J|A|C|S
ARTICLES
Published on Web 04/07/2007

The Origins of Femtomolar Protein−Ligand Binding: Hydrogen-Bond Cooperativity and Desolvation Energetics in the Biotin−(Strept)Avidin Binding Site

Jason DeChancie and K. N. Houk*

Contribution from the Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095-1569

Received September 27, 2006; E-mail: houk@chem.ucla.edu

Abstract: The unusually strong reversible binding of biotin by avidin and streptavidin has been investigated by density functional and MP2 ab initio quantum mechanical methods. The solvation of biotin by water has also been studied through QM/MM/MC calculations. The ureido moiety of biotin in the bound state hydrogen bonds to five residues, three to the carbonyl oxygen and one for each −NH group. These five hydrogen bonds act cooperatively, leading to stabilization that is larger than the sum of individual hydrogen-bonding energies. The charged aspartate is the key residue that provides the driving force for cooperativity in the hydrogen-bonding network for both avidin and streptavidin by greatly polarizing the urea of biotin. If the residue is removed, the network is disrupted, and the attenuation of the energetic contributions from the neighboring residues results in significant reduction of cooperative interactions. Aspartate is directly hydrogen-bonded with biotin in streptavidin and is one residue removed in avidin. The hydrogen-bonding groups in streptavidin are computed to give larger cooperative hydrogen-bonding effects than avidin. However, the net gain in electrostatic binding energy is predicted to favor the avidin−bicyclic urea complex due to the relatively large penalty for desolvation of the streptavidin binding site (specifically expulsion of bound water molecules). QM/MM/MC calculations involving biotin and the ureido moiety in aqueous solution, featuring PDDG/PM3, show that water interactions with the bicyclic urea are much weaker than (strept)-avidin interactions due to relatively low polarization of the urea group in water.

Introduction

The noncovalent binding association of biotin to avidin ($K_a \cong 10^{15}$ M^{-1})[1] and streptavidin ($K_a \cong 10^{13}$ M^{-1})[2] is the paragon of high-affinity host−guest/supramolecular interactions and has led to many practical applications.[3−8] In contrast to these strong associations, corresponding to femtomolar and tenth picomolar dissociation constants, respectively, noncovalent host−guest binding affinities in water are typically in the range of 10^1−10^6 M^{-1}, and protein−ligand binding rarely exceeds 10^{11} M^{-1}.[9] The biotin−(strept)avidin complexes generally exceed estimates of affinities from empirically determined free energy scoring functions and binding surveys.[10−13] The mechanism by which biotin−(strept)avidin exceeds normal highest protein−ligand affinities by 10^2−10^4 M^{-1} (ref 9) has been a subject of much interest. Despite the extraordinary utility of this complex, which has resulted in a plethora of experimental studies,[14−21] and the uniqueness of the binding, which has led to many structural and computational studies,[22−25] the origin of binding energetics and dynamics of association has not yet been clearly understood.

Protein binding motifs have been identified through crystal structures of protein−ligand complexes,[2,26−31] thermodynamic experiments,[14,17−21,23,32,33] and molecular dynamics/free energy

(1) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85−133.
(2) Weber, P. C.; Wendoloski, J. J.; Pantoliano, M. W.; Salemme, F. R. *J. Am. Chem. Soc.* **1992**, *114*, 3197−3200.
(3) Wilchek, M.; Bayer, E. A. *Methods Enyzmol.* **1990**, *184*, 5−45.
(4) Wilbur, D.; Pathare, P.; Hamlin, D.; Stayton, P.; To, R.; Klumb, L.; Buhler, K.; Vessella, R. *Biomol. Eng.* **1999**, *16*, 113−118.
(5) Hamblett, K. J.; Kegley, B. B.; Hamlin, D. K.; Chyan, M. K.; Hyre, D.; Press, O. W.; Wilbur, D. S.; Stayton, P. S. *Bioconjugate Chem.* **2002**, *13*, 588−598.
(6) Demidov, V. V.; Bukanov, N. O.; Frank-Kamenetskii, D. *Curr. Issues Mol. Biol.* **2000**, *2*, 31−35.
(7) Bontempo, D.; Maynard, H. D. *J. Am. Chem. Soc.* **2005**, *127*, 6508−6509.
(8) Howarth, M.; Chinnapen, D. J.-F.; Gerrow, K.; Dorrestein, P. C.; Grandy, M. R.; Kelleher, N. L.; El-Husseini, A.; Ting, A. Y. *Nature Methods* **2006**, *3*, 267−273.
(9) Zhang, X.; Houk, K. N. *Acc. Chem. Res.* **2005**, *38*, 379−385.
(10) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872−4897.
(11) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−605.
(12) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.
(13) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997−10002.
(14) Chilkoti, A.; Tan, P. H.; Stayton, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1754−1758.
(15) Freitag, S.; Le Trong, I.; Chilkoti, A.; Klumb, L. A.; Stayton, P. S.; Stenkamp, R. E. *J. Mol. Biol.* **1998**, *279*, 211−221.
(16) Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. *Protein Sci.* **1997**, *6*, 1157−1166.
(17) González, M.; Bagatolli, L. A.; Echabe, I.; Arrondo, J. L. R.; Argaraña, C. E.; Cantor, C. R.; Fidelio, G. D. *J. Biol. Chem.* **1997**, *272*, 11288−11294.
(18) Hyre, D.; Le Trong, I.; Freitag, S.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **2000**, *9*, 878−885.
(19) Klumb, L. A.; Chu, V.; Stayton, P. S. *Biochemistry* **1998**, *37*, 7657−7663.
(20) Le Trong, I.; Freitag, S.; Chu, V.; Stayton, P. S.; Stenkamp, R. E. *Acta. Crystallogr.* **2003**, *D59*, 1567−1673.
(21) Swamy, M. J. *Biochem. Mol. Biol. Int.* **1995**, *36*, 219−225.
(22) Dixit, S. B.; Chipot, C. *J. Phys. Chem. A* **2001**, *105*, 9795−9799.
(23) Dixon, R. W.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 471−473.
(24) Miyamoto, S.; Kollman, P. A. *Proteins: Struct., Funct., Genet.* **1993**, *16*, 226−245.
(25) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402−8406.
(26) Rosano, C.; Arosio, P.; Bolognesi, M. *Biomol. Eng.* **1999**, *16*, 5−12.

perturbation simulations.[22−25] These studies yielded three hypotheses for the strong binding affinity: (1) Weber and Salemme and co-workers reported the original crystal structure of biotin−streptavidin and suggested that the hydrogen-bonding interactions to the ureido ring system of biotin dominate the stabilization of the complex. (2) Miyamoto and Kollman contended that the energy provided by van der Waals contacts, primarily of the four tryptophan residues that line the binding pocket, is greater than the electrostatic/hydrogen-bonding free energy benefit. (3) Williams and co-workers proposed that the binding of biotin to streptavidin results in a large free energy benefit due to the strengthening of existing noncovalent interactions within streptavidin.[34]

The quantitative importance of both hydrogen-bonding residues (Weber−Salemme) and tryptophan contacts (Miyamoto−Kollman) to the 18−20 kcal/mol binding free energy has been verified by Stayton, Stenkamp, and co-workers through an elegant series of site-directed mutagenesis experiments.[14,18−20,35] Additionally, biotin−streptavidin binding is accompanied by significant flap motions of the 3−4 loop. This facilitates binding by exclusion of solvent molecules from the binding site and increases the contribution of a number of direct hydrogen bonds and hydrophobic interactions with the valerate moiety. Removal of the 3−4 loop in streptavidin results in a reduction of the K_a by 6 orders of magnitude from wild type, nearly half of the binding energy.[16,36]

Molecular dynamics and separate electrostatic and van der Waals free energy calculations of the biotin−streptavidin complex suggested that the largest contribution to the extremely negative free energy of binding is the nonpolar van der Waals contribution of the tryptophan residues Trp79, Trp92, Trp108, and Trp120 rather than electrostatic forces.[24,25] The simulations also predict a large electrostatic contribution inside the protein, supported by site-directed mutagenesis experiments studies on streptavidin, which show that mutation of a residue directly hydrogen-bonded to the ureido moiety of biotin results in a substantial decrease in binding energy.[14,18−20] However, Kollman et al. proposed that this electrostatic stabilization is similar to that in the water solvent, so the net electrostatic contribution to binding of the ureido group is small.[24,25] Therefore, it was contended that the hydrogen-bonding residues merely form a weak recognition pocket and give only small net free energy gain from water to protein.

The hydrogen-bonding interactions involved in biotin−(strept)avidin binding and biotin in aqueous solution have been re-evaluated using density functional and ab initio QM and combined QM/MM statistical mechanical calculations.

Computational Methodology

QM. Model structures were optimized using density functional theory with B3LYP/6-31+G(d,p)[37−39] and MPWB1K/6-31+G(d,p) functionals/basis set. MPWB1K, developed by Truhlar and co-workers,[40] based on the modified Perdew and Wang 1991 functional[41] (MPW) and Becke's 1995 meta correlation functional,[42] gave excellent overall results in recent benchmarks for nonbonded interactions.[40,43] B3LYP has been shown to produce accurate orientations and nonbonding distances compared to protein structures,[44] and B3LYP and MP2 have been shown to produce realistic hydrogen-bond energies for a series of model complexes involving ligands and protein backbone and side-chain residues.[45] Default convergence criteria in Gaussian 03 were met except for the rms displacements for a few structures (see Supporting Information). The failure of rms displacements to converge is a result of the extremely flat potential surface of the hydrogen-bonding complexes for these large models. In such occurrences, changes in electronic energy across several steps were only 10^{-7} hartrees (6×10^{-5} kcal/mol). For all structures, the optimization step-sizes were decreased in order to facilitate location of the stationary points. All model protein systems were characterized by frequency calculations at the level of theory corresponding to the optimized structures. To determine atomic charges, the CHelpG algorithm, developed by Breneman and Wiberg,[46] and natural population analysis (NPA)[47−50] have been used.

MP2/6-31+G(d,p) single-point evaluations[51] were performed on the density functional optimized geometries to better estimate the magnitudes of weak intermolecular interactions. Correction for basis set superposition error (BSSE) was performed using the counterpoise (CP) method of Boys and Bernardi.[52−56]

CPCM-SCRF[57−60] single-point calculations on DFT geometries utilizing the UAKS cavity size at the HF/6-31+G(d,p) level were carried out on the biotin-residue or biotin-water stationary points for calculation of solvation energies. This method has been shown to be the most accurate standard cavity model for aqueous solvation.[61] Free energies of solvation were then added to MP2/6-31+G(d,p) energies. A dielectric constant of 4.33 was used to simulate the environment of multiple tryptophan residues and deeply buried hydrogen-bonding residues

(27) Pugliese, L.; Coda, A.; Malcovati, M.; Bolognesi, M. *J. Mol. Biol.* **1993**, *231*, 698−710.
(28) Hendrickson, W. A.; Pahler, A.; Smith, J. L.; Satow, Y.; Meritt, Y. S.; Phizackerley, R. P. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2190−2194.
(29) Livnah, O.; Bayer, E. A.; Wilchek, M.; Sussman, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5076−5080.
(30) Freitag, S.; Le Trong, I.; Klumb, L.; Stayton, P. S.; Stenkamp, R. E. *Protein Sci.* **1997**, *6*, 1157−1166.
(31) Pahler, A.; Hendrickson, W. A.; Gawinowicz Kolks, M. A.; Argaraña, C. E.; Cantor, C. R. *J. Biol. Chem.* **1987**, *262*, 13933−13937.
(32) Marttila, A. T.; Hytönen, V. P.; Laitinen, O. H.; Bayer, E. A.; Wilchek, M.; Kulomaa, M. S. *Biochem. J.* **2003**, *369*, 249−254.
(33) Stayton, P. S.; Freitag, S.; Klumb, L. A.; Chilkoti, A.; Chu, V.; Penzotti, J. E.; To, R.; Hyre, D.; Le Trong, I.; Lybrand, T. P.; Stenkamp, R. E. *Biomol. Eng.* **1999**, *16*, 39−44.
(34) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596−6616.
(35) Freitag, S.; Chu, V.; Penzotti, J. E.; Klumb, L. A.; To, R.; Hyre, D.; Le Trong, I.; Lybrand, T. P.; Stenkamp, R. E.; Stayton, P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8384−8389.
(36) Chu, V.; Freitag, S.; Le Trong, I.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **1998**, *7*, 848−859.
(37) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(38) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(40) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.
(41) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11.
(42) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040−1046.
(43) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415−432.
(44) Morozov, S. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6946−6951.
(45) Hao, M.-H. *J. Chem. Theory Comput.* **2006**, *2*, 863−872.
(46) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(47) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1988**, *83*, 735−746.
(48) Wiberg, K. B.; Rablen, P. R. *J. Comput. Chem.* **1993**, *14*, 1504−1518.
(49) McAllister, M. A.; Tidwell, T. T. *J. Org. Chem.* **1994**, *59*, 4506−4515.
(50) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.
(51) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618−622.
(52) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873−1885.
(53) Gutowski, M.; van Duijneveldt-van de Rijdt, J. G. C. M. *J. Chem. Phys.* **1993**, *98*, 4728−4737.
(54) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F. *J. Chem. Phys.* **1986**, *84*, 2279−2289.
(55) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418−2426.
(56) Szalewicz, K.; Cole, S. J.; Kolos, W.; Bartlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3662−3673.
(57) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(58) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312−9320.
(59) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(60) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.
(61) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70−77.

encompassing the binding site.[62] A dielectric constant of 78.39 was used for water. All reported B3LYP and MPWB1K structures, MP2 energies, CPCM solvation energies, and CP corrections were obtained using Gaussian 03.[63]

QM/MM. Combined quantum mechanical and molecular mechanics (QM/MM) calculations,[64,65] as implemented in BOSS 4.6,[66,67] were performed to analyze the properties and energetics of biotin in aqueous solution. The solute was treated with the PDDG/PM3 method,[68,69] which has been extensively tested for gas-phase structures and energetics[68,69] and has successfully been applied to solution-phase QM/MM studies of S_N2,[70] S_NAr,[71] Kemp[72] and biotin[73] decarboxylations, and Cope eliminations.[74] The solvent molecules were represented by the TIP4P water model.[75] The Metropolis Monte Carlo (MC) simulations were performed in a periodic box of 750 (minus the number of non-hydrogen atoms of the solute) TIP4P water molecules at 25 °C and 1 atm in the NPT ensemble.[76] Each simulation consisted of 3.2 million configurations of equilibration and 20 million configurations of averaging. The solute energy and energy changes are treated quantum mechanically using PDDG/PM3, and computation of the QM energy and atomic charges was performed for every solute move. The partial charges were obtained from the CM3 charge model,[77] and the PDDG/PM3 wavefunction was unscaled for negatively charged solutes or scaled by 1.14 for neutral charged solutes.[78] Solute−solvent and solvent−solvent intermolecular cutoff distances of 12 Å were employed.

MD. Molecular dynamics were performed on a monomer of unliganded avidin (PDB 1AVE)[27] and the unliganded streptavidin tetramer (PDB 1SWC)[16] using AMBER 8.[79] The crystal structure for unliganded streptavidin contains two subunits adopting an open conformation for the 3−4 loop. The related loop in unliganded avidin, however, remains in the closed conformation for each subunit. From the protein crystal structures, hydrogen atoms and a cubic box of TIP3P[75] water molecules extending 10 Å from the solute were added using the xLEaP module of AMBER. Crystal waters were removed. The Duan et al. (ff03) force field[80] was employed to describe the proteins, and long-range electrostatic forces were treated using the particle mesh Ewald (PME) method.[81,82] Trajectories were integrated with a 1 fs time step. All simulations were carried out under constant temperature (300 K) and pressure (1 atm). The avidin system was equilibrated for 70 ps and the streptavidin for 133 ps. The SHAKE algorithm[83] was used for all hydrogens with a nonbonding pair cutoff of 10 Å. Average structures were generated using the PTRAJ utility for the unliganded avidin over 60−70 ps and for unliganded streptavidin over 123−133 ps.

Results and Discussion

The binding energy of each hydrogen-bonding residue in the (strept)avidin binding site that interacts with the ureido of biotin was evaluated through the application of two model systems. The first model treats the hydrogen-bonding residues as separate entities, and the latter mimics the protein environment where all residues are interacting with the ligand in concert. Binding energies obtained for the isolated residue−biotin complexes are referred to as isolated residue binding energies (IRBEs), and those derived from the protein binding site models are labeled binding site residue binding energies (BSRBEs). The aim of this differentiation is to explore the energetic effect of the hydrogen-bonding array in the protein binding site upon each particular residue−biotin interaction. This is determined quantitatively by the difference in IRBE and BSRBE for a designated residue. The resultant quantity is defined as the hydrogen-bond cooperativity involving specific hydrogen-bonding interactions of one biotin to one subunit. Following the analysis of biotin−(strept)avidin hydrogen-bonding interactions, the binding energetics of biotin- and unliganded (strept)avidin−water complexes are presented.

Isolated Hydrogen-Bonding Residues. Biotin bound to (strept)avidin accepts three hydrogen bonds to the ureido carbonyl and one to each −NH group. The optimized isolated residue−biotin complexes at the B3LYP/6-31+G(d,p) level are shown in Figure 1. Hydrogen-bond complexes, **1**−**7**, represent the five hydrogen bonds that are made with the ureido of biotin in either the streptavidin or avidin binding site. Methanol is used as the mimic for a serine residue, phenol for tyrosine, acetate for aspartate, and formamide for an asparagine residue. These residue models have been used successfully for the investigation of mechanism and stereoselectivity of enzyme-catalyzed reactions.[84,85]

The fully optimized protein binding site models for streptavidin and avidin shown in Figure 2 served as a starting point for the calculation of the isolated complexes. Each individual hydrogen-bonding residue to the ureido moiety was constrained to match the optimized angle and dihedral angle in the streptavidin model, and the remaining variables were allowed to relax. For example, the initial geometry of the Tyr−OH···ureido hydrogen-bond complex was obtained by removing all other hydrogen-bonding residues from the optimized streptavidin−biotin model (Figure 2a); the dihedral angle C═O-(ureido)···O−C(Tyr) was constrained, and the remaining variables were optimized, including the hydrogen-bonding length. All other IR complexes were obtained in an analogous manner. These restraints caused structures **2**, **3**, and **7** not to be true minima in the gas phase as indicated by frequency analysis. The very small imaginary frequency (ca. -20 cm^{-1}) in these structures corresponds to an in-plane bending motion of the residue to form an additional hydrogen bond to the carbonyl oxygen or −NH group of biotin. However, for comparison with the hydrogen-bond energies obtained for the (strept)avidin

(62) Simonson, T.; Brooks, C. L., III. *J. Am. Chem. Soc.* **1996**, *118*, 8452−8458.
(63) Frisch, M. J.; et al. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
(64) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(65) Kaminshi, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787−1796.
(66) Jorgensen, W. L. *BOSS*, version 4.6; Yale University: New Haven, CT, 2004.
(67) Jorgensen, W. L.; Tirado-Rives, J. *J. Comput. Chem.* **2005**, *26*, 1689−1700.
(68) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.
(69) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L.; Brauman, J. I. *J. Am. Chem. Soc.* **2003**, *125*, 138−150.
(70) Vayner, G.; Houk, K. N.; Jorgensen, W. L.; Brauman, J. I. *J. Am. Chem. Soc.* **2004**, *126*, 9054−9058.
(71) Acevedo, O.; Jorgensen, W. L. *Org. Lett.* **2004**, *6*, 2881−2884.
(72) Acevedo, O.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2005**, *127*, 8829−8834.
(73) Acevedo, O.; Jorgensen, W. L. *J. Org. Chem.* **2006**, *71*, 4896−4902.
(74) Acevedo, O.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2006**, *128*, 6141−6146.
(75) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(76) Jorgensen, W. L. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: New York, 1998; Vol. 3, pp 1754−1763.
(77) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2003**, *24*, 1291−1304.
(78) Udier-Blagović, M.; Morales De Tirado, P.; Pearlman, S. A.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 1322−1332.
(79) Case, D. A.; et al. *Amber 8*; University of California: San Francisco, CA, 2004.
(80) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(81) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(82) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(83) van Gusteren, W. F.; Berendesen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.
(84) Tantillo, D. J.; Chen, J.; Houk, K. N. *Curr. Opin. Chem. Biol.* **1998**, *2*, 743−750.
(85) Cannizzaro, C. E.; Ashley, J. A.; Janda, K. D.; Houk, K. N. *J. Am. Chem. Soc.* **2003**, *125*, 2489−2506.

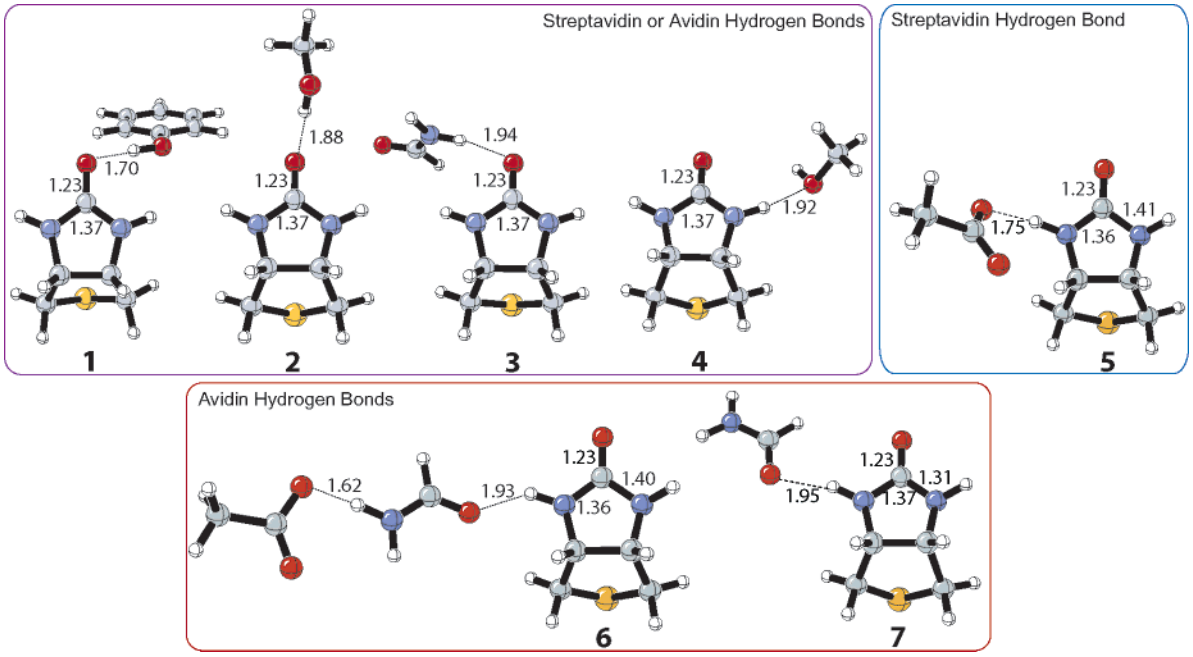

Figure 1. Optimized structures of the isolated hydrogen-bonding residues for streptavidin and avidin model binding sites at the B3LYP/6-31+G(d,p) level.

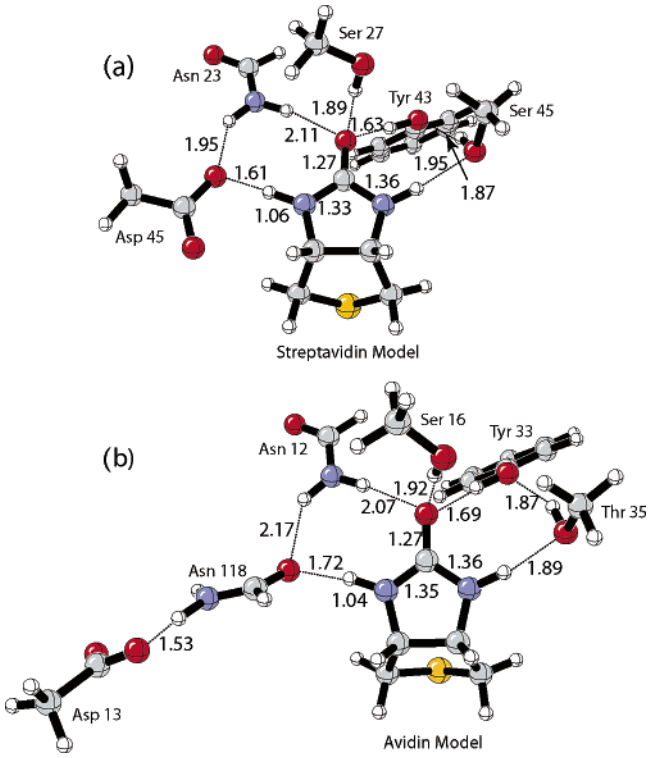

Figure 2. B3LYP/6-31+G(d,p) optimized binding site models: (a) streptavidin; (b) avidin.

complex (Figure 2), these dihedral constraints were necessary in order to maintain only a single hydrogen bond to biotin.

Table 1 gives IRBEs for complexes, **1**−**7**, involving isolated hydrogen-bonding residues and the bicyclic heterocycle of biotin. The IRBEs for the isolated complexes, computed at MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p) and MP2/6-31+G(d,p)//MPWB1K/6-31+G(d,p) levels of theory, show excellent agreement in the gas-phase (Table 1, column 2) and CPCM model (column 3). The gas-phase binding energies are quite high, especially for complexes involving a charged residue. The

Table 1. Isolated Residue Binding Energies (IRBEs) in kcal/mol for Streptavidin and Avidin

isolated complex[a]	$IRBE_{vac}$[b]	$IRBE_{CPCM}$[c]
	streptavidin or avidin	
Tyr43/33 **1**	−12.9 (−11.8)	−5.3 (−5.2)
Ser27/16 **2**	−8.0 (−7.8)	−4.7 (−3.0)
Asn23/12 **3**	−9.6 (−9.7)	−2.7 (−2.4)
Ser45/Thr35 **4**	−8.9 (−9.5)	−3.5 (−2.6)
	streptavidin	
Asp128 **5**	−28.5 (−28.9)	−5.3 (−5.2)
	avidin	
Asn118−Asp13 pair **6**	−16.7	−3.6
Asn118 **7**	−9.6	−2.6

[a] See Figure 2 for residue identification. [b] MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p); MP2/6-31+G(d,p)//MPWB1K/6-31+G(d,p) in parentheses. [c] CPCM/HF/6-31+G(d,p) level corrections to MP2 energies using the dielectric of ether ($\epsilon = 4.33$).

reported values in column 3, involving CPCM solvation corrections to mimic the interior of a protein, are more in line with the expected values for such hydrogen-bonding complexes. Hydrogen bonds by the tyrosine or serine models to the urea carbonyl oxygen, as well as the urea NH bond to the aspartate, are the strongest (∼5 kcal/mol), while the NH···O hydrogen bonds (**4**, **6**, **7**) are all about 3 kcal/mol, in agreement with expected magnitudes.[86] Our discussion concentrates on quantities given by the CPCM model in column 3.

Streptavidin Model. The model systems for the streptavidin and avidin binding sites consist of the ureido of biotin in the presence of the key first- and second-shell hydrogen-bonding residues. The optimized streptavidin−biotin structure at the B3LYP/6-31+G(d,p) level of theory is shown in Figure 2a. The corresponding MPWB1K/6-31+G(d,p) optimized structure is given in Supporting Information.

(86) Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. *Nature* **1985**, *314*, 235−238.
(87) Hyre, D. E.; Le Trong, I.; Merritt, E. A.; Eccleston, J. F.; Green, N. M.; Stenkamp, R. E.; Stayton, P. S. *Protein Sci.* **2006**, *15*, 459−467.

Table 2. Binding Site Residue Binding Energies (BSRBEs) in kcal/mol for Hydrogen-Bonding Residues of the Streptavidin Model and Binding Enthalpy Changes from Site-Directed Mutagenesis Studies

streptavidin residues	$BSRBE_{vac}$[a]	$BSRBE_{CPCM}$[b]	$\Delta\Delta H_{WT-mutant}$ (25 °C)[c]
Tyr43	−19.9 (−20.6)	−10.2 (−11.1)	Y43A = 8.9 ± 1.4[19,20]
Ser27	−12.9 (−12.6)	−4.1 (−3.4)	S27A = 1.6 ± 0.7[19,20]
Asn23	−2.7 (−2.7)	−2.9 (−3.5)	N23A = 6.9 ± 0.7[19,20]
Asp128	−38.2 (−55.7)	−12.3 (−11.6)	D128A = 5.9[87]
Ser45	−11.4 (−16.8)	−5.7 (−5.6)	S45A = 6.1[87]

[a] MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p); MP2/6-31+G(d,p)//MPWB1K/6-31+G(d,p) in parentheses. [b] CPCM/HF/6-31+G(d,p) level corrections to MP2 energies using the dielectric of ether ($\epsilon = 4.33$). [c] Binding energy change upon mutation of wild type to mutant streptavidin (kcal/mol); see refs 19, 20, and 87.

Calculated hydrogen-bonding energies for each of the five residues to the ureido moiety in the streptavidin model at the MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p) and MP2/6-31+G(d,p)//MPWB1K/6-31+G(d,p) levels are reported in Table 2. The calculated individual hydrogen-bonding energy of each residue in the binding site was obtained in the following manner: starting from the optimized model shown in Figure 2a, the residue of interest was removed, and a single-point calculation of the energy of the structure with the missing residue was computed. The energy change resulting from removal of a hydrogen-bonding residue is defined as the residue deletion energy (RDE) (see Supporting Information). This energy involves interaction energies of the residue with biotin and other neighboring groups. In the streptavidin model, Asp128 makes a hydrogen bond to Asn23 and Tyr43 hydrogen bonds with the adjacent Ser45 (Figure 2). The energy of interaction of each residue of the binding site with biotin only is defined as the binding site residue binding energy (BSRBE). The BSRBEs are estimated by subtracting the energy calculated for hydrogen bonding of the residue with neighboring non-biotin residues from the RDE (eq 1).

$$\mathrm{BSRBE} = \mathrm{RDE} - \Delta E_{\mathrm{non-biotin\_residue\_complex}} \qquad (1)$$

For example, the BSRBE for the Tyr43 in streptavidin is computed by subtracting the energy of the Tyr43−Ser45 complex from the RDE_{Tyr}. The BSRBEs for all five hydrogen-bonding residues for the streptavidin model, defined as the energy of interaction between each residue model and the bicyclic heterocycle of biotin, are given in Table 2.

The calculated BSRBEs of column 3 can be compared to experimental site-directed mutagenesis studies reported in column 4. The $\Delta\Delta H_{WT-mutant}$ quantity reported corresponds to the change in enthalpy of binding between wild-type streptavidin and the corresponding mutant. The computed BSRBEs are analogous to residue X-to-null mutations, so perfect agreement of columns 3 and 4 is not expected. A direct comparison between these two energies is most appropriate when the mutant structure shows minimal disruption of the hydrogen-bonding residues and water does not replace the missing side-chain atoms. This situation is applicable for the S45A,[18] Y43A,[20] and N23A.[20]

In S45A, biotin shifts toward Asp128, but only an average variation of 0.1 Å is observed in the hydrogen-bond lengths to the ligand. The calculated $BSRBE_{CPCM}$ for Ser45, −5.7 kcal/mol, is very close to the experimental value, 6.1 kcal/mol (Table 2, column 3).[18]

Y43A results in a 180° rotation of the neighboring Phe29 to fill the void created by deletion of the large tyrosine side chain.[20] Yet, as in the case of S45A, hydrogen-bonding lengths to biotin are moderately perturbed, ±0.2 Å.[20] The experimental value, 8.9 kcal/mol, corresponding to a loss of hydrogen-bonding functionality, is comparable to the $BSRBE_{CPCM}$, 10 kcal/mol (Table 2, column 3).[19]

The $BSRBE_{CPCM}$ for N23A is lower than the observed value by 4 kcal/mol.[20] However, deletion of this residue results in the loss of hydrogen-bonding interactions to Asp128 side-chain and the backbone NH functionalities of Leu25, Gly26, and Ser27.[2] These interactions are expected to be quite favorable as a single NH···O hydrogen bond in formamide dimers has been reported to be approximately 5 kcal/mol in the gas phase at the MP2/aug-cc-pVDZ level,[88] rivaling the hydrogen bond with the urea carbonyl of biotin, −2.7 kcal/mol (Table 2, column 2). As a consequence, the mutated unbound protein structurally alters to compensate for the removal of this residue.[20] In contrast, the biotin-bound mutant structure displays little deviation compared to the native. It is likely that the destabilization of the mutant unliganded protein is a significant factor contributing to the relatively higher binding enthalpy obtained from the mutation experiment.

The overestimations of BSRBE for Asp128 and Ser27 (Table 2, column 3) compared to experiment are reasonable due to structural changes and substitution of water for the mutated residue.[20,35] In D128A, biotin shifts away from the side chain of alanine and, presumably, strengthens interactions with residues of the opposing side of the binding pocket.[35] In addition, a water molecule is observed within 0.5 Å of the Asp128 carboxylate in the wild type.[35] Similarly, the water observed in the S27A mutant crystal structure resides in the position occupied by the side chain in the wild type.[20] Therefore, the computational binding enthalpies of Asp128/Ser27−biotin interactions are likely to be higher than those obtained from mutation experiments, particularly due to the water−biotin hydrogen bonds gained in the mutants.

Overall, the calculated BSRBEs in streptavidin are quite large, in agreement with experimental mutagenesis studies, in particular for the Tyr43, Asp128, and Ser45 models, which give among the largest energy changes reported for removing single hydrogen-bonding residues.[18,19,35]

Avidin Model. The optimized avidin model is shown in Figure 2b. There are two important differences between the streptavidin and avidin models: in avidin, (1) Asn118 replaces Asp128; (2) the charged aspartic acid residue (Asp13) is not in direct contact with the ligand but is in the second shell, as opposed to Asp128 in streptavidin.

The calculated BSRBEs for the avidin model are shown in Table 3. These can be compared to IRBEs, given in Table l. The trends are similar to the streptavidin results (Table 2), and we concentrate on column 3, the CPCM corrected values, in the following discussion. The hydrogen-bonding energies for Tyr33 and Asn12, which form hydrogen bonds to the carbonyl of the ureido moiety, and the Thr35 hydrogen bond to the −NH group remain almost identical between the streptavidin (Table 2, column 3) and avidin models (Table 3, column 3). Only a slight difference is observed for the streptavidin Ser27 hydrogen bond to the carbonyl of the bicyclic urea, which is calculated

(88) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 4963−4968.

Table 3. Binding Site Residue Binding Energies (BSRBEs) in kcal/mol for the Hydrogen-Bonding Residues of the Avidin and Avidin with Asp13 Removed Models

protein residues	$BSRBE_{vac}$[a]	$BSRBE_{CPCM}$[b]
avidin		
Tyr33	−18.1	−9.4
Ser16	−11.3	−3.2
Asn12	−4.9	−3.1
Asn118−Asp13 pair	−20.5	−9.0
Thr35	−13.1	−6.0
avidin with Asp13 removed		
Tyr33	−15.3	−8.1
Ser16	−9.8	−2.7
Asn12	−8.3	−1.4
Asn118	−5.0	−3.5
Thr35	−13.5	−6.7

[a] MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p). [b] CPCM/HF/6-31+G(d,p) level corrections to MP2 energies using the dielectric of ether ($\epsilon = 4.33$).

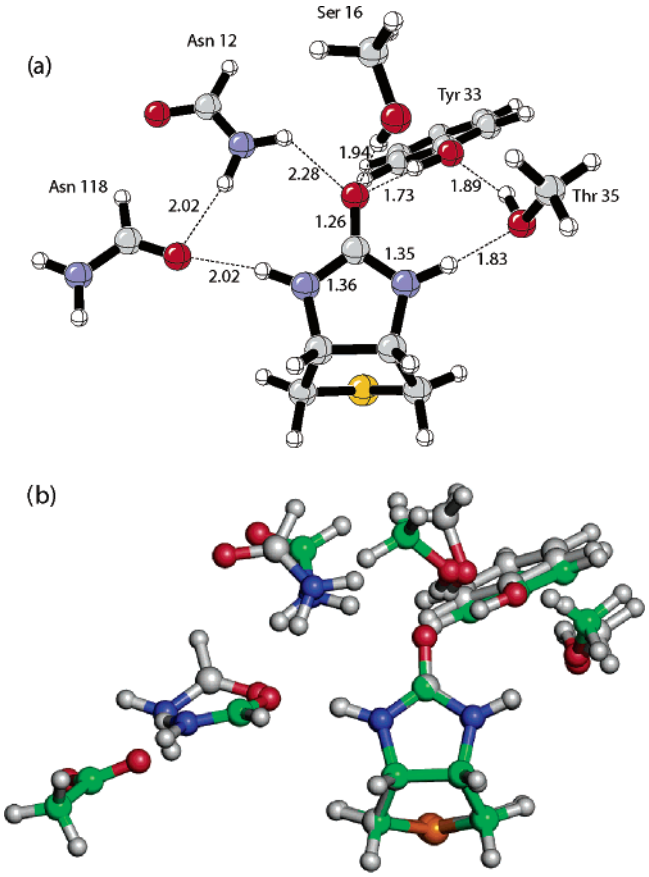

Figure 3. (a) B3LYP/6-31+G(d,p) optimized avidin model with Asp13 removed. (b) Superposition of the urea heavy atoms in avidin (green carbon atoms) and avidin with Asp13 removed (gray carbon atoms) models.

to be approximately 1 kcal/mol higher than that for Ser16 in the avidin model. The greatest difference, however, is observed in the interaction energy for the Asp13−Asn116−NH(ureido) hydrogen bond, −9.0 kcal/mol, compared to −12.3 kcal/mol for the Asp128−NH(ureido) hydrogen bond in the streptavidin model. It is apparent that the magnitude of the hydrogen-bonding interaction is quite sensitive to the position of the aspartic acid.

Avidin Model (Asp13 Removed). To further assess the influence of the aspartate residue on the binding energy, the avidin model with the Asp13 removed was optimized with B3LYP (Figure 3a). The geometrical changes resulting from the deletion of Asp13 in avidin are displayed by the superposition of the bicyclic urea bound to either of the avidin models (Figure 3b). The most significant deviations occur with Asn118 and Asn12 residues, while Tyr33, Ser16, and Thr35 remain relatively unperturbed. To understand the binding energetics of this model structure, the calculated BSRBEs are reported in Table 3. As expected, the largest reductions in binding energy results from Asn118 of 5.5 kcal/mol, which is now devoid of the second-shell aspartate hydrogen bond, and Asn12 of 1.7 kcal/mol (Table 3, column 3). Small decreases in BSRBE are predicted for Tyr33 and Ser16 of 1.3 and 0.4 kcal/mol, respectively.

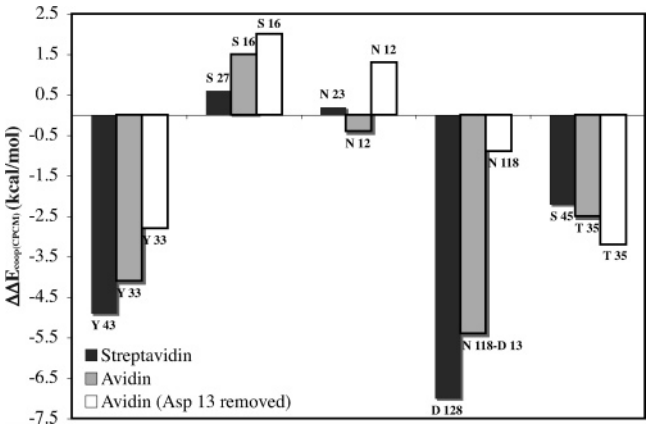

Figure 4. Computed hydrogen-bond cooperativity (IRBE − BSRBE) for streptavidin, avidin, and avidin with Asp13 removed models. Negative values correspond to increased (cooperative) hydrogen bonding, while positive values correspond to negative (anti-cooperative) hydrogen bonding. Values correspond to CPCM model ($\epsilon = 4.33$).

Another important consequence of the removal of the aspartate in avidin is an increase in hydrogen-bond lengths for the three residues surrounding the ureido carbonyl ranging from 0.05 to 0.17 Å and the significant lengthening of the Asn118 hydrogen bond to the −NH group by 0.3 Å. Thr35 hydrogen bond to the opposing −NH group is shortened by 0.06 Å, resulting in the only, albeit modest, increase in BSRBE compared to avidin of 0.7 kcal/mol (Table 3, column 3).

(Strept)Avidin Hydrogen-Bond Cooperativity. The difference in energy between the IRBE (Table 1) and BSRBE (Tables 2 and 3) values is defined as the cooperative hydrogen-bonding energy ($\Delta\Delta E_{coop-x}$) (eq 2).

$$\Delta\Delta E_{coop} = BSRBE_x - IRBE_x \qquad (2)$$

Cooperativity is the sum in total binding energy versus the sum of individual hydrogen bonds. The cooperative hydrogen-bonding energies with the CPCM model for all of the residues in the streptavidin and avidin binding sites to the bicyclic urea are shown in Figure 4. Gas-phase and MPWB1K cooperative hydrogen-bonding energies parallel the results shown in Figure 4 and are given in the Supporting Information.

The cooperative hydrogen-bonding energy, $\Delta\Delta E_{coop}$, is large for the hydrogen-bond acceptors, Asp128 and Ser45, interacting with the urea NHs, and for the strongest hydrogen bond donor, Tyr43, to the carbonyl oxygen. While this work was in progress, Stenkamp and Stayton and co-workers characterized the S45A/D128A double mutant in streptavidin. The free energy change ($\Delta\Delta G_{298}$) between wild type and the double mutant was 1.4 kcal/mol higher than the sum of the $\Delta\Delta G_{298}$ values for S45A and D128A single mutations.[87] Our results support that non-

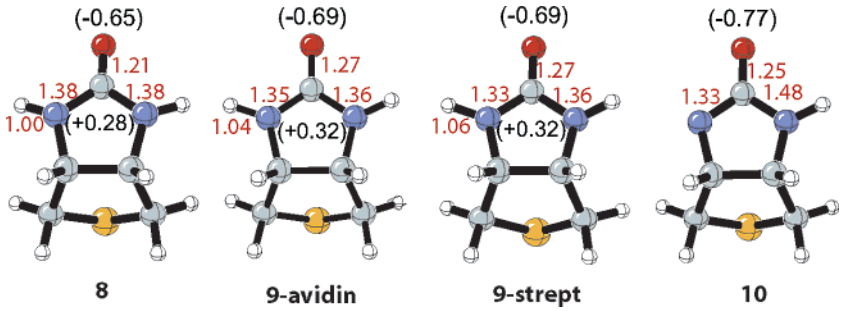

Figure 5. Comparison of bond lengths (Å) and charges (NPA) of isolated bicyclic urea unit of biotin (**8**), the urea group when bound in the (strept)avidin model binding sites (**9**), and the conjugate base (**10**). Structures were optimized by B3LYP/6-31+G(d,p).

additivity in this case is due to the large hydrogen-bond cooperativity associated with these two residues.

The hydrogen-bond cooperativity values ($\Delta\Delta E_{coop}$) for avidin and avidin with Asp13 removed are also reported in Figure 4. Streptavidin has ~4 kcal/mol larger cooperative binding energy than avidin. More than half of the difference in cooperative hydrogen-bond energy is due to the more favorable BSRBE of the aspartate directly interacting with the bicyclic urea in streptavidin.

When the Asp13 is removed from avidin, $\Delta\Delta E_{coop(CPCM)}$ is significantly reduced by 7 kcal/mol. The remaining neutral residues still provide cooperative hydrogen-bonding energy of −3.6 kcal/mol. However, a decrease in cooperativity is observed for all residues except for the Thr35 hydrogen bond to the −NH group of the bicyclic urea (Figure 4). The larger cooperative hydrogen-bond energy for this residue of −0.7 kcal/mol is likely due to strengthened interactions reflected in a shorter hydrogen-bond length (Figure 3). The Ser16 and Asn12 hydrogen bonds to the carbonyl of the biotin become significantly anti-cooperative at approximately +1.5 kcal/mol. Although some cooperativity is maintained with only neutral residues, the aspartate residue is critical for the large non-additive binding energies ($\Delta\Delta E_{coop(CPCM)}$) predicted for avidin and streptavidin of −11 and −14 kcal/mol, respectively.

Origin of Hydrogen-Bond Cooperativity. Cooperative interactions are important for the stability of many hydrogen-bonded systems.[89−94] Predictions of hydrogen-bond cooperativity involving systems containing urea functionality have been reported in previous theoretical work.[95−97] Masunov and Dannenberg applied ab initio and density functional theory calculations to a contiguous chain of urea molecules (Figure 5a) and demonstrated that hydrogen bonding became stronger as the number of molecules increased, due to mutual polarization of the ureas as a function of chain length.[95]

The origin of biotin−(strept)avidin hydrogen-bond cooperativity is similar; the induced polarization of the urea function of biotin causes strong interactions with the first and second contact-shell hydrogen-bonding residues. The strongly polarized urea allows interactions to work in a cooperative (non-additive) way.[98] This mechanism to achieve hydrogen-bond cooperativity has been shown to enhance binding in other hydrogen-bond complexes, such as peptides and proteins,[95,99−113] nucleic acid pairs,[114] β-diketone fragments,[115] and carboxylic acid dimers.[116,117]

A polarized urea function resulting from hydrogen bonds was first reported by Blessing through analysis of more than 100 crystal structures.[118] This study revealed that with increasing strong hydrogen bonding to the urea oxygen there is a progressive lengthening of the C−O bond and corresponding shortening of the C−N bond. The computed structure of biotin surrounded by the hydrogen-bonding residues in the (strept)-avidin binding site is in agreement with this observation. Figure 5 shows optimized geometries for the bicyclic urea **8**, the bicyclic urea bound in (strept)avidin **9** (Figure 2), and deprotonated bicyclic urea, **10**. The C−O bond of the urea increases from 1.21 to 1.27 Å upon binding both the (strept)avidin models, **9**. Yet, the C−N bond decreases from 1.38 to 1.33 Å for streptavidin (**9-strept**), while the avidin bicyclic urea (**9-avidin**) is not as dramatically perturbed. This difference in geometry is likely manifested in the differing $\Delta\Delta E_{coop}$ between the streptavidin and avidin models, as shown in Figure 4. It is important to note that the C−O bond length change of 0.06 Å is quite dramatic since the C−O bond length of 1.25 Å in the conjugate base, **10**, is actually shorter. The augmented negative charge to

(89) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.
(90) Scheiner, S. *Hydrogen Bonding*; Oxford University Press: New York, 1997.
(91) Guo, H.; Sirois, S.; Proynov, E. I.; Salahub, D. R. In *Theoretical Treatments of Hydrogen Bonding*; Hadzi, D., Ed.; John Wiley & Sons: New York, 1997; Chapter 3.
(92) Çarçabal, P.; Jockusch, R. A.; Hünig, I.; Snoek, L. C. T. K. R.; Davis, B. G.; Gamblin, D. P.; Compagnon, I.; Oomens, J.; Simons, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 11414−11425.
(93) Williams, D. H.; Maguire, A. J.; Tsuzuki, W.; Westwell, M. S. *Science* **1998**, *280*, 711−714.
(94) Suhai, S. *J. Quantum Chem.* **1994**, *52*, 395−412.
(95) Masunov, A.; Dannenberg, J. J. *J. Phys. Chem. B* **2000**, *104*, 806−810.
(96) Dovesi, R.; Causa', M.; Orlando, R.; Roetti, C. *J. Chem. Phys.* **1990**, *92*, 7402−7411.
(97) Belosludov, R. V.; Li, Z.-Q.; Kawazoe, Y. *Mol. Eng.* **1999**, *8*, 105−120.
(98) Steiner, T. *Chem. Commun.* **1997**, 727−734.
(99) Ludwig, R. *J. Mol. Liq.* **2000**, *84*, 65−75.
(100) Ludwig, R.; Reis, O.; Winter, R.; Weinhold, F.; Farrar, T. C. *J. Phys. Chem. B* **1998**, *102*, 9312−9318.
(101) Ludwig, R.; Weinhold, F.; Farrar, T. C. *J. Phys. Chem. A* **1997**, *101*, 8861−8870.
(102) Kobko, N.; Paraskevas, L.; del Rio, E.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2001**, *123*, 4348−4349.
(103) Guo, H.; Gresh, N.; Roques, B. P.; Salahub, D. R. *J. Phys. Chem. B* **2000**, *104*, 9746−9754.
(104) Guo, H.; Salahub, D. R. *Angew. Chem., Int. Ed.* **1998**, *37*, 2985−2990.
(105) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104−7105.
(106) Zhao, Y.-L.; Wu, Y.-D. *J. Am. Chem. Soc.* **2002**, *124*, 1570−1571.
(107) Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125*, 8124−8129.
(108) Moisan, S.; Dannenberg, J. J. *J. Phys. Chem. B* **2003**, *107*, 12842−12846.
(109) Salvador, P.; Kobko, N.; Wieczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2004**, *126*, 14190−14197.
(110) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2004**, *108*, 9205−9212.
(111) Sheridan, R. P.; Lee, R. H.; Peters, N.; Allen, L. C. *Biopolymers* **1979**, *18*, 2451−2458.
(112) Van Duijnen, P. T.; Thole, B. T. *Biopolymers* **1982**, *21*, 1749−1761.
(113) Morozov, A. V.; Tsemekhman, K.; Baker, D. *J. Phys. Chem. B* **2006**, *110*, 4503−4505.
(114) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441−6443.
(115) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989**, *111*, 1023−1028.
(116) Subramanian, K.; Lakshmi, S.; Rajagopalan, K.; Koellner, G.; Steiner, T. *J. Mol. Struct.* **1996**, *384*, 121−126.
(117) Turi, L.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1994**, *116*, 8714−8721.
(118) Blessing, R. H. *J. Am. Chem. Soc.* **1983**, *105*, 2776−2783.

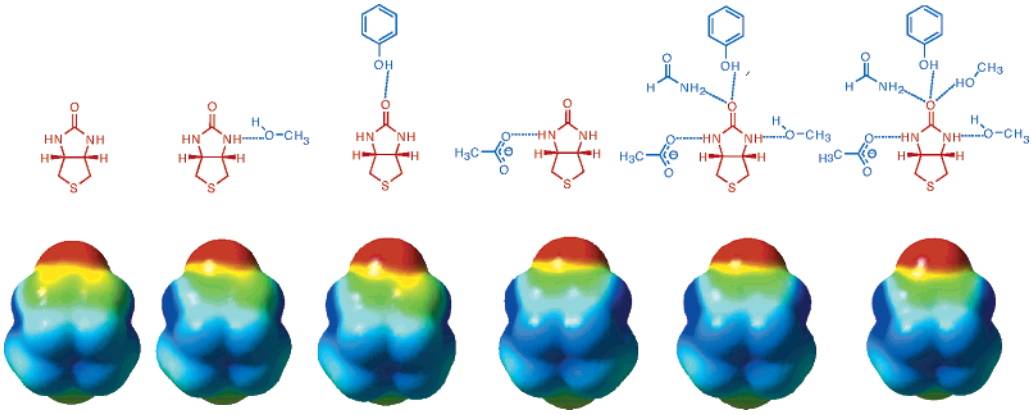

Figure 6. Electrostatic potential surfaces of bicyclic urea from B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d,p) calculations. From left (isolated ureido moiety) to right (streptavidin model), there is an increase in red (oxygen becoming more negative) and an increase in blue surface area (nitrogens becoming more positive).

the carbonyl oxygen and positive charge to the remaining urea atoms mirror the geometrical perturbations from the isolated to bound (strept)avidin bicyclic urea (Figure 5).

Electrostatic potential surfaces in various hydrogen-bonding complexes are shown in Figure 6. This shows a gradual increase in electron density on the carbonyl and a gradual decrease of electron density on the nitrogen and carbon atoms of the urea functionality with growing number of hydrogen-bond donors and acceptors to the biotin model. The electrostatic potential of the bicyclic urea hydrogen-bonded to acetate (aspartic acid model) is nearly identical to the streptavidin model system (far right) supporting the critical role of a charged species in generating significant polarization of the ureido group of biotin.

The evidence for significant polarization of the ureido moiety in Figure 6 is reflected in the calculated dipole moment of the ureido moiety in various cases. The computed dipole moment of the bicyclic molecule is exceptionally large (15.72 D) in the geometry it has in the streptavidin model as compared to 5.40 D of the unperturbed ureido structure (MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p)).[119] The hydrogen-bonding residues in streptavidin interacting with the ureido moiety lead to a nearly 300% increase in dipole moment. The structural relationship between the three residues and a polarized ureido oxygen is indeed reminiscent of an oxyanion hole, an important component of catalysis in proteases[120] and esterases,[121] as initially postulated by Weber et al.[2]

Biotin in Aqueous Solution. The strong oxyanion hole-like binding interaction in streptavidin should be compared to solvation of biotin and the predicted binding site in water. To obtain information about aqueous solvation, QM/MM/MC calculations were applied to various electronic structures of the bicyclic urea of biotin and multiple conformations of D-(+)-biotin using the TIP4P potential for water and the semiempirical PDDG/PM3 for the solutes.[68,69]

Five separate simulations have been carried for various bicyclic urea ground state solutes: (1) the bicyclic urea from the streptavidin model (Figure 2a) with internal coordinates fixed at those values (**9-con-strept**), (2) the bicyclic urea from the avidin model (Figure 2b) with internal coordinates fixed at those values (**9-con-avidin**), (3) the bicyclic urea geometry optimized in the gas phase by B3LYP/6-31+G(d,p) with internal coordinates of the solute fixed at those values (**8-con**), (4) the urea optimized by PDDG/PM3 in water (**8-opt**), (5) the bicyclic urea conjugate base optimized by PDDG/PM3 in the liquid phase (**10-opt**).

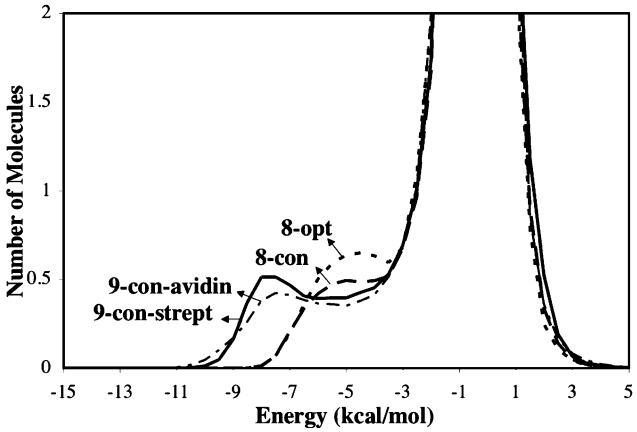

Figure 7. Solute−solvent energy pair distributions for different structures of the ureido moiety. The ordinate records the number of solvent molecules that interact with the solute with interaction energies on the abscissa. Units for the ordinate are number of molecules per kcal/mol.

Solute−solvent energy pair distributions for the simulations involving the various geometries of the ureido moiety, **9-con-strept**, **9-con-avidin**, **8-con**, and **8-opt**, are shown in Figure 7. The solute−solvent energy pair distributions record the average number of solvent molecules that interact with the solute with an associated energy. The most favorable hydrogen-bonding interactions are reflected in the left-most region, with energies more attractive than −3 kcal/mol. The large bands near 0 kcal/mol result from the many distant water molecules in outer shells.

The outstanding feature of Figure 7 is the presence of waters with especially large energies of interaction for the **9-con** polarized structures compared to **8-con** and **8-opt**, arising from stronger hydrogen bonds with the solute. The differing distribution of **8-con** and **8-opt** compared to those of the **9-con** structures suggests that water does not strongly polarize the solute unlike the (strept)avidin binding sites. The solvation of the water-perturbed bicyclic urea is thus not particularly unusual.

The coordination of water to the carbonyl oxygen of the bicyclic urea is better quantified by radial distribution functions (rdfs). Shown in Figure 8a are the rdfs for the four structures

(119) The dipole for the bicyclic urea in the streptavidin model binding site is derived by subtracting the dipole vector contributions of the surrounding hydrogen-bonding residues from the total dipole moment.
(120) Menard, R.; Storer, A. C. *Biol. Chem. Hoppe-Seyler* **1992**, *373*, 393−400.
(121) Nachon, F.; Asojo, O. A.; Borgstahl, G. E.; Masson, P.; Lockridge, O. *Biochemistry* **2005**, *44*, 1154−1162.

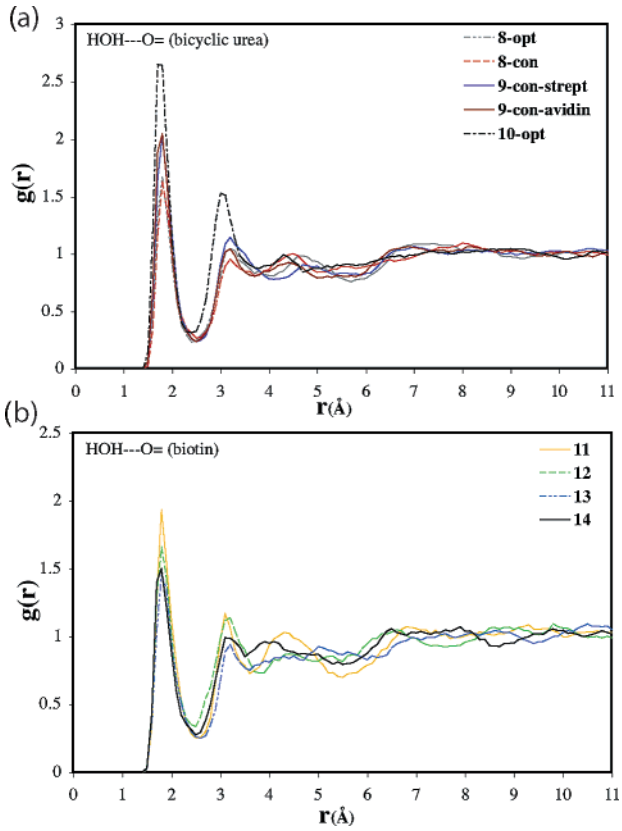

Figure 8. Computed HOH···O═ radial distribution functions for different structures of (a) bicyclic urea and (b) D-(+)-biotin.

discussed above. The number of contact waters to the carbonyl oxygen is determined by integrating the first peak of the rdf from 1 to 2.5 Å. The polarized structures of **9-con-avidin** and **9-con-strept** involve 2.6 and 2.7 hydrogen bonds to the carbonyl, respectively. While **8-con** and **8-opt** involve only 2.3 or 2.2 hydrogen bonds to water, respectively. On average, the bicyclic urea forms two hydrogen bonds between the carbonyl oxygen and water, but the artificially polarized bicyclic ureas can form up to three hydrogen bonds to its carbonyl, more like an alkoxide. The **9-con-avidin** is slightly less polarized than **9-con-strept**, which results in a slightly smaller probability to form a three hydrogen-bonded system. The sensitivity of solvent hydrogen bonds to the ureido carbonyl oxygen was also found in a recent study by Acevedo and Jorgensen on the aqueous solvent effects upon of biotin decarboxylation.[73] The ureido conjugate base carbonyl oxygen **10-opt** occupies 5.0 hydrogen bonds. The similar geometries between the **9-con** bicyclic ureas and the conjugate base (Figure 5) do not translate to comparable ranges of hydrogen-bond accepting ability of the carbonyl oxygen. Nevertheless, the rdfs support that the bicyclic urea is uniquely polarized when bound to the (strept)avidin binding sites (Figure 8a).

The critical factors for hydrogen-bond cooperativity in the (strept)avidin binding sites are the charged aspartate residues present in each biotin binding site. Biotin has been shown to adopt multiple conformations in the gas phase[122] and in aqueous solution.[123] On possible conformation is an intramolecular hydrogen bond involving the near −NH group and the carboxylate tail **11** shown in Figure 9. However, this conformation must compete with other conformers, such as folded **12**, semi-folded **13**, and water-mediated intramolecular H-bond **14**. These four representative biotin conformers were initially optimized in the gas phase then subject to QM/MM/MC simulations to gain insight into the energetics in water.

Shown in Figure 8b are the rdfs for each of the biotin conformers. Similar radial distribution functions around the carbonyl oxygen of the ureido moiety are predicted for the intramolecular hydrogen-bond conformer of biotin and the polarized structures, **9-con**. This is reasonable since the ureido moiety in **11** is polarized by the intramolecular hydrogen bond involving the carboxylate tail of biotin. As expected, integration yields 2.6 water molecules hydrogen-bonded to the carbonyl oxygen for conformer **11** compared to 2.2−2.3 waters for conformers **12**−**14**. However, the computed relative free energies of solvation, $\Delta\Delta G_{solv}$, for the various conformations of biotin reported in Table 4 show the intramolecular hydrogen-bond conformer will not contribute significantly to the overall ensemble of solvated geometries.[123] The relative biotin solvation energies strongly disfavor conformation **11** compared to **12**−**14** (Table 4). The energy gained by the intramolecular hydrogen bond in **11** does not exceed the penalty for disrupting the apparently more favorable solvation of the carboxylate by water. This is in agreement with work by Li and co-workers where the intramolecular hydrogen-bonded conformation was observed only fleetingly during a 15 ns simulation trajectory of neutral biotin.[123] The predicted most favorable conformations, **12** and **14**, are in qualitative agreement with NMR experiments that predict a compact structure of biotin in aqueous solution.[124,125]

To assess the net binding energy in water, the ureido group with four coordinating waters (two to the carbonyl oxygen and one to each −NH group), in accordance with QM/MM/MC predictions, was optimized at the B3LYP/6-31+G(d,p) level (see Supporting Information). The energies for each water molecule interacting with the ureido of biotin were calculated in an analogous fashion to BSRBEs using MP2/6-31+G(d,p)//B3LYP/6-31+G(d,p) level with CPCM corrections using a dielectric of water (ϵ = 78.39). A total binding energy of −14.2 kcal/mol was calculated for the four water molecules hydrogen-bonded to the ureido of biotin. This is considerably lower than the total BSRBE of streptavidin (−35.2 kcal/mol) and avidin (−30.7 kcal/mol) reported in Tables 2 and 3, column 3. Furthermore, the calculated dipole moment on the ureido moiety with four coordinating water molecules is only 9.27 D, while the induced dipole moment of the ureido in the streptavidin model is 15.72 D. Biotin in aqueous solution is polarized much less than biotin in the (strept)avidin binding sites, and the hydrogen bonds to the ureido group are much stronger when bound to the protein.

Solvation of Unliganded (Strept)Avidin: Why Avidin is a Better Binder than Streptavidin. The total BSRBEs for streptavidin or avidin minus −14.5 kcal/mol for biotin−water interactions provide estimates of the binding energies for the protein−ligand complex. Free energy contributions from residues providing favorable hydrophobic/packing interactions have been shown by Lazaridis et al. to be nearly identical in streptavidin and avidin.[126] Thus, the BSRBE quantities predict

(122) Strzelczyk, A. A.; Dobrowolski, J. C.; Mazurek, A. P. *J. Mol. Struct. (THEOCHEM)* **2001**, *541*, 283−290.
(123) Lei, Y.; Li, H.; Zhang, R.; Han, S. *J. Phys. Chem. B* **2004**, *108*, 10131−10137.
(124) Fry, D. C.; Fox, T. L.; Lane, M. D.; Mildvan, A. S. *J. Am. Chem. Soc.* **1985**, *107*, 7659−7665.
(125) Tonan, K.; Adachi, K.; Ikawa, S. *Spectrochim. Acta, Part A* **1998**, *54*, 989−997.

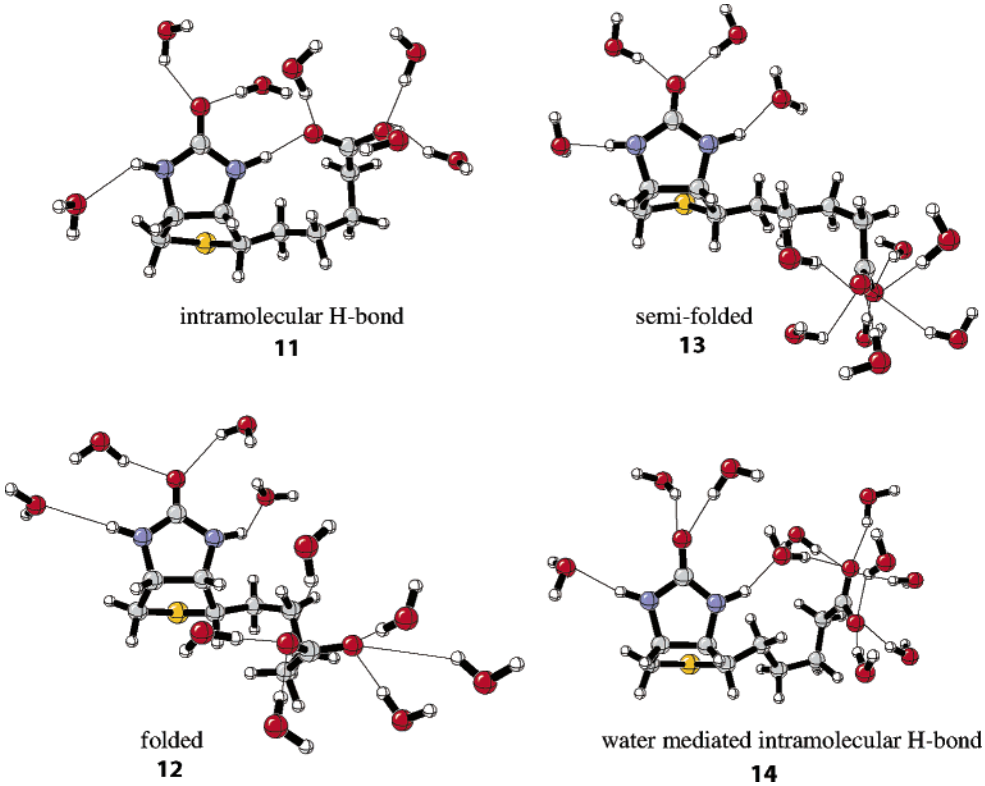

Figure 9. Representative conformations of biotin in aqueous phase from QM/MM/MC simulations.

Table 4. Relative Free Energy of Solvation, $\Delta\Delta G_{solv}$ (kcal/mol), D-(+)-Biotin Conformations from QM/MM/MC Simulations

solute	E_{QM} (solute)	E_{MM} (solute–solvent van der Waals)	$E_{QM/MM}$ (solute–solvent electrostatic)	$\Delta\Delta G_{solv}$[a]
11	−154.5	−12.7	−177.7	+37.3 ± 2.2
12	−141.2	−9.3	−202.3	+2.8 ± 2.0
13	−149.4	−10.9	−193.9	+17.6 ± 1.2
14	−141.8	−9.0	−205.9	0.0

[a] $\Delta G_{solv} = (E_{solute-solvent\ electrostatic} + E_{solute-solvent\ van\ der\ Waals}) - E_{QM}$. $\Delta\Delta G_{solv} = \Delta G_{solv(\mathbf{11},\mathbf{12},\mathbf{13}\ or\ \mathbf{14})} - \Delta G_{solv(\mathbf{14})}$.

that streptavidin is the better binder, while experiment has shown the free energy of binding is 2−3 kcal/mol better with avidin.[1,21] However, the interaction of bound water molecules with the hydrogen-bonding residues in the unliganded protein needs to be taken into account.

In the absence of a ligand bound to (strept)avidin, the binding site is occupied by water.[2,26] The solvation energy of the streptavidin and avidin binding sites were obtained by determining the number and geometric placement of occupying waters that are interacting with the bicyclic urea hydrogen-bonding residues using classical molecular dynamic simulations (see Computational Methodology). Average structures for unliganded (strept)avidin are shown in Supporting Information.

The generally proposed open conformation of the 3−4 loop in unliganded streptavidin[16] is predicted to facilitate an influx of water molecules to line the binding cavity. Molecular dynamics show that for the subunits with an open conformation, B and D, there are five waters: two interacting with the oxyanion hole residues (Ser27, Asn23, Tyr45), three waters hydrogen-bonded to Asp128 with two hydrogen bonding to the inner-binding-site oxygen and one to the outer, exposed to bulk solvent. Additionally, there are two waters interacting with the side chain of Ser45, a member of the 3−4 loop. In subunit D, however, a disordered rotamer conformation of Ser45 is observed, such that it is not arranged to bind biotin, unlike in subunit B. Therefore, we choose to focus on subunit B where the hydrogen-bonding residues are best prearranged to bind biotin. The placement and number of waters is consistent with crystallographic studies[2] and previous calculations on unliganded streptavidin,[24] which predict five to six waters in the binding site.

Contrary to streptavidin, the 3−4 in avidin remains in a closed conformation in absence of biotin.[27] Dynamics studies show that unliganded avidin contains two water molecules hydrogen-bonded to each other. One of the waters is found in the deepest part of the binding cavity bound to Asn118 and the oxyanion hole side-chain residue atoms of Ser16 and Tyr33 and loosely to Asn12. The other water is hydrogen-bonded to the side chain of Thr35. This structural depiction for the solvated binding site is in excellent accord with deductions from the crystal structure.[27]

B3LYP/6-31+G(d,p) optimizations were performed starting from the computed average atomic coordinates of the (strept)-avidin binding sites and bound waters, as shown in Figure 10. Due to dramatic distortions resulting from unconstrained optimizations, constraints were designed to impose the geometric integrity of the protein scaffold to locate realistic minima corresponding to nature. Fixing interatomic distances between the residue side-chain and intraresidue side-chain angles and dihedral angles and then optimizing the remaining degrees of freedom determined the streptavidin−7H_2O and avidin−2H_2O complexes.

(126) Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 194−208.
(127) Gonçalves, P. F. B.; Stassen, H. *Pure Appl. Chem.* **2004**, *76*, 231−240.
(128) Korndörfer, I. P.; Skerra, A. *Protein Sci.* **2002**, *11*, 883−893.

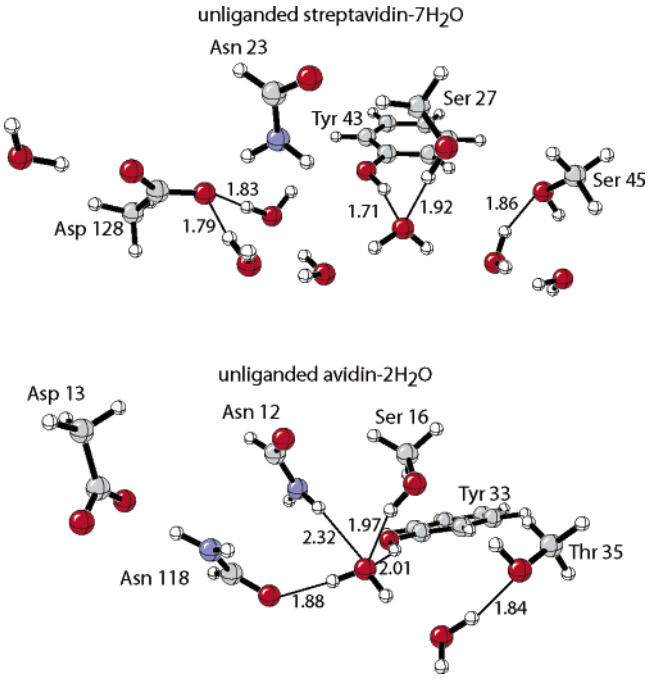

Figure 10. Model protein binding sites for streptavidin and avidin in complex with water at the B3LYP/6-31+G(d,p) level.

The binding energies of two and seven water molecules in the avidin and streptavidin binding sites were determined from the energy to transfer the H_2O molecules from the unliganded (strept)avidin$-H_2O$ complexes into bulk aqueous solution. The resulting binding energies include energetic contributions from the hydrogen-bonding interactions between the individual water molecules and the protein side chains. To account for the energy corresponding to the H_2O-H_2O hydrogen bonds in the protein, the binding energy of the water complex was subtracted from this energy difference. The binding energy of the unliganded (strept)avidin$-n H_2O$ complexes, $\Delta E_{\text{protein-water\_binding}}$, was calculated by eq 3.

$$\Delta E_{\text{protein-water\_binding}} = [E_{\text{unliganded\_(strept)avidin-}n\text{H}_2\text{O\_complex}} - (E_{n\text{H}_2\text{O(aq)}} + E_{\text{unliganded\_(strept)avidin}})] - \Delta E_{\text{H}_2\text{O\_cluster}} \quad (3)$$

Only six waters were transferred out of the streptavidin binding sites, in accord with the crystal structure (PDB 1STP), which shows that the water molecule hydrogen-bonded to the outside oxygen of Asp128 remains bound with biotin. The energy corresponding to the transfer of a water molecule into bulk water is taken as the experimental value of the free energy of solvation for water, -6.3 kcal/mol.[127] Different dielectric constants ($\epsilon = 4.33, 78.39$) were applied to the CPCM-SCRF treatment of the unliganded (strept)avidin$-n H_2O$ complexes due to the uncertainty of the dielectric of the unliganded protein binding pocket. Nevertheless, the binding energy ($\Delta E_{\text{protein-water\_binding}}$) for streptavidin remains approximately 8 kcal/mol larger than that for avidin. Hydrogen-bonding distances shown in Figure 10 indicate that the larger desolvation energies for the streptavidin binding site are derived from the stronger H_2O-Asp128/Tyr 43 hydrogen bonds compared to the analogous sites in avidin.

Further sampling may yield alterative configurations for the water binding site interactions, in addition to the consideration of other crystal structures of unliganded streptavidin.[128] However, it is clear that the key factors giving rise to the large number of occupying waters in streptavidin are the first-shell aspartate residue, which renders the binding pocket more polar, and the open conformation of the 3−4 loop. Avidin will bind biotin more strongly than streptavidin because the placement of Asp13 reduces the large desolvation penalty resulting from a first-shell charged residue and provides good hydrogen-bond cooperativity via juxtaposition to Asn118. The evolutionary driving force for avidin to bind biotin with greater affinity is not achieved by strengthening interactions with the ligand, but by weakening binding interactions with water.

Conclusion

Quantitative investigation of the binding interactions within the streptavidin and avidin binding sites demonstrates that the hydrogen-bonding residues induce a polarized electronic structure of the ureido moiety of biotin, which causes the noncovalent hydrogen-bonding interactions with the ligand to be unusually strong. This relationship between the residues and ligand creates a cooperative effect: the binding energy inside the protein is larger than expected based on the sum of individual hydrogen bonds. The biotin−(strept)avidin complexes that form strong stabilizing oxyanion holes provide superior hydrogen-bonding interactions compared to the electrostatic interactions in aqueous solution. Streptavidin is predicted to have larger cooperative hydrogen-bonding energy compared to that of avidin, but the large inherent binding energy is counteracted by the higher desolvation energy of streptavidin.

Acknowledgment. We are grateful to the National Science Foundation (CHE-0548209) and DARPA for financial support of this research. This research was also facilitated through the Partnerships for Advanced Computational Infrastructure (PACI) through the support of the National Science Foundation. The computations were performed on the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputing Center (PSC) and on the UCLA Academic Technology Services (ATS) Hoffman Beowulf cluster. J.D. would like to acknowledge the National Institutes of Health, UCLA Chemistry−Biology Interface training program, for financial support through a traineeship.

Supporting Information Available: Energies and Cartesian coordinates of all structures calculated at B3LYP/6-31+G(d,p) and MPWB1K/6-31+G(d,p) including counterpoise corrections, solute−solvent energy pair distributions for biotin, MD average structures of unliganded (strept)avidin, and complete refs 63 and 79. This material is available free of charge via the Internet at http://pubs.acs.org.

JA066950N