

Charge Distribution and Hydrogen Bonding of a Collagen α_2 -Chain in Vacuum, Hydrated, Neutral, and Charged Structural Models

Jay Eifler,^[a] Rudi Podgornik,^[b,c,d] Nicole F. Steinmetz,^[e,f,g,h] Roger H. French,^[g]
V. Adrian Parsegian,^[d] and Wai-Yim Ching*^[a]

A challenging task in computational biophysics is to ascertain the solvent effect on the electronic structure and interatomic bonding at the atomistic level. Simulations must be carried out on reasonably large biomolecules for accurate calculations to yield valid results. We report the results of a calculation on collagen model in the form of a peptide under three different environments: vacuum, solvated and with neutral and charged sites. Quantitative results and analysis of the partial charge (PC) distribution on each amino acid are discussed. A significant charge transfer of more than 1 electron from protein to water molecules is found with similar results when the model contains charged sites. The main contributions to the interatomic bonding are from hydrogen bonds (HBs) between

water-water and water-protein pairs. A connection between PC and HBs can be established since the nonpolar amino acids form no HBs and have the smallest PC and vice versa. The *ab initio* PC obtained are used in the NAMD simulation showing significant improvement over the default values as reflected in the root mean square deviation of atomic positions in the MD steps and the total free energy in energy minimization. These results could facilitate the interpretation of data on interaction of various ligands in charged proteins in relation to isoelectric points. © 2016 Wiley Periodicals, Inc.

DOI: 10.1002/qua.25089

Introduction

In recent years, computational biophysics has started to focus on understanding the structure and interactions of biomolecules at the atomistic level by using a combination of quantum mechanical calculations and classical molecular dynamics (MD) simulations.^[1–3] Although *ab initio* calculations can provide more accurate results on energy levels, charge distributions, and inter-atomic bonding that are generally not feasible for large proteins because of severe computational limitations. A viable approach is to use more efficient *ab initio* methods on systems of intermediate complexity, and then use these results to feed back into MD to increase the accuracy of simulations on much larger systems. However, detailed information for large biomolecular systems is very limited and research focuses mostly on either small fragments of a large molecular structure or well-known structural subunits.^[4] Moreover, most relevant biomolecular systems are always bathed in the aqueous environments, further exacerbating the complexity of computational challenges.^[5] To address such issues, knowledge of accurate partial charge (PC) distributions on protein molecules is essential for determining the electrostatic interactions [including hydrogen bonding (HB)]. The landscape started to change recently due to the development of more accurate and efficient methods.^[6–8]

In proteins, amino acids on the surface are in close proximity to water molecules. Charged surface sites are considered to be the most important due to their ability to interact with various ligands and water molecules. Surface amino acids have charged sites depending on the number of hydrogen atoms

bonded to the terminal side-chains. The protonation of an amino acid is related to the pH value of the aqueous solvent, and the acid dissociation constant, or pKa of the side-group.^[9] The pKa is proportional to the dissociation energy G , that is,

[a] J. Eifler, Wai-Yim Ching
Department of Physics and Astronomy, University of Missouri–Kansas City,
Missouri 64110
E-mail: ChingW@umkc.edu

[b] R. Podgornik
Department of Physics, University of Massachusetts, Amherst,
Massachusetts 01003

[c] R. Podgornik
Department of Theoretical Physics, J. Stefan Institute, Ljubljana SI-1000,
Slovenia

[d] R. Podgornik, V. A. Parsegian
Department of Physics Faculty of Mathematics and Physics, University
of Ljubljana, SI-1000, Ljubljana Slovenia

[e] N. F. Steinmetz
Department of Biomedical Engineering, Case Western Reserve University,
Cleveland, Ohio 44106

[f] N. F. Steinmetz
Department of Radiology, Case Western Reserve University, Cleveland, Ohio
44106

[g] N. F. Steinmetz, R. H. French
Department of Materials Science and Engineering, Case Western Reserve
University, Cleveland, Ohio 44106

[h] N. F. Steinmetz
Department of Macromolecular Science and Engineering, Case Western
Reserve University, Cleveland, Ohio, 44106

Contract grant sponsor: Division of Materials Science and Engineering, US
DOE-Office of BES; contract grant numbers: DE-SC008176 and DE-SC8068.
Contract grant sponsor: Office of Science of DOE; contract grant number:
DE-AC03-76SF00098.

© 2016 Wiley Periodicals, Inc.

$pK_a = -\log K_a = -\log(\exp(-\Delta G)/RT)$. It is these charged surface sites, which predominately control the surface charge distribution of the protein.^[10] However, more recent evidence suggests that amino acids other than the charged sites on the surface can also influence the overall charge distribution of the protein^[11] and could significantly change the local environment. Specifically, polar amino acids, which have more electronegative atoms, may play a pivotal role. To avoid any confusion, we follow the generally accepted practice of dividing amino acids into three groups^[12]: polar, nonpolar, and charged. Polar amino acids contain electronegative atom(s) either O or N in their side-group. They prefer to be in close contact with water, or hydrophilic. Nonpolar amino acids do not contain electronegative atom(s) and consist of only C and H, and do not prefer to be in close contact with water molecules. Charged amino acids are polar amino acids with electronegative atoms in their side-groups but they can also become charged by attaching or removing protons (H atoms). The charge differences between these groups, or the charge transfer among them is mediated by water and influenced by other nearby amino acids in the local environment. It is highly desirable that a detailed information on the charge distribution of a protein due to interactions of polar and charged sites be thoroughly investigated rather than merely assigning an electron charge (positive or negative) to the charged amino acids as routinely done in the literature. In this article, we present the results of such a study which provide information on PC, interatomic bonding, especially the HB, in a specific model of a typical α_2 -chain of collagen.

Collagen is one of the most important structural proteins in connective tissues; it is also a structural component in tissues such as bone, teeth and the cornea in animals.^[13] No complete atomic-level structural model has yet been established for the collagen molecules using methods such as crystallographic X-ray diffraction or NMR, due to the difficulty of extracting the whole fibrous protein, except for a rather small segment.^[14–20] However, the collagen molecule is known to be approximately 300 nm long, and made up of three individual chains twisted together into a triple-helical rope-like aggregate, though some controversy still exist on the details of this collagen model.^[21] The homotrimer form of the collagen triple-helix consists of three α_1 -chains, while the heterotrimer form consists of two α_1 -chains and one α_2 -chain with a slightly different amino acid composition. A unique feature of collagen molecules is that every third residue in the amino acid sequence is Gly, and there is also a preponderance of Pro which can be modified post-translationally to Hyp.^[22]

Here, we report on a detailed computational study on the α_2 -chain model of collagen as a representative peptide to study the effect of solvation and protonation/deprotonation on some of its side chains. This α_2 -chain model is similar to the one in a previous study, which has a triplet of two α_1 -chains and one α_2 -chain in the development of a computational scheme, the so-called Amino Acid Potential Method (AAPM).^[23] In this work, we start with this model to examine collagen α_2 -chain as a representative protein of reasonable size in three different environments and modifications, namely in vacuum, the solvated form with water molecules, and a model with modifications of certain charged sites due to

addition or removal of a proton at the site. It should be emphasized in the beginning that the goal of this paper is NOT to investigate the detailed structures of the collagen molecule and its various biomedical or biophysical properties and their implications on health related issues. Our goal is to use a protein model of reasonable size and recognition for the computational development that can be applied to other more complex biomolecular systems. For each model under different environments, we calculate the PC for all atoms in the protein and its constituent amino acids. Moreover, we investigate the bonding between atomic pairs, especially the HB under different environments. We then compare the results of the three models to assess the effect of solvation and protonation/deprotonation on PC distribution and HB. We present quantitative results from the atomic level up, and examine the effect of the charged sites on other sites in relation to polar, charged, and nonpolar amino acids. These *ab initio* quantum mechanically derived PC results are used as input into the Nanoscale Molecular Dynamics (NAMD) to ascertain any improvement on the accuracy of the simulation for the stability and dynamics of the suggested structural model.^[24–26] Such results may eventually be used to compare with experiments in identifying possible location of the isoelectric point (IEP).^[27] The IEP experiments measure mobility of the protein under the applied electric field in a medium as part of the long range electrostatic interaction, not the overall global protein charge. Hence, different experimental groups usually disagree on the IEP values for a protein. Thus, more accurate estimates of the actual charge distribution on the protein could help in analyzing the IEP results.

The layout of the article is as follows. In Section “Three models for the collagen α_2 -chain” we briefly describe the construction of the three structural models used in the calculations. In Section “Computational Methods and Steps,” the computational methods and steps used are described. In the main Section, Section “Results and Discussion,” the calculated results on these three models are compared and discussed. We end with some conclusions and our vision for future work with emphasis on the importance of using explicit solvation for biomolecular systems.

Three Models for the Collagen α_2 -Chain

There are many studies on different structure models for collagen.^[28–33] The three models based on the α_2 -chain of collagen used in the present work has the same origin as used in our earlier study.^[23,34] We stress that this model is used solely for the investigation of a typical protein under different environments, not for any specific properties of collagen per se. Here we used NAMD as the starting tool^[35] to build the three α_2 -chain models with increasing complexity. Figure 1 illustrates all the three models used in this study.

Model 1: nonhydrated collagen model

The purpose of model 1 is to first establish the distribution of PC and interatomic bonding in a dry molecule (peptide) in vacuum and to be used as a base-line in comparing with results of model 2 to assess the effect of solvation on PC and

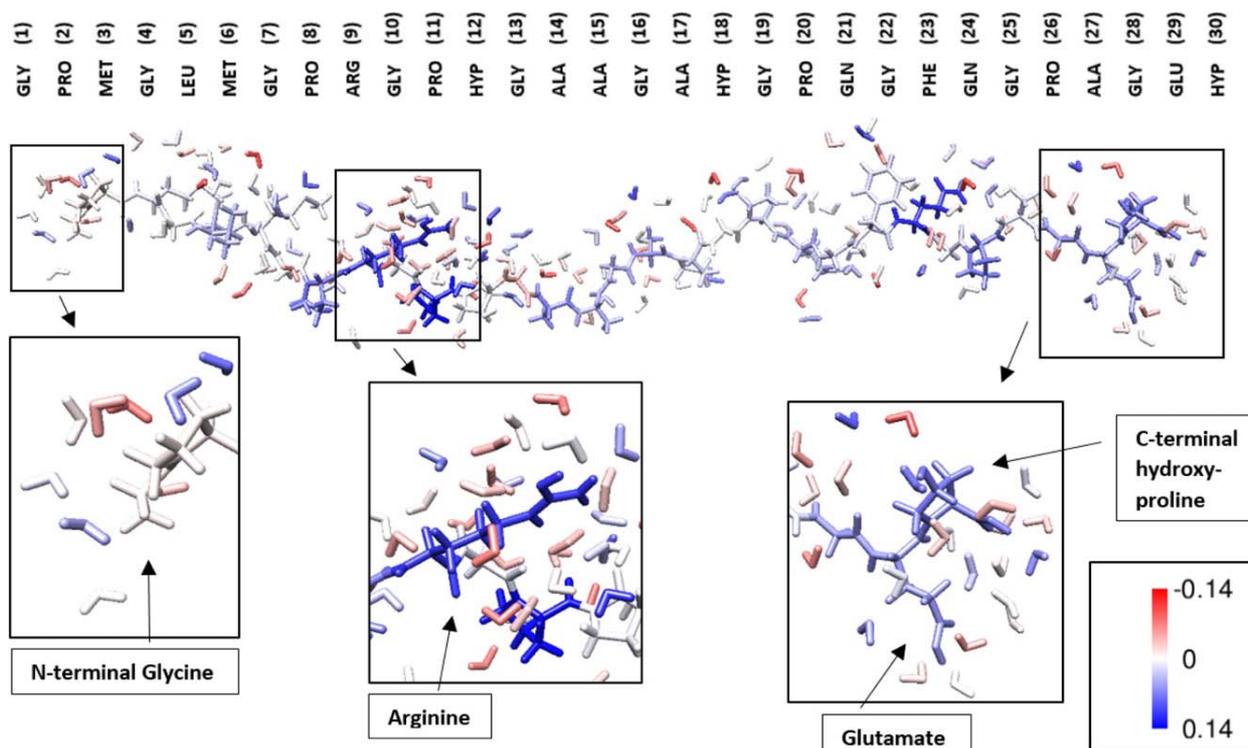


Figure 1. Structural model for the collagen α_2 -chain showing all the amino acids and water molecules (model 2a). Model 1 is the same model with all water molecules removed. The enlarged box shows the 2 protonated sites and 2 deprotonated sites for model 3. The color bar indicates the magnitudes of the PC in model 2a discussed in section “PC and HB in model 2a and 2b.” The sequence of the 30 amino acids is indicated at the top of the figure.

bond order (BO) results. It is the same collagen α_2 -chain as model 2 but with all the surrounding water molecules removed.

Model 2: hydrated models using NAMD solvate

Model 2 starts with the collagen α_2 -chain model used in previous study.^[23] We first added the H atoms by using NAMD,^[36] which relies on a topology file for the relative coordinates of the H atoms in relation to the atoms they are bonded to, consistent with the CHARMM force-field as implemented in NAMD. The “charged sites” in this model are Gly(1), Arg(9), Glu(29), and Hyp(30) which are all made charge neutral. The number in the parenthesis (1), (9), (29), (30) denotes the sequential number of the amino acids in the α_2 -chain model as listed in Table 1. The parameters used for charge neutral amino acids were those specified in CHARMM. Additionally, the C-terminal of the α_2 -chain which was terminated by adding a H in the previous work^[23] was made consistent with NAMD C-terminals by adding an O atom to the terminal carboxylic carbon.

Next, water molecules were added to the α_2 -chain using the “Solvate” command in NAMD in a pre-equilibrated box of water. The α_2 -chain is placed in a periodic box ($a = 11$ nm, $b = 4$ nm, $c = 4$ nm), which includes water molecules within 1.2 nm from the collagen chain and with overlapping water molecules deleted.^[37] To obtain more accurate positions of water molecules with respect to the collagen chain, we opti-

mize the structure of the peptide in the solvated model using the energy minimization procedure based on molecular mechanics as implemented in NAMD. The α_2 -chain and the added water molecules were allowed to relax their positions and orientations to minimize the total empirical force-field energy. Finally, we removed the water molecules that are further than 0.4 nm from the α_2 -chain, which resulted in 152 water molecules. This allows us to capture all water molecules that can possibly interact with the α_2 -chain through HB and significantly reduces the computational burden. It is generally accepted that it is only necessary to include water molecules in the proximity of the protein for calculation of the electronic structure of the solvated protein.^[38] With this model, we are able to quantitatively evaluate HB strength through BO calculations to be described in Section “Computational Methods and Steps.” Model 2 is illustrated in Figure 1. The removal of all water molecules in model 2 gives model 1.

To investigate the degree of variations of the solvated model described above (model 2a) on the statistical fluctuations in the distribution of water molecules, we build another solvated model (model 2b), which has a different number of water molecules with different positions and orientations relative to the peptide structure for comparison. To this end, we added water using the Chimera package which uses an AmberTools TIP3P waterbox.^[39] The same peptide of model 2a is held fixed and put into in this new water box from Chimera with a cutoff of 12 Å from the peptide. The structure is then minimized (with the peptide fixed) so that the water molecules can readjust

Table 1. Calculated PC (electrons) and number of HB in Region I for each amino acid in three models.

Amino Acid	Seq. #	Polar	Charged	Nonpolar	Model 2a # of HB	Model 2b # of HB	Model 3 # of HB	Model 1 ΔQ^*	Model 2a ΔQ^*	Model 2b ΔQ^*	Model 3 ΔQ^*
GLY	1		X		2	0	2	-0.082	-0.019	-0.092	0.692
PRO	2			X	0	1	0	-0.008	-0.014	0.007	0.016
MET	3			X	0	0	1	0.011	0.016	0.063	0.001
GLY	4			X	1	0	1	0.001	-0.003	0.020	0.009
LEU	5			X	2	0	2	-0.027	0.040	-0.008	0.001
MET	6			X	1	0	1	0.010	0.015	0.024	0.043
GLY	7			X	1	0	2	0.004	0.002	0.008	0.006
PRO	8			X	2	1	2	0.019	0.076	0.053	0.059
ARG	9		X		5	2	6	-0.024	0.110	0.098	0.836
GLY	10			X	1	0	2	-0.021	0.015	0.017	0.027
PRO	11			X	1	0	0	0.050	0.135	0.034	0.103
HYP	12	X			3	1	3	-0.007	0.007	0.027	0.027
GLY	13			X	1	1	1	-0.005	-0.043	0.053	-0.050
ALA	14			X	2	0	2	0.018	0.070	0.052	0.048
ALA	15			X	2	1	2	-0.018	0.049	0.000	0.044
GLY	16			X	1	0	2	-0.005	0.048	-0.001	0.035
ALA	17			X	1	0	1	0.006	0.063	0.033	0.106
HYP	18	X			3	1	3	-0.003	0.029	-0.010	0.043
GLY	19			X	0	1	0	-0.030	-0.010	0.033	0.012
PRO	20			X	0	0	0	0.032	0.035	0.055	0.029
GLN	21	X			5	1	5	0.023	0.052	0.088	0.069
GLY	22			X	1	0	1	-0.018	0.028	0.001	0.009
PHE	23			X	2	1	2	-0.012	0.025	0.029	0.019
GLN	24	X			2	1	4	0.005	0.124	0.039	0.140
GLY	25			X	1	0	1	-0.018	0.021	0.013	0.019
PRO	26			X	1	0	1	0.024	0.073	0.006	0.062
ALA	27			X	1	1	1	-0.006	0.021	-0.016	-0.023
GLY	28			X	2	0	1	0.001	0.052	0.028	0.017
GLU	29		X		3	2	6	0.004	0.058	0.088	-0.564
HYP	30		X		4	1	7	0.076	0.072	0.142	-0.554
Totals		4	4	22	51	21	62	0.000	1.146	0.883	1.281

their positions. We intentionally using a different procedure for model 2b than used for model 2a. We end up with 123 water molecules in model 2b, as opposed to 152 in model 2a, but with different positions and orientations relative to the amino acids in the peptide. In the discussions below, model 2a is used when we refer to model 2.

Model 3: protonated/deprotonated solvated model

Model 3 uses the same starting point as model 2, except that the α_2 -chain is now protonated at two amino acids (Gly(1) and Arg(9)) and deprotonated at other two amino acids [Glu(29) and Hyp(30)], as illustrated in the highlighted portion of Figure 1. Essentially, the N-terminal Gly(1) now has an extra H on the amine group and Arg(9) has an extra H on one of its side-group terminal N. Simultaneously, the H atom on the carboxylic O of the side-group in Glu(29) and on the carboxylic O of the C terminal of Hyp(30) are now removed. At physiological pH values, charged groups such as the side-chains of Arg and Glu and the terminals of the α_2 -chain, participate in acid-base equilibria, so most of the "charged sites" are protonated/deprotonated.^[40] We have constructed model 3 so that all "charged sites" are charged.^[41] By adding and removing 2 protons simultaneously at 4 different amino acids. In this way, we maintain the overall charge neutrality of the system and enables us to study the effect of protonation/deprotonation. Next, water molecules were added to the protonated/deprotonated model using the same procedure as used in

model 2, followed again by a total energy minimization. The difference in the atomic positions of the α_2 -chain and water molecules between model 3 and model 2 are very small, except for those atoms close to the protonated/deprotonated sites. Again, only the water molecules within 0.4 nm from the protein are included for the calculation of the electronic structure and bonding. Model 3 contains 151 water molecules, only one less than in model 2 with 152 water molecules.

Computational Methods and Steps

NAMD^[35] is used in the construction of the three models as described above. This is a very common and efficient package particularly suitable for computationally demanding simulations on biomolecular systems with large numbers of atoms. We used NAMD to add H atoms and water molecules to the α_2 -chain model and to perform structural optimization based on the CHARMM force-field model. The force-field calculation in NAMD involves two sets of parameters: structural and electrostatic. The structural parameters preserve the structure of the molecule and include parameters for bond length (BL), bond angle, dihedral angle and improper dihedrals. The electrostatic parameters include atomic PC and van der Waals (vdW) parameters between certain atomic types. The contribution from electrostatic forces and vdW forces are calculated explicitly up to a cutoff distance. We adopted the default

values in the package, which are all available within the NAMD package.

The electronic structure and bonding properties of the three models are calculated using the orthogonalized linear combination of atomic orbitals (OLCAO) method. OLCAO is an *ab-initio* package developed in-house for the calculation of electronic structure, bonding and physical properties applicable to both crystalline and non-crystalline systems including biomolecular ones.^[42] The method is based on the density functional theory (DFT) within the local density approximation for exchange and correlation potential. It uses localized atomic orbitals in the basis expansion where the radial part of the basis function is expanded in terms of Gaussian-type orbitals. The use of atomic orbitals makes the calculation of complex biomolecular systems particularly efficient especially for the PC and BO values. The method has been applied recently to a large organic crystal,^[43] a model of super-cooled water network,^[44] and an increasing number of complex biomolecular systems.^[45–47] It has also been used in the previous study of the triple-helix model of collagen and the development of a computationally efficient scheme called the AAPM applicable to very large proteins.^[23]

The use of atomic orbitals enables us to quantify interatomic bonding and charge transfer efficiently and naturally. We use the Mulliken scheme^[48] to obtain effective charge (Q^*) on each atom:

$$Q_{\alpha}^* = \sum_{n,occ} \sum_I \sum_{j,\beta} C_{i,\alpha}^{n*} S_{i\alpha,j\beta} \quad (1)$$

and BO $\rho_{\alpha\beta}$ between pairs of atoms, which is a quantitative measure of the bond strength:

$$\rho_{\alpha\beta} = \sum_{ij} C_{i\alpha}^{n*} C_{j\beta}^n S_{i\alpha,j\beta} \quad (2)$$

Here, i is a collective index for the principal and orbital quantum numbers, α and β designate different atoms in the model, and n represents energy levels (band index in inorganic crystals). C is the eigenvector coefficient of wave function and $S_{i\alpha,j\beta}$ are the overlap integrals. $\Delta Q_{\alpha} = Q_{\alpha}^0 - Q_{\alpha}^*$ is the PC on the atom α which is the deviation of charge or the charge transfer from the neutral atom as a result of interatomic interactions based on quantum mechanics. Hence, a negative PC implies a gain in electron charge. The summation of all PC values of the atoms in an amino acid gives the PC for that amino acid. The summation of BO values within each model gives the total bond order (TBO). When normalized with the volume of the cell, it gives the total BO density (TBOD), a single quantum mechanical metric particularly useful for assessing internal cohesion of complex crystals with well optimized cell volume.^[49]

It should be made clear that Mulliken scheme is basis dependent and its calculation must be confined to a specific method and a well-defined basis set. It is most effective when a more localized minimal basis (MB) set is used. In the present work, the PC ΔQ_{α} and BO values are calculated using a MB set

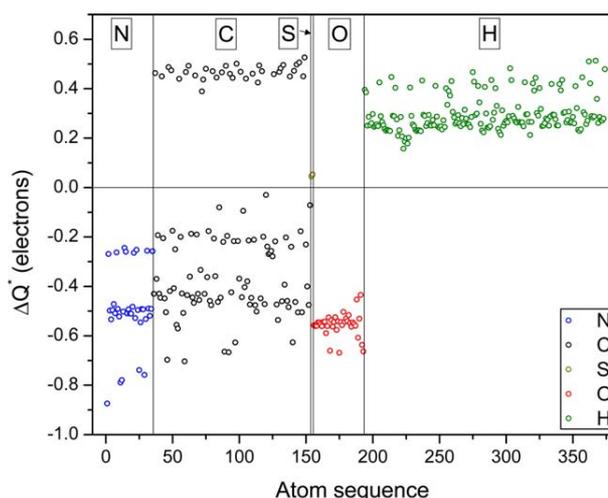


Figure 2. Calculated atomic PC for model 1.

whereas for the self-consistent potential and the electronic structure calculation, a full basis (FB) set is used. These are carefully constructed and well-tested basis set for each atom within the data base of the OLCAO package.^[50,51] The MB consists of the core orbitals of the atom plus the orbitals of the valence shell and the FB has one more shell of the higher unoccupied orbitals. As an example, the MB of atom sulfur S consists of 9 atomic orbitals ($1s, 2s, 2p_x, 2p_y, 2p_z, 3s, 3p_x, 3p_y, 3p_z$) and the FB has 18 orbitals (MB plus $4s, 4p_x, 4p_y, 4p_z, 3d_{xy}, 3d_{yz}, 3d_{zx}, 3d_{2-y_2}, 3d_{3z^2-r^2}$).^[51] Although there are other alternative methods to calculate PC and BO values, they are usually limited to small molecules or simple crystals based on numerical evaluations on 3-dimensional mesh. Some of these approaches used include numerically fitting the atom-centered PCs from molecular electrostatic potential instead of directly obtained from *ab initio* calculations on exact atomic geometry used in our approach. Such methods are usually more onerous and less efficient when applying to large complex biomolecules.

Results and Discussion

PC and HB in model 1

In a closed and charge-neutral system, the net PC for the system is zero but with different local PC values on each atom due to charge transfer. The calculated PC on each atom in model 1 is displayed in Figure 2. The 5 panels represent the 5 different types of atomic species (N, C, S, O, and H). The 374 atoms in the α_2 -chain are numbered sequentially as implemented in the OLCAO method. As can be seen, all N and O atoms have negative PC and all H atoms have positive PC. The C atoms have positive or negative PC depending on their local bonding configurations with the negatively charged C atoms having a wider distribution. There are no positively charged C atoms that have H atoms bonded to them except the γ -C of Phe(23) which is slightly negative and is bonded to three other C atoms. The large distribution of negatively charged C is related to the number of H atoms it bonds to. The more it

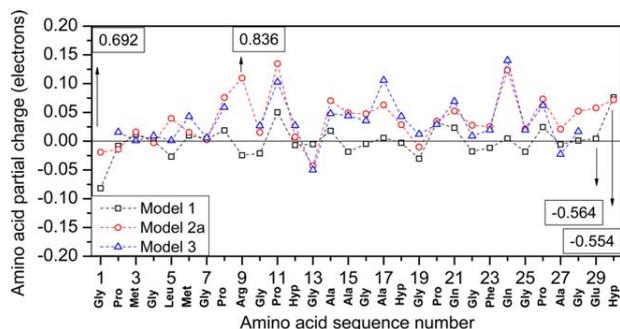


Figure 3. Calculated PC on each amino acid in models 1, 2a, and 3.

bonds to H, the more it is negatively charged. All H atoms are positively charged with those bonded to O having higher positive charge than those bonded to N since the former is more electronegative. The 2 S atoms from Met(3) and Met(6) have a surprisingly small positive charge of 0.045 and 0.053, respectively, which is very different from O even though they are both from column VI of the Periodic Table. This could be due to the fact that the 2 C atoms attached to S have PC of -0.500 and -0.700 . By tracing the PC on atoms to the amino acids they belong to, we find that the highest positive ΔQ comes almost entirely from carboxylic carbons. The most negative ΔQ comes from the O atom of the C-terminal and γ -O of Hyp; the amine nitrogen of the N-terminal, the terminal nitrogen of the Arg(9) side-group and Gln side-chains; the β -C of Ala and ε -C of Met. Overall, the most negatively (positively) charged atom with $\Delta Q = -0.874$ (0.526) is from amine N-terminal of Gly(1) [carboxylic C-terminal of Hyp(30)].

By summing up the atomic PC for all the atoms within a given amino acid, we have the PC for each of the 30 amino acids in model 1. These are listed in Table 1 and shown in Figure 3 (which also include the results for model 2 and model 3 to be discussed later). Table 1 indicates which amino acids are polar, charged, or nonpolar as classified according to reference.^[41] However, in models 1 and in model 2, the charged sites are made charge neutral in our calculation, hence they can be considered as polar. Of the 30 amino acids listed in Table 1, half are positively charged and half are negatively charged. In model 1 for the neutral isolated α_2 -chain, none of the amino acids have any significantly large PC. Gly(1) at the N-terminal has the largest negative PC of -0.082 electron and Hyp(30) at the C-terminal has the largest positive PC of 0.076 e⁻.

The 4 polar and 4 charged amino acids in the α_2 -chain of model 1 are the N-terminal Gly, C-terminal Hyp, two additional Hyp, Arg, Glu, and two Gln. The remaining 22 nonpolar amino acids (Gly, Met, Pro, Ala, Leu, and Phe) make up the 30 amino acids in α_2 -chain. They are all initially assumed to be charge neutral. Thus, the ratio of nonpolar to polar amino acids in model 1 is 11:4. Table 2 lists the charge transferred to charged, polar, and nonpolar amino acids for all models, as well as the total BO and the number of HBs of these groups for models 2 and 3 (to be discussed below). The total charge on the polar plus charged and nonpolar amino acids in model 1 are equal and opposite, or -0.008 and 0.008 , respectively, which is consistent with the fact that model 1 is charge neutral.

We have calculated the BO values for the H-bonding in all three models according to eq. (2). Since Model 1 contains no water molecules, any HB in model 1 are weak intra-molecular HB with large O...H or N...H bond distance larger than 0.24 nm. We will discuss it in comparison with the solvated model 2 in the following subsection.

PC and HB in model 2a and 2b

Model 2 is the solvated model with 152 water molecules. Although the overall net PC of the system is zero, we found that the protein transfers 1.146 electron (e⁻) charge to water molecules. This is consistent with our recent finding in the case of the solvated model of doxorubicin, which also shows that the molecule transfers charge of $0.123e^-$ to the water molecules.^[52] In Figure 4, we display the calculated PC of all atoms in model 2. The PC for O and H from water molecules are included and shown in separate panels. As can be seen, the PC for O (H) in water molecules are more negative (positive) than those in the α_2 -chain. The PC in the solvated model 2 vary little compared to those of model 1 except in those cases where the water molecules are close to some particular amino acids, mostly the carboxylic carbons and some H atoms in water molecules. The highest PC are $-0.849e^-$ in N-terminal Gly(1) and $+0.579e^-$ in C-terminal Hyp(30), respectively. The corresponding values in model 1 are $-0.874e^-$ and $0.526e^-$. Thus part of the solvation effect is to reduce PC from the highest partially charged N and O atoms on the terminal amino acids.

The PC for each amino acid in model 2 is listed in Table 1 and illustrated in Figure 3. It can be seen that there are now fairly large variations in ΔQ from model 1 with some amino

Table 2. PC and protein-water HB totals for BO and number of HBs for polar, nonpolar, and charged amino acids for Regions I, II, and III including N...H and O...H HBs only.

Amino Acid type	# of Amino Acids	Model 1	Model 2a (Model 2b)			Model 3		
		Total PC	Total PC	Total BO	# of HBs	Total PC	Total BO	# of HBs
Polar	4	0.017	0.213 (0.144)	0.312 (0.152)	14 (10)	0.279	0.374	15
Nonpolar	22	0.008	0.717 (0.503)	0.710 (0.346)	26 (19)	0.591	0.681	27
Charged	4	-0.025	0.219 (0.236)	0.305 (0.158)	19 (9)	0.410	0.669	21
Total	30	0	1.146 (0.883)	1.328 (0.656)	59 (38)	1.281	1.725	63

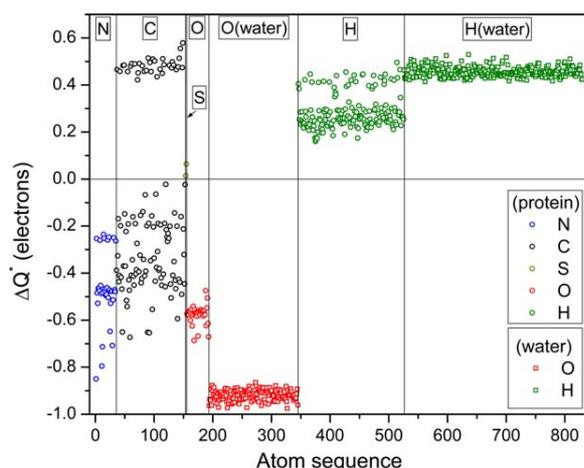


Figure 4. Calculated atomic PC for model 2a.

acid PC changing sign due to the presence of water molecules. Gly(13) is the most negatively charged amino acid ($-0.043 e^-$) and Pro(11) is the most positively charged amino acid ($0.135 e^-$). There are now more highly positively charged amino acids such as Arg(9) ($0.110 e^-$) and Gln(24) ($0.124 e^-$). This large shift towards high positive charge is due to the transfer of electrons to the nearby water molecules. Figure 1 also shows the positions of the water molecules relative to the affected amino acids in the protein and their PC. The sum of amino acid PC of charged, polar and nonpolar amino acids are listed in Table 2. All these groups lose electrons to water molecules.

The main purpose for constructing model 2b described in the last section is to estimate the variation in PC distribution on amino acids due to different distribution of water molecules. They are illustrated in Figure 5 and listed in Tables 1 and 2. Due to presence of a different number of water molecules, their positions and orientations relative to the peptide structure, the PC on amino acid PC are altered. The most negative and positive charged amino acids are now the Gly (1) N-terminal ($-0.092e$) and Hyp (30) C-terminal ($0.142e$), respectively, at the ends of the peptide rather than near the middle as in model 2a. The other two most highly charged amino acids are Arg (9) ($0.098e$) and Gln (21) ($0.088e$). The overall charge of model 2b is $0.883e$, which is less than in

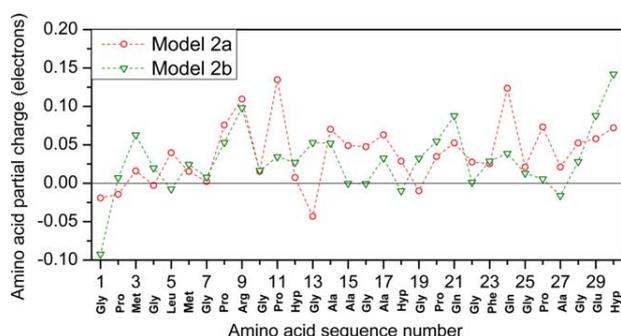


Figure 5. Comparison of calculated PC on amino acids of models 2a and 2b.

model 2a ($1.146e$). Part of these is due to less number of water molecules in model 2b. In Table 2, we see that the charged amino acids groups are more positively charged in model 2b than in model 2a, but that the polar and nonpolar amino acid groups are equally less positively charged. On the whole, we conclude that the charge transfer from protein to water molecules can vary in magnitude depending on their distribution, but the qualitative features remain the same as can be seen in Figure 5.

By feeding back the *ab initio* PC calculated from OLCAO method for all atoms instead of using the default values into the NAMD package, we can investigate the changes in the stability of the structure of the protein model. It should again be made crystal clear that the purpose of this test is to demonstrate the effect of using more accurate PC from *ab initio* calculation and not to address any issues related to the nonequilibrium structure of our peptide model. The result of our test is shown in Figure 6. First, the minimized model 2a is slowly heated from 0 to 300 K over 300 ps. Next, an NPT MD simulation for 10ns is performed using either NAMD or OLCAO atomic PC of model 2a. The result for root mean square deviation (RMSD) versus the MD steps over the 10 ns run is shown in Figure 6a. The RMSD is defined using the first time step t_1 as a reference according to eq. (3) below where N_α is the number of α carbon atoms in the backbone of the model, t_j is the time step and $r_\alpha(t_j)$ is the position of atom at time t_j .

$$\text{RMSD}_\alpha(t_j) = \sqrt{\frac{\sum_{\alpha=1}^{N_\alpha} (r_\alpha(t_j) - r_\alpha(t_1))^2}{N_\alpha}}, \quad (3)$$

It is clear that the average of the two types of the NAMD runs are similar after sufficient long steps but the use of OLCAO PC has less fluctuations than using the default PC. This is particularly true in the initial region of up to 3 ns. The use of default charge has RMSD fluctuating around 5 Å whereas the use of OLCAO PC fluctuates around 9 Å and both stabilize around 9 Å. This can be interpreted that the later represent a more stable structure.^[53] This conclusion is further supported by the 2D map of the relative free energy obtained from Weighted Histogram Analysis Method (WHAM) scheme^[53] as a function of collective variables of RMSD and radius of gyration (Rg), which is an indicator of protein structure compactness.^[54] These are shown in Figures 6b and 6c for using the default PC and the OLCAO PC, respectively. It can be seen that in Figure 6b, there are showing two separate regions of lower free energy indicative of a less stable structure.^[53] The same plot using the OLCAO PC (Fig. 6c) shows only 1 region of lower free energy hence a more stable structure. Thus our test calculation indicates that the use of quantum mechanically derived PC from OLCAO in the NAMD code could potentially improve the accuracy of dynamic studies in large proteins.

We now focus on the discussion of the HB in the solvated model 2, which is an important subject recognized by many but seldom discussed in a quantitative manner. We have

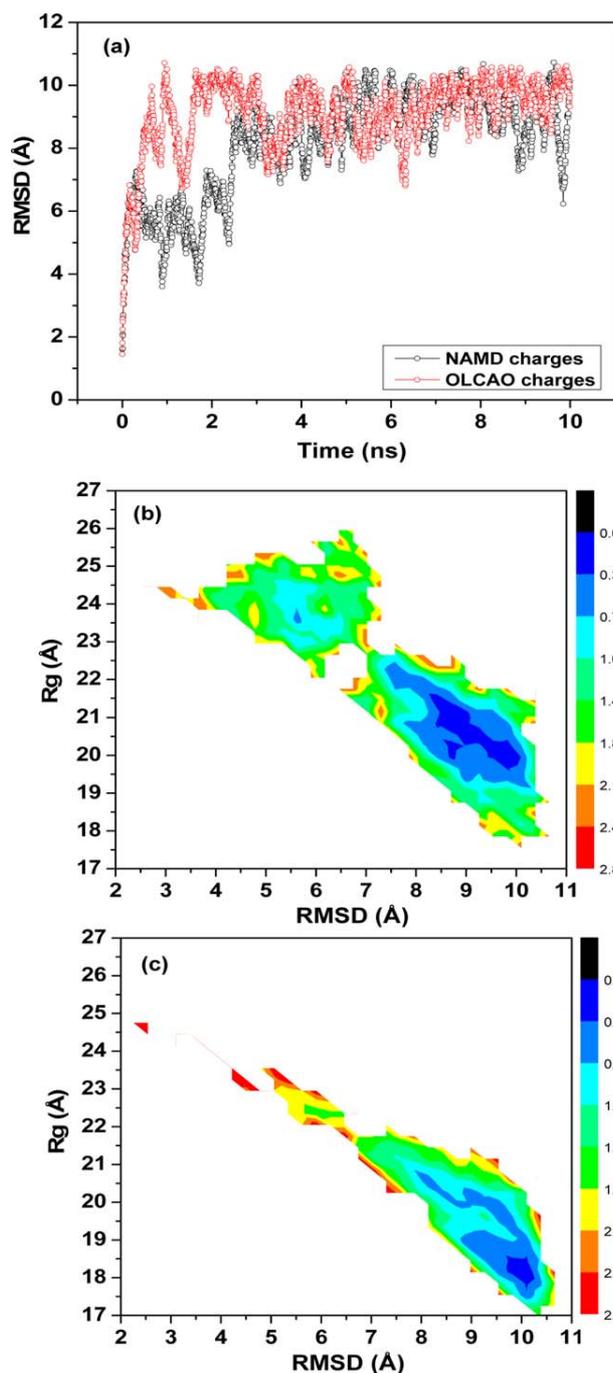


Figure 6. a) Plot of RMSD deviation of the carbon backbone for model 2a using the default PC in NAMD (black) or the OLCAO PC (red) in a MD time simulation of 10 ns in steps of 2fs. b) Plot 2D-map of the relative free energy (in unit of kcal/mol) as a function of gyration (Rg) and RMSD for the case of using default NAMD PC showing 2 separate regions of lower free energy indicative of a less stable structure. c) Same plot using the OLCAO PC showing only 1 region of lower free energy hence a more stable structure.

previously studied a HB network in a model of super-cooled water using the same computational procedure.^[44] The BO values are calculated according to eq. (2). Figure 7 shows the plot of bond BO vs BL for HBs in model 2. The distribution of the HBs can be roughly divided into three regions with differ-

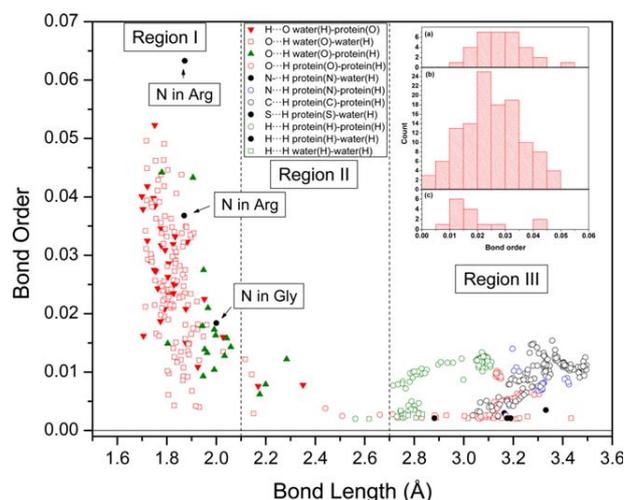


Figure 7. BO versus BL for HB in model 2a. Different symbols and colors are used to represent different types of HBs. HBs related to N atom in region 1 are marked by arrow. The inset shows the histogram distribution of BO values for HBs in region I: a), b), and c) correspond to the three cases of HBs described in the text.

ent ranges of H BL: region I (BL < 0.21 nm), region II (0.21 nm < BL < 0.27 nm), and region III (BL > 0.27 nm). To fully understand the contribution from different types of HBs (water-water, protein-water, protein-protein), we categorize the HB into 10 types as shown in the inset of Figure 7, which reveal the following interesting facts:

- In region I, the HBs are from: (1) between H in water and O in protein, (2) between water molecules, (3) between O in water and H in protein, plus from HBs between N in protein and H in water (highlighted). The distribution of BO values of the HBs in (1), (2) and (3) are shown in histogram plots in the inset [marked as (a), (b), and (c) of Figure 7. It is apparent that the majority of the HBs are O...H and H...O in region I (case (2)) with a near Gaussian distribution with maximum centered at BO \sim 0.025. The HBs in the other two cases (1) and (3) are much less in number with HB between H in water and O in protein having a near-normal distribution, and the HBs between O in water and H in protein having much weaker BO.
- There are relatively few HBs in region II. Their BO values are generally < 0.015.
- In region III with larger BL, there are many HBs, mostly from those involving atoms from the α_2 -chain. Although the BO values in region III are small, the preponderance of their numbers indicates their contribution to total BO is not negligible, and may affect the stability of the peptide structure.^[23]
- There are a few isolated HBs involving N, S and H from the protein to water molecules. Of particular significance is the three HBs between N to H in water in region I (highlighted). As a matter of fact, the highest BO for HB is from N...H in Arg(9) with a BO of 0.063.
- We have noticed that individual polar and charged amino acids have more HBs than the nonpolar amino acids as listed in Table 1, with the exception of the N-terminal

Table 3. HB statistics for models 2a and 3.

HB result	Water (H)- protein (O) Red down triangle		Water (O)- water (H) Red open square		Water (O)- protein (H) Green up triangle	
	Model 2a	Model 3	Model 2a	Model 3	Model 2a	Model 3
BL avg	1.825	1.790	1.988	1.993	1.999	1.980
BO total	0.955	1.262	3.130	3.055	0.318	0.463
BO avg	0.027	0.029	0.022	0.024	0.017	0.023
Number of HBs	35	43	140	126	18	20

Gly(1), which has only 2 HBs. We also note that nonpolar amino acids have stronger HBs than those in polar or charged amino acids, and that there are more HBs from nonpolar amino acids as a whole. Since the charged amino acids in model 2 are protonated or deprotonated to be charge-neutral, they are also considered as polar amino acids. By grouping the polar and charged groups together, we note they form more HBs collectively but with smaller total BO than the nonpolar group.

To facilitate a more focused discussion on the nature of HB involving water molecules in both model 2 and model 3, we list the average BL, total BO, average BO and number of HBs for the stronger HBs involving water molecules in Table 3.

PC and HB in model 3

Model 3 includes the protonated and deprotonated charged sites and with surrounding water molecules. Our calculation shows that the protein molecule transfers $1.281e^-$ to the water, which is slightly more than in model 2 ($1.146e^-$). Other than the protonated and deprotonated sites, the atomic PC are similar to model 2 (not shown here) and the various atomic species of N, C, O, H, and S of model 3 all have atomic PC values that cluster into groups of higher or lower values corresponding to the local bonding configurations. Compared to model 2, the N of Arg(9) and Gly(1) have PC of $-0.125e^-$ and $-0.116e^-$, respectively, whereas for the O of Glu(29) and Hyp(30), the PC are $0.113e^-$ and $0.126e^-$, respectively. Other atoms close to these 4 sites also have PC changed but less significantly. At the amino acid level, Figure 3 shows that Gly(1) ($\Delta Q = 0.692e^-$), Arg(9) ($\Delta Q = 0.836e^-$), Glu(29) ($\Delta Q = -0.564e^-$), and Hyp(30) ($\Delta Q = -0.554e^-$) are the most significantly partially charged ones and have to be marked off-scale. The other amino acids with large PC are similar to model 2 except Ala (17) ($\Delta Q = 0.106e^-$). Thus, the protonated/deprotonated amino acid sites in model 3 have the largest PC in the presence of surrounding water molecules. They are most likely to interact with ions that are close to them.^[54] The PC values for each of the 30 amino acids are also listed in Table 1.

Referring to Table 2, the total PC of the polar amino acids increases from $0.213e^-$ in model 2 to $0.279e^-$ in model 3. The charged group also increases in positive PC from $0.219e^-$ to $0.410e^-$. However, a large portion of the PCs of these charged sites are mostly cancelled out in the summation. Conventionally, the sum of these charged sites is thought of having a zero net charge,^[41] but here they sum up to $0.410e^-$. The

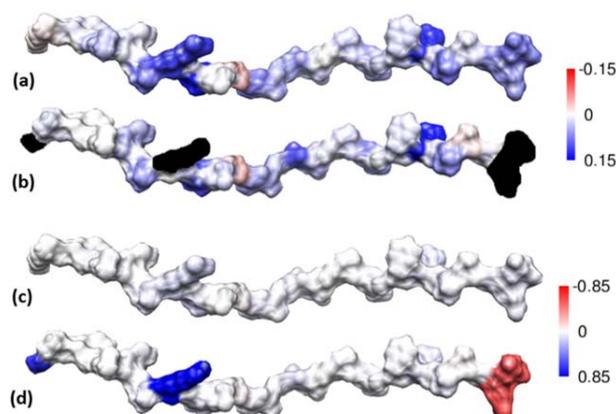


Figure 8. Plot of PC on the solvent-excluded surface of amino acids in α_2 -chain in two different scales: a) and b) are for model 2a and model 3 in smaller scale. The blackened areas are for protonated and deprotonated amino acids, which are off-scale. c) and d) are for model 2 and model 3 in larger scale.

nonpolar amino acids decrease significantly in total positive PC. This indicates that the protonation/deprotonation of particular charge sites have only a modulatory effect on the charges of other amino acids.

To show the effect of protonated/deprotonated sites on amino acids, we show in Figure 8, the comparison of amino acid PC on the solvent excluded surface (SES) for models 2 and 3 in two different scales. The top portion (a) and (b) are on a reduced color scale and the bottom portion on (c) and (d) on the full color-scale. The PC on the charged sites in (b) are blackened out since they are off the scale. On the reduced scale, the differences in amino acid PC of polar and nonpolar sites between the two models can be easily visualized. For example, in Ala(17) [near the middle section in (b)], there is a more bluish or higher positive PC in model 3 while near the right, C-terminal end Ala (27) changes to light reddish indicating a slightly negative PC. The IEP of a protein is the location where the net charge is zero. Its experimental determination depends on pKa value of the protein. How to use the changes in PC due to the protonation/deprotonation effect is a challenging task since direct calculation of pKa values on large protein is unfeasible. This is a long sought but still unreachable goal of protein biophysics.

The distribution of HBs in model 3 with a similar break down into different types and in various regions as in model 2 is presented in Figure 9. In model 3, the HBs at the charged sites become shorter and stronger overall. The main difference is that there are no longer any $N\cdots H$ HBs in region I. Apparently, the extra H bonded to these nitrogenous sites where the $N\cdots H$ HBs locate are now replaced by protein-H and water-O HBs with longer BL as highlighted in region II of Figure 9. Similar to Figure 7, we display the distribution of HBs in region I for model 3 for three different groups of HBs in the inset. These distributions of HBs in region I are slightly different between model 3 and model 2, accentuating the effect of protonation/deprotonation on the HB distributions. In particular, the middle panel (b) for HB distribution between water

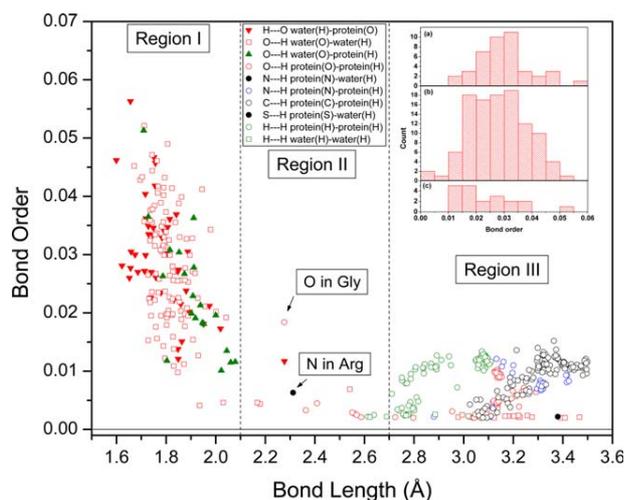


Figure 9. BO versus BL for HB in model 3. The inset shows the histogram distribution of BO values for HBs in region 1.

molecules is reduced in number and of less Gaussian type distribution. The HB BO values for both the charged and polar amino acids are listed in Table 2. Collectively, the polar and charged groups have larger total BO compared with model 2, and less total BO for the nonpolar group. However, only the charged groups have a significantly larger total BO.

It is interesting to note that within the same polar amino acid type [Gln(21) and Gln(24)] can have different PC in models 2 and 3 with Gln(24) being significantly more charged. Gln(21) has more HBs in model 2 but they are almost equal in model 3. However, the number of water molecules surrounding Gln(21) is still greater than Gln(24) and this may account for the lower PC. It appears that HB alone cannot account for the solvation effect but the number of surrounding water molecules near the amino acid should also be taking into account as demonstrated in the comparison between model 2a and 2b, since PC is influenced by the local environment of the amino acid.

Conclusions

We have studied the electronic structure and H-bonding in three models of a peptide derived from the α_2 -chain of the collagen: dried, solvated and with protonated/deprotonated sites. Quantitative results and critical analysis of PC on the amino acids and distribution of H-bonds in these three models are presented and discussed. We find that a significant charge transfer of more than one electron from the protein to the water occurs, giving the protein an overall net positive charge. The amount of charge transfer depends upon, to a large extent, the number of water molecules included in the simulation and their local distributions and orientations. A similar situation exists for the model where there are charged sites due to protonation and deprotonation of dissociable amino acids. The PC on specific amino acids identified as charged, polar, and nonpolar are obtained from atomic PC and their linkage to the solvation effect is assessed, yielding a direct evidence that the interaction with the aqueous solvent affects the PC of

amino acids significantly and essentially, mostly on polar and/or charged amino acids. The importance of the vicinal solvent has been argued for many years based on thermodynamic measurements^[55] and therefore the connection with the microscopic picture is thus highly desirable. Large PCs for the protonated or deprotonated charged sites are observed as expected, and they dominate the overall charge properties of the biomolecule. They should certainly play a significant role in protein interactions with various ligands and in helping the identification of the IEP of solvated proteins. This is illustrated by plotting the PC on the solvent excluded surface of the solvated α_2 -chain model with and without the protonated/deprotonated sites.

Quantitative analysis of the HB in the three models of increasing complexity are carried out by plotting the calculated BO values versus BL for different types of atom pairs in water-water, water-protein, and protein-protein. It is shown that the main contributors are HBs from water-water and water-protein pairs with a large spread in the BO values and with BL < 0.21 nm. The specific changes in the HB distribution in three different models are discussed in detail. The stronger HBs between the protein and the water molecules mostly involve the H atom in the water and the backbone carboxylic O from nonpolar amino acids. There is an increase in HBs due to protonation/deprotonation for polar and charged amino acids with a concomitant increase in their BO values. For the nonpolar amino acids, they may increase or decrease in the HBs and their total BO values. A loose connection between PC and HB can thus be made in the sense that nonpolar amino acids that form no HBs have the smallest PC and vice versa. Nonpolar amino acids that form no HBs have the smallest PC and strongly polar amino acids have the highest PC.

The main conclusion of this work is that molecular details of the aqueous solvent as well as the interaction between vicinal water and protein solvent accessible surface at atomistic level can be obtained from full quantum mechanical calculations. While a full scale *ab initio* simulation of IEP and pKa values for large proteins is computationally still out of research, *ab initio* PC calculation on protonated and deprotonated models can complement experimental measurements and in probing long range electrostatic interactions in and between the biomolecules. We also demonstrate that accurate *ab initio* PC from DFT calculations can be incorporated into NAMD package for improved simulation results contingent upon the availability of sufficient computing resources.

In dealing with solvation effect, there are generally two approaches,^[56,57] the explicit solvent model with include explicit number of water molecules as we try to do in this paper, and the implicit solvent model (or the continuum solvent model) by treating water as a continuous medium having average properties for the solvents surround the biomolecule (the solute). The former is more rigorous and specific but computationally very demanding since ideally, a large number of solvent molecules may be required to get conclusive results. The latter is more associated with the generalized Born model in solving the Poisson-Boltzmann Equation in traditional biophysics theory.^[58] Both models are facing some technical

difficulties and uncertainty in the justification using some parameters and simplification.^[57] Only the explicit solvent model can directly address the HB but has difficulties with hydrophobic effects. These are certainly nontrivial issues but at the frontiers of the biophysical research. Our work could be a significant step forward in using the explicit solvation model in modern computation biophysics research.

Acknowledgments

This research used the resources of NERSC. Use of the cluster computer at the University of Missouri Bioconsortium is greatly appreciated.

Keywords: collagen model · *ab initio* calculation · partial charge · solvent effect · hydrogen bonding

How to cite this article: J. Eifler, R. Podgornik, N. F. Steinmetz, R. H. French, V. A. Parsegian, W.-Y. Ching. *Int. J. Quantum Chem.* **2016**, *116*, 681–691. DOI: 10.1002/qua.25089

- [1] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198.
 [2] T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung, X. Xiao, *Q. Rev. Biophys.* **2011**, *44*, 191.
 [3] K. Park, A. W. Götz, R. C. Walker, F. Paesani, *J. Chem. Theory Comput.* **2012**, *8*, 2868.
 [4] R. A. Friesner, V. Guallar, *Annu. Rev. Phys. Chem.* **2005**, *56*, 389.
 [5] D. M. Leitner, M. Gruebele, M. Havenith, *HFSP J.* **2008**, *2*, 314.
 [6] S. C. Kamerlin, S. Vicatos, A. Dryga, A. Warshel, *Annu. Rev. Phys. Chem.* **2011**, *62*, 41.
 [7] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269.
 [8] A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 527.
 [9] G. Kieseritzky, E. W. Knapp, *Proteins: Struct. Funct. Bioinformatics* **2008**, *71*, 1335.
 [10] G. R. Grimsley, J. M. Scholtz, C. N. Pace, *Protein Sci.* **2009**, *18*, 247.
 [11] I. Gitlin, J. D. Carbeck, G. M. Whitesides, *Angew. Chem. Int. Ed.* **2006**, *45*, 3022.
 [12] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*; Garland Science: New York **2002**, *4*.
 [13] Fratzl, P. Collagen: structure and mechanics. Springer, New York, NY, **2008**.
 [14] R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari, H. M. Berman, *J. Mol. Biol.* **1998**, *280*, 623.
 [15] R. Z. Kramer, J. Bella, B. Brodsky, H. M. Berman, *J. Mol. Biol.* **2001**, *311*, 131.
 [16] R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, *Biopolymers* **2000**, *56*, 8.
 [17] R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, *Protein Sci.* **2002**, *11*, 262.
 [18] K. Okuyama, M. Takayanagi, T. Ashida, M. Kakudo, *Polym. J.* **1977**, *9*, 341.
 [19] K. Okuyama, T. Kawaguchi, M. Shimura, K. Noguchi, K. Mizuno, H. P. Bächinger, *Biopolymers* **2013**, *99*, 436.
 [20] K. Okuyama, K. Miyama, K. Mizuno, H. P. Bächinger, *Biopolymers* **2012**, *97*, 607.
 [21] K. Okuyama, X. Xu, M. Iguchi, K. Noguchi, *Peptide Sci.* **2006**, *84*, 181.
 [22] J. J. Hutton, A. Kaplan, S. Udenfriend, *Arch. Biochem. Biophys.* **1967**, *121*, 384.
 [23] J. Eifler, P. Rulis, R. Tai, W. Y. Ching, *Polymers* **2014**, *6*, 491.
 [24] Y. Tong, Y. Mei, Y. L. Li, C. G. Ji, J. Z. Zhang, *J. Am. Chem. Soc.* **2010**, *132*, 5137.
 [25] C. Ji, Y. Mei, J. Z. Zhang, *Biophys. J.* **2008**, *95*, 1080.
 [26] T. Ishida, *J. Chem. Phys.* **2008**, *129*, 125105.
 [27] P. G. Righetti, *J. Chromatogr. A* **2004**, *1037*, 491.
 [28] T. E. Klein, C. C. Huang, *Biopolymers* **1999**, *49*, 167.
 [29] S. D. Mooney, C. C. Huang, P. A. Kollman, T. E. Klein, *Biopolymers* **2001**, *58*, 347.
 [30] B. Madhan, P. Thanikaivelan, V. Subramanian, J. R. Rao, B. U. Nair, T. Ramasami, *Chem. Phys. Lett.* **2001**, *346*, 334.
 [31] R. Improta, F. Mele, O. Crescenzi, C. Benzi, V. Barone, *J. Am. Chem. Soc.* **2002**, *124*, 7857.
 [32] A. Gautieri, S. Vesentini, F. M. Montevercchi, A. Redaelli, *J. Biomech.* **2008**, *41*, 3073.
 [33] A. Gautieri, A. Russo, S. Vesentini, A. Redaelli, M. J. Buehler, *J. Chem. Theory Comput.* **2010**, *6*, 1210.
 [34] S. Vesentini, C. F. Fitié, F. M. Montevercchi, A. Redaelli, *Biomech. Model. Mechanobiol.* **2005**, *3*, 224.
 [35] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
 [36] A. D. MacKerell, D. Bashford, M. L. D. R. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. A. Ha, D. Joseph-McCarthy, *J. Phys. Chem. B* **1998**, *102*, 3586.
 [37] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33.
 [38] M. Peng, H. Zheng, *J. Mol. Model.* **2011**, *17*, 111.
 [39] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
 [40] J. C. Masini, O. E. Godinho, L. M. Aleixo, *Fresenius J. Anal. Chem.* **1998**, *360*, 104.
 [41] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell*; Garland Science: New York, **2002**.
 [42] W.-Y. Ching, P. Rulis, *Electronic Structure Methods for Complex Materials: The orthogonalized linear combination of atomic orbitals*; Oxford University Press, Oxford, UK, **2012**.
 [43] L. Liang, P. Rulis, B. Kahr, W. Ching, *Phys. Rev. B* **2009**, *80*, 235132.
 [44] L. Liang, P. Rulis, L. Ouyang, W. Ching, *Phys. Rev. B* **2011**, *83*, 024201.
 [45] P. Adhikari, A. M. Wen, R. H. French, V. A. Parsegian, N. F. Steinmetz, R. Podgornik, W. Y. Ching, *Sci. Rep.* **2014**, *4*.
 [46] L. Poudel, P. Rulis, L. Liang, W. Ching, *Phys. Rev. E* **2014**, *90*, 022705.
 [47] L. Poudel, A. M. Wen, R. H. French, V. A. Parsegian, R. Podgornik, N. F. Steinmetz, W. Y. Ching, *Chem. Phys. Chem.* **2015**, *16*, 1451.
 [48] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833.
 [49] C. Dharmawardhana, A. Misra, W. Y. Ching, *Sci. Rep.* **2014**, *4*.
 [50] W. Y. Ching, *J. Am. Ceramic Soc.* **1990**, *73*, 3135.
 [51] W. Y. Ching, P. Rulis, *Electronic Structure Methods for Complex Materials: The orthogonalized linear combination of atomic orbitals* 24; Oxford University Press, Oxford, UK, **2012**.
 [52] L. Poudel, A. M. Wen, R. H. French, V. A. Parsegian, R. Podgornik, N. F. Steinmetz, W. Y. Ching, *Chem. Phys. Chem.* **2015**, *16*, 1451.
 [53] L. L. Duan, Y. Mei, Q. G. Zhang, B. Tang, J. Z. Zhang, *J. Theor. Comput. Chem.* **2014**, *13*, 1440005.
 [54] D. Ben-Yaakov, D. Andelman, R. Podgornik, *J. Chem. Phys.* **2011**, *134*, 074705.
 [55] V. Parsegian, T. Zemb, *Curr. Opin. Colloid Interface Sci.* **2011**, *16*, 618.
 [56] B. Xia, V. Tsui, D. A. Case, H. J. Dyson, P. E. Wright, *J. Biomol. NMR* **2002**, *22*, 317.
 [57] R. Zhou, *Proteins: Struct. Funct. Bioinformatics* **2003**, *53*, 148.
 [58] V. Tsui, D. A. Case, *Biopolymers* **2000**, *56*, 275.

Received: 5 December 2015
 Accepted: 4 January 2016
 Published online 1 February 2016