Vol 18 No 5, May 2009
1674-1056/2009/18(05)/1968-11

Chinese Physics B

© 2009 Chin. Phys. Soc.
and IOP Publishing Ltd

Simulation of water potential for the electronic structure of serine*

Wang Xing-Rong(王兴荣) and Zheng Hao-Ping(郑浩平)

Pohl Institute of Solid State Physics, Tongji University, Shanghai 200092, *China*

(Received 28 July 2008; revised manuscript received 5 November 2008)

First-principles, all-electron, *ab initio* calculations have been performed to construct an equivalent water potential for the electronic structure of serine (Ser) in solution. The calculation is composed of three steps. The first step is to search for the configuration of the Ser + nH$_2$O system with a minimum energy. The second step is to calculate the electronic structure of Ser with the water molecule potential via the self-consistent cluster-embedding method (SCCE), based on the result obtained in the first step. The last step is to calculate the electronic structure of Ser with the dipole potential after replacing the water molecules with dipoles. The results show that the occupied states of Ser are raised by about 0.017 Ry on average due to the effect of water. The water effect can be successfully simulated by using the dipole potential. The obtained equivalent potential can be applied directly to the electronic structure calculation of protein in solution by using the SCCE method.

Keywords: serine, electronic structure, water, self-consistent cluster-embedding calculation

PACC: 7115M

1. Introduction

The structural taxonomy of proteins has been intensely studied by crystallography and nuclear magnetic resonance (NMR) spectroscopy for decades. More than 30 000 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 acquire some knowledge of the electronic structure (ground state and excited states) of a protein which shows particular properties, while the time-dependent Schrödinger equation and quantum statistical mechanics can be used to understand the processes that facilitate the explanation of the biological functions of a protein. Computational research on proteins by first-principles calculation, however, has not yet been well performed because of the huge number of atoms contained in a protein molecule and the limitation of both computational conditions and traditional calculation methods. To the best of our knowledge, no successful first-principles, all-electron, *ab initio* calculation of electronic structure of any protein existed before 2000. From this perspective, a feasible computational technique, which can greatly reduce the computational effort, 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.[1−15] Self-consistent cluster-embedding calculation (SCCE),[16,17] a first-principles, all-electron, *ab initio* calculation method, is developed based on density functional theory (DFT). Unlike the traditional calculation method, the one-electron wavefunctions obtained by using the SCCE method are localized, i.e. each one-electron wavefunction is localized only in a part of the region occupied by the system. This approach possesses two advantages. (1) The SCCE method can be applied to any complex system and the localized valence electrons in material can be better described by the localized wave functions obtained in the SCCE calculation. (2) The computational effort is reduced greatly while the calculation accuracy is kept unchanged. This makes the electronic structure calculation of a protein realizable. SCCE calculations have been successfully applied to several insulators, semiconductors, metals, crystals with defects and impurities, and surfaces.[18−24] The results are in good agreement with those obtained from other methods.[25,26]

*Project supported by the National Natural Science Foundation of China (Grant No 30470410) and the Science and Technology Development Foundation of Shanghai, China (Grant No 03JC14070).

Up to now, the electronic structures of three proteins with four three-dimensional structures have been obtained by using the SCCE method.[27−30] Another protein was calculated by Sato *et al*[31] and Yoshihiro *et al*[32] by using their own all-electron method. However, the former calculations did not include solvent influence due to the limited computational conditions and the following reasons: the hydrogen bonds between a protein and a water molecule, and those between water molecules, break easily at room temperature. Water molecules move ceaselessly, and there is no fixed relative position between water molecules and a protein. The effect of water molecules on a protein is a kind of average force, which mainly influences the three-dimensional structures of protein. So the electronic structure of a protein in solution with a certain geometric structure may be roughly approximated by the electronic structure of the same protein with the same geometric structure but no surrounding water molecules. This viewpoint has been demonstrated by our previously 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, for a protein containing thousands of atoms the computational effort is still so large that the capability of a computer is required to reach almost the upper limit of a supercomputer. It is impossible to add a large number of water molecules into the calculation. For the electronic structure calculation of protein, it is necessary to construct an equivalent water potential, which must be simple in the form of expression and easy in use and, more importantly, require little additional computational effort.

The computational study of protein molecules, say, protein folding and molecule dynamics calculations, has 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 were developed.[33] The energy function including implicit solvation, applied to the protein folding calculation, was established by Lazaridis and Karplus in 1998.[34] There are some typical models for the solvation effect, such as Onsager[35] and conductor-like screening models.[36] The influence of solvent (water, ethanol, ethyleneglycol, and trifluoroethanol) on the stability of the peptide hydrogen bond was investigated by Guo and Karplus[37] through first-principles calculation in 1994. In 1995, the solvent effect on several small protein molecules such as the bovine pancreatic trypsin inhibitor (BPTI)[38] was also studied. In recent years, some other treatments have been developed, such as the COSMO-RS approach,[39] and solvent reaction field models[40] (the continuum model, PCM, IPCM). However, they are all designed for the calculation of geometric structure, but not for electronic structure. Because the first-principles, all-electron, *ab initio* calculation of the electronic structure of protein is at its primary stage, there is no suitable equivalent water potential available for the electronic structure calculation of protein using the SCCE method.

There are more than a hundred thousand proteins. But we only need to construct the equivalent potentials of 20 amino acids, i.e. the building blocks of proteins. Six amino acids have already been investigated by our group: cysteine (Cys),[41] lysine (Lys^+),[42] glutamic acid (Glu),[43] histidine (His),[44] aspartic acid (Asp),[45] and alanine (Ala).[46] In the present paper, we are concerned with serine (Ser). Ser has 14 atoms with a hydroxyl on the tip of a lateral chain. It is a polar amino acid and is hydrophilic. In a neutral solution, the Ser N-terminal gains a proton to become H_3N^+, the C-terminal loses a proton to become COO^-, and the lateral chain is polarized but neither loses nor gains a proton. So the Ser in solution stays neutral.

The present work is based on two considerations. (1) There is no fixed hydrogen bond between water molecules and protein in a solvent, and so no fixed relative position between water molecules and protein. For any given geometrical structure at any moment, there is always one electron distribution (electronic structure) that leads the total energy of system to reach a minimum value, because electrons moves much faster than the nucleus. In our calculation, the protein is already of a solution structure which does not change during the calculation. The water molecules are, with the highest probability, at the position where the total energy of the system reaches a minimum value. The electronic structure of protein with water molecules at these positions can be considered approximately as the electronic structure of protein in water. (2) The electronic structure of protein is calculated by using the SCCE method. Dipoles composed of point charges can be easily added to the SCCE calculation with almost no additional computational effort. On the other hand, the average potential of polar wa-

ter molecules can be reasonably simulated by dipoles. So we choose dipoles composed of point charges to simulate the water potential, acting on the electronic structure of Ser. The present work consists of three steps. First, we determine the geometric structure of Ser + $6H_2O$ system by using the free cluster calculation method. Second, based on the obtained geometric structure, the electronic structure of Ser with the potential of water molecules is calculated by using the self-consistent cluster-embedding method. Finally, by replacing water molecules with dipoles, the electronic structure of Ser with the dipole potential is calculated. The dipoles are adjusted so that Ser with the dipole potential and a water molecule are very much alike in electronic structure.

2. Theoretical model

The 'free cluster calculation' and the 'self-consistent cluster-embedding (SCCE) calculation' methods, which are based on density functional theory (DFT),[47,48] have been described in detail elsewhere (see Refs.[16, 49, 50] and website www.esprotein.org.cn). In particular, we refer readers to Ref.[45], in which the calculation of glutamic acid (Glu) is described, for further details concerning the DFT, free cluster calculation, and the SCCE calculation.

3. System, computational procedure and results

Both the geometric and electronic structures of a Ser+$6H_2O$ system have been found by using the free cluster calculation. Considering the degree of freedom of the geometric structure of Ser in solution, six water molecules are apparently not enough to describe the water effect on the geometric structure of Ser in solution. However, our concern is the electronic structure rather than the geometric structure. We believe that six water molecules are appropriate to describe the water effect on the electronic structure of Ser in solution. The reasons are as follows.

(1) Both the property and the function of an amino acid are determined mainly by electronic states near the Fermi level, which are easily affected by the solution environment. In contrast it is reasonable to assume that electronic states whose energies are much lower than the Fermi energy are hardly affected. So we actually use six water molecules to simulate the water solvent effect on Ser's electronic states near the Fermi level. If we choose ten electronic states near the Fermi level, the degrees of freedom are small, and six water molecules should be reasonable, because atomicity of them (18 atoms) is larger than that of Ser (14 atoms).

(2) The valence electrons in an amino acid are localized electrons except for the PI electrons of a benzene type ring which are shared with all of the 6 carbon atoms. The electronic states of Ser near the Fermi level are mainly localized electrons of the amido (H_3N^+), carboxyl (COO^-), and hydroxyl (OH) on the tip of a lateral chain. So only the water molecules surrounding the three parts need considering. From the viewpoint of hydrogen-bonding, we use the minimum number of water molecules: each H atom of the N-terminal H_3N^+ (or hydroxyl OH) may form a hydrogen bond with an O atom of water molecule, and each O atom of the C-terminal COO- (or hydroxyl OH) may form a hydrogen bond with an H atom of water molecule. Since a water molecule has one O atom and two H atoms, the minimum numbers of water molecules near the three parts are as follows: three for the N-terminal H_3N^+, one for the C-terminal COO^-, and two for the hydroxyl OH. Therefore, six water molecules will perform well if they are at the right geometric positions.

(3) Without question, the more molecules that are used, the better approximation will be achieved. Choosing only six water molecules may be acceptable and doable due to the following two reasons. A) A water molecule near the amino acid will impede other ones approaching the amino acid, and the potential of a removed dipole attenuates as $1/r^2$. So the minimal number of water molecules (they are all located close to and can form hydrogen bonds with the Ser), at least to a first order approximation, plays a major part in the solvent effect on Ser's electronic structure. B) If we increase the number of water molecules from the minimal number 6 to 9, for example, there may be more than 30 geometric structures, all with almost the same total energy. It is this degeneracy that makes the calculation impossible: searching the 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 to anybody. Our aim in the present study is to construct a simple and easy-to-use potential. Limited by the computa-

tional capacity, we choose six water molecules.

3.1. Determination of the geometric structure of the Ser+6H_2O system

3.1.1. Initial geometric structure of the Ser+6H_2O system

The coordinates of the 14 atoms of Ser in solution, given in Table 1, originate from a PDB structure file given by the Laboratory of Mass Spectrometry and Gaseous Ion Chemistry at the Rockefeller University (http://prowl.rockefeller.edu/aainfo/struct.htm).

Table 1. Atomic coordinates of Ser.

No.	atom	X/nm	Y/nm	Z/nm
1	C_α	0.0431	0.0077	–0.0230
2	C	–0.0896	0.0794	0.0013
3	O	–0.1971	–0.0143	–0.0100
4	N	0.0606	–0.1006	0.0786
5	C	0.1566	0.1066	–0.0111
6	O	0.2306	0.1262	–0.1060
7	O	0.1751	0.1693	0.0978
8	H	0.0432	–0.0358	–0.1227
9	H	–0.1023	0.1582	–0.0726
10	H	–0.0897	0.1228	0.1010
11	H	–0.0172	–0.1684	0.0704
12	H	0.1505	–0.1492	0.0621
13	H	0.0605	–0.0593	0.1735
14	H	–0.2818	–0.0305	0.0053

In a neutral solvent (pH=7), the Ser N-terminal gains a proton to become H_3N^+, and the C-terminal loses a proton to become COO^-. The polar water molecules mainly affect Ser's charged H_3N^+, COO^- and the polar tip OH of the lateral chain. At the beginning, the 6 water molecules are distributed semi-randomly around the Ser according to our previous experience, i.e. three water molecules were near the H_3N^+, one near the COO^- and two near the OH. Each water molecule, with its initial position and orientation, can form a hydrogen bond with an atom H or O of Ser, so the total energy of the Ser+6H_2O system should be lower.

The linear combination of Gaussian orbits is used as a basis function. The optimized Gaussian bases are the same as those used in the electronic structure calculation of proteins,[27−30] i.e. C — 8s6p, 26 Gaussian bases; N — 8s7p, 29 Gaussian bases; O — 8s7p, 29 Gaussian bases; H — 8s1p, 11 Gaussian bases; S — 9s9p1d, 41 Gaussian bases. The total number of Gaussian bases is 577. In order to calculate the exchange–correlation energy, the space occupied by the Ser+6H_2O system is divided into 352000 grid points.

The software 'free cluster calculation' was developed by Callaway.[49] The electronic structures of many molecules and clusters have been calculated using this software.[51−57] By solving the Kohn–Sham equation (4) in Ref.[41] self-consistently, we obtain the electronic structure, the total energy, and the force applied to each atomic nucleus.

3.1.2. Adjustment of six water molecules

In the Ser+6H_2O system, the geometric structure of Ser should stay unchanged, so that for each water molecule only the relative positions between the water molecules and Ser should be adjusted. Thus, during the optimization, all the nuclei of Ser are fixed and the nuclei of the water molecules are movable under the applied forces while the geometric structure of each water molecule is kept unchanged. Then the electronic structure of the Ser+6H_2O system is self-consistently calculated by using the first-principles, all-electrons, *ab initio* calculation method.

The six water molecules are adjusted one by one. Each time, the nuclei of five water molecules are fixed, only the nuclei of one water molecule under the strongest applied force move in the direction of applied force: eight different sets of positions along the force direction are chosen, and eight jobs, each with the nuclei of the water molecule on one set of positions, are sent to and done by a supercomputer. After convergence of all eight jobs, the set of positions of the nuclei of the water molecule with the lowest total energy is found. Then we move the nuclei of another water molecule which experiences the strongest applied force now. After the nuclei of six water molecules have all been adjusted, one course is finished. Then the second course begins. After hundreds of adjustments, the geometric structure of the Ser +6H_2O system with the lowest total energy is obtained. The total energy of the final structure is –1707.1346 Ry, 0.6477 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 Ser+6H_2O system is shown in Fig.1, where most of atoms are numbered according to their numbers given in Tables 1 and 2.

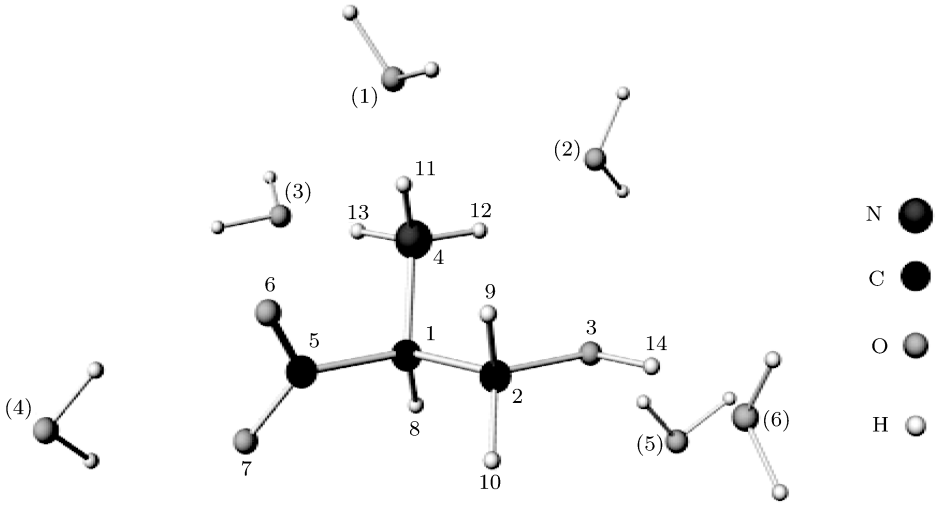


Fig.1. Final geometric structure of the Ser+6H_2O system.

Table 2. Final atomic coordinates of the six water molecules.

No.	atom	X/nm	Y/nm	Z/nm
(1)	O	0.06202	–0.02614	0.32203
	H	–0.00141	0.03466	–0.36001
	H	0.12208	–0.04622	0.39946
(2)	O	–0.14667	–0.36452	0.10460
	H	–0.17803	–0.40624	0.18484
	H	–0.18083	–0.42016	0.03393
(3)	O	0.29834	–0.25409	0.02712
	H	0.33948	–0.33953	0.04336
	H	0.37127	–0.19211	0.02814
(4)	O	0.40549	0.32842	0.00522
	H	0.34445	0.30649	0.07628
	H	0.37957	0.27099	–0.06684
(5)	O	–0.30360	–0.14778	–0.19717
	H	–0.25887	–0.10549	–0.12152
	H	–0.36912	–0.20316	–0.15734
(6)	O	–0.40779	0.15677	0.02129
	H	–0.45757	0.21571	–0.03538
	H	–0.43694	0.17942	0.11015

Of course, it is not absolutely certain that we have achieved the best structure, i.e. the best relative positions between water molecules and Ser, where the total energy is the lowest (global minimum), even another hundred adjusting steps are performed. But we are sure that either the room for further adjustment is very small, or the total energy of our final geometric structure is very close to the lowest energy of the Ser+6H_2O system. So, using the final structure, the calculated electronic structure of Ser in the water potential will be a good approximation to that in a real aqueous solution. The reasons are as follows. (1) We do not study the relative positions between the water molecules and Ser. (2) There is no fixed hydrogen bond between water molecules and protein in a solution, so there exists no fixed relative position between water molecules and protein. The water molecules are, only with the highest probability, at the position where the total energy of the system reaches a minimum. (3) In order to reduce the computational effort, charge-density fitting is used in both the free cluster calculation and the band structure calculation: a pseudo charge-density, which differs from the real one but can give a total energy very close to that calculated by using the real charge-density, is used to calculate the electronic structure. It is deemed that the electronic structure calculated by using the pseudo charge-density is a good approximation to the real electronic structure of the system.

3.2. Electronic structure of Ser in the potential of water molecules

Based on the geometric structure of the Ser+6H_2O system determined in Section 3.1, the system is now divided into seven embedded clusters for the self-consistent cluster-embedding calculation (SCCE). The Ser is the first cluster, and each water molecule is treated as one cluster. In the SCCE calculation, the total potential is the same as that in the free cluster calculation, and only the one-electron

wave functions are changed from extended to localized: each one-electron-wave function is now localized in the region of a cluster. So the electronic structure of Ser is separated from that of the water molecules.

The calculation contains two kinds of iterations: (i) intra-cluster iteration and (ii) inter-cluster iteration. After ten inter-cluster iterations, we achieve converged results. Given in Table 3 are the eigenvalues and Mulliken populations of ten orbitals near the Fermi level of Ser in the potential of six water molecules.

Table 3. Some of the eigenvalues and Mulliken populations of Ser with the potential of six water molecules.

state	energy/Ry	Mulliken population							
		C		N		O		H	
		s	p	s	p	s	p	s	p
30	–0.1426	–0.5488	0.0433	0.2798	0.5806	0.3059	1.4730	–1.1376	0.0039
29	–0.1809	0.2692	–0.0403	3.6301	0.8094	0.0944	0.1811	–3.7641	0.0090
The above states are unoccupied									
28	–0.3610	0.0020	0.0035	0.0077	0.0000	0.0072	0.9747	0.0046	0.0002
27	–0.3663	0.0178	0.0356	0.0238	–0.0038	0.0498	0.9323	–0.0562	0.0006
26	–0.4146	0.3633	–0.0840	–0.0162	–0.0069	0.0134	0.7969	–0.0399	0.0002
25	–0.5456	–0.0127	0.0486	0.0012	0.0115	0.0048	0.7496	0.1829	0.0140
24	–0.6196	–0.0227	0.4565	–0.0024	0.0348	0.0034	0.4202	0.1070	0.0033
23	–0.6622	–0.0286	0.3642	0.0142	0.0119	0.0191	0.5322	0.0841	0.0030
22	–0.6856	–0.0068	0.2400	0.0014	0.0129	0.1170	0.5748	0.0574	0.0035
21	–0.7527	0.0339	0.2541	0.0006	0.0027	0.2064	0.4686	0.0314	0.0023

3.3. Equivalent water potential on Ser simulated by dipoles

Each water molecule in Subsection 3.2 is now replaced by a dipole: the O atom of the water molecule is replaced by a negative point charge and the two H atoms are replaced by a positive charge located in the middle of the line connecting the two H atoms. Adjusting the point charges and positions of the dipoles, the electronic structure of Ser is recalculated using the SCCE calculation. There is now no inter-cluster iteration, however: only the intra-cluster iteration of the Ser cluster needs to be performed because the dipoles have no electron.

It is important that the free cluster calculation of Ser+6H_2O cannot be used as the fitting criteria to adjust the SCCE calculation of Ser+6 dipole, because the two systems have different numbers of electrons, and the distribution regions of the electrons of two systems are not the same. On the other hand, the SCCE calculation of Ser+6H_2O, i.e. the electronic structure of cluster Ser obtained in Subsection 3.2, is considered approximately as the electronic structure of Ser in water, and it is suitable for the criteria. Accordingly, the dipoles are adjusted. In order to evaluate the difference between the calculated electronic structure and that obtained in Subsection 3.2, two quantitative criteria are 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 $\varepsilon_{n0}^{\sigma}$ are the eigenvalues of the n'th electronic state with spin σ calculated in this Subsection and in Subsection 3.2, respectively. N^{σ} is the number of electrons with spin σ.

(2) The equivalent mean square deviation of

charge density.

$$\overline{\Delta C^{\sigma}} = \frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}(\sum_{n=1}^{N^{\sigma}} C_{ni}^{\sigma *}C_{nj}^{\sigma} - \sum_{n=1}^{N^{\sigma}} C_{ni0}^{\sigma *}C_{nj0}^{\sigma})^2\right]^{1/2},$$

where the C_{ni}^{σ} and C_{ni0}^{σ} are the expansion coefficients of the eigenfunctions of the n'th electronic state with spin σ calculated in this Subsection and in Section 3.2, respectively. N is the number of Gaussian bases used to expand the one-electron wavefunction. (Please note the charge density:

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

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

The intra-cluster iteration of the Ser cluster is performed using the SCCE method, and the electronic structure of Ser with the potential of the six dipoles is obtained. Then the values of the two criteria are calculated and compared. Here are the initial values of the two criteria: the mean square deviation of the eigenvalues is 2.1393×10^{-3}, and the equivalent mean square deviation of the charge density is 6.8954×10^{-2}. After hundreds of adjustments, the two values both reach their minima. This means that the electronic structure of Ser with the dipole potential is close to that of the water molecules. The final values of the two criteria are 1.7126×10^{-3} and 6.3121×10^{-2}, respectively.

There exist a number of studies in the literature addressing the problem of simplified discrete solvent representations by point charges: each of them has its own advantages and disadvantages. From the website http://www.lsbu.ac.uk/water/models.html 23 water models and more than 1200 papers can be found. Listed in the review paper written by Guillot[58] are 46 distinct water models, which indirectly indicate that they are unsuccessful in quantitatively reproducing the properties of real water. In our previous studies of constructing equivalent potentials,[44,45] we tried two popular 3-charge water models, TIP4P-FQ[59] and SPC,[60,61] as well as a dipole model. However, for the calculation of the electronic structure of an amino acid, the more complicated 3-charge water models did not give a better fitting than the simple dipole model (this will be discussed in another paper). Two facts should be emphasized. (i) We are not aiming to construct a general water model suitable for all cases, which is a complex problem and still poses a lot of questions. What we are aiming to do is construct a special equivalent water potential suitable only for the calculations of electronic structures of proteins in solution. For example, the dipoles constructed for Ser will be applied only to Ser which is in the surface of a protein in solution. (ii) We are not aiming to construct an 'exact' special equivalent water potential, which actually does not exist: it is impossible to make the electronic structure of Ser with the dipole potential identical to that of water molecules. Our aim is to construct a simple and easy-to-use potential, which, at least to the first order approximation, plays a major part in the solvent effect on the electronic structure of protein. So we choose the dipole model.

The initial charges of all six dipoles are set to the same value, 0.5 e. The distance between the positive and negative charges is 0.05858 nm, which remains unchanged in calculation. The charges and positions of the six dipoles are adjusted one by one. After hundreds of such adjustments, the two criteria reach their minima. The charges of the last six dipoles are as follows: the charges of three dipoles around the amidogen are $0.515e$, $0.52e$ and $0.52e$, separately. The charge of one dipole near the carboxyl is $0.665e$. The charges of two dipoles approaching the hydroxyl OH are $0.22e$ (near the O atom) and $0.5e$ (near the H atom), respectively. The charges and coordinates of the last six dipoles are listed in Table 4, and the geometric structures of the Ser+6 dipoles are shown in Fig.2. Indicated in Table 5 are the eigenvalues and Mulliken populations of ten orbits near the Fermi level of Ser in the potential of six dipoles.

Table 4. Point charges and coordinates of 6 dipoles.

No.	charge (e)	X/nm	Y/nm	Z/nm
(1)	–0.515	0.06442	–0.04093	0.32909
	0.515	0.06323	–0.02376	0.38538
(2)	–0.52	–0.14667	–0.36452	0.10460
	0.52	–0.17943	–0.41320	0.10938
(3)	–0.52	0.29834	–0.25409	0.02712
	0.52	0.35538	–0.26582	0.03575
(4)	–0.665	0.39545	0.31951	0.00445
	–0.665	0.35197	0.27983	0.00395
(5)	–0.22	–0.30360	–0.14778	–0.19717
	0.22	–0.31399	–0.15432	–0.13943
(6)	–0.50	–0.40779	0.15677	0.02129
	0.50	–0.44725	0.19757	0.03739

Table 5. Some of the eigenvalues and Mulliken populations of Ser with the potential of 6 dipoles.

state	energy/Ry	Mulliken population							
		C		N		O		H	
		s	p	s	p	s	p	s	p
30	–0.0488	–0.2651	0.0839	0.2332	0.2638	0.1515	0.2653	0.2631	0.0043
29	–0.1011	0.4481	–0.0040	1.2158	0.2103	–0.0309	0.0369	–0.8845	0.0083
The above states are unoccupied.									
28	–0.3563	0.0007	0.0028	0.0123	0.0007	0.0016	0.9815	0.0004	0.0001
27	–0.3605	0.0041	0.1097	–0.0013	–0.0020	–0.0372	0.9363	–0.0102	0.0006
26	–0.4149	0.2309	–0.0094	–0.0154	–0.0041	0.0092	0.8095	–0.0208	0.0002
25	–0.5424	–0.0005	0.0450	0.0013	0.0059	0.0010	0.7818	0.1510	0.0145
24	–0.6227	–0.0034	0.4123	–0.0005	0.0179	0.0452	0.4337	0.0910	0.0037
23	–0.6599	–0.0205	0.3303	0.0067	0.0157	0.0901	0.5284	0.0470	0.0023
22	–0.6734	–0.0012	0.1611	0.0084	0.0085	0.1813	0.5752	0.0613	0.0054
21	–0.7417	0.0404	0.1954	0.0000	0.0039	0.2767	0.4727	0.0097	0.0013

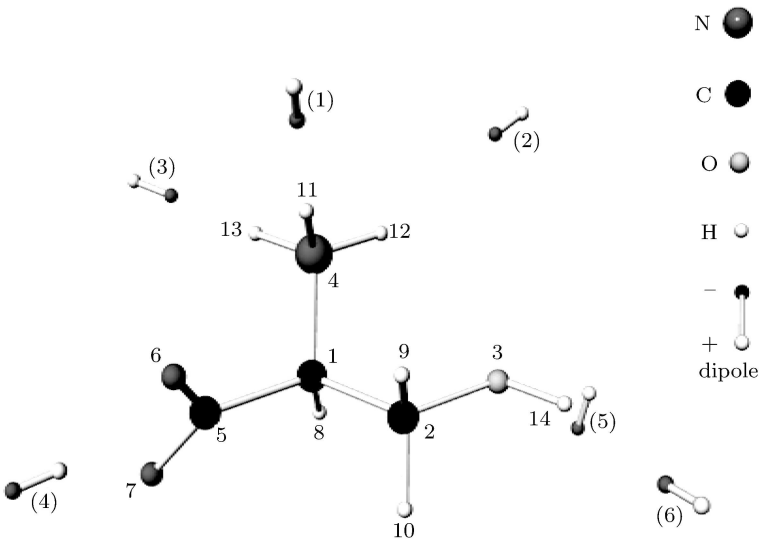


Fig.2. Final geometric structure of the Ser +6 dipoles system.

4. Discussion

In order to acquire knowledge of the water solvent effect on the electronic structure of Ser, the electronic structure of an isolated Ser is calculated by using the free cluster calculation. The total energy of this isolated Ser is –793.8821 Ry. Shown in Table 6 are the eigenvalues and Mulliken populations of 10 states near the Fermi level of the isolated Ser.

Shown in Table 7 are the eigenvalues of orbits 21 to 30 of Ser in three cases, i.e. with dipole potential, with potential of water molecules, and without potential. The last row in Table 7 indicates the energy gap between states 28 and 29. A sketch map comparing the three sets of eigenvalues of Ser is shown in Fig.3.

The properties of Ser are determined mainly by the electronic states near the Fermi level. In Tables 3, 5 and 6, Mulliken population analysis shows that in all three cases, the electronic states below the Fermi level are similar. States 28 and 27 are occupied mainly by oxygen 2p electrons of backbone COO^-. State 26 is a hybridized state, to which the contribution is made mainly by oxygen 2p electrons of backbone COO^- and by the carbon 2s and 2p electrons of atom C_α. State 25 is occupied mainly by oxygen 2p electrons of the hydroxyl OH of the lateral chain. State 24 is a hybridized state, to which the contribution is made mainly by the oxygen 2p and carbon 2p electrons of backbone COO^- and by carbon 2p electrons of the lateral chain and carbon 2p electrons of atom C_α.

States 23 and 21 are occupied mainly by oxygen 2p and carbon 2p electrons of backbone COO^-. State 22 is a hybridized state, to which the contribution is made mainly by the oxygen 2p electron of the hydroxyl OH of the lateral chain, and by the oxygen 2p electron of backbone COO^-.

Table 6. Some of the eigenvalues and Mulliken populations of an isolated Ser.

state	energy/Ry	Mulliken population							
		C		N		O		H	
		s	p	s	p	s	p	s	p
30	–0.0989	–0.5536	0.1612	–0.0507	0.0496	–0.0096	–0.1130	1.5073	0.0087
29	–0.1802	0.4420	0.0562	–0.8375	0.1735	–0.0272	0.0097	1.1710	0.0123
The above states are unoccupied.									
28	–0.3695	0.0053	0.0473	–0.0137	0.0007	–0.0117	0.9655	0.0062	0.0003
27	–0.3703	0.0049	0.0790	0.0028	–0.0042	–0.0144	0.9302	0.0011	0.0006
26	–0.4232	0.1089	0.1150	–0.0234	–0.0022	–0.0073	0.8021	0.0066	0.0002
25	–0.5904	–0.0004	0.0508	0.0020	0.0002	0.0002	0.7761	0.1517	0.0148
24	–0.6471	–0.0009	0.4208	–0.0032	0.0136	0.0518	0.4676	0.0485	0.0018
23	–0.6737	–0.0053	0.2631	–0.0015	0.0134	0.1796	0.5244	0.0255	0.0009
22	–0.7082	–0.0135	0.2422	0.0138	0.0077	0.1006	0.5235	0.1173	0.0084
21	–0.7586	0.0544	0.1708	–0.0027	0.0015	0.2841	0.4905	0.0005	0.0008

By comparing the second column with the third one in Table 7 and Fig.3(a) with Fig.3(b), it is found that water molecules do not markedly change the relative positions of the occupied electronic states. The main effect of water molecules is the narrowing of the energy gap by about 4.6%, and raising all eigenvalues about 0.0167 Ry on average. Among the eight states, the eigenvalues of states 25, 24 and 22 are raised by 0.0448 Ry, 0.0275 Ry and 0.0226 Ry, respectively, much larger than those of other states.

Table 7. Three sets of eigenvalues of Ser.

state	eigenvalues/Ry		
	no potential	water molecules	dipoles
30(unoccupied)	–0.0989	–0.1426	–0.0488
29(unoccupied)	–0.1802	–0.1809	–0.1011
28 (E_F)	–0.3695	–0.3610	–0.3563
27	–0.3703	–0.3663	–0.3605
26	–0.4232	–0.4146	–0.4149
25	–0.5904	–0.5456	–0.5424
24	–0.6471	–0.6196	–0.6227
23	–0.6737	–0.6622	–0.6559
22	–0.7082	–0.6856	–0.6734
21	–0.7586	–0.7527	–0.7417
$E_g = E_{29} - E_{28}$	0.1893	0.1801	0.2552

Comparison between the third and fourth columns in Table 7 and between Figs.3(b) and 3(c) reveal that below the Fermi level, the eigenvalues of Ser in the dipole potential are very close to those of Ser in the water molecule potential except for states 21 and 22. The eigenvalues of states 21 and 22 are raised by 0.0110 Ry and 0.0122 Ry, respectively, a bit larger than those of other states. But above the Fermi level, the energy gap calculated using the dipole potential is broadened by 42% compared with that using the water molecule potential. This suggests that unoccupied states in the dipole potential are overestimated. Because unoccupied states make no contribution to charge density, it can be concluded that the calculated dipole potential gives a good simulation of the effect of water on the electronic structure of Ser.

In addition, our dipole potential is transferable, although it is a special equivalent water potential. We use the dipoles to construct the equivalent water potentials for the electronic structures of 20 amino acids. When this is finished, the obtained equivalent potentials can be applied to the SCCE calculation as well as any first-principles, all-electron, *ab initio* calculation method, as long as they are used for calculating the electronic structure of a protein in solution. Amino acids lose their water molecules to turn into amino acid residues, and then they are combined into several polypeptide chains which are interlaced with each other, 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 in solution with a known geometric

structure, a water solvent does not affect the molecular orbitals of the amino acid residues in the interior of the protein because of the screening effect. Thus, the dipoles are located (according to our dipole potential) only near the N-terminal and C-terminal of a chain, and on the tips of the lateral chains of the amino acid residues which are on the surface of the protein. Both the charges and the locations of dipoles around those amino acid residues can be obtained by using our program, inputting the locations of these amino acid residues. This should help determine more reliably the active sites and reactive sites of a protein in solution, which are usually at the N-terminal and C-terminal of a chain, and on the tips of the lateral chains of amino acid residues which are on the surface of the protein.

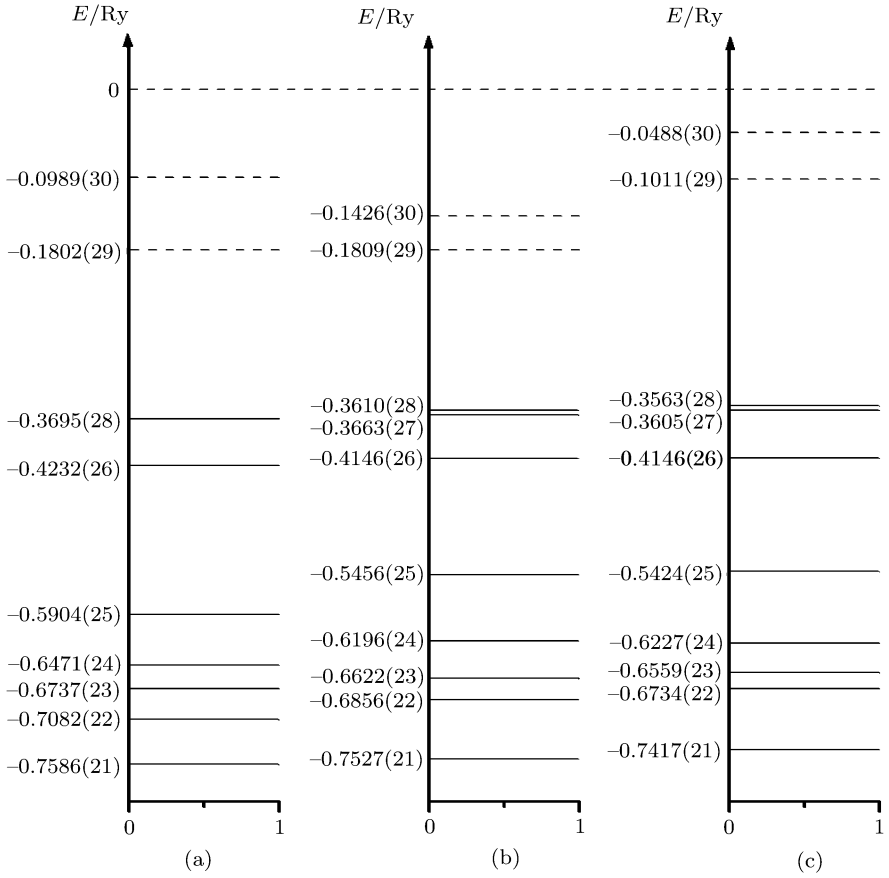


Fig.3. Comparison of eigenvalues among cases with no potential (a), with a water molecule potential (b), and with a dipole potential (c), where a solid line represents an occupied state and a dashed line refers to an unoccupied state.

5. Conclusions

We have performed first-principles, all-electron, *ab initio* calculations to construct an equivalent water potential for the electronic structure of Ser in solution. The calculations are composed of three steps. Firstly, we use the free cluster calculation to obtain the geometric structure of the Ser+$6H_2O$ system with the lowest total energy. Secondly, based on the obtained geometric structure, we use the self-consistent cluster-embedding (SCCE) calculation to acquire the electronic structure of Ser in the water molecule potential. Finally, we replace the six water molecules with six adjustable dipoles. The dipoles are adjusted until the electronic structure of Ser with the dipole potential is close to that of the water molecules.

We have made a comparison of the electronic structure of Ser in cases with a water potential, with a dipole potential, and without any external potential. The main effect of water on Ser's electronic structure is the narrowing of the energy gap by about 4.6%, and the raising of all eigenvalues about 0.0167 Ry on

average. A simple dipole potential can be used to simulate the effect of water on the electronic structure of Ser: for electronic states under the Fermi level, their eigenvalues in the first two cases above are very close to each other. So a simple, easy-to-use dipole potential has been established by making little additional computational effort. More equivalent potentials for other amino acids will be constructed. The results of this work, together with those of the previous studies and the coming studies, will be directly applied to the electronic structure calculation of a protein in solution.

Acknowledgement

This work was supported in part by Supercomputing Center, the Computer Network Information Center of Chinese Academy of Science, Beijing, China.

References

[1] Yang W T 1991 *Phys. Rev. Lett.* **66** 1438
[2] Cortona P 1991 *Phys. Rev.* B **44** 8454
[3] Galli G and Parrinello M 1992 **69** 3547
[4] Mauri F, Galli G and Car R 1993 *Phys. Rev.* B **47** 9973
[5] Li X P, Nunes R W and Vanderbilt D 1993 *Phys. Rev.* B **47** 10891
[6] Ordejon P, Drabold D A, Martin R M and Grumbach M P 1995 *Phys. Rev.* B **51** 1456
[7] Yang W T and Lee T S A 1995 *Chem. Phys.* **103** 5674
[8] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168
[9] Ordejón P, Artacho P E and Soler J M 1996 *Phys. Rev.* B **53** 10441
[10] Baer R and Gordon M H 1997 *Phys. Rev. Lett.* **79** 3962
[11] Klessinger M and Mcweeny R 1965 *J. Chem. Phys.* **42** 3343
[12] Li J B and McWeeny R 2001 *Int. J. Quant. Chem.* **89** 208
[13] Wesolowski T A and Warshel A 1993 *J. Phys. Chem.* **97** 8050
[14] Wesolowski T A 2006 *In Computational Chemistry: Reviews of Current Trends* (World Scientific: Singapore) p1
[15] Govind N, Wang Y A, Silva A J R Da and Carter E A 1998 *Chem. Phys. Letters* **295** 129
[16] Zheng H P 1997 *Phys. Lett.* A **226** 223
[17] Zheng H P 1997 *Phys. Lett.* **231** 453
[18] Zheng H P 1993 *Phys. Rev.* B **48** 14868
[19] Zheng H P 1995 *Physica* B **212** 125
[20] Zheng H P, Rao B K and Khanna S N P 1997 *Phys. Rev.* B **55** 4174
[21] Zheng H P, Wang Y and Ma G 2002 *Eur. Phys. J.* B **29** 61
[22] He J and Zheng H P 2002 *Acta Phys. Sin.* **51** 2580 (in Chinese)
[23] Lin S J and Zheng H P 2005 *Acta Phys. Sin.* **54** 4680 (in Chinese)
[24] Zhen H P and Lin S J 2006 *J. Phys. Conf. Ser.* **29** 129
[25] Guo J Y, Zheng G, He K H and Chen J Z 2008 *Acta. Phys. Sin.* **57** 3740 (in Chinese)
[26] Chen Q and Cao H H 2004 *Chin. Phys.* **13** 2121
[27] Zheng H P 2000 *Phys. Rev.* E **62** 5500
[28] Zheng H P 2000 *Progress in Physics* **20** 291
[29] Zheng H P 2002 *Mod. Phys. Lett.* B **16** 1151
[30] Zheng H P 2003 *Phys. Rev.* E **68** 051908
[31] Sato F, Yoshihiro T, Era M and Kashiwagi H 2001 *Chem. Phys. Lett.* **341** 645
[32] Yoshihiro T, Sato F and Kashiwagi H 2001 *Chem. Phys. Lett.* **346** 313
[33] Lazaridis T and Karplus M 1999 *Proteins: Structure, Function and Genetics* **35** 133
[34] Lazaridis T and Karplus M 1998 *J. Mol. Biol.* **288** 477
[35] Onsager L 1936 *J. Am. Chem. Soc.* **58** 1486
[36] Klamt A and Schuurmann G 1993 *J. Chem. Soc. Perkin. Trans.* **2** 799
[37] Guo H and Karplus M 1994 *J. Phys. Chem.* **98** 7104
[38] Schaefer M and Karplus M A 1996 *J. Phys. Chem.* **100** 1578
[39] Eckert F and Klamt A 2002 *AiChE. J.* **48** 369
[40] Foresman J B, Keith T A and Wiberg K B 1996 *J. Phys. Chem.* **100** 16098
[41] Wang X M, Zheng H P and Li C J 2006 *Eur. Phys. J.* B **52** 255
[42] Li C J, Zheng H P and Wang X M 2007 *Science in China* G **50** 15
[43] Zhang T, Zheng H P and Yan S 2007 *J. Comput. Chem.* **28** 1848
[44] Li C J, Zheng H P and Wang X M 2007 *J. Phys.: Condensed Matter* **19** 116102
[45] Zhang T, Zheng H P and Yan S 2008 *J. Comput. Chem.* **29** 1780
[46] Yan S, Zheng H P and Zhang T 2008 *Molecular Physics* **106** 1427
[47] Hohenberg P and Kohn W 1964 *Phys. Rev.* B **136** 864
[48] Kohn W and Sham L 1965 *J. Phys. Rev.* A **140** 1133
[49] Chen H 1988 *Electronic Structure of Clusters: Applications to High-T_c Superconductors* Ph. D. Dissertation, Louisiana State University, America
[50] Zheng H P 1993 *Self-Consistent Cluster-Embedding Calculation Method and the Electronic Structure of NiO and CoO* Ph.D. Dissertation, Louisiana State University, America
[51] Chen H, Callaway J and Misra P K 1988 *Phys. Rev.* B **38** 195
[52] Chen H and Callaway J 1991 *Phys. Rev.* B **44** 2289
[53] Zheng H and He J 2001 *J. Tongji Univ.* **29** 593
[54] Xu W H and Zheng H P 2003 *J. Tongji Univ.* **31** 374
[55] Lin S J and Zheng H P 2004 *J. Tongji Univ.* **32** 551
[56] Hao J A and Zheng H P 2004 *Acta Phys. Sin.* **53** 1044 (in Chinese)
[57] Zheng H P and Hao J A 2005 *Chin. Phys.* **14** 529
[58] Guillot B 2002 *J. Mol. Liquids* **101** 219
[59] Rick S W 2001 *J. Chem. Phys* **114** 2276
[60] Berendsen H J C, Postma J P M, van Gunsteren W F and Hermans J 1981 *Intermolecular Forces* (Reidel: Dordrecht) p331
[61] Robinson G W, Zhu S B, Singh S and Evans M W 1996 *Water in Biology, Chemistry and Physics: Experimental Overviews and Computational Methodologies* (World Scientific: Singapore)