

Equivalent potential of water for the electronic structure of glycine

Min Peng · Haoping Zheng

Received: 19 December 2009 / Accepted: 1 March 2010 / Published online: 10 April 2010
© Springer-Verlag 2010

Abstract First-principles, all-electron, ab initio calculations have been performed to construct an equivalent potential of water for the electronic structure of glycine (Gly) in solution. The calculation involved three steps. The first step was to search for the minimum-energy geometric structure of the Gly + nH₂O system. The second step was to calculate the electronic structure of Gly with the potential of water molecules via the self-consistent cluster-embedding method (SCCE), based on the result obtained in the first step. The last step was to calculate the electronic structure of Gly with the potential of dipoles after replacing the water molecules with dipoles. The results show that the occupied molecular orbitals of Gly are raised by about 0.0524 Ry on average due to the effect of water. The effect of water can be simulated well using the dipole potential. The equivalent potential obtained can be applied directly to electronic structure calculations of proteins in solution using the SCCE method.

Keywords Glycine · Electronic structure · Water · Free-cluster calculation · Self-consistent cluster-embedding calculation

Introduction

The structural taxonomy of proteins has been intensely studied by crystallography and nuclear magnetic resonance (NMR) spectroscopy for decades. More than thirty thousand protein structures are known. Recently, a great deal of attention has been paid to the properties and biological functions of various proteins in solution. According to quantum mechanics, the stationary Schrödinger equation can be used to elucidate the electronic structure (ground state and excited states) of a protein, which explains its properties, while the time-dependent Schrödinger equation and quantum statistical mechanics can be used to understand the processes that lead to the biological functions of a protein. However, computational research into proteins by first-principles calculations is still rather scarce due to the huge computational effort required when attempting to perform calculations for the large numbers of atoms contained in protein molecules, as well as the limitations on available computational power and limits on traditional calculation methods. To the best of our knowledge, there was no successful first-principles, all-electron, ab initio calculation of the electronic structure of any protein before the year 2000. Thus, a feasible computational technique that can greatly reduce the computational effort required to investigate proteins is needed.

In the last two decades, in both computational condensed matter physics and quantum chemistry, there has been a great deal of interest in developing so-called O(N) methods, for which the computational effort scales linearly with the number (N) of atoms [1–17]. Self-consistent cluster-embedding calculation (SCCE) [16, 17], a first-principles, all-electron, ab initio calculation method, was developed based on density functional theory (DFT). Unlike the traditional calculation method, one-electron wavefunctions obtained using the SCCE method are localized (i.e., each

M. Peng · H. Zheng
Physics Department, Tongji University,
Shanghai 200092, People's Republic Of China

H. Zheng (✉)
Shanghai Key Laboratory of Special Artificial Microstructure
Materials and Technology,
Shanghai 200092, People's Republic Of China
e-mail: zhenghp@tongji.edu.cn

one-electron wavefunction is localized to only part of the region occupied by the system). This approach possesses two advantages. First, the SCCE method can be applied to any complex system, and the localized valence electrons in a material can be better described by the localized wavefunctions obtained in the SCCE calculation. Second, the computational effort is reduced greatly while the accuracy of the calculation remains unchanged. This makes calculations of the electronic structures of proteins achievable. SCCE calculations have been successfully applied to several insulators, semiconductors, metals, crystals with defects and impurities, and surfaces [17–23]. The results obtained are in good agreement with the corresponding experimental data. So far, the electronic structures of three proteins with four three-dimensional structures have been obtained using the SCCE method [24–27]. Another protein was calculated by Sato et al. [28] and Yoshihiro et al. [29] using their own all-electron method.

However, the former calculations did not include the influence of the solvent. This was due to the limited computational power available and the following reasons: the hydrogen bonds between the protein and water molecules and those between water molecules easily break at room temperature. Water molecules move ceaselessly, and there is no fixed relative position between the water molecules and the protein. The water molecules exert a kind of average force on the protein, which mainly influences the three-dimensional structures of the protein. Thus, the electronic structure of a protein with a certain geometric structure in solution may be roughly approximated by the electronic structure of the protein when it has the same geometric structure but is not surrounded by water molecules. The validity of this approach has been demonstrated by our previous calculated results, which are in agreement with experimental data. In order to make the calculation more reliable, however, the effect of water should be considered.

Although the computational effort is reduced greatly in the SCCE method, the computational effort required for a protein containing thousands of atoms is still so large that it approaches the limits of supercomputers. Thus, it is impossible to add a large number of water molecules to the calculation. To calculate the electronic structure of a protein, it is necessary to construct an equivalent potential of water, which must be expressed simply, be easy to use, and most importantly, require little additional computational effort.

Computational studies of protein molecules, such as those focusing on protein folding and molecular dynamics calculations, have until now focused mainly on their geometric structures. An effective energy function for proteins and a Gaussian solvent-exclusion model for the solvation free energy have been developed [30]. The energy function, including implicit solvation, was first applied to protein folding calculations by Lazaridis and Karplus in

1998 [31]. There are some well-known models of the solvation effect, such as that of Onsager [32] and conductor-like screening models [33]. The influence of the solvent (water, ethanol, ethyleneglycol, and trifluoroethanol) on the stability of the peptide hydrogen bond was investigated by Guo and Karplus [34] using a first-principles calculation in 1994. In 1995, solvent effects on several small protein molecules, such as the bovine pancreatic trypsin inhibitor (BPTI) [35], were also studied. In recent years, some other treatments have been developed, such as the COSMO-RS approach [36] and solvent reaction field models [37] (the continuum model, PCM, IPCM). However, these are all designed for calculations of geometric structure, not electronic structure.

For a macro system and its macro properties, the effect of the water molecules can be reasonably treated as the effect of a continuous medium. However, for the lateral chain of a residue (which is comparable with a water molecule in size) and its localized molecular orbitals, a conductor-like polarizable continuous medium is clearly not a good approximation. We really need to consider the effect of the nearest individual water molecules. However, because the first-principles, all-electron, *ab initio* calculation of the electronic structure of protein is in its infancy, no suitable equivalent potential of water is available for electronic structure calculations of proteins using the SCCE method.

There are more than a hundred thousand proteins. However, we only need to construct the equivalent potentials of water for 20 amino acids—the building blocks of proteins. Nine amino acids have already been investigated: cysteine (Cys), lysine (Lys+), histidine (His), glutamic acid (Glu−), alanine (Ala), aspartic acid (Asp−), serine (Ser), threonine (Thr) and asparagine (Asn) [38–46]. In the present paper, we are concerned with glycine (Gly). Gly has ten atoms, and a hydrogen atom as its lateral chain. It is a nonpolar amino acid, and is hydrophobic. In a neutral solution, the N-terminus of Gly obtains a proton to become H_3N^+ , its C-terminus loses a proton to become COO^- , while the lateral chain is unpolarized and neither loses nor gains a proton. Therefore, Gly remains neutral in solution.

The present work is based on two considerations. First, we were not aiming to mimic the situation found in water at room temperature, but rather to mimic the potential that is exerted on the electronic structure of Gly by water. From the viewpoint of the electronic structure of Gly, it is reasonable to consider only the nearest water molecules that form hydrogen bonds with Gly and minimize the total energy. The reasons for this are as follows: (i) in our calculation, the Gly has already adopted its structure in solution, which does not change; (ii) there are only a limited number of water molecules close enough to form hydrogen bonds with the Gly; (iii) although they fluctuate significantly, the water molecules surrounding the Gly are

most probably at the positions that minimize the total energy of the Gly + nH₂O system; (iv) the potential of a removed dipole (such as a polar water molecule located far from the Gly) attenuates as 1/r², so the minimum number of water molecules (which are all the nearest water molecules that form hydrogen bonds with Gly and minimize the total energy, at least in a first-order approximation) exert most of the solvent’s potential on the electronic structure of Gly, no matter how the other water molecules are distributed—thus the electronic structure of Gly with water molecules at these positions can be considered to approximate the electronic structure of Gly in water. Second, the electronic structure of the protein can be calculated via the SCCE calculation. Dipoles consisting of point charges can easily be added to the SCCE calculation, and very little extra computational effort is required. On the other hand, the average potential of the polar water molecules can be reasonably simulated by dipoles. Thus, we choose dipoles consisting of point charges to simulate the potential of water on the electronic structure of Gly.

The present work consists of three steps. First, we determine the geometric structure of the Gly + 6H₂O system using the free-cluster calculation method. Second, based on the geometric structure obtained, the electronic structure of Gly with the potential of water molecules is calculated using the self-consistent cluster-embedding method. Finally, by replacing the water molecules with dipoles, the electronic structure of Gly with the potential of dipoles is calculated. The dipoles are adjusted so that Gly with the potential of dipoles and Gly with the potential of water molecules are very much alike in terms of electronic structure.

Theoretical model

The “free-cluster calculation” and the “self-consistent cluster-embedding (SCCE) calculation” methods have been described in detail elsewhere (see [16] and the website <http://www.esprotein.org.cn>), so we will only provide a brief overview of them here for completeness.

According to DFT [47, 48], the total energy of a system containing *N* electrons and *M* fixed nuclei can be written as (no relativistic effects are included; atomic units are used: e² = 2, h = 1, and 2m_e = 1, the unit of energy is the Rydberg constant Ry = e²/2a₀ = 13.6049eV):

$$E_v[\rho] = T_{ni}[\rho] + E_{xc}[\rho] + \iint \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' - 2 \times \sum_{l=1}^M \int \frac{\rho(\vec{r})Z_l}{|\vec{r} - \vec{R}_l|} d\vec{r} + \sum_{l \neq m}^M \frac{Z_l Z_m}{|\vec{R}_l - \vec{R}_m|}, \tag{1}$$

where *T_{ni}*[ρ] is the kinetic energy of a noninteracting electron system, and *E_{xc}*[ρ] is the exchange-correlation energy. In

deriving Eq. 1, Kohn and Sham have assumed the existence of a noninteracting electron system with the same ground-state charge density ρ(⃗r) as that of the real interacting system [48]. Each noninteracting electron can then be represented by a stationary state one-electron wavefunction ψ_{*n*}^σ(⃗r). Thus, the charge density ρ(⃗r) and kinetic energy *T_{ni}*[ρ] of the noninteracting system can be written as:

$$\rho(\vec{r}) = \rho^\uparrow(\vec{r}) + \rho^\downarrow(\vec{r}) = \sum_{occupied...i} |\psi_i^\uparrow(\vec{r})|^2 + \sum_{occupied...j} |\psi_j^\downarrow(\vec{r})|^2, \tag{2}$$

$$T_{ni}[\rho] = \sum_{occupied...i} \int \psi_i^{\uparrow*}(\vec{r})(-\nabla^2)\psi_i^\uparrow(\vec{r})d\vec{r} + \sum_{occupied...j} \int \psi_j^{\downarrow*}(\vec{r})(-\nabla^2)\psi_j^\downarrow(\vec{r})d\vec{r}. \tag{3}$$

Using Eqs. 2 and 3, the single-electron Schrödinger equation, i.e., the well-known Kohn–Sham equation [48], is obtained from the variation of functional (1) with respect to ψ_{*n*}^{σ*}(⃗r) under the normalization rule ∫ ψ_{*n*}^{σ*}(⃗r)ψ_{*n*}^σ(⃗r)d⃗r = 1 (*nσ* = 1, 2, … *N*) :

$$\left\{ -\nabla^2 + 2 \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' - 2 \sum_{l=1}^M \frac{Z_l}{|\vec{r} - \vec{R}_l|} + V_{xc}^\sigma(\vec{r}) \right\} \psi_n^\sigma(\vec{r}) = \varepsilon_n^\sigma \psi_n^\sigma(\vec{r}) \tag{4}$$

(*nσ* = 1, 2, … *N*),

where the exchange-correlation potential is

$$V_{xc}^\sigma(\vec{r}) = \frac{\partial E_{xc}(\vec{r})}{\partial \rho^\sigma(\vec{r})}. \tag{5}$$

The exact solution of Eq. 1 can be obtained if the following two conditions are satisfied: (i) *E_{xc}*[ρ] is exact; (ii) the trial one-electron wavefunctions ψ_{*n*}^σ(⃗r) are unconstrained when solving the Kohn–Sham equation (4) (as required by the variational principle). Then we should obtain a uniquely exact set of {ψ_{*n*}^σ(⃗r)} which corresponds to a uniquely correct ρ(⃗r). The Kohn–Sham equation (4), with an exact *V_{xc}*^σ(⃗r) and a correct ρ(⃗r), describes a situation in which one electron (represented by a ψ_{*n*}^σ(⃗r)) moves under the average potentials of all other electrons and nuclei. Because the potential produced by an electron (|ψ_{*n*}^σ(⃗r)|²) does not act on itself, the exact single-electron Hamiltonian in Eq. 4 includes only the potentials produced by ρ'(⃗r) = ρ(⃗r) - |ψ_{*n*}^σ(⃗r)|². This leads to two results: (a) different ψ_{*n*}^σ(⃗r) will correspond to different single-electron Hamiltonians (or *V_{xc}*^σ(⃗r)) in Kohn–Sham equation (4); (b) in general, the Hamiltonian acting on a ψ_{*n*}^σ(⃗r) has no symmetry of ρ(⃗r). As an example, consider a perfect periodic crystal. Although its charge density ρ(⃗r) is related

to the lattice periodicity, in general, the exact single-electron Hamiltonian of Kohn–Sham equation (4) will have no lattice periodicity.

In practical calculations, however, the two conditions above cannot be satisfied. Two approximations are made: (1) the exchange-correlation potential $V_{xc}^\sigma(\vec{r})$ is averaged over all one-electron states with spin σ (such as LDA); (2) the unconstrained condition is set aside, and each $\psi_n^\sigma(\vec{r})$ is constrained to satisfy a certain boundary condition. The following facts should be emphasized. The first approximation means that all $\psi_n^\sigma(\vec{r})$ in Eq. 4 now correspond to the same $V_{xc}^\sigma(\vec{r})$ (and single-electron Hamiltonian), which have symmetry in $\rho^\sigma(\vec{r})$, and all $\psi_n^\sigma(\vec{r})$ are now constrained by this added symmetry. However, this added constraint is not physically essential, and can be removed by a boundary condition. The second approximation means that all wavefunctions that do not satisfy the boundary condition are thrown away, even though they may be the true solutions of the system. When one kind of $\{\psi_n^\sigma(\vec{r})\}$ is chosen (a boundary condition is applied), this means that one kind of noninteracting electron is used to describe the real system approximately. By choosing different boundary conditions, we can use different kinds of $\{\psi_n^\sigma(\vec{r})\}$ to get different approximate $\rho(\vec{r})$. This raises a question: among the different kinds of $\{\psi_n^\sigma(\vec{r})\}$, which is the best one for describing a real system? The answer is: according to the variational principle, the calculated energy will be close to the true ground-state energy only if the trial one-electron wavefunctions $\psi_n^\sigma(\vec{r})$ describe real electrons well. For example, a set of Bloch functions can provide a good description of quasi-free electrons, but not localized electrons. The latter are best described by a set of localized one-electron wavefunctions.

We now assume that the first approximation has been made, so all $\psi_n^\sigma(\vec{r})$ correspond to the same single-electron Hamiltonian $H^\sigma(\vec{r})$ (without a boundary condition), and are constrained by the added symmetry in $\rho^\sigma(\vec{r})$. We now discuss the second approximation. There are two kinds of noninteracting electrons: extended and localized. They satisfy different boundary conditions, and correspond to different calculation methods.

Extended noninteracting electron model

Each one-electron wavefunction $\psi_n^\sigma(\vec{r})$ is constrained to spread over the whole region occupied by the system. In this model, Eq. 4 with a periodic boundary condition can be used for a perfect periodical crystal. The single-electron Hamiltonian with a periodic boundary condition has a periodicity of $\rho(\vec{r})$, the Bloch theorem is valid, and the band structure calculation is performed. For a free

cluster, Eq. 4 is solved with the finite boundary condition $\psi_n^\sigma(\vec{r}) \xrightarrow{\vec{r} \text{ go away from cluster}} 0$, and the single-electron Hamiltonian now has the point symmetry of the free cluster.

Localized noninteracting electron model

Each one-electron wavefunction $\psi_n^\sigma(\vec{r})$ is constrained to be distributed in part of the region occupied by the system. In this model, Eq. 4 is used for the self-consistent cluster-embedding (SCCE) calculation [16]. The system is divided into P embedded clusters, and N $\psi_n^\sigma(\vec{r})$ are divided into P groups. The $\psi_n^\sigma(\vec{r})$ in the i th group satisfy the i th set of special boundary conditions, and localize in the i th region (embedded cluster).

The details are as follows. Consider the i th embedded cluster. In this case, the other $P-1$ embedded clusters are regarded as the fixed environment, and their atoms are the surrounding atoms. Each surrounding atom has a core region. The electronic density of the i th embedded cluster is represented by $\rho_1(\vec{r})$, and the electronic density of the other $P-1$ embedded clusters is $\rho_2(\vec{r})$, which has a small overlap with $\rho_1(\vec{r})$. Because all N $\psi_n^\sigma(\vec{r})$ are localized, we have ($N = N_1 + N_2$)

$$\begin{aligned} \rho(\vec{r}) &= \sum_{\text{occupied} \dots n\sigma}^N |\psi_n^\sigma(\vec{r})|^2 \\ &= \sum_{\text{occupied} \dots n_1\sigma_1}^{N_1} |\psi_{n_1}^{\sigma_1}(\vec{r})|^2 + \sum_{\text{occupied} \dots n_2\sigma_2}^{N_2} |\psi_{n_2}^{\sigma_2}(\vec{r})|^2 \\ &\equiv \rho_1(\vec{r}) + \rho_2(\vec{r}) \end{aligned} \quad (2')$$

$$\begin{aligned} T_{ni}[\rho] &= \sum_{\text{occupied} \dots n\sigma}^N \int \psi_n^{\sigma*}(\vec{r}) (-\nabla^2) \psi_n^\sigma(\vec{r}) d\vec{r} \\ &= \sum_{\text{occupied} \dots n_1\sigma_1}^{N_1} \int \psi_{n_1}^{\sigma_1*}(\vec{r}) (-\nabla^2) \psi_{n_1}^{\sigma_1}(\vec{r}) d\vec{r} \\ &\quad + \sum_{\text{occupied} \dots n_2\sigma_2}^{N_2} \int \psi_{n_2}^{\sigma_2*}(\vec{r}) (-\nabla^2) \psi_{n_2}^{\sigma_2}(\vec{r}) d\vec{r} \\ &\equiv T_{ni}[\rho_1] + T_{ni}[\rho_2]. \end{aligned} \quad (3')$$

A zero-value term $\int \rho_1(\vec{r}) V_{or}(\vec{r}) d\vec{r}$ is added to the right hand side of Eq. 1. For fixed $\rho_2(\vec{r})$, using Eqs. 2' and 3', the variational principle now leads to the basic equation of the

SCCE method (the i th cluster has $N_{i\sigma}$ electrons with spin σ) [16]:

$$\left\{ \begin{aligned} &-\nabla^2 + 2 \int \frac{\rho_1(\vec{r}') + \rho_2(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' - 2 \sum_{l=1}^M \frac{Z_l}{|\vec{r} - \vec{R}_l|} + V_{xc}^\sigma(\vec{r}) + V_{or}(\vec{r}) \\ &\psi_n^\sigma(\vec{r}) = \varepsilon_n^\sigma \psi_n^\sigma(\vec{r}) \end{aligned} \right\}$$

$$(n\sigma = 1, 2, \dots, N_{i\sigma}), \tag{4'}$$

where the $\psi_n^\sigma(\vec{r})$ represent only the noninteracting electrons localized in and around the i th embedded cluster.

Apparently, Eq. 4' is exactly the same as Kohn–Sham equation (4) except for $V_{or}(\vec{r})$. The $V_{or}(\vec{r})$ is defined as

$$V_{or}(\vec{r}) = \begin{cases} 2 \sum_{l=1}^{M_2} \frac{Z_l}{|\vec{r} - \vec{R}_l|} & \text{if } \vec{r} \text{ is in the core regions} \\ 0 & \text{of surrounding atoms} \\ & \text{otherwise} \end{cases}, \tag{6}$$

where M_2 is the number of surrounding atoms. $V_{or}(\vec{r})$ cancels the nuclear Coulomb potential in the core regions of all surrounding atoms. The cluster electrons will only feel a positive electron–electron Coulomb potential in these regions, and will be forced out. Thus, the $\psi_n^\sigma(\vec{r})$ in Eq. 4' satisfy a special boundary condition caused by the $V_{or}(\vec{r})$:

$$\psi_n^\sigma(\vec{r})|_{\vec{r} \text{ is in the core regions of surrounding atoms}} = 0. \tag{7}$$

The physical reasons for boundary condition (7) are given in [16]. Consider Eqs. 2', 6 and 7. It is easy to see that, as long as boundary condition (7) is satisfied, we have $\int \rho_1(\vec{r}) V_{or}(\vec{r}) d\vec{r} = 0$. Thus, the $V_{or}(\vec{r})$ in Eq. 4' does not change energy functional (1). The only effect of $V_{or}(\vec{r})$ is that it causes the special boundary condition (7). In fact, what we have done is simply to transform boundary condition (7) into an equivalent potential $V_{or}(\vec{r})$ in Eq. 4'. Thus, Eq. 4' is just Kohn–Sham equation (4) with the special boundary condition (7). In addition, there is a special finite boundary condition for $\psi_n^\sigma(\vec{r})$ because of its locality:

$$\psi_n^\sigma(\vec{r}) \xrightarrow{\vec{r} \text{ go away from the embedded cluster}} 0 \tag{8}$$

The boundary conditions (7) and (8) are different for different embedded clusters. By calculating all P embedded clusters one by one, Eq. 4' gives a complete set of one-electron wavefunctions for the whole system that minimizes the total energy in Eq. 1 [16].

The optimum values of the core radii of the surrounding atoms are determined via two criteria: (i) there is no collapse disaster; (ii) the number of cluster electrons that remain in the surrounding core regions is minimized. In general, boundary condition (7) can be satisfied to a high precision, and it is found that the results are not sensitive to the core radii if they are close to the optimal values.

System, computational procedure and results

Both the geometric and the electronic structures of the Gly + 6H₂O system were found using the free-cluster calculation. Considering the degrees of freedom of the geometric structure of Gly in solution, six water molecules are clearly not enough to describe the effect of water on the *geometric* structure of Gly in solution. However, our concern is the *electronic* structure rather than the geometric structure. We believe that six water molecules are appropriate for describing the effect of water on the electronic structure of Gly in solution, due to the following reasons:

- (1) Both the properties and the functions of an amino acid are determined mainly by the molecular orbitals near the HOMO, which are easily affected by the solution environment. On the other hand, it is reasonable to assume that the molecular orbitals that have energies which are much lower than the HOMO are hardly affected. So we can actually use six water molecules to simulate the effect of water on the molecular orbitals of Gly near the HOMO. If we choose ten molecular orbitals near the HOMO, the number of degrees of freedom is small, and six water molecules should be reasonable, because their atomicity (18 atoms) is larger than that of Gly (ten atoms).
- (2) The valence electrons in the amino acid are localized electrons except for the π -electrons of the benzene-type ring, which are shared with all of the six carbon atoms. The molecular orbitals of Gly near the HOMO are mainly localized electrons from the amido (H₃N⁺) and carboxyl (COO⁻) groups. Thus, only the water molecules that surround these two groups need to be considered. From the viewpoint of hydrogen bonding, we use the minimum number of water molecules: each H atom of the N-terminal H₃N⁺ group may form a hydrogen bond with an O atom of a water molecule, each O atom of the C-terminal COO⁻ group may form a hydrogen bond with an H atom of a water molecule, and each H atom near C α may form a hydrogen bond with an O atom of a water molecule. Since each water molecule has one O atom and two H atoms, the minimum number of water molecules that must be near the three groups are as follows: three for the N-terminal H₃N⁺ group, one for the C-terminal COO⁻ group, and two for H atoms near C α . Therefore, six water molecules will perform well if they are at the correct geometric positions.
- (3) Without question, if more molecules are used, a better approximation will be achieved. However, consider the following two facts which indicate that six water molecules may be the only choice that is both acceptable and achievable. First, when a water

molecule is near the amino acid, it will impede other water molecules from approaching the amino acid. The potential of a removed dipole attenuates as $1/r^2$. Therefore, the minimum number of water molecules (i.e., those located closest to the Gly, and which form hydrogen bonds with it), at least to a first-order approximation, exert most of the effect of the solvent on the electronic structure of Gly. Second, if we increase the number of water molecules from the minimal number of six to nine, for example, there can be more than 30 geometric structures, all of which have almost the same total energy. This degeneracy makes calculations impossible: investigating 30 geometric structures requires a huge calculation effort. This also makes a quantitative analysis of convergence impossible. On the other hand, the complex interactions between water molecules are of no interest here. Our aim is to construct a simple and easy-to-use potential. Thus, since we are limited by our computational capacity, we choose six water molecules.

Determination of the geometric structure of the Gly + 6H₂O system

Initial geometric structure of the Gly + 6H₂O system

The coordinates of the ten atoms of Gly in solution (given in Table 1) originate from a PDB structure file provided by the Laboratory of Mass Spectrometry and Gaseous Ion Chemistry at the Rockefeller University (<http://prowl.rockefeller.edu/aainfo/struct.htm>).

In a neutral solvent (pH 7), the N-terminus of Gly gains a proton to become H₃N⁺, and the C-terminus loses a proton to become COO⁻. The polar water molecules mainly affect the charged H₃N⁺ and COO⁻ groups of Gly, and the H in its lateral chain. At the start, the six water molecules

were distributed semi-randomly around the Gly, based on our previous experience; i.e., three water molecules were near the H₃N⁺, one was near the COO⁻, and two were near the two H atoms. Each water molecule, depending on its initial position and orientation, can form a hydrogen bond with an H or O atom of Gly, lowering the total energy of the Gly + 6H₂O system.

In all calculations performed in this study, we used the Von Barth and Hedin [49] form of the exchange-correlation potential, as parametrized by Rajagopal and coworkers [50]. The optimized linear combination of the Gaussian basis sets of C, N, O, and H atoms was used [51–55], parts of the original bases were uncontracted, several diffuse bases were inserted, and two polarization functions were added. These were the same as those used in other electronic structure calculations for proteins [24–27]; i.e., C, 8s6p, 26 Gaussian bases; N, 8s7p, 29 Gaussian bases; O, 8s7p, 29 Gaussian bases; H, 8s1p, 11 Gaussian bases. The total number of Gaussian bases was 500. In order to calculate the exchange-correlation energy, the space occupied by the Gly + 6H₂O system was divided into 308,002 grid points.

The software “free cluster calculation” was developed by the group of Prof. Callaway in the Department of Physics and Astronomy, Louisiana State University (USA) [56]. The electronic structures of many molecules and clusters have been calculated using this software [57–63]. By solving Kohn–Sham equation (4) self-consistently, we obtained the electronic structure, the total energy, and the force exerted on each atomic nucleus.

Adjusting the six water molecules

In the Gly + 6H₂O system, the geometric structures of Gly and the water molecules were kept fixed. Only the relative positions between the water molecules and Gly were adjusted. Thus, during the optimization, all of the nuclei in Gly were fixed, and the nuclei in the water molecules were moved according to the forces applied, while the geometric structure of each water molecule remained unchanged. The electronic structure of the Gly + 6H₂O system was self-consistently calculated using the first-principles, all-electron, ab initio calculation method.

The six water molecules were adjusted one by one. In other words, each time, the nuclei of five of the water molecules were fixed, while the nuclei of the other water molecule were moved in the direction of the strongest force applied. When adjusting the nuclei of this one water molecule, we used eight step lengths, ranging from small to very large, to stop the adjustment from falling into a local minimum. So, in all probability, our final optimized structure does not depend on the initial geometry. The new positions of the nuclei of this one water molecule that yield the lowest total energy were then found. Then we

Table 1 Atomic coordinates of Gly

No.	Atom	X (Å)	Y (Å)	Z (Å)
1	C _α	0.066	-0.181	0.855
2	N	-1.335	0.173	0.469
3	C	1.036	0.600	0.002
4	O	2.237	-0.457	0.158
5	O	0.601	1.407	-0.877
6	H	0.225	-1.247	0.703
7	H	0.225	0.063	1.903
8	H	-2.000	-0.363	1.053
9	H	-1.487	1.186	0.614
10	H	-1.487	-0.060	-0.528

moved the nuclei of another water molecule that experienced the strongest applied force. After the nuclei of all six water molecules had been adjusted, one course was finished. The second course was then started. After hundreds of adjustments, the geometric structure of the Gly + 6H₂O system with the lowest total energy was obtained. The total energy of the final structure was -1479.1572 Ry, 0.4503 Ry lower than that of the initial structure. The final atomic coordinates of the six water molecules are given in Table 2. The final geometric structure of the Gly + 6H₂O system is shown in Fig. 1, where most of the atoms are numbered according to their numbers in Tables 1 and 2.

Of course, we cannot be absolutely certain that we have achieved the best structure—i.e., the best relative positions between the water molecules and Gly which lead to the lowest total energy (global minimum)—even if we were to perform another hundred adjustment steps. However, we can be sure that the room for further adjustment is very small, and that the total energy of our final geometric structure is very close to the lowest energy of the Gly + 6H₂O system. Therefore, if we use our final structure, the calculated electronic structure for Gly in the potential of water will be a good approximation to the real electronic structure of Gly in a real aqueous solution. This is due to the following reasons. First, we do not study the relative positions between the water molecules and Gly. Second, there are no fixed hydrogen bonds between the water molecules and the protein in solution, so there are no fixed

relative positions between the water molecules and the protein. The water molecules are, most probably, at the positions that cause the total energy of the system to be minimized. Third, in order to reduce the computational effort, charge-density fitting is used in both the free-cluster and the band structure calculations: a pseudo charge density that differs from the real one but gives a total energy that is very close to that calculated using the real charge density is used to calculate the electronic structure. It is considered that the electronic structure calculated using this pseudo charge density is a good approximation to the real electronic structure of the system.

Electronic structure of Gly in the potential of water molecules

Based on the geometric structure of the Gly + 6H₂O system determined in “Initial geometric structure of the Gly + 6H₂O system,” the system was then divided into seven embedded clusters for the self-consistent cluster-embedding calculation (SCCE). Gly was the first cluster, and each water molecule was treated as one cluster. In the SCCE calculation, the total potential was the same as that in the free cluster calculation; however, the one-electron wavefunctions were changed from being extended to localized: each one-electron wavefunction was localized in the region of a cluster. Thus, the electronic structure of Gly was separated from that of the water molecules. The software used for the SCCE calculation was developed by our group [64], and has been applied to several insulators, semiconductors, metals, crystals with defects and impurities, surfaces, and proteins [17–27].

The calculation involved two kinds of iterations: (i) *intracluster* iterations and (ii) *intercluster* iterations. Converged results were achieved after ten intercluster iterations. Table 3 shows the eigenvalues and Mulliken populations of ten orbitals near the HOMO of Gly in the potential of six water molecules.

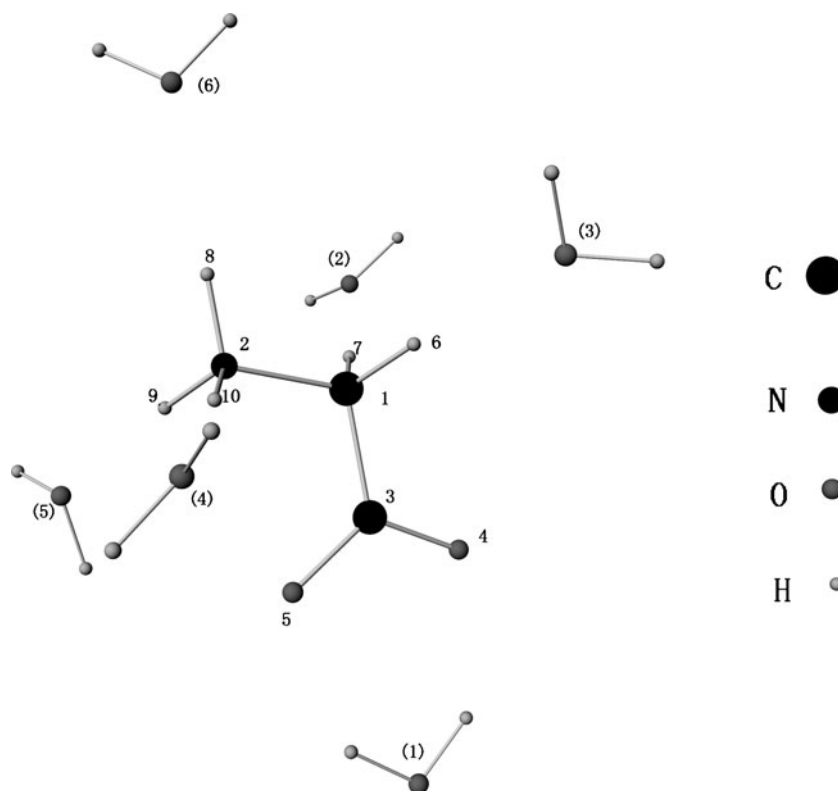
Equivalent potential of water on Gly simulated by dipoles

Each water molecule was then replaced by a dipole: the O atom was replaced by a negative point charge, while the two H atoms were replaced by a positive charge located in the middle of the line connecting the two H atoms. Adjusting the point charges and the positions of dipoles, the electronic structure of Gly was recalculated using the SCCE calculation. There were no *intercluster* iterations this time; only the *intracluster* iterations of the Gly cluster were required, because the dipoles had no electrons.

Table 2 Final atomic coordinates of the six water molecules

No.	Atom	X (Å)	Y (Å)	Z (Å)
[1]	11–O	3.0658	2.2373	–1.7807
	12–H	3.2693	1.6047	–1.0918
	13–H	2.1093	2.2625	–1.8084
[2]	14–O	0.3622	0.4175	4.1080
	15–H	0.2766	1.2381	4.5934
	16–H	0.6549	–0.2182	4.7611
[3]	17–O	0.5601	–3.4100	0.2116
	18–H	–0.1654	–4.0315	0.2728
	19–H	1.3443	–3.9583	0.1868
[4]	20–O	–1.8806	–0.4901	–2.5152
	21–H	–2.0843	–1.3612	–2.8556
	22–H	–2.1367	0.1079	–3.2174
[5]	23–O	–1.7491	3.0273	0.5691
	24–H	–1.0210	3.5585	0.2467
	25–H	–2.1720	3.5756	1.2230
[6]	26–O	–3.3764	–1.4555	2.3111
	27–H	–3.1936	–2.2446	2.8212
	28–H	–4.1551	–1.0795	2.7214

Fig. 1 Final geometric structure of the Gly + 6H₂O system



It is important to note that the results from the free-cluster calculation of Gly + 6H₂O could not be used as fitting criteria to adjust the dipoles in the SCCE calculation of Gly + six dipoles, because the two systems had different numbers of electrons and the distribution regions of the electrons in the two systems were not the

same. On the other hand, the SCCE calculation of Gly + 6H₂O (i.e., the electronic structure of cluster Gly obtained above) can be considered to approximate the electronic structure of Gly in water, and thus was suitable for use as fitting criteria. The dipoles were therefore adjusted accordingly. In order to evaluate the difference

Table 3 Some of the eigenvalues and Mulliken populations of Gly in the potential of six water molecules

State	Energy (Ry)	Mulliken population							
		C		N		O		H	
		<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>
22	-0.0967	3.4580	0.6858	0.0449	0.2961	-0.1065	-0.1280	-3.2538	0.0037
21	-0.1685	0.5276	0.0337	2.5597	1.0919	-0.0034	0.0832	-3.3021	0.0093
The above states are unoccupied									
20	-0.3420	-0.0044	0.0052	0.0489	0.0127	-0.0041	0.9751	-0.0338	0.0004
19	-0.3463	-0.0030	0.0018	0.1052	0.0263	0.0137	0.9367	-0.0813	0.0006
18	-0.4075	0.2562	0.1521	0.0208	-0.0064	-0.0520	0.7061	-0.0774	0.0005
17	-0.5902	-0.0001	0.3780	0.0000	0.0008	0.0001	0.5014	0.1169	0.0028
16	-0.6485	0.0154	0.2123	0.0003	0.0118	0.2097	0.5532	-0.0033	0.0006
15	-0.7158	0.0388	0.2220	0.0026	0.0018	0.2465	0.4684	0.0183	0.0015
14	-0.7753	0.0000	0.5454	0.0000	0.0283	0.0000	0.0774	0.3347	0.0141
13	-0.8696	0.1428	0.4209	0.0279	0.0707	0.0413	0.1265	0.1578	0.0122

Note: because of the small values associated with them, *d*-electrons are not given in this table or in subsequent tables

between the electronic structure calculated here and that obtained in the previous section, two quantitative criteria were established:

(1) The mean square deviation of eigenvalues

$$\overline{\Delta E^\sigma} = \frac{1}{N^\sigma} \left[\sum_{n=1}^{N^\sigma} (\varepsilon_n^\sigma - \varepsilon_{n0}^\sigma)^2 \right]^{1/2},$$

where ε_n^σ and ε_{n0}^σ are the eigenvalues of the n th molecular orbital with spin σ calculated in this section and in the previous section, respectively. N^σ is the number of electrons with spin σ .

(2) The equivalent mean square deviation of charge density

$$\overline{\Delta C^\sigma} = \frac{1}{N} \left[\sum_{i=1}^N \sum_{j=1}^N \left(\sum_{n=1}^{N^\sigma} C_{ni}^{\sigma*} C_{nj}^\sigma - \sum_{n=1}^{N^\sigma} C_{n0}^{\sigma*} C_{n0}^\sigma \right)^2 \right]^{1/2},$$

where C_{ni}^σ and C_{n0}^σ are the expansion coefficients of the eigenfunctions of the n th molecular orbitals with spin σ calculated in this section and in the previous section, respectively. N is the number of Gaussian bases used to expand the one-electron wavefunction. (Note the charge density:

$$\begin{aligned} \rho^\sigma(\vec{r}) &= \sum_{n=1}^{N^\sigma} |\psi_n^\sigma(\vec{r})|^2 \\ &= \sum_{n=1}^{N^\sigma} \left[\sum_{i=1}^N C_{ni}^{\sigma*} U_i^*(\vec{r}) \right] \left[\sum_{j=1}^N C_{nj}^\sigma U_j(\vec{r}) \right] \\ &= \sum_{i=1}^N \sum_{j=1}^N \left[\sum_{n=1}^{N^\sigma} C_{ni}^{\sigma*} C_{nj}^\sigma \right] U_i^*(\vec{r}) U_j(\vec{r}), \end{aligned}$$

where the Gaussian bases $U_i(\vec{r})$ are identical in the two calculations.)

The *intracluster* iterations of the Gly cluster were performed using the SCCE method, and the electronic structure of Gly in the potential of six dipoles was obtained. The values of the two criteria were then calculated and compared. Initially, the mean square eigenvalue deviation was 1.0115×10^{-2} , and the equivalent mean square charge density deviation was 3.2947×10^{-2} . After hundreds of adjustments, both values reached minima of 3.7131×10^{-3} and 3.1830×10^{-2} , respectively. This means that the electronic structure of Gly in the potential of dipoles is close to that obtained in the potential of water molecules.

There are a number of studies in the literature that address the problem of simplified discrete solvent representation via point charges; each has its own advantages and disadvantages. According to the website <http://www.lsbu.ac.uk/water/models.html>, 23 water models and more than 1200 papers can be found on this subject. The review written by Guillot [65] listed 46 distinct water models, which indirectly indicates their lack of success in quantitatively reproducing the properties of real water. In our previous studies of the construction of equivalent potentials [40, 42], we tried two popular three-charge water models, TIP4P-FQ [66] and SPC [67, 68], as well as a dipole model. However, when calculating the electronic structure of an amino acid, the more complicated three-charge water

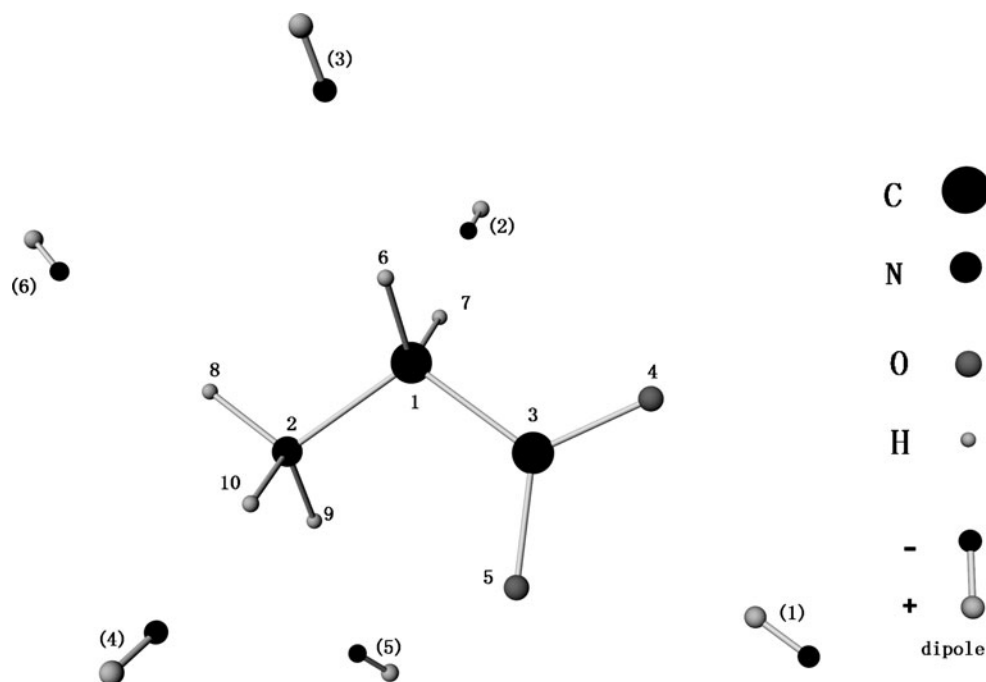
models did not give a better fit than the simple dipole model (this will be discussed in another paper). Two facts should be emphasized here. First, we are not constructing a general model of water that is suitable for all cases here, since this is a complex problem that has still a lot of issues. We are constructing a special equivalent potential of water that is only suitable for calculations of the electronic structures of proteins in solution. For example, the dipoles constructed for Gly will only be applied to the Gly present on the surface of a protein in solution. Second, we are not aiming to construct an exact special equivalent potential of water. Actually, this does not exist: it is impossible to make the electronic structure of Gly in the potential of dipoles identical to the electronic structure of Gly in the potential of water molecules. Our aim is to construct a simple and easy-to-use potential, which, at least to a first-order approximation, exerts most of the effect of the solvent on the electronic structure of the protein. Therefore, we chose the dipole model.

The initial charges of all six dipoles were set to the same value, 0.500e. The distance between a positive charge and a negative charge was 0.5858 Å and did not change during the calculations. The charges of the six dipoles were adjusted one by one. After hundreds of adjustments, the two criteria reached their minima. The final charges of the six dipoles were as follows: the charges of the three dipoles around the amidogen were 0.550e, 0.670e and 0.615e, respectively; the charge of the one dipole near the carboxyl was 1.030e; the charges of the two dipoles that approach the hydrogen H were 0.760e and 0.800e, respectively. The charges and coordinates of the six dipoles are listed in Table 4, and the geometric structures of the Gly + six dipoles are shown in Fig. 2. Table 5 shows the eigenvalues and Mulliken populations of ten orbitals near the HOMO of Gly in the potential of six dipoles.

Table 4 Point charges and coordinates of six dipoles

No.	Charge (e)	X (Å)	Y (Å)	Z (Å)
(1)	-1.030	3.0658	2.2373	-1.7807
	1.030	2.6893	1.9336	-1.4501
(2)	-0.760	0.3569	0.4111	4.0689
	0.760	0.4604	0.5036	4.6381
(3)	-0.800	0.5601	-3.4100	0.2116
	0.800	0.5895	-3.9949	0.2298
(4)	-0.550	-1.8807	-0.4901	-2.5152
	0.550	-2.1105	-0.6266	-3.0365
(5)	-0.670	-1.7491	3.0273	0.5691
	0.670	-1.5965	3.5670	0.7383
(6)	-0.615	-3.3764	-1.4555	2.3111
	0.615	-3.6744	-1.6620	2.7713

Fig. 2 Final geometric structure of the Gly + six dipoles system



Discussion

In order to acquire a knowledge of the effect of water solvent on the electronic structure of Gly, the electronic structure of an isolated Gly was calculated using a free-cluster calculation. The total energy of the isolated Gly was -565.9391 Ry. The eigenvalues and Mulliken populations of ten orbitals near the HOMO of the isolated Gly are shown in Table 6.

Shown in Table 7 are the eigenvalues of orbitals 13 to 22 of Gly for three cases: in the potential of dipoles, in the potential of water molecules, and in no applied potential.

The last row in Table 7 gives the energy gap between states 20 and 21. A sketch map comparing the three sets of eigenvalues of Gly is shown in Fig. 3.

The properties of Gly are determined mainly by the molecular orbitals near the HOMO. In Tables 3, 5 and 6, Mulliken population analysis shows that, in all three cases, the molecular orbitals below the HOMO are similar. Orbitals 20 and 19 are occupied mainly by the $2p$ electrons of the oxygen in the backbone COO^- . Orbital 18 is a hybridized state contributed to mainly by the $2p$ electrons of the oxygen in the backbone COO^- and the $2p$ and $2s$ electrons of C_α . Orbital 17 is a hybridized state which is

Table 5 Some of the eigenvalues and Mulliken populations of Gly in the potential of six dipoles

State	Energy (Ry)	Mulliken population							
		C		N		O		H	
		<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>
22	-0.0122	0.1770	0.1227	0.0713	0.2252	0.0225	-0.0615	0.4397	0.0031
21	-0.0901	0.3152	-0.0465	1.4481	0.2485	-0.0024	0.0307	-1.0032	0.0077
The above states are unoccupied									
20	-0.3352	0.0165	0.0305	0.0431	0.0031	-0.0264	0.9405	-0.0077	0.0004
19	-0.3390	0.0001	0.0035	0.0008	-0.0013	-0.0003	0.9939	0.0032	0.0003
18	-0.4166	0.0853	0.1329	0.0088	0.0023	0.0052	0.7776	-0.0126	0.0006
17	-0.6049	0.0000	0.4041	0.0000	-0.0004	0.0001	0.4937	0.0993	0.0032
16	-0.6331	0.0077	0.1997	0.0026	0.0197	0.2421	0.5208	0.0064	0.0009
15	-0.7196	0.0565	0.1491	0.0022	0.0050	0.3008	0.4839	0.0018	0.0007
14	-0.8011	0.0000	0.5624	0.0000	0.0409	0.0000	0.0581	0.3215	0.0171
13	-0.8740	0.0916	0.4120	0.0176	0.1785	0.0554	0.0481	0.1825	0.0143

Table 6 Some of the eigenvalues and Mulliken populations of an isolated Gly

State	Energy (Ry)	Mulliken population							
		C		N		O		H	
		<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>	<i>s</i>	<i>p</i>
22	−0.0716	0.0140	0.0454	0.0012	−0.5054	0.0387	−0.0076	1.4049	0.0087
21	−0.1904	0.1076	0.0016	−0.4496	0.0449	−0.0117	0.0295	1.2649	0.0129
The above states are unoccupied									
20	−0.3669	0.0208	0.0492	0.0205	0.0104	−0.0263	0.9066	0.0183	0.0005
19	−0.3782	0.0000	0.0047	0.0000	−0.0035	0.0000	0.9902	0.0081	0.0004
18	−0.4438	−0.0445	0.1958	0.0361	0.0113	−0.0024	0.7889	0.0140	0.0008
17	−0.6530	0.0000	0.4108	0.0000	−0.0022	0.0000	0.5075	0.0819	0.0019
16	−0.6681	0.0047	0.1803	0.0032	0.0158	0.2772	0.5145	0.0036	0.0006
15	−0.7636	0.0607	0.1370	0.0011	0.0045	0.2994	0.4945	0.0022	0.0005
14	−0.8895	0.0000	0.5518	0.0000	0.0347	0.0000	0.0341	0.3620	0.0173
13	−0.9510	0.1137	0.4913	0.0082	0.0689	0.0681	0.0486	0.1901	0.0112

mainly contributed to by the $2p$ electrons of oxygen and carbon in the backbone COO^- . Orbitals 16 and 15 are occupied mainly by the $2p$ electrons of oxygen and carbon, and the $2s$ electrons of carbon in the backbone COO^- . Orbitals 14 and 13 are hybridized states that are mainly contributed to by $2p$ electrons from the carbon and the $1s$ electrons of the hydrogens in $\text{C}_\alpha\text{H}_2$, as well as the $2p$ electrons of the carbon in the backbone COO^- .

Upon comparing the second column with the third one in Table 7, and Fig. 3a with Fig. 3b, it is found that water does not markedly change the relative positions of the occupied molecular orbitals. The main effect of water is to narrow the energy gap by about 1.7%, and raise all eigenvalues by about 0.0524 Ry on average. Among the eight orbitals, the eigenvalues of orbitals 20, 19 and 18 are raised by 0.0249 Ry, 0.0319 Ry and 0.363 Ry, respectively, much less than those of the other orbitals.

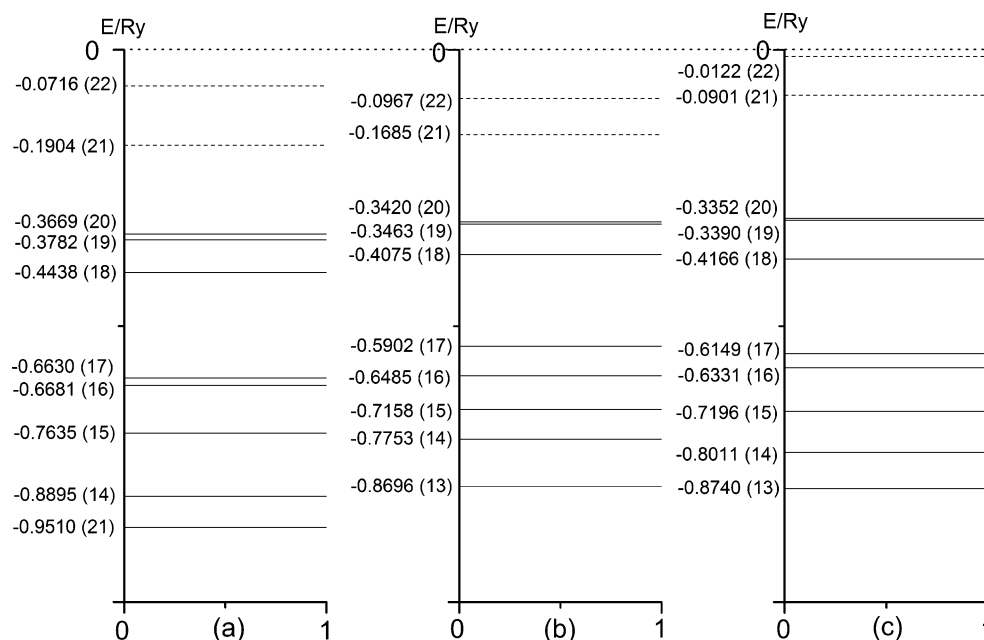
Comparisons between the third and fourth columns in Table 7 and between Figs. 3b and 3c reveal that, below the HOMO, eigenvalues of Gly in the potential of dipoles are very close to those of Gly in the potential of water, except for orbital 14, which decreases by 0.0258 Ry. However, above the HOMO, the energy gap calculated using the potential of dipoles is broadened by 41% compared with that calculated using the potential of water molecules. This suggests that, in the potential of dipoles, unoccupied orbitals are overestimated. Because unoccupied orbitals make no contribution to the charge density, it can be concluded that the calculated potential of dipoles gives a good simulation of the effect of water on the electronic structure of Gly.

Although it is a special equivalent potential of water, our dipole potential is transferable. We will use dipoles to construct the equivalent potentials of water for the

Table 7 Three sets of eigenvalues of Gly

State	Eigenvalues (Ry)		
	No potential	Potential of water molecules	Potential of dipoles
22 (unoccupied)	−0.0716	−0.0967	−0.0122
21 (unoccupied)	−0.1904	−0.1685	−0.0901
20 (E_F)	−0.3669	−0.3420	−0.3352
19	−0.3782	−0.3463	−0.3390
18	−0.4438	−0.4075	−0.4166
17	−0.6530	−0.5902	−0.6049
16	−0.6681	−0.6485	−0.6331
15	−0.7636	−0.7158	−0.7196
14	−0.8895	−0.7753	−0.8011
13	−0.9510	−0.8696	−0.8740
$E_g = E_{21} - E_{20}$	0.1765	0.1735	0.2451

Fig. 3 Comparison of eigenvalues among the cases with no potential (a), with the potential of water molecules (b), and with the potential of dipoles (c), where a *solid line* represents an occupied orbital and a *dashed line* refers to an unoccupied orbital



electronic structures of 20 amino acids. When this is finished, it will be possible to apply the obtained equivalent potentials to the SCCE calculation, as well as any first-principles, all-electron, ab initio calculation method, as long as they are used to calculate the electronic structure of a protein in solution. Amino acids lose their water molecules when they turn into amino acid residues, and they are then combined into several interlaced polypeptide chains, thereby forming a protein. Each chain has only one N-terminal H_3N^+ , one C-terminal COO^- , and many lateral chains. For the electronic structure of a protein with a known geometric structure in solution, water solvent does not affect the molecular orbitals of amino acid residues in the interior of the protein because of screening effects. Thus, we only need to put the dipoles (according to our dipole potentials) near the N-terminus and C-terminus of a chain, and near the tips of lateral chains of the amino acid residues that are on the surface of the protein. Both the charges and the locations of the dipoles around those amino acid residues can be obtained using our program if the locations of those amino acid residues are inputted into the program. This should help to more reliably determine the active sites and the reactive sites of a protein in solution, which are usually at the N-terminus and C-terminus of a chain, and on the tips of lateral chains of amino acid residues that are on the surface of the protein.

Conclusions

We have performed first-principles, all-electron, ab initio calculations to construct the equivalent potential of water for the electronic structure of Gly in solution. The

calculations involve three steps. First, we used the free-cluster calculation to obtain the geometric structure of the Gly + $6\text{H}_2\text{O}$ system that has the lowest total energy. Second, based on the geometric structure obtained, we used the self-consistent cluster-embedding (SCCE) calculation to acquire the electronic structure of Gly in the potential of water molecules. Finally, we replaced the six water molecules with six adjustable dipoles. The dipoles were adjusted until the electronic structure of Gly with the potential of dipoles was close to the electronic structure of Gly with the potential of water molecules.

We compared the electronic structures of Gly in three cases: with the potential of water, with the potential of dipoles, and without any external potential. The main effect of water on the electronic structure of Gly is to narrow the energy gap by about 1.7% and raise all eigenvalues by about 0.0524 Ry on average. A simple potential of dipoles can be used to simulate the effect of water on the electronic structure of Gly: for the molecular orbitals under the HOMO, the eigenvalues in the first two cases above are very close to each other. Thus, a simple dipole potential that is easy to use and requires little additional computational effort was established. More equivalent potentials for other amino acids will now be constructed. The results of this work, together with those from previous studies and forthcoming studies, will be directly applied to electronic structure calculations of proteins in solution.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. 30970694). The work was also supported by the Shanghai Supercomputer Center. The calculations were performed on the supercomputer DAWN 5000A of the Shanghai Supercomputer Center of China.

References

1. Yang WT (1991) Direct calculation of electron density in density-functional theory. *Phys Rev Lett* 66:1438–1441
2. Cortona P (1991) Self-consistently determined properties of solids without band structure calculations. *Phys Rev B* 44:8454–8458
3. Galli G, Parrinello M (1992) Large scale electronic structure calculations. *Phys Rev Lett* 69:3547–3550
4. Mauri F, Galli G, Car R (1993) Orbital formulation for electronic-structure calculations with linear system-size scaling. *Phys Rev B* 47:9973–9976
5. Li XP, Nunes RW, Vanderbilt D (1993) Density-matrix electronic-structure method with linear system-size scaling. *Phys Rev B* 47:10891–10894
6. Ordejón P, Drabold DA, Martin RM, Grumbach MP (1995) Linear system-size scaling methods for electronic-structure calculations. *Phys Rev B* 51:1456–1476
7. Yang WT, Lee TS (1995) A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *Chem Phys* 103:5674–5678
8. Kohn W (1996) Density functional and density matrix method scaling linearly with the number of atoms. *Phys Rev Lett* 76:3168–3171
9. Ordejón P, Artacho PE, Soler JM (1996) Self-consistent order-N density-functional calculations for very large systems. *Phys Rev B* 53:10441–10444
10. Baer R, Gordon MH (1997) Sparsity of the density matrix in Kohn–Sham density functional theory and an assessment of linear system-size scaling methods. *Phys Rev Lett* 79:3962–3965
11. Klessinger M, Mcweeny R (1965) Self-consistent group calculations on polyatomic molecules. *J Chem Phys* 42:3343–3354
12. Li JB, McWeeny R (2002) VB2000: pushing valence bond theory to new limits. *Int J Quantum Chem* 89:208–216
13. Wesolowski TA, Warshel A (1993) Frozen density functional approach for ab initio calculations of solvated molecules. *J Phys Chem* 97:8050–8053
14. Wesolowski TA (2006) One-electron equations for embedded electron density: challenge for theory and practical payoffs in multi-level modeling of soft condensed matter. In: Leszczynski J (ed) *Computational chemistry: reviews of current trends*, vol X. World Scientific, Singapore, pp 1–82
15. Govind N, Wang YA, da Silva AJR, Carter EA (1998) Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment. *Chem Phys Lett* 295:129–134
16. Zheng H (1997) One-electron approach and the theory of the self-consistent cluster-embedding calculation method. *Phys Lett A* 226:223–230
17. Zheng H (1993) Self-consistent cluster-embedding calculation method and the calculated electronic structure of NiO. *Phys Rev B* 48:14868–14883
18. Zheng H (1995) Electronic structure of CoO. *Physica B* 212:125–138
19. Zheng H, Rao BK, Khanna SN, Jena P (1997) Electronic structure and binding energies of hydrogen-decorated vacancies in Ni. *Phys Rev B* 55:4174–4181
20. Zheng H, Wang Y, Ma G (2002) Electronic structure of LaNi₅ and its hydride LaNi₅H₇. *Eur Phys J B* 29:61–69
21. He J, Zheng H (2002) The electronic structure of GaN and a single Ga-vacancy in GaN crystal. *Acta Phys Sin* 51:2580–2588
22. Lin S, Zheng H (2005) Electronic structure of the surface of LaNi₅ crystal. *Acta Phys Sin* 54:4680–4687
23. Zhen H, Lin S (2006) First-principles calculation of LaNi₅ surface. *J Phys Conf Ser* 29:129–140
24. Zheng H (2000) Electronic structure of trypsin inhibitor from squash seeds in aqueous solution. *Phys Rev E* 62:5500–5508
25. Zheng H (2000) First principle ab initio calculation of the electronic structure of protein molecule. *Prog Phys* 20:291–300
26. Zheng H (2002) Ab initio calculations of the electronic structures and biological functions of protein molecules. *Mod Phys Lett B* 16:1151–1162
27. Zheng H (2003) Electronic structures of ascaris trypsin inhibitor in solution. *Phys Rev E* 68:051908
28. Sato F, Yoshihiro T, Era M, Kashiwagi H (2001) Calculation of all-electron wavefunction of hemoprotein cytochrome c by density functional theory. *Chem Phys Lett* 341:645–651
29. Yoshihiro T, Sato F, Kashiwagi H (2001) Distributed parallel processing by using the object-oriented technology in proteinDF program for all-electron calculations on proteins. *Chem Phys Lett* 346:313–321
30. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35:133–152
31. Lazaridis T, Karplus M (1998) Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *J Mol Biol* 288:477–487
32. Onsager L (1936) Electric moment of molecules in liquids. *J Am Chem Soc* 58:1486–1493
33. Klamt A, Schuurmann G (1993) COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans* 2:799–803
34. Guo H, Karplus M (1994) Solvent influence on the stability of the peptide hydrogen bond: a supramolecular cooperative effect. *J Phys Chem* 98:7104–7105
35. Schaefer M, Karplus MA (1996) Comprehensive analytical treatment of continuum electrostatics. *J Phys Chem* 100:1578–1599
36. Eckert F, Klamt A (2002) Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J* 48:369–385
37. Foresman JB, Keith TA, Wiberg KB (1996) Solvent effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation on ab initio reaction field calculations. *J Phys Chem* 100:16098–16104
38. Wang X, Zheng H, Li C (2006) The equivalent potential of water molecules for electronic structure of cysteine. *Eur Phys J B* 52:255–263
39. Li C, Zheng H, Wang X (2007) The equivalent potential of water molecules for electronic structure of lysine. *Sci China Ser G* 50:15–30
40. Li C, Zheng H, Wang X (2007) The equivalent potential of water molecules for the electronic structure of histidine. *J Phys Condens Mat* 19:116102
41. Zhang T, Zheng H, Yan S (2007) Equivalent potential of water molecules for electronic structure of glutamic acid. *J Comput Chem* 28:1848–1857
42. Yan S, Zheng H, Zhang T (2008) The equivalent potential of water molecules for electronic structure of alanine. *Mol Phys* 106:1427–1439
43. Zhang T, Zheng H, Yan S (2008) Equivalent potential of water molecules for electronic structure of aspartic acid. *J Comput Chem* 29:1780–1787
44. Wang X, Zheng H (2009) Simulation of water potential for the electronic structure of serine. *Chin Phys B* 18:1968–1178
45. Shen X, Gao Y, Zheng H (2009) The equivalent dipole potential of water for the electronic structure of threonine. *Mol Phys* 107:1393–1405
46. Gao Y, Shen X, Zheng H (2010) Equivalent potential of water for electronic structure of asparagines. *Int J Quantum Chem* 110:925–938
47. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev B* 136:864–871
48. Kohn W, Sham L (1965) Self-consistent equations including exchange and correlation effects. *J Phys Rev A* 140:1133–1138

49. von Barth U, Hedin L (1972) A local exchange-correlation potential for the spin polarized case: I. *J Phys C* 5:1629–1637
50. Rajagopal AK (1979) Theory of inhomogeneous electron systems: spin-density-functional formalism. *Adv Chem Phys* 41:59–153 [quote of Rajagopal AK, Singhal S, Kimball J (1979)]
51. van Duijneveldt FB (1971) Gaussian basis sets for the atoms HNe for use in molecular calculations. *IBM J Res Dev* 945:16437
52. Lie GL, Clementi E (1974) Study of the electronic structure of molecules. XXI. Correlation energy corrections as a functional of the Hartree–Fock density and its application to the hydrides of the second row atoms. *J Chem Phys* 60:1275–1287
53. Poirier RA, Daudel R, Mezey PG, Csizmadia IG (1982) Ab initio calculations on sulfur-containing compounds. I. Uniform quality basis sets for sulfur: total energies and geometries of H₂S. *Int J Quantum Chem* 21:799–811
54. Huzinaga S (1965) Gaussian-type functions for polyatomic systems. I. *J Chem Phys* 42:1293–1302
55. Poirier R, Kari R, Csizmadia IG (1985) *Handbook of Gaussian basis sets*. Elsevier, New York
56. Chen H (1988) *Electronic structure of clusters: applications to high-T_c superconductors* (Ph.D. dissertation) Louisiana State University, Baton Rouge
57. Chen H, Callaway J, Misra PK (1988) Electronic structure of Cu–O chains in the high-T_c superconductor YBa₂Cu₃O₇. *Phys Rev B* 38:195–203
58. Chen H, Callaway J (1991) Local electronic structure and magnetism of 3d transition-metal impurities (Cr, Mn, Fe, Co, and Ni) in La_{2-x}Sr_xCuO₄. *Phys Rev B* 44:2289–2296
59. Zheng H, He J (2001) Limitations of conventional one electron approximation methods. *J Tongji Univ* 29:593–597
60. Xu W, Zheng H (2003) Theoretic calculations of Co and Ni clusters with different sizes. *J Tongji Univ* 31:374–378
61. Lin S, Zheng H (2004) Electronic structure of new oxygen molecule O₄. *J Tongji Univ* 32:551–555
62. Hao J, Zheng H (2004) Theoretical calculation of structures and properties of Ga₆N₆ cluster. *Acta Phys Sin* 53:1044–1049
63. Zheng H, Hao J (2005) Ab initio study of the electronic properties of the planar Ga₅N₅ cluster. *Chin Phys* 14:529–532
64. Zheng H (1993) *Self-consistent cluster-embedding calculation method and the electronic structure of NiO and CoO* (Ph.D. dissertation). Louisiana State University, Baton Rouge
65. Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. *J Mol Liq* 101:219–260
66. Rick SW (2001) Simulations of ice and liquid water over a range of temperatures using the fluctuating charge model. *J Chem Phys* 114:2276–2283
67. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) *Intermolecular forces*. Reidel, Dordrecht, p 331
68. Robinson GW, Zhu SB, Singh S, Evans MW (1996) *Water in biology, chemistry and physics (Experimental Overviews and Computational Methodologies)*. World Scientific, Singapore