egg whites origin with the same binding strength to biotin. Recently, NeutrAvidin proteins were addressed to biotin-functionalized sites on DNA origami constructs to use as probes for biotinylated antibodies, the green fluorescent protein (GFP), and other molecular species.[9] Therefore, the detailed knowledge of the physical and chemical properties of noncovalent complexes such as biotin−(strept)avidin is important for gaining an insight into biomolecular recognition as well as for engineering bionanomaterials and bionanoelectronic components. The characterization of bonded biotin-containing complexes is especially important since they could strongly influence the functionality of the biochemical sensors due to the high flexibility and optical activity of biotin.

In this work, biotin−streptavidin complex was studied with density functional theory (DFT), molecular mechanical methods (MM), and a hybrid DFT/MM approach in order to obtain equilibrium atomic and electronic structures, binding, optical transitions (with time-dependent (TD) DFT), harmonic vibrations, and absorption spectra. Similar theoretical descriptions were used in our recent work to successfully predict low-energy conformations of a double-stranded DNA fragment TGCGCA capped with trimethoxystilbene carboxamide (TMS) and interpret THz absorption spectra of biotin in polyethylene.[10,11] This work presents the study of the full tetrameric biological assembly of the streptavidin and biotin−streptavidin complexes. Previously, a binding strength of a single molecule of biotin to a single streptavidin monomer was calculated ab initio using the Hartree−Fock (HF) approximation with the 3-21g basis in the molecular fractionation with conjugate caps (MFCC) approximation.[12] The obtained HF/3-21g energies can have large errors because of the small basis size and a lack of the electron correlation treatment.[12]

Received: July 31, 2012
Revised: November 28, 2012
Published: December 6, 2012

   dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

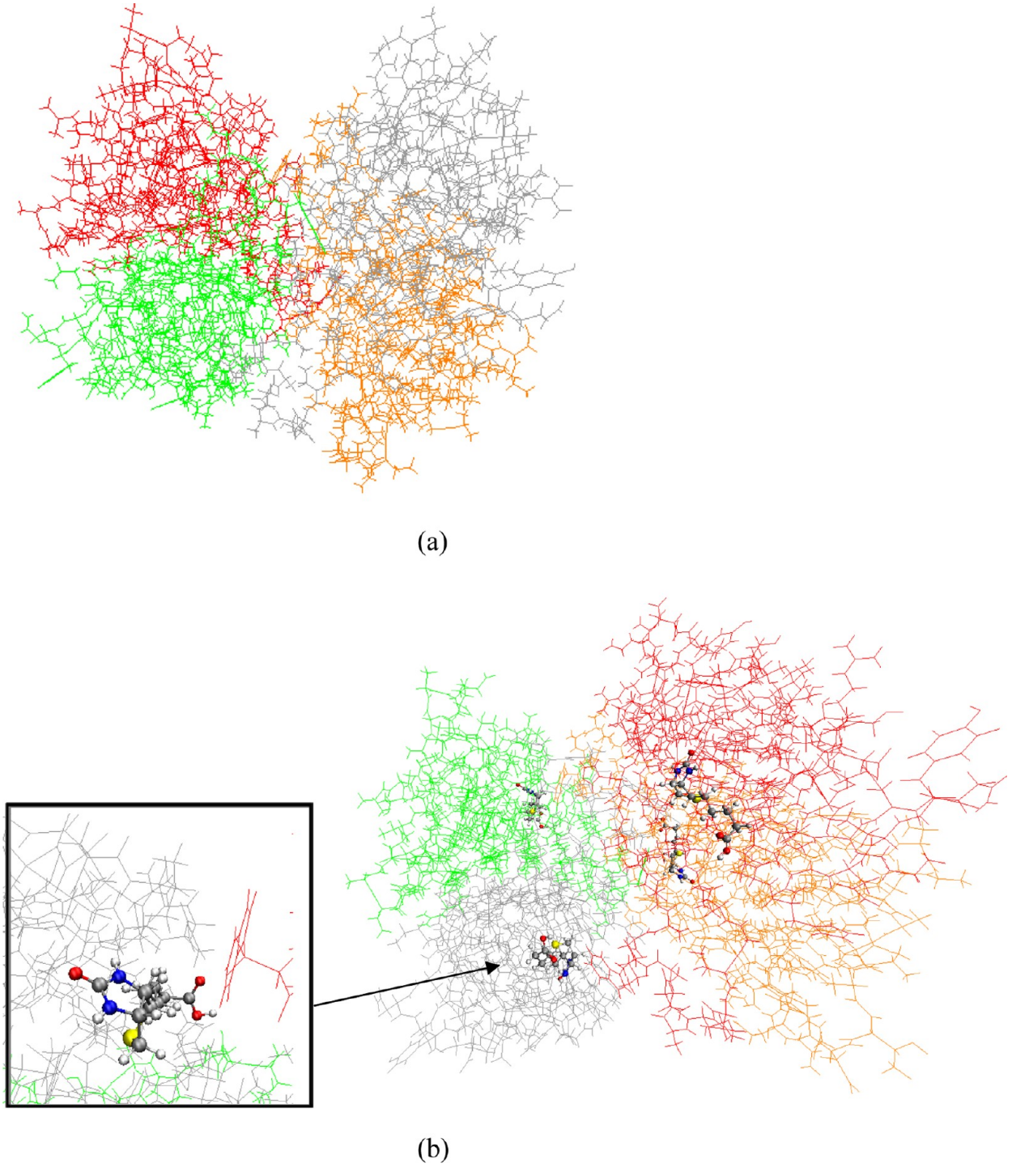


Figure 1. (a) Complex **S4** with each monomer distinguished by a different color. Here monomer one is silver, monomer two is green, monomer three is orange, and monomer four is red. (b) Complex **B4S4** (on right) where the components for each biotin are represented by colored spheres. An exploded view (on left) has been included to illustrate the amino acid binding to the biotin located closest to the bottom of **B4S4**.

In the MFCC approach, the streptavidin monomer is split into separate amino acids that are capped to saturate broken bonds. Then, the biotin binding energy is calculated for rigid molecules as a sum of all biotin−amino acid pair energies with energies of the caps subtracted. The geometry optimization was not performed and the biotin binding energy of the biological assembly was not calculated in ref 12. In the present work, the entire biological assembly of the complex consisting of four streptavidin monomers and up to four biotins was optimized at the DFT/MM level of theory using a larger split-valence basis than in ref 12 as described in the next section. The effects of wave function polarization, solvation in water, and protonation are considered.

THEORETICAL AND MODELING METHODOLOGY

In order to simulate the electronic properties of the complex containing between 6880 and 7008 atoms depending on the number of biotins (i.e., which can be up to four), the hybrid approach was applied to obtain the ground and excited electronic states. In this

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

Table 1. Energy Contributions of Simulated Structures[a]

			energies for different MM models and protonation states			
structure	contribution type		MM1 state1	MM2 state1	MM1 state2	MM2 state2
B4S4	DFT partition	vacuum (solution)	−2231.529 (−2232.078)	−2231.519 (−2232.073)	−2231.460 (−2232.047)	−2231.487 (−2232.077)
		charge embedding	−2441.195	−2441.382	−2441.316	−2441.475
	MM energy of DFT partition		−0.558	−0.487	−0.509	−0.430
	MM energy of entire structure from QM/MM model and independently minimized (in parentheses)		−17.865 (−17.900)	−16.910 (−16.943)	−17.935 (−17.939)	−16.990 (−17.058)
	hybrid QM/MM energy	mechanical embedding	−2248.836	−2247.942	−2248.886	−2248.046
		charge embedding	−2248.887	−2247.984	−2248.90	−2248.041
B3S4	DFT partition	vacuum (solution)	−2087.488 (−2088.038)	−2087.464 (−2088.032)	−2087.439 (−2088.034)	−2087.496 (−2088.054)
		charge embedding	−2297.078	−2297.530	−2297.421	−2297.549
		MM energy	−0.274	−0.161	−0.217	−0.213
	MM energy of entire structure obtained from QM/MM and independently minimized (in parentheses)		−17.519 (−17.819)	−16.745 (−16.799)	−17.819 (−17.927)	−16.749 (−16.787)
	Hybrid QM/MM energy	mechanical embedding	−2104.733	−2104.048	−2105.042	−2104.033
		charge embedding	−2104.771	−2104.072	−2105.027	−2104.024
B+B3S4−B4S4	DFT energy, eV vacuum (solution)		5.8 (2.9)	6.2 (3.0)	5.2 (2.2)	4.4 (2.46)
B+B3S4−B4S4	MM, eV		−2.72	−1.01	−4.60	2.45

[a]Binding energy, eV; all other energies, hartree.

(TD)DFT/MM approach, a long-range and van der Waals (VDW)-corrected functional was employed as implemented in the Gaussian 09 system.[13,14] The Stevens/Basch/Krauss effective-core potential (ECP) split-valence basis was used in the DFT part of the simulations, and Amber-based protein force fields were used in the MM part.[15,16] The geometrical details of the biotin−streptavidin complex of interest are illustrated in Figure 1. Figure 1a depicts the basic structure of the streptavidin molecule without any biotins and utilizes a color-coding scheme to illustrate the fact that it has a tetramer form (denoted as **S4**) with four identical monometric units. Figure 1b depicts the same streptavidin molecule when it has acquired a maximum number of biotins (i.e., denoted as **B4**) and also provides a magnified view showing the amino acid binding of the lowest streptavidin monomer in the complex to biotin. Due to the inherent four-component form of the **S4** complex and large separation between the individual biotin binding sites and/or biotin bindings, it is possible to derive all the quantum mechanical dependent properties with a partitioning where only one of the active binding sites (or biotin-containing sites) is described by the DFT model and the three-component remainder by the MM model. This hybridization was made consistent using a method known as ONIOM.[17] In this approach, the total extrapolated energy is a sum of the MM energy of the entire structure (E_{MM}) and a difference between a DFT (E_{DFT}) energy of the selected model partition and its MM energy (E_{mm}), thus eliminating double counting of energy contributions, $E_{extrap} = E_{MM} - E_{mm} + E_{DFT}$. The geometry is optimized to obtain the minimum of the extrapolated energy. This approach requires the knowledge of the MM force field in the entire structure, not just outside the DFT region. Therefore, it incorporates all interactions between DFT atoms and MM atoms outside the DFT partition at a MM level of theory. A simple measure of this interaction is the DFT and MM energy increases, $E_{DFT} - E_{DFT0}$ and $E_{MM} - E_{MM0}$, where the energies with "0" subscripts denote an unconstrained (i.e., no extrapolation) energy minimum at a respective level of theory. Hence, if the DFT and MM partitions were completely uncoupled, $E_{DFT} = E_{DFT0}$ and $E_{MM} = E_{MM0}$. Because of the coupling, the energy minimization is performed under a set of constraints and obtained energies E_{DFT} and E_{MM} are higher, $E_{DFT} > E_{DFT0}$ and $E_{MM} > E_{MM0}$. This point is illustrated in the simulation results given in Tables 1−2 (see, for example, the third row of structures **B4S4** and **B3S4** in Table 1), where the resulting MM energies of the complex are presented both for the hybrid optimization, E_{MM}, and for the independent MM minimization, E_{MM0} (in parentheses). Two Amber's protein force fields, ff99SB (MM1) and ff99SBnmr with NMR-based modifications to backbone torsions[18] (MM2), were used to provide the MM description for the streptavidin.

Table 2. Energy Contributions of S4, Protonation State 1, hartree

	MM model	
contribution type	MM1	MM2
DFT partition: vacuum (solution)	−2087.481 (−2088.029)	−2087.461
MM energy of DFT partition	−0.267	−0.182
MM energy of entire structure from QM/MM model and independently minimized (in parentheses)	−16.783 (−16.912)	−15.875 (−15.933)
hybrid QM/MM energy	−2103.997	−2103.149

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

There are two different ways in which E_{DFT} and E_{mm} can be calculated for the DFT partition. The first approach does not involve any interactions with the surrounding atoms and is known as mechanical embedding. Therefore, all interactions with the surroundings are included at a MM level, as was already mentioned. In the second approach that is known as electronic or charge embedding, electrostatic interactions between DFT atoms and surrounding MM atoms are included in E_{DFT} and E_{mm}. Therefore, the wave function becomes polarized, but the net change in the total extrapolated energy is generally small because the electrostatic interactions are incorporated in both approaches through the MM energy of an entire structure and because DFT partition-related energy terms tend to cancel each other out. In this work, the initial structures were based on X-ray diffraction data[19] and optimized using the mechanical embedding approach to obtain geometries corresponding to the energy minima. The wave function polarization effect caused by the electrostatic interactions with the atoms outside the DFT partition was calculated for the obtained geometries using the electronic embedding for the ground and excited electronic states. Solvation with water was simulated for biotin and the complex using a continuum model.[20] In the simulations, the system was considered to be electrically neutral. Electrical neutrality is important to avoid problems that can be caused by nonspecific binding to a charged protein. The neutrality was maintained by protonating four of the eight histidine residues. To evaluate the influence of the protonation distribution on the complex, two different protonation states were simulated: one with a protonated histidine 114 (protonation state 1) and another with a protonated histidine 74 (protonation state 2). This protonation scheme was repeated for all four monomers of the streptavidin. For the biotin, the force field parameters were generated using restrained electrostatic potential (RESP) charges[21] that were extracted from the DFT optimization.

In order to test the generated MM force field for biotin, atomic vibrations were calculated for the optimized biotin, and its THz-to-IR light absorption spectrum was simulated on both MM and DFT levels. Since simulated atomic vibrations are extremely sensitive to the choice of force field parameters, such a comparison

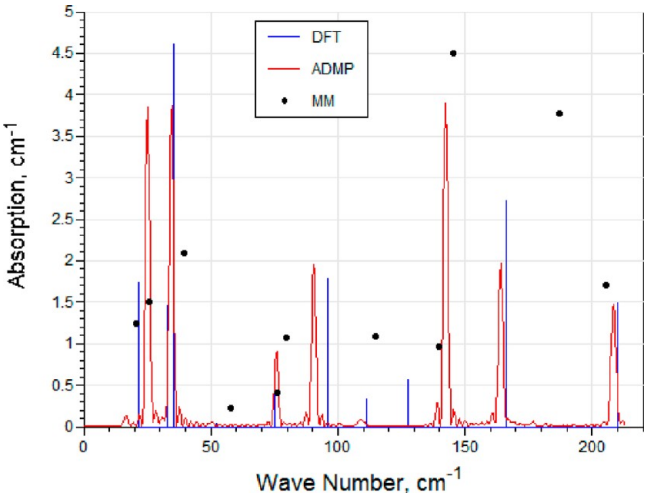


Figure 2. Low-THz biotin absorption obtained with DFT-based and MM-based harmonic analysis (vertical lines and dots, respectively) and with ADMP MD (continuous curve).

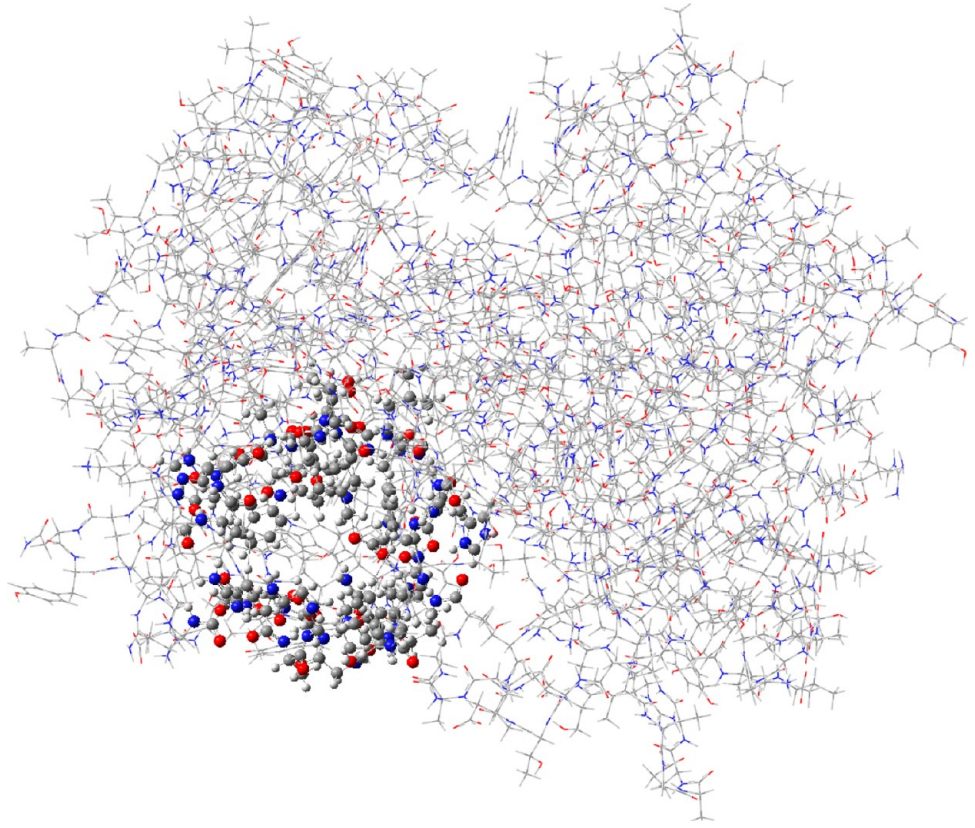


Figure 3. Streptavidin tetramer optimized with DFT/MM1. Colored spheres represent the DFT partition which includes contributions from the first and the fourth monomers.

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

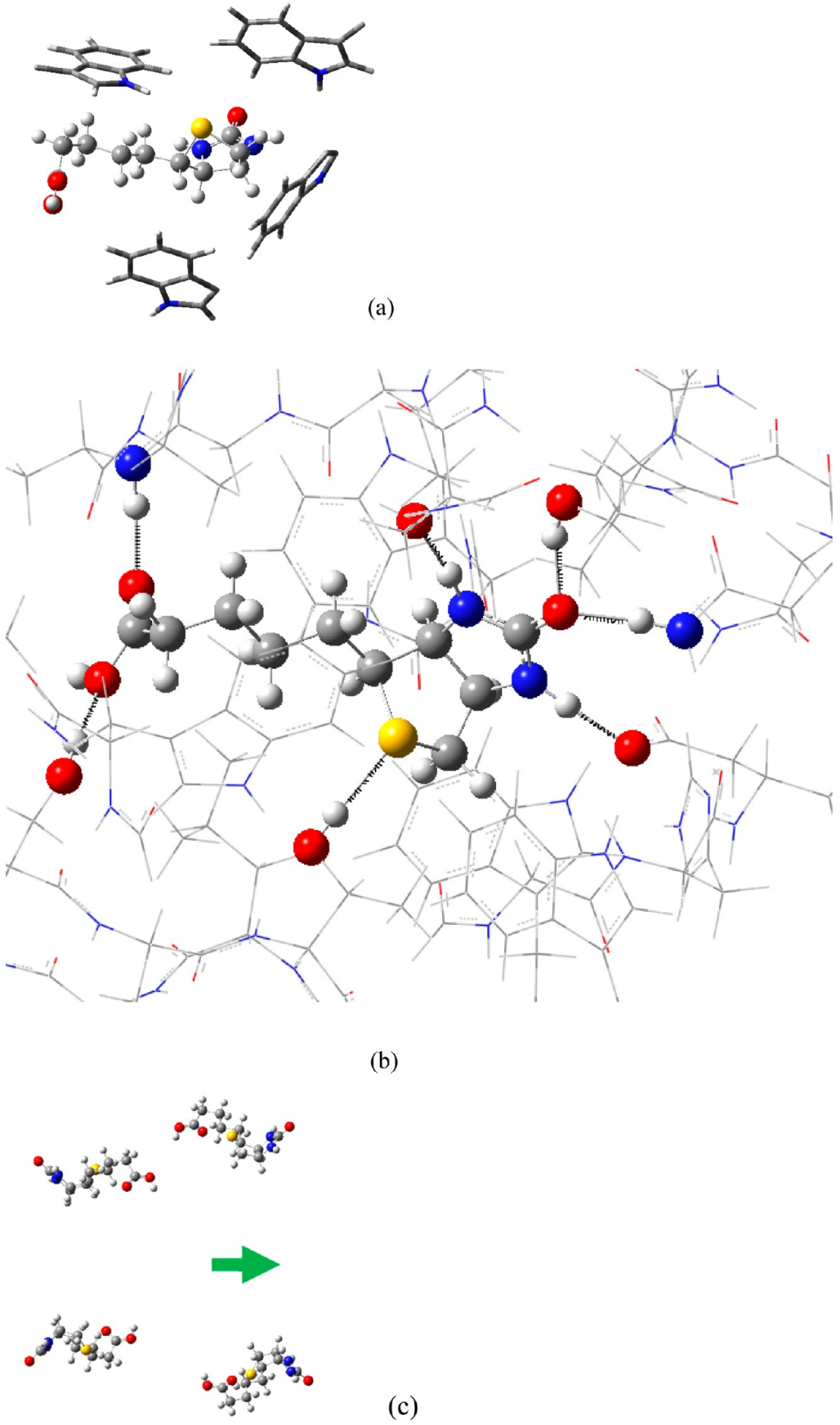


Figure 4. Biotin (colored spheres) inside the DFT partition with (a) TRP residues (tubes) around biotin, (b) hydrogen bonds (dotted lines), and (c) **B4S4** dipole (green arrow).

dx.doi.org/10.1021/jp3075833 | *J. Phys. Chem. B* 2013, 117, 25−37

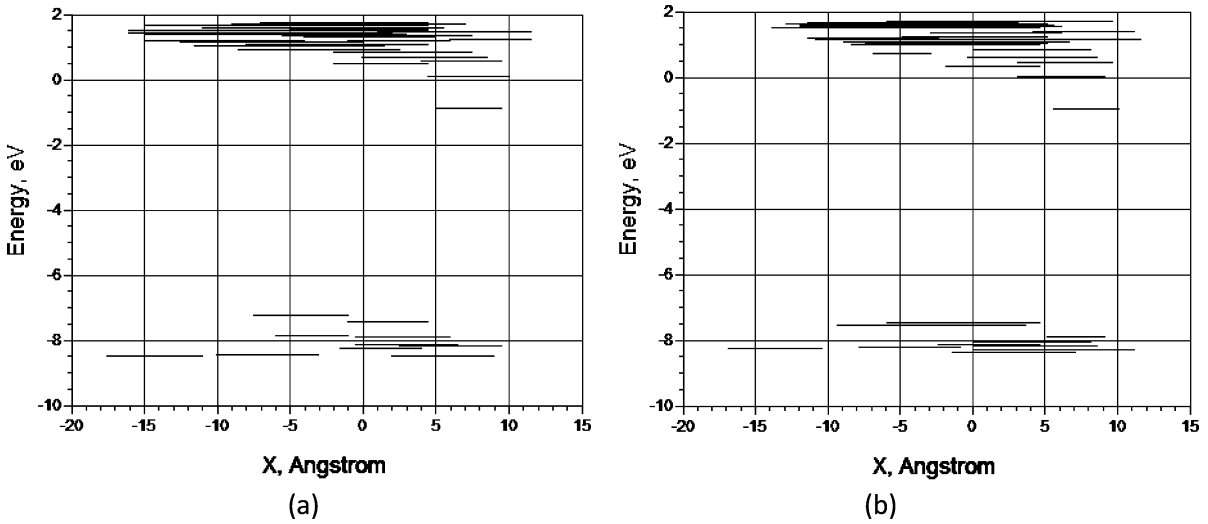


Figure 5. Energy−spatial extent diagrams of molecular orbitals obtained using DFT/MM1 with mechanical embedding for the protonation state 1: (a) **S4**; (b) **B4S4**. Coordinate system is centered on biotin.

provides a check of the force field accuracy. The MM-based absorption was found to be consistent with DFT-based absorption results that were obtained both from the harmonic analysis and from the atom-centered density matrix propagation molecular dynamics model (ADMP MD).[22] In the DFT-based ADMP MD simulations, a low-temperature (4 K) dipole moment time history of biotin was calculated for about 15 ps with a 1 fs time step.[11] At each step, the dipole moment was computed quantum-mechanically using the wave function that depends on atomic positions. The ADMP-based biotin absorption spectrum was computed within the framework of the linear response theory using the relationship between the equilibrium dipole moment fluctuations and the light absorption cross section.[23] The predicted MM- and DFT-based low-THz spectra of biotin are plotted in Figure 2. The MM absorption reproduces the overall THz fingerprint and predicts the strongest line in 0−220 cm^{-1} interval to be around 150 cm^{-1}, which is close to the ADMP-based prediction. Also, the MM-based simulation predicts that the strongest biotin line within 0−120 cm^{-1} is at 40 cm^{-1}, which is close both to the DFT results and to the experimental result of 43 cm^{-1}.[5] While the MM results are not highly accurate for the weak absorptions in the midrange, the low-THz and shorter wavelength results are quite good.

According to the X-ray crystallographic data, each streptavidin monomer consists of a long covalent chain that has the ability to loop around and bind to biotin.[19] The state of this loop is strongly dependent on the presence of biotin. In fact, when biotin interacts with the surrounding amino acids the local binding site becomes more rigid, as has been confirmed by experimental studies in solution.[24] Moreover, in the absence of biotin, the binding loop changes conformation from the "closed" state that clasps biotin to an "open" state so mobile that it is often not observable in the electron density mapping.[24,25] Therefore, a proper DFT partition must incorporate a sufficient biotin binding loop (i.e., with or without biotin) to accurately describe the ground and excited electronic states for open and closed molecular conformations. In order to derive information on biotin binding, this type of partition must be applied to the **B3S4** and **B4S4** complexes to generate their individual total energies, which is then used together with the calculated total energy of an isolated biotin molecule. Similarly, an accurate partitioning must be applied to both the open state (i.e., **B3S4**) and closed state (i.e., **B4S4**) to describe changes in the allowed optical transitions caused by the change in the binding loop state.

The biotin binding loop poses a challenge for the DFT/MM partitioning. If one introduces a shell boundary around biotin that is made of overlapping spheres centered on the biotin atoms, it can cut both the streptavidin backbone and its side chains. This is not desirable because each cut across a covalent bond introduces an open valence that needs to be saturated. Dangling bond saturation, which is usually accomplished by adding hydrogen atom, is an approximation that can introduce structural distortions. On the other hand, from a partitioning standpoint, a DFT partition has to be compact and include most atoms within the active site in order to describe its electronic structure and avoid potential optimization problems. Therefore, the DFT partition was optimized to satisfy these requirements, in particular, to minimize the number of cuts while keeping the partition as compact as possible. Also, no cuts across the protein side chains and double bonds were allowed. The resulting DFT partition contains biotin along with a number of the streptavidin polypeptide clusters surrounding it. Each cluster consists of an electrically neutral fragment of the main chain (i.e., protein backbone) with all respective side chains. Since each biotin is enveloped almost entirely by a single monomeric unit of streptavidin, most of the DFT atoms belong to a single monomer. The remainder of the DFT atoms is located in a cluster from another monomer. Specifically, in the pdb file order generated by X-ray diffraction studies, the first biotin is trapped by the first and fourth monomers. The resulting DFT partition is located within an ~0.7 nm shell surrounding biotin and contains 7 DFT clusters chains with a total of 474 atoms (see Figure 3). The small DFT cluster that belongs to the fourth monomer consists mostly of tryptophan (TRP) and has only five protein backbone atoms and 28 atoms overall. In Figure 4a, this TRP is shown at the bottom together with biotin and other three TRP residues surrounding biotin inside the streptavidin complex. The remaining six DFT clusters of the first monomer have at least five protein backbone atoms with an overall atom count 72, 72, 28, 100, 87, and 41, as well as 12 link atoms. In Figure 4b, biotin is shown with hydrogen bonds inside the DFT partition (wires), and in Figure 4c, the **B4S4** dipole moment is shown based on the MM simulation.

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

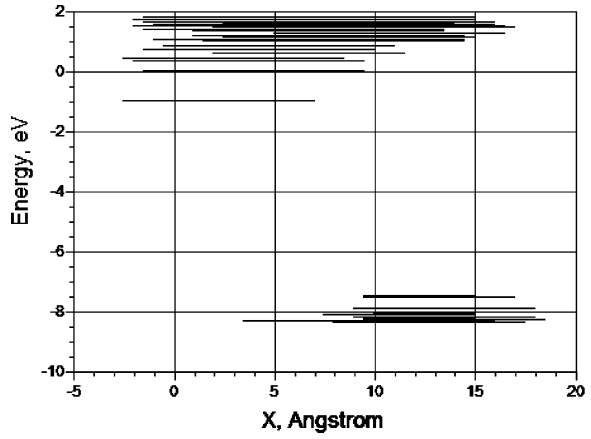


Figure 6. Energy−spatial extent diagram of molecular orbitals for **B4S4** obtained using DFT/MM1 with charge embedding for the protonation state 1. Coordinate system is shifted ~1.2 nm with respect to the one for the mechanical embedding case.

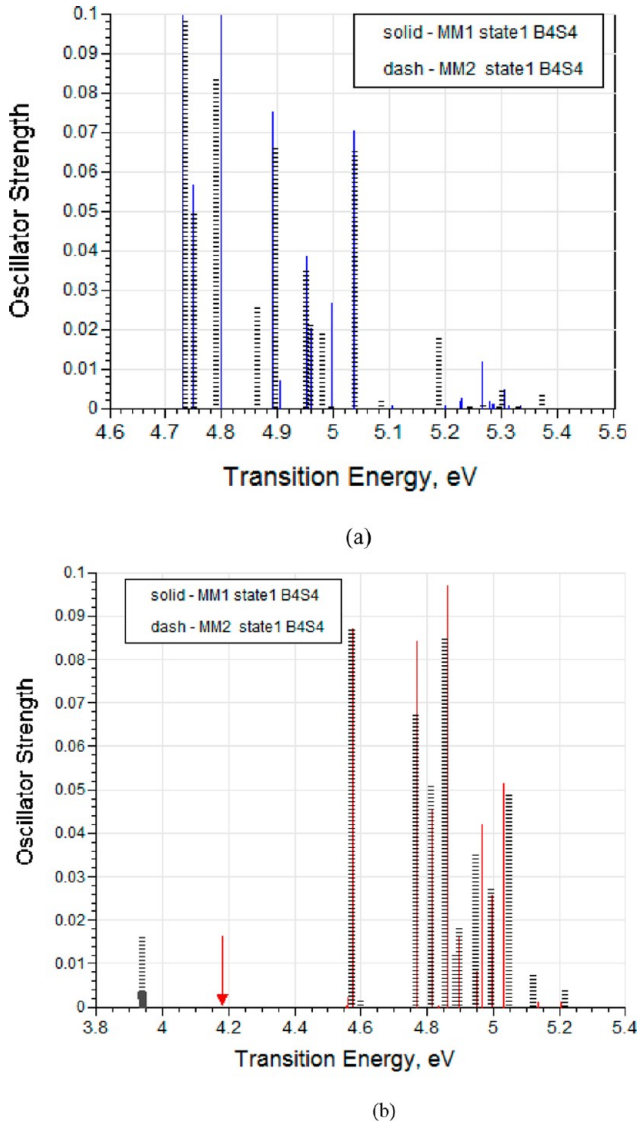


Figure 7. Lowest 20 transitions for **B4S4** in the protonation state 1, obtained using DFT/MM with (a) mechanical and (b) electronic embedding. Solid lines, MM1; dashes, MM2. Arrows point to weakly absorbing states that correspond to HOMO−LUMO transitions.

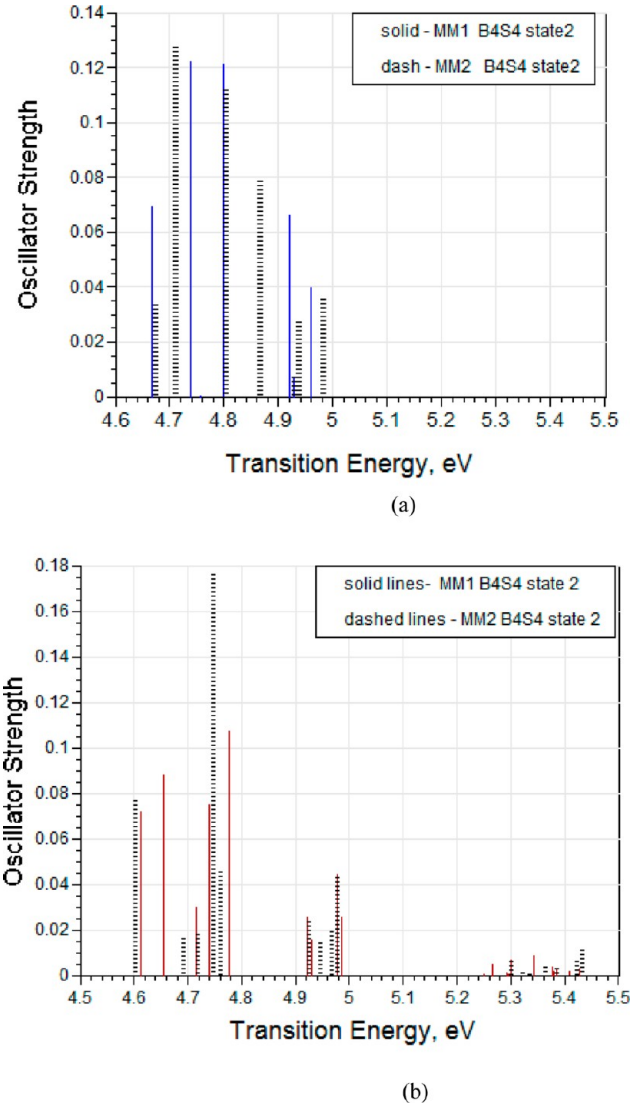
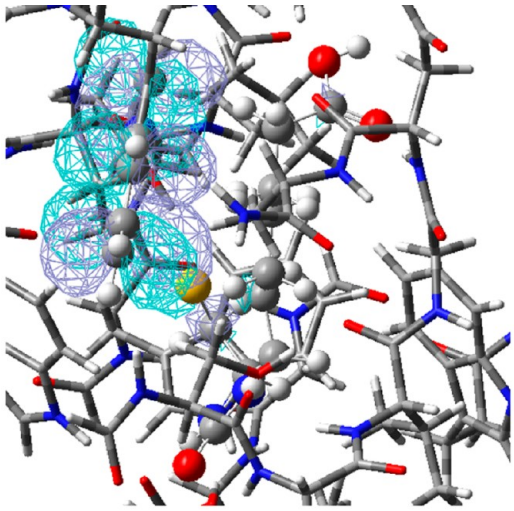


Figure 8. Lowest 20 transitions for **B4S4** in the protonation state 2, obtained using DFT/MM with (a) mechanical and (b) electronic embedding. Solid lines, MM1; dashes, MM2.

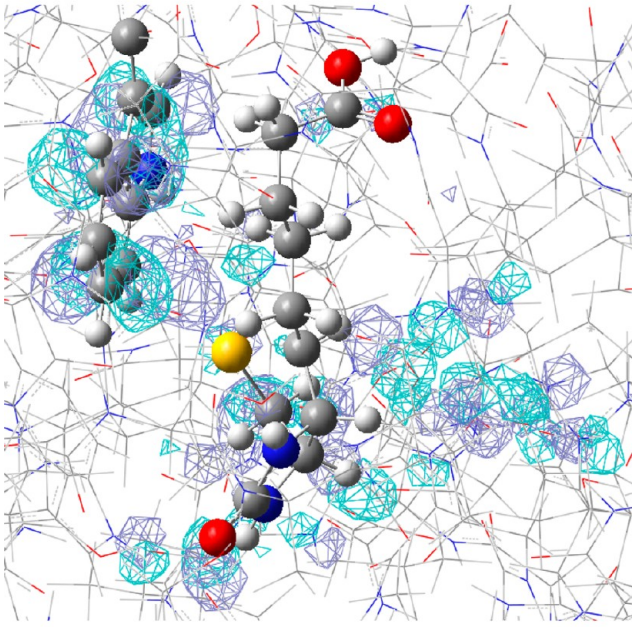
GROUND ENERGY STATES

The structures with and without biotin in the DFT region were optimized using DFT/MM to find their energy minima. The optimization results including all energy contributions for the ground electronic state are listed in Table 1 for **B4S4**, **B3S4** and Table 2 for **S4**. For **B4S4** and **B3S4**, the results are given for the two protonation states. Specifically, DFT energies are given both in vacuum and in water solution (in parentheses), as well as for the case of the electronic embedding. MM energies are given both for the entire system and for the DFT partition. For the entire system, two MM energies are given; one is a result of the DFT/MM optimization and another (in parentheses) is a result of an independent MM optimization. For **B4S4**, the resulting MM energies of the entire complex and extrapolated total QM/MM energies point to the protonation state 2 as a lower energy state of the system in accordance with both DFT/MM1 and DFT/MM2. Notice that the DFT and MM energies of the DFT partition that were calculated with the mechanical embedding do not reflect this result because they do not include electrostatic interactions with the rest of the system. When the

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37



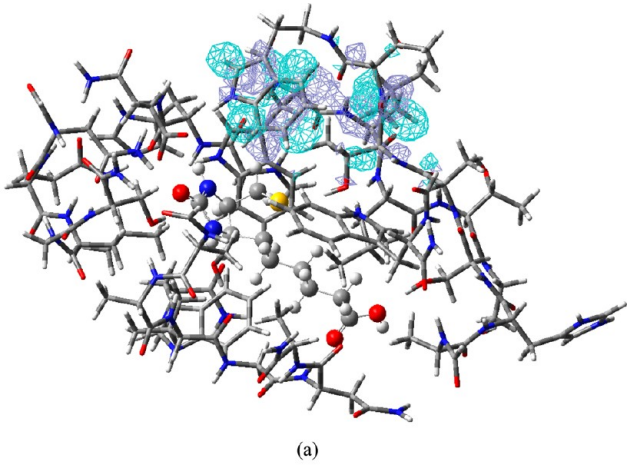
(a)



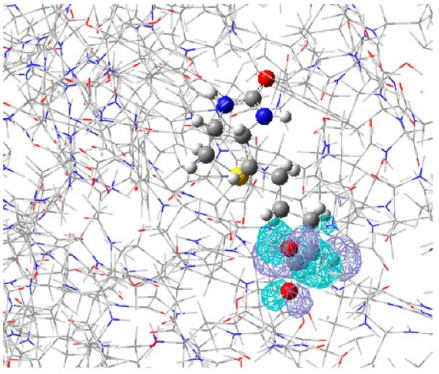
(b)

Figure 9. Unoccupied molecular orbitals with the highest contribution into the predicted absorption edge for **B4S4** in the protonation state 1. Theory, DFT/MM1. Colored spheres, highlighted DFT atoms (biotin and TRP-65). Tubes, other DFT atoms. Wires, DFT and MM atoms. Mesh, orbital surface. (a) Mechanical embedding; (b) electronic embedding.

charge embedding is included, those energies also follow the total energy trend, which is reflected in the DFT results for **B4S4**. Therefore, electrostatic interactions between the different parts of the **B4S4** complex determine the lowest energy state. For **B3S4**, the situation is more complicated. DFT/MM1 results again point to the protonation state 2 as a lower energy state; however, DFT/MM2's total extrapolated energy is lower for the state 1. Since all the energy contributions in the DFT/MM2 case including the MM energy of the entire complex are lower for the state 2, the extrapolated energy is higher for this state only because a decrease in E_{mm} that enters the extrapolated energy with a negative sign overcompensates a smaller decrease in E_{DFT}.



(a)



(b)

Figure 10. Unoccupied molecular orbitals with the highest contribution into the predicted absorption edge for **B4S4** in the protonation state 2. Theory, DFT/MM1. Colored spheres, biotin in DFT partition. Tubes, other DFT atoms. Wires, DFT and MM atoms. Mesh, orbital surface. (a) Mechanical embedding; (b) electronic embedding.

SOLVATION AND BIOTIN BINDING

The biotin binding was estimated for the reaction **1B** + **B3S4** ↔ **B4S4**. The energy decrease induced by the biotin binding to **B3S4** in the protonation state 1 in vacuum is estimated at 5.76 eV for the DFT/MM1 model from the difference between the DFT energies in vacuum: $E_{binding} = E_B + E_{B3S4} - E_{B4S4}$. Simulations of the system in vacuum overestimate the binding energy since biotin's energy in vacuum is substantially higher than in solution. Also, the **B3S4** solvation energy might be larger than that of **B4S4** because water might have a better access to streptavidin when it has a cavity without biotin inside. In this work, the solvation energy was estimated using a model that is based on a solute electron density obtained with DFT and a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions (SMD model).[20] A substantial solvation energy was obtained for biotin in water (−2.81 eV) by including the SMD model into the DFT-based energy minimization process. Additional contributions from the solvation of **B3S4** and **B4S4** were obtained with the SMD model for their DFT partitions. The DFT partitions were taken from the vacuum optimization results. No further optimization in solution was performed for **B3S4** and **B4S4**. Results for all models are presented in Table 1. For the lower energy state (i.e., protonation state 2), the biotin binding in water was estimated at 2.2 eV or 50.7 kcal/mol, which is 2.4 times smaller than in vacuum (see DFT/MM1 results

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

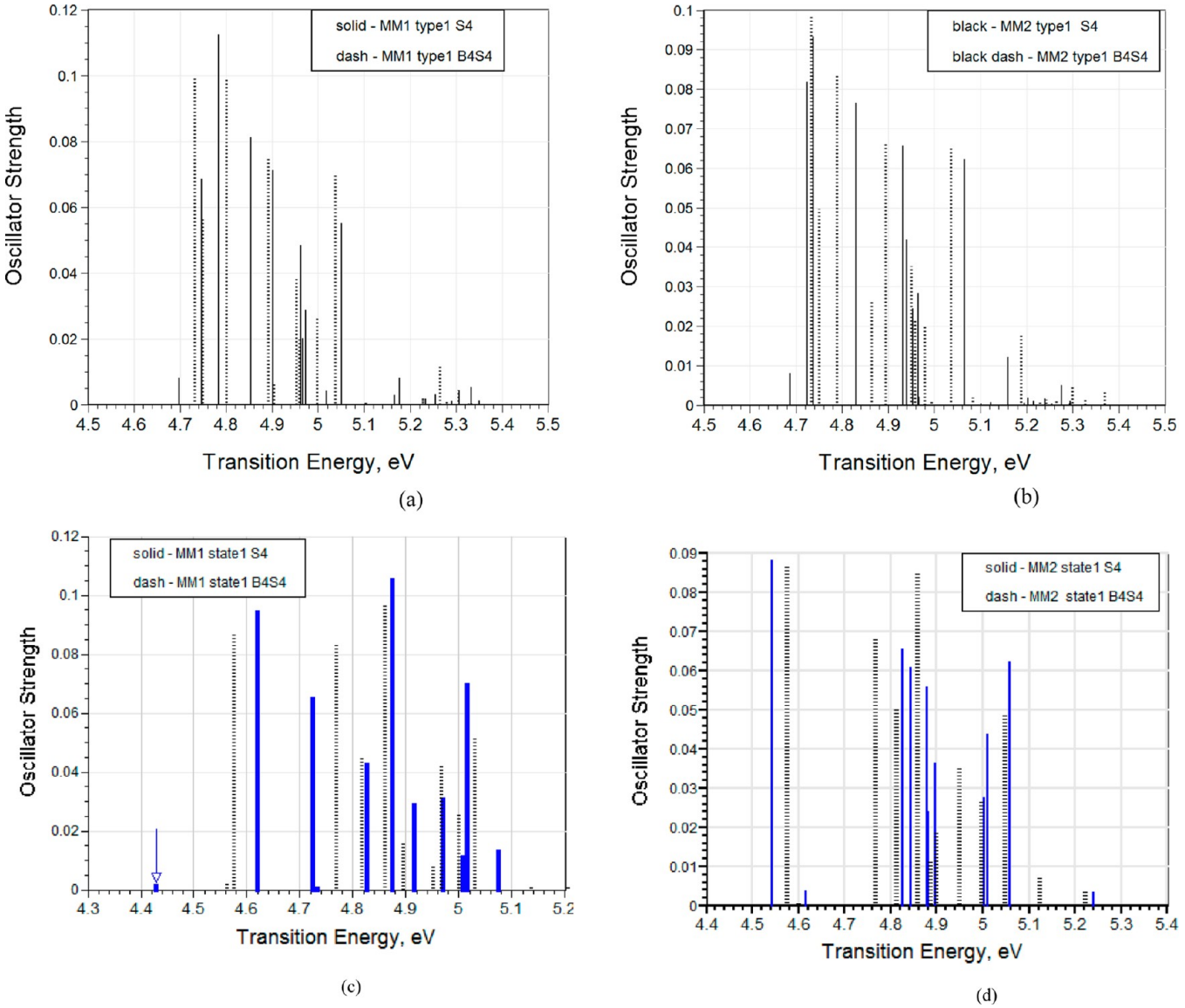


Figure 11. Lowest 20 transitions. Solid lines, **S4**; dashed lines, **B4S4** in the protonation state 1. DFT/MM with mechanical (a, b) and electronic (c, d) embedding; a, c—MM1, b, d—MM2.

in the Table 1). For the higher energy protonation state 1, the result was 2.93 eV or 67.6 kcal/mol. Notice that these energies represent a low-temperature limit for the potential energy and are substantially higher than room temperature results of about 1 eV that were obtained experimentally and by the MM MD for the free energy of binding.[26] The structures that were simulated with DFT/MM1 have consistently predicted more effective solvation for **B3S4** vs **B4S4**, as expected. For the protonation state 2, the **B3S4**'s energy has decreased in solution by 0.19 eV more than **B4S4**'s. Therefore, the simulation results indicate that binding of biotin to streptavidin is strongly protonation state dependent. This is consistent with the observation that biotin binding to the structurally similar protein avidin is pH dependent.[27] The present work confirms that the biotin binding is irreversible and is on the order of a strength of a chemical bond in the low-temperature limit. Explicit inclusion of the solvent molecules in the DFT partition instead of—or in addition to—the continuum model is required for a more accurate determination of the binding strength in solution. Notice that separate MM simulations predicted biotin binding for the reaction only for the protonation state 2 with MM2 model.

In ref 12, biotin binding was estimated in vacuum for the reaction **1B** + **1S** ↔ **B1S1** using MFCC at the HF/3-21G level of theory and an Amber force field. Notice that this reaction does not correspond to the tetrameric biological assembly of streptavidin.[19] According to the MFCC calculation, the binding energy is 5.030 eV. The obtained Amber-based binding is 3.816 eV.[12] For the protonation state 1, the DFT/MM1 simulations of this work have produced for the same (**1B** + **1S**) reaction 5.9 eV (this reaction is not included in the table). Notice that this is 0.1 eV (2.3 kcal/mol) stronger binding than the binding predicted for the **B3S4**/**B4S4** case (see the first column of the Table 1). Overall, the DFT/MM simulations in vacuum yield for the biotin binding to **B3S4** from 4.4 to 6.2 eV depending on the force field and the protonation state (see the Table 1). In particular, for the protonation state 2 in vacuum, the DFT/MM2-based binding is 4.4 eV and the corresponding Amber MM2-based binding energy is 2.45 eV. The closest to the MFCC binding calculation is the result obtained for the protonation state 2 with

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

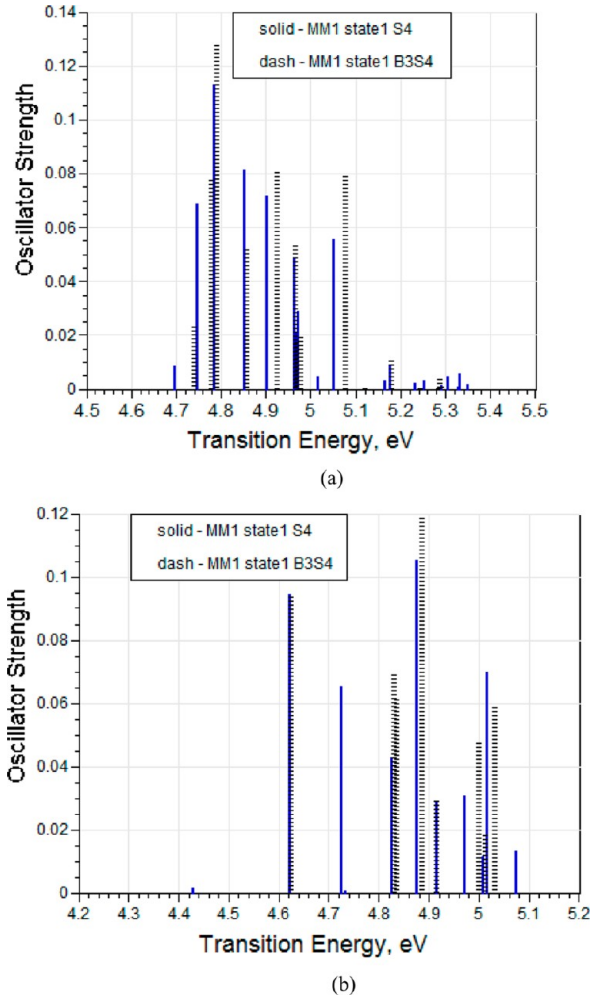


Figure 12. Lowest 20 transitions obtained with DFT/MM1 for **S4** (solid lines) and **B3S4** (dashed lines) in the protonation state 1: (a) mechanical; (b) electronic embedding.

DFT/MM1 (5.2 eV). As was discussed in this section, a much weaker binding, which is substantially closer to the room temperature results, is predicted by the present work for the system in water.

EXCITED ENERGY STATES

The molecular orbitals for **S4** and **B4S4** in the protonation state 1 that were obtained using DFT/MM1 with the mechanical embedding are plotted in Figures 5a–b in the energy–spatial extent representation. The coordinate system in this figure is centered on biotin. The corresponding electronic embedding result for **B4S4** is plotted in Figure 6. The coordinate system in Figure 6 is shifted to the center of **B4S4**. Notice that in all cases, HOMO and LUMO are centered on the atomic groups that are spatially separated by approximately 1 nm, and, therefore, HOMO–LUMO transitions do not contribute into the UV light absorption. The TDDFT/MM simulations were performed on optimized structures with and without biotin to study light-induced electronic excitations in **B4S4**, **B3S4**, and **S4**. The importance of electronic embedding and biotin's influence on light absorption were evaluated. In Figures 7a–b, oscillator strengths for the lowest 20 transitions in **B4S4** (protonation state 1) are plotted for the two hybrid models, DFT/MM1, DFT/MM2 with the mechanical (Figure 7a) and electronic (Figure 7b) embedding.

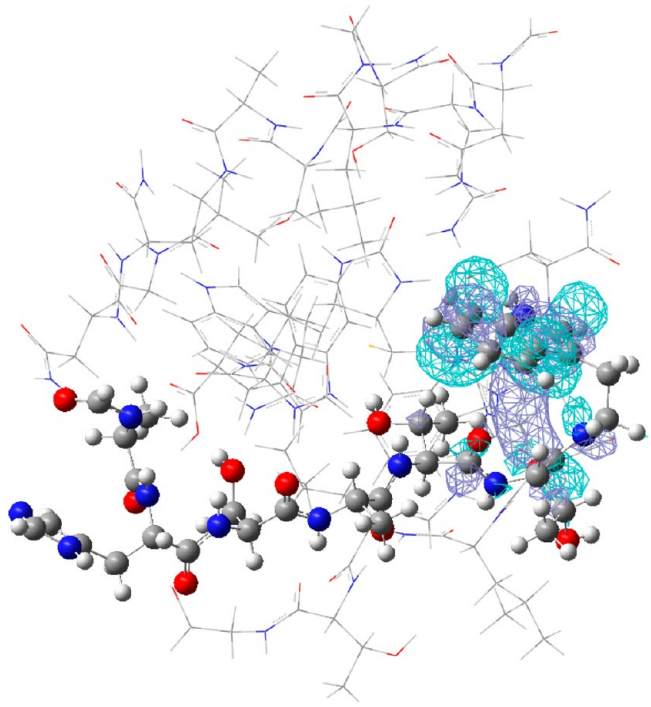


Figure 13. Reduced DFT/MM model. Wires, MM atoms; spheres, DFT atoms. Mesh, unoccupied orbital with partial localization on TRP (rings with mesh) that is responsible for the lowest absorption line within this model.

The predicted absorption edge is virtually insensitive to the choice of the force field. Moreover, for **B4S4** in the protonation state 1, most lines are insensitive to the force field choice. The absorption edge for the protonation state 1 is predicted to be at 4.56 eV for the electronic embedding, 0.17 eV lower than that for the mechanical embedding. Lower energy states are predicted that correspond to HOMO–LUMO transitions, but they do not contribute to absorption (see arrows in Figure 7b). For the protonation state 2, the mechanical and electronic embedding absorption edge predictions are 4.67 and 4.60–4.61 eV, respectively (see Figure 8 a–b). Therefore, the predicted UV light absorption edge has a small dependence on the protonation state of the streptavidin. In Figure 9 a–b, orbital surfaces of the most important (i.e., largest transition weighting) unoccupied molecular orbitals for defining the absorption edge (i.e., lowest allowed excited state) in **B4S4** (protonation state 1) are plotted for the mechanical and electronic embedding, respectively. In both instances, the electron density is partially localized on the TRP-65 residue that is a part of the biotin cage (both biotin and TRP are shown as colored spheres). Also in both cases, there is a delocalized contribution that stretches from TRP toward biotin's sulfur. In the case of the electronic embedding, the orbital contains many additional contributions that are spread over different chain fragments (see Figure 9 b). For the protonation state 2 (see Figure 10), there is also a substantial change in the predicted molecular orbitals depending on the type of embedding. In the mechanical embedding case, the most important unoccupied orbitals (for the absorption edge) are localized on two TRP residues (Figure 10 a). In the electronic embedding case, the main unoccupied orbital is now exclusively localized on and around the O=C–OH (carboxyl) group of biotin (Figure 10 b). The corresponding occupied orbital dominates the region around TRP, but it is not illustrated in Figure 10 b. For the structurally similar avidin, the dominant role of the TRP residues in the UV

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25–37

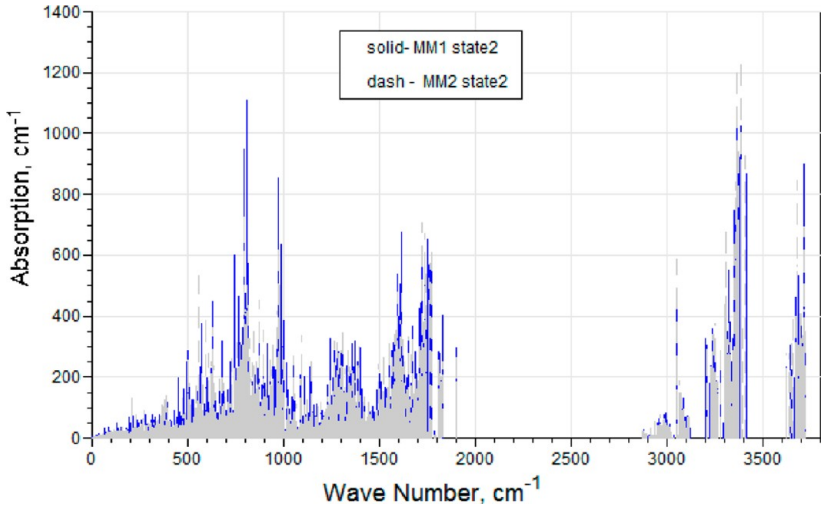


Figure 14. Absorption spectra for **B4S4** in the protonation state 2: solid lines, MM1; dashed lines, MM2.

spectra was deduced from the experimental observations.[27] Also, the conclusions about the importance of the sulfur atom and the carboxyl group for the UV absorption of biotin−streptavidin are consistent with the experimental observations for biotin−avidin.[27] For avidin, the UV absorption was measured at approximately the same energies as predicted for streptavidin, specifically at about 4.15−4.4 eV (280−300 nm) and at higher energies. The fluorescence was measured at 3.67 eV (338 nm) for avidin without biotin and at 3.78 eV (328 nm) for biotin−avidin.[27] The lower energies for the fluorescence imply that the structure relaxes before emitting light.

The TDDFT/MM1 and TDDFT/MM2 simulations were also performed for streptavidin (**S4**) in the protonation state 1, i.e., without biotins. For both types of embedding, the absorption edge for **S4** is predicted to shift to a slightly lower energy as compared to **B4S4** by both DFT/MM models, i.e., to just under 4.7 eV for the mechanical embedding (see Figure 11a and b) and to 4.42−4.54 eV for the electronic embedding (see Figure 11c and d). However, the 4.42 eV line that is obtained for the electronic embedding (see Figure 11c) is very weak. Neglecting this weak line, the DFT/MM1 state 1 results point to a red shift for the UV absorption with biotin by ∼50 meV. This is close to the experimental UV absorption results for avidin where a red shift of about 70 meV was observed in the UV absorption spectra in the presence of biotin. Based on the UV absorption study of avidin with biotin and biotin analogues, that red shift was attributed to biotin's interaction with TRP.[27]

An evidence of the structural changes in the biotin site can be seen in the changes in the predicted optical transition lines when **S4** is compared with **B3S4** that has biotins in the MM partition only (see Figure 12a and b for the mechanical and electronic embedding cases). Therefore, biotin sites in the "empty" streptavidin and in the streptavidin that is partially filled with biotins are not equivalent.

Since the TDDFT simulations with 500 DFT atoms are rather complex, possible model simplifications for light-induced transition simulations were also explored. The goal was to obtain the simplest hybrid model that best reproduces the absorption edge prediction that was obtained for the 500-atom model. First, TDDFT was applied to each separate DFT cluster to find the cluster that gives the lowest UV transition energy. The calculated absorption edges for separate clusters ranged from 4.89 to 5.24 eV. The lowest transition energy belongs to the 87-atom

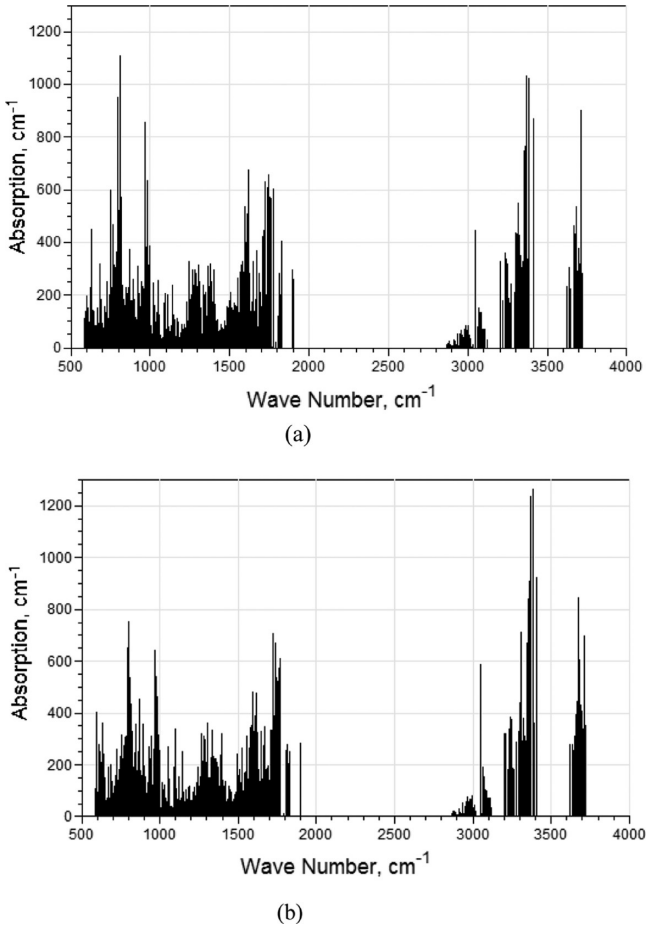


Figure 15. Biotin−streptavidin, protonation state 2. Predicted IR integral line intensities that were obtained with (a) MM1 and (b) MM2.

cluster that contains TRP, leucine (LEU), threonine (THR), and serine (SER) (see Figure 13). It was selected as a new DFT partition in a reduced QM/MM model. All higher transition energy DFT clusters were put into a new MM partition; i.e., they were replaced with point charges. TDDFT/MM simulations with the electronic embedding using this reduced model produced an absorption edge at 4.815 eV, which means that electrostatic effect of the MM part on the wave function of the DFT part have

dx.doi.org/10.1021/jp3075833 | J. Phys. Chem. B 2013, 117, 25−37

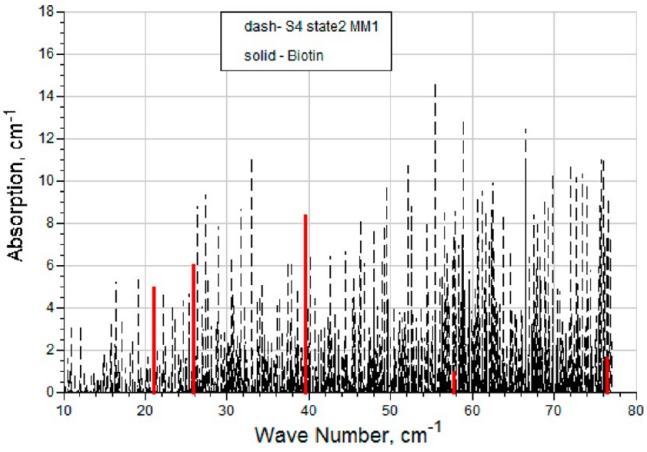


Figure 16. Simulated low-THz absorption lines. Solid lines, biotin; dashed lines, **S4**. Protonation state 2. MM1.

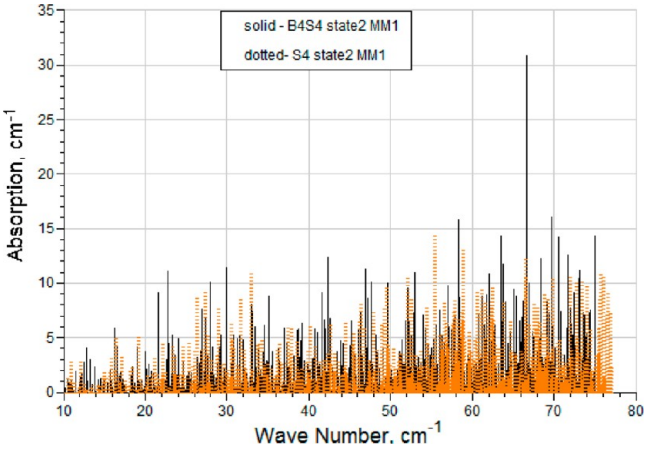


Figure 17. Simulated low-THz absorption lines. Solid lines, **B4S4**; dotted lines, **S4**. Protonation state 2. MM1.

produced a −0.075 eV shift of the absorption edge. The exact result for the absorption edge as calculated with TDDFT on the entire 500-atom partition is 4.73 eV (for **B4S4** in the protonation state 1). Hence, the electronic interactions between the chain fragments within the DFT partition are responsible for introducing a much larger −0.16 eV shift of the absorption edge. These facts prove that the energy shift is not exclusively a result of electrostatic polarization because only about a half of this shift (−0.075 eV) is recovered by the reduced QM/MM model as described above. Therefore, all the identified molecular clusters must be included into the DFT model to obtain more accurate energy transition results. The predicted red shift of −0.16 eV for the TRP based cluster is close to the measured red shift of −0.174 eV for TRP in avidin (based on the 219 nm line of TRP).[27]

VIBRATIONS AND ABSORPTION SPECTRA

Normal mode analysis and light absorption spectra simulations demonstrated a consistent biotin−streptavidin THz-IR absorption fingerprint for the two molecular mechanical force fields used in this study in terms of both line positions and integral line intensities (see Figures 14 and 15a and b). For the structures in the protonation state 2, the IR portions (i.e., over 500 cm^{-1}) of the spectra are plotted in Figure 15a and b for MM1 and MM2, respectively. One can observe that line positions that are generated by the two force fields are very close; however, there are some shifts in the relative balance of integral line intensities between lower and higher frequencies. In particular, MM1 result predicts almost 50% higher intensity around 800 cm^{-1} and 30% higher intensity close to 1000 cm^{-1}. There is a considerable spectral overlap between biotin and streptavidin (**S4**), in particular, at low THz (see Figure 16). However, the influence of biotin on absorption is predicted to be significant and leading to identifiable spectral changes. In Figure 17, these changes can be seen in the low-THz interval, where **B4S4** (solid lines) has stronger intensities beginning as low as 12−13 and 21−23 cm^{-1} and at higher THz frequencies.

CONCLUSIONS

Biotin binding estimates in a low-temperature limit were performed for the reaction **B** + **B3S4** ↔ **B4S4** within the DFT framework both in vacuum and in water using a continuum solution model. It was predicted that the solvation of the biotin−streptavidin complex in water reduces the binding strength by more than half. All studied properties, including the biotin binding and the absorption edge of the biotin−streptavidin complex, are predicted to be protonation state dependent. The UV absorption edge of the complex calculated with TDDFT/MM was found to be virtually insensitive to the choice of the MM force field and strongly dependent on the type of embedding. For **B4S4**, the electronic embedding shifts the predicted absorption edge to a lower energy by 0.07−0.17 eV depending on its protonation state. The predicted UV spectra of streptavidin are dominated by tryptophan and are sensitive to biotin, in particular, to its carboxyl group and sulfur, which is consistent with the experimental data that were obtained previously for the UV absorption and fluorescence in the biotin−avidin complex.[27] Both MM force fields that were used in this work (ff99SB and ff99SBnmr with NMR-based corrections) lead to very similar THz-to-IR spectra for the complex, but there are noticeable differences in relative intensities of lines in different spectral bands, especially around 800 and 1000 cm^{-1}. Despite a considerable spectral overlap between biotin and **S4**, there are many significant changes in the predicted absorption lines at the low-THz frequencies when the two are combined in the complex, which points to the possibility of detecting the presence or absence of biotin in the tetramer using THz spectroscopy measurements.

AUTHOR INFORMATION

Corresponding Author

*Phone (919) 593-4400; fax (919) 515-5523; e-mail abykhov@ncsu.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a U.S. Army Research Laboratory (ARL) award and by a National Research Council (NRC) fellowship.

REFERENCES

(1) Wilchek, M.; Miron, T.; Bayer, E. A. *Chem. Isr.* **2005**, *20*, 23−31.
(2) Trotter, J.; Hamilton, J. A. *Biochemistry* **1966**, *5*, 713−714.
(3) Giuliano, E. *Proteomics* **2008**, *8*, 4012−4024.
(4) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(5) Korter, T. M.; Plusquellic, D. F. *Chem. Phys. Lett.* **2004**, *385*, 45−51.
(6) Elia, G. *Proteomics* **2008**, *8*, 4012−4024.
(7) Norton, M. *Int. J. High Speed Electron. Syst.* **2007**, *17*, 311−326.

(8) Eskelinen, A.-P.; Kuzyk, A.; Kaltiaisenaho, T. K.; Timmermans, M. Y.; Nasibulin, A. G.; Kauppinen, E. I.; Törmä, P. *Small* **2011**, *7*, 746−750.
(9) Rahman, M.; Zhong, H.; Neff, D.; Norton, M. L. Presented at Nanoelectronic Devices for Defense & Security Conference (2011 NANO-DDS), Brooklyn, NY, August 30, 2011; ID 58.
(10) Bykhovski, A.; Woolard, D. *IEEE Trans. Nanotechnol.* **2010**, *9*, 565−574.
(11) Bykhovski, A.; Woolard, D. In *Biological and Medical Sensor Technologies*; Iniewski, K., Ed.; CMOS Emerging Technologies Inc., CRC Press, Taylor & Francis Group: Boca Raton, FL, 2012; pp 39−70.
(12) Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. *J. Phys. Chem. B* **2003**, *107*, 12039−12041.
(13) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
(15) Stevens, W.; Basch, H.; Krauss, J. *J. Chem. Phys.* **1984**, *81*, 6026−6033.
(16) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(17) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **2000**, *113*, 2969−2975.
(18) Li, D.-W.; Brüschweiler, R. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778−6780.
(19) Weber, P. C.; Wendolski, J. J.; Pantoliano, M. W.; Salemme, F. R. *J. Am. Chem. Soc.* **1992**, *114*, 3197−3200.
(20) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(21) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(22) Iyengar, S. S.; Schlegel, H. B.; Millam, J. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2001**, *115*, 10291−10302.
(23) Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1981**, *74*, 4872−4882.
(24) Cerutti, D. S.; Trong, I.; Le; Stenkamp, R. E.; Lybrand, T. P. *J. Phys. Chem. B* **2009**, *113*, 6971−6985.
(25) Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. *Protein Sci.* **1997**, *6*, 1157−1166.
(26) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.
(27) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85−133.