Eur. Phys. J. B **52**, 255–263 (2006)
DOI: 10.1140/epjb/e2006-00284-8

THE EUROPEAN PHYSICAL JOURNAL B

The equivalent potential of water molecules for electronic structure of cysteine

X.-M. Wang[a], H.-P. Zheng, and C.-J. Li

Pohl Institute of Solid State Physics, Tongji University, Shanghai 200092, P.R. China

Received 14 April 2006

Abstract. In order to get more reliable electronic structure of protein in aqueous solution, it is necessary to construct a simple, easy-use equivalent potential of water molecules for protein's electronic structure calculation. The first-principles, all-electron, ab initio calculations have been performed to construct the equivalent potential of water molecules for the electronic structure of Cys. The process consists of three steps. First, the electronic structure of the cluster containing Cys and water molecules is calculated. Then, based on the structure, the electronic structure of Cys with the potential of water molecules is calculated using the self-consistent cluster-embedding method. Finally, the electronic structure of Cys with the potential of dipoles is calculated. The dipoles are adjusted so the electronic structure of Cys with the potential of dipoles is close to that of water molecules. The calculations show that the major effect of water molecules on Cys' electronic structure is lowering the occupied electronic states by about 0.032 Ry, and broadening energy gap by 16%. The effect of water molecules on the electronic structure of Cys can be simulated by dipoles potential.

PACS. 31.15.Ar Ab initio calculations – 31.70.Dk Environmental and solvent effects – 71.15.-m Methods of electronic structure calculations

1 Introduction

The knowledge of electronic structure is essential for understanding the properties and biological functions of a protein according to the quantum mechanics. But it is difficult to calculate the electronic structure of a protein because of its incredible huge computational effort, which is too large to be affordable by any supercomputer in the present and near future. To the best of our knowledge, there is no successful first-principles, all-electron, ab initio calculation of any protein before 2000.

Self-consistent cluster-embedding calculation (SCCE), a first-principles, all-electron, ab initio calculation method, is developed based on the density functional theory [1]. Different from the traditional calculation method, the obtained one-electron wave functions by the SCCE method are localized — each one-electron wave function is localized only in a part of the region occupied by the system. This results in 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 precision is kept. This makes the electronic structure calculation of protein a reality. The SCCE calculations have been successfully applied to several insulators, semi-conductors, metals, crystals with defects and impurities, and surfaces [2–8].

For the first time, the first-principles, all-electron, ab initio calculation of electronic structure of a real protein was completed in 2000 [9]. Up to day, the electronic structures of three proteins with four three-dimensional structures have been obtained [10–12]. However, the former calculations did not include solvent influence due to limited computational conditions and the following reasons: water molecules are usually quivering, turning and fungible at high speed in action with protein. There is no fixed chemical binding between water molecules and protein in solution. The effect of water molecules on 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 having a certain geometric structure can be approximated by the electronic structure of the same protein having the same geometric structure but no water molecules around. This viewpoint has been demonstrated by our previous calculated results, which agree with experimental data. In order to make the calculation more reliable, however, the effect of water molecules should be considered.

---

[a] e-mail: `xmwang1981@163.com`

Although the computational effort is reduced greatly in the SCCE method, for a protein containing thousands of atoms, however, the computational effort is still so large which almost approaches the upper limit of a supercomputer. So it is impossible to add a large amount of water molecules to the calculation. For the electronic structure calculation of protein, it is necessary to construct an equivalent potential of water molecules, which must be simple, easy-use and with almost no additional computational effort.

The computational study of protein molecules, say, protein folding and molecule dynamics calculation, has until now focused mainly on their geometric structure. Effective energy function for proteins and a Gaussian solvent-exclusion model for the solvation free energy are developed [13]. The energy function including implicit salvation, applied for the protein folding calculation, had been established by Lazaridis and Karplus in 1998 [14]. There are some typical models about the salvation effect, such as onsarger [15] and the conductor-like screening models [16]. The solvent's influence (water, ethanol, ethyleneglycol, and trifluoroethanol) on the stability of the peptide hydrogen bond have been calculated by Karplus and Hong Guo by the first-principles calculation in 1994 [17]. In 1995, solvent's effect on several small protein molecules such as the bovine pancreatic trypsin inhibitor (BPTI) [18] has also been studied. In recent years, some other treatments are developed such as COSMO-RS approach [19], the solvent reaction field models [20] (the continuum model, PCM, IPCM). However, they are all designed for the calculation of geometric structure, not electronic structure. Because the first-principles, all-electron, ab initio calculation of the electronic structure of protein is on its primary stage, there is no suitable equivalent potential of water molecules for the electronic structure calculation of protein. The equivalent potential developed in this paper can be applied directly to the electronic structure calculation of protein using the SCCE method.

There are more than a hundred thousand of proteins, but all are made up of 20 amino acids. So for the electronic structure calculation of proteins, we only need to construct the equivalent potentials of 20 amino acids. As the first sample, we choose Cysteine (Cys) which is relatively small. Cys has 14 atoms with a lateral chain of mercaptan. In a real protein, two nearby Cys usually combined as a Cystine through a disulfide bond formed by the oxidation of two mercaptans. Disulfide bond is of importance for stabilizing three-dimensional structure of protein. For some proteins, disulfide bond connects different polypeptides, but its major effect is to stabilize the folding of peptide chain. A small protein molecule usually needs several disulfide bonds to stabilize its three-dimensional structure.

This work is based on two considerations: (1) there is no fixed chemical binding between water molecules and protein in solution, so no fixed relative position between water molecules and protein. The water molecules are with the most probability at the position which makes the total energy of the system minimum. The electronic structure of protein with water molecules at these positions can be considered approximately as the electronic structure of protein in solution; (2) the electronic structure of protein is calculated by the SCCE calculation. The dipoles made up 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 water molecules can be reasonably simulated by dipoles. So we choose dipoles made up of point charges to simulate the potential of water molecules on the electronic structure of Cys. The work consists of three steps. First, the geometric structure of the Cys + nH_2O system is determined using the "free cluster calculation". Second, based on the geometric structure obtained in the first step, the electronic structure of Cys with the potential of water molecules is calculated using the "self-consistent cluster-embedding (SCCE) method". Third, the water molecules of the second step are replaced by dipoles made up of point charges. The dipoles are adjusted so the electronic structure of Cys with the potential of dipoles is close to that obtained in the second step. Thus the equivalent potential of water molecules for the electronic structure of Cys is established using the dipoles made up of point charges.

2 Theoretical model and computational procedure

The "free cluster calculation" and the "self-consistent cluster-embedding (SCCE) calculation" methods have been described in detail elsewhere (see Refs. [1,23] and website `www.esprotein.org.cn`), here we only give a brief over view for completeness.

According to the density functional theory (DFT) [21,22], the total energy of a system containing N electrons and M fixed nuclei can be written as (no relativistic effect is included; atomic units are used: $e^2 = 2$, $\hbar = 1$, and $2m_e = 1$, the unit of energy is the Rydberg constant $\mathrm{Ry} = e^2/2a_0 = 13.6049$ eV):

$$E_V[\rho] = T_{ni}[\rho] + E_{xc}[\rho] + \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' - 2\sum_{j=1}^{M}\int \frac{\rho(\mathbf{r})Z_j}{|\mathbf{r}-\mathbf{R}_j|}dr + \sum_{i\neq j}^{M}\frac{Z_iZ_j}{|\mathbf{R}_i-\mathbf{R}_j|}, \quad (1)$$

where $T_{ni}[\rho]$ is the kinetic energy of a non-interacting electron system, $E_{xc}[\rho]$ is the exchange-correlation energy. In deriving equation (1), Kohn and Sham have assumed the existence of a non-interacting electron system having the same ground-state charge density $\rho(\mathbf{r})$ as that of the real interacting system [22]. Each non-interacting electron can now be represented by a stationary state one-electron wave function $\Phi_n^\sigma(\mathbf{r})$. So the charge density $\rho(\mathbf{r})$ and kinetic energy $T_{ni}[\rho]$ of the non-interacting system can be written as:

$$\rho\left(\mathbf{r}\right)=\rho^{\uparrow}\left(\mathbf{r}\right)+\rho^{\downarrow}\left(\mathbf{r}\right)=\sum_{occupied...l}\left|\Phi_{l}^{\uparrow}\left(\mathbf{r}\right)\right|^{2}+\sum_{occupied...m}\left|\Phi_{m}^{\downarrow}\left(\mathbf{r}\right)\right|^{2},\quad(2)$$

$$T_{ni}[\rho]=\sum_{occupied...l}\int\Phi_{l}^{\uparrow *}\left(\mathbf{r}\right)\left(-\nabla^{2}\right)\Phi_{l}^{\uparrow}\left(\mathbf{r}\right)d\mathbf{r}+\sum_{occupied...m}\int\Phi_{m}^{\downarrow *}\left(\mathbf{r}\right)\left(-\nabla^{2}\right)\Phi_{m}^{\downarrow}\left(\mathbf{r}\right)d\mathbf{r}.\quad(3)$$

Using formulas (2) and (3), a single-electron Schrodinger equation, i.e., the well known Kohn-Sham equation [22], is obtained by the variation of functional (1) with respect to $\Phi_n^{\sigma *}(\mathbf{r})$ under conservation rule $\int\rho\left(\mathbf{r}\right)d\mathbf{r}=N$:

$$\left\{-\nabla^{2}+2\int\frac{\rho\left(\mathbf{r}'\right)}{\left|\mathbf{r}-\mathbf{r}'\right|}d\mathbf{r}'-2\sum_{i=1}^{M}\frac{Z_{i}}{\left|\mathbf{r}-\mathbf{R}_{i}\right|}+V_{xc}^{\sigma}\left(\mathbf{r}\right)\right\}\times\Phi_{n}^{\sigma}\left(\mathbf{r}\right)=\varepsilon_{n}^{\sigma}\Phi_{n}^{\sigma}\left(\mathbf{r}\right),\quad(4)$$

where the exchange-correlation potential is

$$V_{xc}^{\sigma}(\mathbf{r})=\frac{\partial E_{xc}[\rho]}{\partial\rho^{\sigma}(\mathbf{r})}.\quad(5)$$

The exact solution of equation (1) could be obtained if the following two conditions were satisfied: (i) $E_{xc}[\rho]$ is exact; (ii) the trial one-electron wave functions $\Phi_n^{\sigma}(\mathbf{r})$ are unconstrained in solving the Kohn-Sham equation (4) (required by the variational principle). Then we should obtain an uniquely exact set of $\{\Phi_n^{\sigma}(\mathbf{r})\}$ which corresponds to an uniquely correct $\rho(\mathbf{r})$. The Kohn-Sham equation (4), with an exact $V_{xc}^{\sigma}(\mathbf{r})$ and a correct $\rho(\mathbf{r})$, describes such a situation in which one electron (represented by an $\Phi_n^{\sigma}(\mathbf{r})$) moves under the average potentials of all other electrons and nuclei. Because the potential produced by an electron ($|\Phi_n^{\sigma}(\mathbf{r})|^2$) does not act on itself, the single-electron Hamiltonian in equation (4) includes only the potentials produced by $\rho'\left(\mathbf{r}\right)=\rho\left(\mathbf{r}\right)-|\Phi_n^{\sigma}(\mathbf{r})|^2$. This leads two results: (a) different $\Phi_n^{\sigma}(\mathbf{r})$ will correspond to different single-electron Hamiltonian (or $V_{xc}^{\sigma}(\mathbf{r})$) in Kohn-Sham equation (4); (b) in general, the Hamiltonian acting on an $\Phi_n^{\sigma}(\mathbf{r})$ has no symmetry of $\rho(\mathbf{r})$. As an example, consider a perfect periodic crystal. Although its charge density $\rho(\mathbf{r})$ is of lattice periodicity, in general, the exact single-electron Hamiltonian of Kohn-Sham equation (4) will have no lattice periodicity.

In practical calculation, however, the two conditions above can not be satisfied. Two approximations are taken: (1) the exchange-correlation potential $V_{xc}^{\sigma}(\mathbf{r})$ is averaged over all one-electron states with spin σ (such as LDA); (2) each $\Phi_n^{\sigma}(\mathbf{r})$ is constrained to satisfy a certain boundary condition. Please note, the first approximation means that all $\Phi_n^{\sigma}(\mathbf{r})$ in equation (4) now correspond to a same $V_{xc}^{\sigma}(\mathbf{r})$ (and single-electron Hamiltonian) which having the symmetry of $\rho(\mathbf{r})$, and all $\Phi_n^{\sigma}(\mathbf{r})$ are now constrained by this added symmetry. But this added constraint is not physically essential, and can be removed by boundary condition. The second approximation means that all wave functions, which do not satisfy the boundary condition, are thrown away, although they may be the true solutions of the system. When one kind of $\{\Phi_n^{\sigma}(\mathbf{r})\}$ is chosen (a boundary condition is applied), it means that a kind of non-interacting electrons is used to describe the real system approximately. So by choosing different boundary conditions, we can use different kinds of $\{\Phi_n^{\sigma}(\mathbf{r})\}$ to get approximate $\rho(\mathbf{r})$. According to variational principle, the calculated energy will be close to the true ground-state energy, only if the trial one-electron wave functions $\Phi_n^{\sigma}(\mathbf{r})$ describe real electrons well. For example, a set of Bloch functions can give a good description to quasi-free electrons, but can not do it for localized electrons. The latter can be best described by a set of localized one-electron wave functions.

We now assume the first approximation being taken, so all $\Phi_n^{\sigma}(\mathbf{r})$ correspond to a same single-electron Hamiltonian $\mathrm{H}^{\sigma}(\mathbf{r})$ (without boundary condition), and are constrained by an added symmetry of $\rho^{\sigma}(\mathbf{r})$. We discuss the second approximation. There are two kinds of non-interacting electrons, extended and localized. They satisfy different boundary conditions, and correspond to different calculation methods.

2.1 Extended non-interacting electron model

Each one-electron wave function $\Phi_n^{\sigma}(\mathbf{r})$ is constrained to spread over the whole region occupied by the system. Under this model, the equation (4), with periodic boundary condition, can be used to a perfect periodical crystal. The single-electron Hamiltonian with periodic boundary condition has the periodicity of $\rho(\mathbf{r})$, the Bloch theorem is valid and band structure calculation is performed. For a free cluster, equation (4) is solved with the natural finite boundary condition $\Phi_n^{\sigma}(\mathbf{r})\xrightarrow{r\to\infty}0$, and the single-electron Hamiltonian has the point symmetry of the free cluster.

2.2 Localized non-interacting electron model

Each one-electron wave function $\Phi_n^{\sigma}(\mathbf{r})$ is constrained to distribute in a part of the region occupied by the system. Under this model, equation (4) is used for the self-consistent cluster-embedding (SCCE) calculation [1]: the system is divided into k embedded clusters, and $N\Phi_n^{\sigma}(\mathbf{r})$ are divided into k groups. The $\Phi_n^{\sigma}(\mathbf{r})$ in the ith group satisfy the ith set of special boundary conditions, and localize in the ith region (embedded cluster).

The details are as follows. Consider the ith embedded cluster whose electronic density is represented by $\rho_1(\mathbf{r})$. The rest of the system is treated as environment with electronic density $\rho_2(\mathbf{r})$ which has small overlap with $\rho_1(\mathbf{r})$.

Because all $N \Phi_n^\sigma(\mathbf{r})$ are localized, we have ($N = N_1 + N_2$)

$$\begin{aligned}\rho(\mathbf{r}) &= \sum_{occupied \ldots n,\sigma}^{N} \left|\Phi_n^\sigma(\mathbf{r})\right|^2 \\ &= \sum_{occupied \ldots n_1,\sigma}^{N_1} \left|\Phi_{n_1}^\sigma(\mathbf{r})\right|^2 \\ &+ \sum_{occupied \ldots n_2,\sigma}^{N_2} \left|\Phi_{n_2}^\sigma(\mathbf{r})\right|^2 \equiv \rho_1(\mathbf{r}) + \rho_2(\mathbf{r}) \quad (2')\end{aligned}$$

$$\begin{aligned}T_{ni}[\rho] &= T_{ni}[\rho_1 + \rho_2] \\ &= \sum_{occupied \ldots n,\sigma}^{N} \int \Phi_n^{\sigma *}(\mathbf{r}) \left(-\nabla^2\right) \Phi_n^\sigma(\mathbf{r})\, d\mathbf{r} \\ &= \sum_{occupied \ldots n_1,\sigma}^{N_1} \int \Phi_{n_1}^{\sigma *}(\mathbf{r}) \left(-\nabla^2\right) \Phi_{n_1}^\sigma(\mathbf{r})\, d\mathbf{r} \\ &+ \sum_{occupied \ldots n_2,\sigma}^{N_2} \int \Phi_{n_2}^{\sigma *}(\mathbf{r}) \left(-\nabla^2\right) \Phi_{n_2}^\sigma(\mathbf{r})\, d\mathbf{r} \\ &\equiv T_{ni}[\rho_1] + T_{ni}[\rho_2]. \quad (3')\end{aligned}$$

A zero-value term $\int \rho_1(\mathbf{r}) V_{or} \mathrm{d}\mathbf{r}$ is added to the right side of formula (1). For fixed $\rho_2(\mathbf{r})$, using formulas ($2'$) and ($3'$), the variational principle now leads to the basic equation of the SCCE method [1]:

$$\left\{ -\nabla^2 + 2\int \frac{\rho_1(\mathbf{r}') + \rho_2(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' - 2\sum_{i=1}^{M} \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|} + V_{xc}^\sigma(\mathbf{r}) + V_{or}(\mathbf{r}) \right\} \Phi_n^\sigma(\mathbf{r}) = \varepsilon_n \Phi_n^\sigma, \quad (4')$$

where the $\Phi_n^\sigma(\mathbf{r})$ represent only the non-interacting electrons localized in and around the ith embedded cluster.

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

$$V_{or} = \begin{cases} 2\sum\limits_{j=1}^{M_2} \frac{Z_j}{|\mathbf{r} - \mathbf{R}_j|} & \text{if } \vec{r} \text{ is in the core regions of surrounding atoms} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

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

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

The physical reasons of boundary condition (7) were given in reference [1]. Consider the formula ($2'$), (6) and (7), it is easy to see that as long as the boundary condition (7) is satisfied, we have $\int \rho_1(\mathbf{r}) V_{or}(\mathbf{r}) \mathrm{d}\mathbf{r} = 0$. So the $V_{or}(\mathbf{r})$ in equation ($4'$) does not change the energy functional (1). Thus, the only effect of $V_{or}(\mathbf{r})$ is causing the special boundary condition (7). What we have done is only transforming the boundary condition (7) into an equivalent potential $V_{or}(\mathbf{r})$ in equation ($4'$). So the equation ($4'$) is just the Kohn-Sham equation (4) with the special boundary conditions (7). In addition, there is a special finite boundary condition for $\Phi_n^\sigma(\mathbf{r})$ because of its locality:

$$\Phi_n^\sigma(r) \xrightarrow{\text{r go away from the embedded cluster}} 0. \quad (8)$$

The boundary conditions (7) and (8) are different for different embedded clusters. By calculating all k embedded clusters one by one, equation ($4'$) gives a complete set of one-electron eigenfunctions of whole system which makes the total energy in formula (1) minimum [1].

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

3 Geometric structure of Cys + 5H$_2$O system

The geometric structure and electronic structure of Cys + 5H$_2$O system are determined by free cluster calculation. Considering the degrees of freedom of the geometric structure of Cys in solution, five water molecules are apparently not enough in describing the water molecule effect on the geometric structure of Cys in solution. However, our concern is the electronic structure rather than the geometric structure. We believe that five water molecules are enough to describe the water molecule effect on the electronic structure of Cys in solution. The reasons are as follows:

(1) it is reasonable to assume that the Cys' electronic states, whose energies being much lower than the Fermi energy, do not be changed by water molecules. Besides, the properties and functions of a protein are mainly determined by the electronic states near the Fermi level. So we are actually use five water molecules to simulate the water solvent effect on the Cys' electronic states near the Fermi level. If we choose ten electronic states near the Fermi level, the degrees of freedom are small, and five water molecules (containing 15 atoms, more than the 14 atoms of Cys) should be reasonable;

(2) the valence electrons in amino acid residue are localized electrons. The protein's electronic states near the Fermi level are mainly the localized electrons of the N-terminal H_3N^+, the C-terminal COO^-, and the tip of lateral chain (neutral or charged). So only the water molecules surrounding the three parts need to be considered. Our calculations show that three water

Table 1. Atomic coordinates of Cys cluster.

No.	Atom	X (Å)	Y (Å)	Z (Å)
1	C	0.7190	–0.1540	0.0100
2	N	0.1250	–1.0540	1.0460
3	C	1.9970	0.4500	0.5410
4	O	3.0520	0.2520	–0.0400
5	O	1.9750	1.1630	1.5910
6	C	–0.2710	0.9620	–0.3220
7	S	–1.8120	0.2350	–0.9620
8	H	0.9340	–0.7270	–0.8890
9	H	0.1600	1.6170	–1.0760
10	H	–0.4870	1.5350	0.5770
11	H	–0.7510	–1.4680	0.6820
12	H	0.7960	–1.8100	1.2710
13	H	–0.0800	–0.5090	1.9010
14	H	–2.4800	1.3660	–1.1670

molecules are needed near N-terminal H_3N^+, one water molecule is needed near C-terminal COO^-, and one water molecule is needed near the neutral tip of lateral chain. Thus, only five water molecules are needed for Cys;

(3) of course, the more water molecules are used, the better approximation the results get. However, more water molecules would make the adjusting of geometric structure and the dipoles much more difficult. Limited by the computational capacity, we choose five water molecules. The software of "free cluster calculation" is developed by the group of Prof. Callaway in Department of Physics and Astronomy, Louisiana State University (USA) [23]. Electronic structures of many molecules and clusters have been calculated using this software [24–31].

3.1 Initial geometric structure of Cys + $5H_2$ O system

The coordinates of 14 atoms of Cys in solution are come from Protein Data Bank (PDB). In neutral solvent (pH = 7), the Cys' N-terminal gets a proton to become H_3N^+, and the C-terminal loses a proton to become COO^-. So polar water molecules mainly influence Cys' charged H_3N^+, COO^-, and the tip of lateral chain. At beginning, the 5 water molecules are put randomly around the Cys: one water molecule near the H_3N^+, one near the COO^-, and three near the tip of lateral chain. The atomic coordinates of Cys in solution are given in Table 1.

The linear combination of Gaussian orbits is used as basis function. The optimized Gaussian bases are the same as that used in the electronic structure calculation of proteins [9–12]: 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 538. There are 319 002 grid points used for numerical calculation of exchange-correlation energy.

By solving the Kohn-Sham equation (4) self-consistently, we get the electronic structure, total energy, and force applied to each atom.

Table 2. Final atomic coordinates of five water molecules.

No.	Atom	X (Å)	Y (Å)	Z (Å)
1	O	–5.5352	1.3298	–1.7041
	H	–5.2262	1.2353	–0.8031
	H	–4.7643	1.6175	–2.1933
2	O	3.3646	1.7794	–2.5639
	H	3.0630	1.1295	–1.9291
	H	3.6670	2.5141	–2.0301
3	O	2.1098	–5.3434	–0.8761
	H	2.3048	–4.5721	–0.3439
	H	1.6990	–4.9923	–1.6661
4	O	–2.3037	–1.3067	1.0990
	H	–2.9388	–1.1917	1.8059
	H	–2.5326	–0.6329	0.4588
5	O	–0.9277	–3.8950	0.0000
	H	0.0202	–3.9850	0.0989
	H	–1.1813	–4.6287	–0.5600

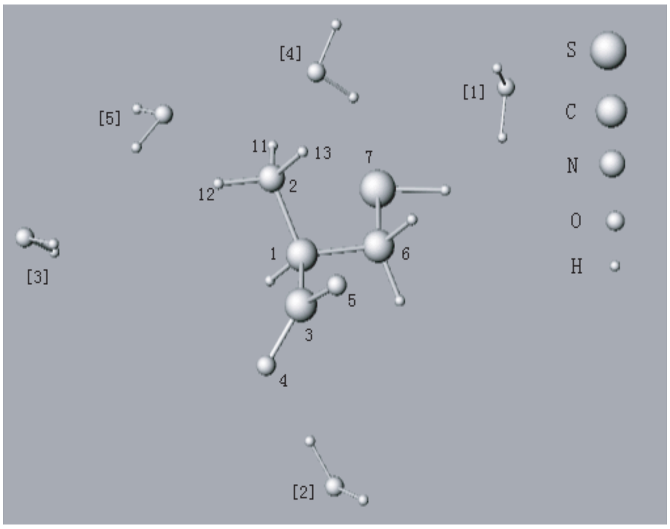

Fig. 1. Final geometric structure of Cys + $5H_2O$ system.

3.2 Adjusting of five water molecules

In the Cys + $5H_2O$ system, the geometric structure of Cys should remain unchanging, so that for each water molecule. Only the relative positions between water molecules and Cys should be adjusted. In the calculation, the water molecules are moved according to the applied forces, while the geometric structure of each water molecule is kept. The five molecules are adjusted one by one. After hundreds of adjustments, the geometric structure of Cys + $5H_2O$ system with the lowest total energy is got. The final atomic coordinates of five water molecules are given in Table 2. The final geometric structure of Cys + $5H_2O$ system is shown in Figure 1. The total energy of the final structure is –2199.2309 Ry, 3.9229 Ry lower than that of initial structure. As shown in the Figure 1, there are now three water molecules near H_3N^+, one near COO^-, and one near the S-H tip of lateral chain.

Table 3. Part of eigenvalues and Mulliken populations of Cys with the potential of 5 water molecules.

State	Energy (Ry)	Mulliken populations								
		C		N		O		H	S	
		s	*p*	*s*	*p*	*s*	*p*	*s*	*s*	*p*
34	–0.1607	–0.3275	0.1874	–0.3415	–0.2230	–0.0114	0.0075	1.7504	–0.0165	–0.0471
33	–0.2041	2.6006	0.2603	–3.2808	0.2594	–0.0701	0.0166	1.0954	–0.0666	0.1655
				Above are unoccupied states						
32	–0.4062	0.0301	0.1404	–0.0462	–0.0056	–0.0655	0.9381	0.0043	–0.0006	0.0044
31	–0.4122	0.0055	0.0514	–0.0205	0.0012	–0.0027	0.9551	0.0032	0.0007	0.0056
30	–0.4655	0.1001	0.1443	–0.0180	0.0001	–0.0124	0.7489	–0.0058	0.0010	0.0405
29	-0.5158	–0.0645	–0.0310	0.0100	0.0000	0.0130	0.0301	0.1493	0.0903	0.7898
28	–0.6887	0.0051	0.3720	–0.0329	0.0051	0.0806	0.4034	0.0203	–0.0048	0.1480
27	–0.6999	0.0447	0.2399	–0.0451	–0.0008	0.1763	0.3378	0.0387	–0.0108	0.2154
26	–0.7145	0.0484	0.3045	0.0076	0.0108	0.0190	0.2471	0.0540	0.0080	0.2953
25	–0.7872	0.0998	0.1866	–0.0192	–0.0007	0.2060	0.4369	0.0305	0.0111	0.0467

Note: because of little amount, *d*-electrons and *p* electrons of H atom are not given in this table and the following tables.

4 The electronic structure of Cys in the potential of water molecules

Based on the geometric structure of Cys + $5H_2O$ system determined in Section 3, the system is now divided into six clusters for the "self-consistent cluster-embedding calculation (SCCE)". The Cys molecule is the first cluster, and each water molecule is one cluster. The software of the SCCE calculation is developed by our group, and has been applied to several insulators, semi-conductors, metals, crystals with defects and impurities, surfaces, and proteins [2–12].

As proven in Section 2, for the SCCE calculation, the potential is not changed, only the one-electron wave functions are chosen to be localized: each one-electron wave function belongs to one cluster, and localizes in the region of the cluster. So the electronic structure of Cys is separated from that of water molecules, i.e., the electronic structure of Cys in the potential of water molecules is obtained.

The calculation contains two kinds of iterations: (i) intra-cluster iteration. For each embedded cluster, equation ($4'$) is calculated self-consistently: $\rho_1(\mathbf{r})$ of the embedded cluster is self-consistently changed during the iterations, while the rest of the system is served as fixed environment $\rho_2(\mathbf{r})$; (ii) inter-cluster iteration. The 6 embedded clusters are synchronously calculated by 6 CPUs, respectively. After the convergence of intra-cluster iterations of all 6 embedded clusters, the results are used for constructing new environments $\rho_2(\mathbf{r})$ for each embedded cluster, and a new inter-cluster iteration begins. After ten inter-cluster interactions, we get converged result. Table 3 gives the information of orbits 25 to 34.

5 The equivalent potential of water molecules on Cys simulated by dipoles

Each water molecule in Section 4 is now replaced by a dipole: the O atom is replaced by a negative point charge, and two H atoms are replaced by a positive point charge located in the middle of line connecting the two H atoms. Adjusting the point charges and positions of dipoles, the electronic structure of Cys is recalculated using the SCCE calculation. There is now no inter-cluster iteration, however, only the intra-cluster iteration of the Cys cluster needs to be performed because the dipoles have no electron.

The electronic structure obtained in Section 4 is considered approximately as the electronic structure of Cys in water solvent. According to that, the dipoles are adjusted. In order to evaluate the difference between the calculated electronic structure and that obtained in Section 4, two 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 the ε_n^{σ} and $\varepsilon_{n0}^{\sigma}$ are the eigenvalues of nth electronic state with spin σ calculated in this section and in Section 4 respectively. N^{σ} is the numbers of electrons with spin σ.

(2) *The equivalent mean square deviation of charge density*

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

where the C_{ni}^{σ} and C_{ni0}^{σ} are the expansion coefficients of the eigenfunctions of the nth electronic states with spin σ calculated in this section and in Section 4 respectively. N is the number of Gaussian bases used to expand the one-electron wave function. (please pay

Table 4. Point charges and coordinates of 5 dipoles.

No.	Charge	X (Å)	Y (Å)	Z (Å)
1	–0.29 e	–5.9401	1.2573	–1.8585
	+0.29 e	–5.4002	1.3539	–1.6526
2	–0.9 e	3.3648	1.8006	–2.2718
	+0.9 e	3.3652	1.8430	–1.6874
3	–0.692 e	1.6783	–3.0984	–1.3918
	+0. 692 e	1.5704	–2.5371	–1.5208
4	–0. 692 e	–2.9517	–0.7151	1.1490
	+0. 692 e	–3.3837	–0.3208	1.1823
5	–0. 692 e	–1.5351	–3.1742	0.4035
	+0. 692 e	–1.1880	–3.5861	0.1729

attention to the formula (2):

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

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

The initial charge of dipoles is set to 0.5 e. The distances between positive charge and negative charge is 0.5858 Å, witch remains unchanging. After the convergence, the values of the two criteria are calculated. Then the point charges and the positions of the 5 dipoles are adjusted one by one, until the two criteria reach the minimum. Thus, the potential of 5 dipoles simulates the potential of water molecules on the electronic structure of Cys. The final charges of dipoles are as follows: the charges of dipoles near H_3N^+, COO^-, and S-H tip of lateral chain are 0.692 e, 0.29 e, and 0.9 e, respectively. The final charges and coordinates of five dipoles are given in Table 4, and the structure is shown in Figure 2. Table 5 gives the eigenvalues and Mulliken populations of ten orbits near Fermi level of Cys in the potential of five dipoles.

6 Discussion

In order to study the water solvent effect on the electronic structure of Cys, the electronic structure of an isolated Cys is calculated by free cluster calculation. The total energy of an isolated Cys is –1438.2737 Ry. The results are given in Table 6.

The eigenvalues of orbits 25 to 34 of Cys with three potentials are given in Table 7. The last row gives the energy gap: $Eg = E_{33} - E_{32}$. Figure 3 shows the comparison between the three sets of eigenvalues of Cys in the potentials of dipoles, water molecules and no potential, respectively.

The properties of Cys are mainly determined by the electronic states near the Fermi level. In the Tables 3, 5 and 6, the Mulliken population analysis shows that in all three cases, the electronic states near the Fermi level are similar: three oxygen 2*p* electrons of COO^- below the Fermi level (states 32, 31 and 30), one sulphur 3*p* electron of S-H tip of lateral chain (state 29), and a hybridized state of carbon 2*p* electron and oxygen 2*p* electron of COO^- (state 28).

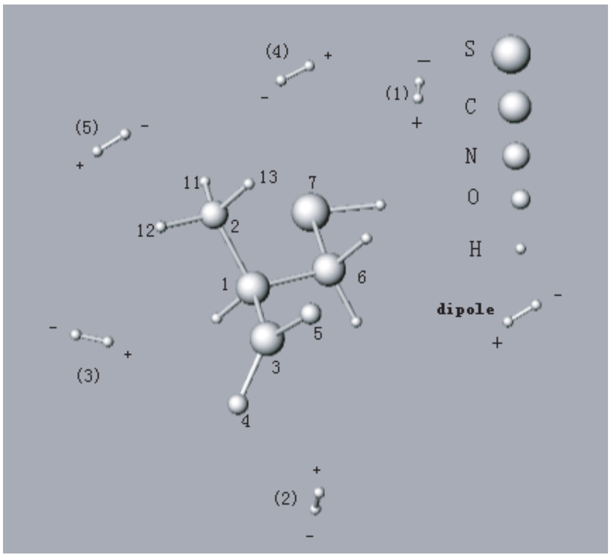

Fig. 2. The final geometric structure of Cys + 5 dipoles system.

By comparing the second and third columns of Table 7, as well as Figures 3b and 3c, 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 to lower all the eigenvalues of Cys by about 0.032 Ry, and to broaden the energy gap by 16%. In Figure 3, the energy origin of 3c has been moved down so the Fermi levels of the Figures 3a–3c being in the same horizontal line. It is easily found that in Figures 3b and 3c, the relative positions of the occupied electronic states are very close.

The comparison of the second and third columns of Table 7, as well as Figures 3a and 3b, reveals that below the Fermi level, eigenvalues of Cys in the potential of dipoles are close to that of Cys in the potential of water molecules. But above the Fermi level, the energy gap is broadened by 14%.

7 Conclusion

The geometric structure of Cys + $5H_2O$ system with the lowest total energy is determined by free cluster calculation. Based on the geometric structure above, the electronic structure of Cys with the potential of water molecules is calculated using the self-consistent cluster-embedding (SCCE) calculation. Then the water molecules are replaced by adjustable dipoles. The dipoles are adjusted so the electronic structure of Cys with the potential of dipoles is close to that of water molecules. The calculations show that the major effect of water molecules on

Table 5. Part of eigenvalues and Mulliken populations of Cys with the potential of 5 dipoles.

State	Energy (Ry)	Mulliken populations								
		C		N		O		H	S	
		s	*p*	*s*	*p*	*s*	*p*	*s*	*s*	*p*
34	–0.1126	–0.6677	0.3159	–0.0098	–0.1205	–0.0004	–0.0075	1.2558	0.0449	0.1641
33	–0.1765	0.3186	0.0578	–0.5871	0.0736	0.0262	0.0204	0.9692	0.0121	0.0917
		Above are unoccupied states								
32	–0.4078	0.0081	0.0447	–0.0168	0.0000	–0.0054	0.9536	0.0063	0.0003	0.0087
31	–0.4142	0.0125	0.1187	–0.0004	–0.0016	–0.0632	0.9351	–0.0054	0.0005	0.0033
30	–0.4714	0.0711	0.1625	–0.0187	–0.0002	–0.0183	0.7746	0.0097	0.0003	0.0180
29	–0.5200	–0.0076	0.0426	–0.0231	–0.0035	0.0008	0.0143	0.0864	0.0027	0.8783
28	–0.6735	–0.0262	0.2447	0.0028	–0.0036	0.0103	0.0604	0.0815	0.0988	0.5228
27	–0.6904	0.0180	0.3745	–0.0014	0.0073	0.0584	0.4172	0.0482	0.0205	0.0538
26	–0.7133	–0.0020	0.2397	–0.0019	0.0199	0.1848	0.5132	0.0241	0.0069	0.0144
25	–0.7902	0.0355	0.2164	0.0042	0.0027	0.2234	0.3883	0.0421	0.0132	0.0702

Table 6. Part of eigenvalues and Mulliken populations of an isolated Cys.

State	Energy (Ry)	Mulliken populations								
		C		N		O		H	S	
		s	*p*	*s*	*p*	*s*	*p*	*s*	*s*	*p*
34	–0.1090	–0.6627	0.2629	–0.0758	–0.0452	0.0003	–0.0087	1.2974	0.0849	0.1232
33	–0.2002	0.2782	–0.0317	–0.7013	0.0852	–0.0143	0.0110	1.3039	0.0371	0.0144
		Above are unoccupied states								
32	–0.3738	0.0052	0.0217	–0.0029	0.0011	–0.0056	0.9714	0.0051	0.0003	0.0035
31	–0.3807	0.0061	0.1417	–0.0092	–0.0060	–0.0542	0.9124	0.0013	0.0005	0.0064
30	–0.4377	0.0909	0.1317	–0.0165	–0.0004	–0.0137	0.7872	0.0100	0.0009	0.0092
29	–0.5221	–0.0008	0.0329	–0.0010	0.0017	–0.0019	0.0045	0.0679	0.0001	0.8881
28	–0.6530	–0.0038	0.3986	–0.0041	0.0081	0.0502	0.4472	0.0158	0.0028	0.0822
27	–0.6723	–0.0312	0.2522	0.0159	–0.0039	0.0540	0.2050	0.1026	0.0741	0.3252
26	–0.6857	0.0003	0.2322	–0.0003	0.0240	0.1315	0.3621	0.0407	0.0498	0.1564
25	–0.7609	0.0484	0.1692	0.0019	0.0023	0.2760	0.4409	0.0180	0.0108	0.0309

Table 7. Three sets of eigenvalues of Cys.

Eigenvalues	Dipoles Energy (Ry)	Water molecules Energy (Ry)	No potential Energy (Ry)
34 (unoccupied)	–0.1126	–0.1607	–0.1090
33 (unoccupied)	–0.1765	–0.2041	–0.2002
32 (E_F)	–0.4078	–0.4062	–0.3738
31	–0.4142	–0.4122	–0.3807
30	–0.4714	–0.4655	–0.4377
29	–0.5200	–0.5158	–0.5221
28	–0.6735	–0.6887	–0.6530
27	–0.6904	–0.6999	–0.6723
26	–0.7133	–0.7145	–0.6857
25	–0.7902	–0.7872	–0.7609
Eg	0.2313	0.2021	0.1736

Cys's electronic structure is to lower all occupied eigenstates by about 0.032 Ry, and broadening energy gap by 16%. The effect of water molecules on the electronic structure of Cys can be simulated by the potential of dipoles, the eigenvalues and Mulliken populations calculated using two kinds of potentials are very close. So we established a simple, easy-use, with almost no additional computational effort, dipoles potential which simulated the effect of water molecules on Cys' electronic structure.

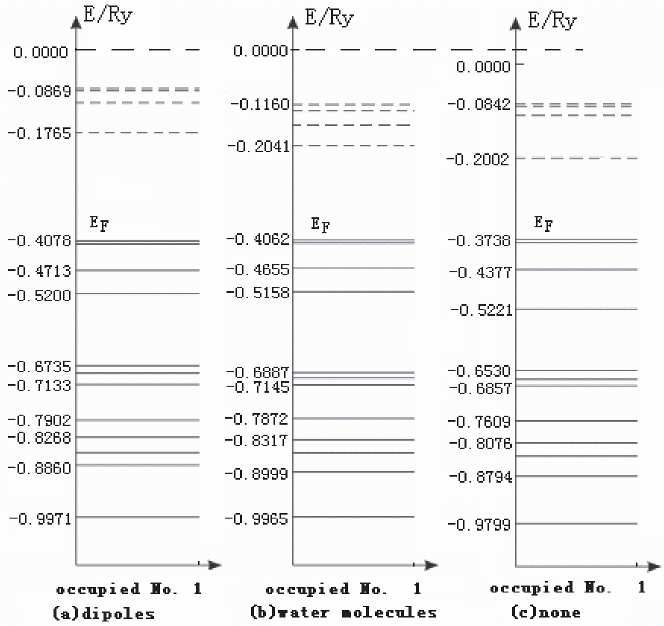

Fig. 3. Comparison of three sets of eigenvalues (Real line represents occupied state, dashed line represents unoccupied state).

A single chain protein molecule in solution has a N-terminal H_3N^+, a C-terminal COO^-, and many lateral chains which can be sorted into three kinds, hydrophilic, hydrophobic and neutral. This work shows that simple dipole potential can be used to simulate the effect of water molecules on the electronic structure of Cys. More dipole potentials simulating the potentials of water molecules on other amino acids will be constructed. All results will be directly applied to the electronic structure calculation of protein in solution. In order to make easy use of the dipole potential, it may be worth to classify the dipole potentials into five kinds suited for N-terminal H_3N^+, C-terminal COO^-, hydrophilic lateral chain, hydrophobic lateral chain and neutral lateral chain, respectively.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 30470410), and by the Science and Technology Development Foundation of Shanghai (Grant No. 03JC14070). The work was also supported by Shanghai Supercomputer Center. The calculations were performed on the super computer DAWN 4000A of Shanghai Supercomputer Center of China.

References

1. Haoping Zheng, Phys. Lett. A **226**, 223 (1997); **231**, 453 (1997)
2. Haoping Zheng, Phys. Rev. B **48**, 14868 (1993)
3. Haoping Zheng, Physica B **212**, 125 (1995)
4. H. Zheng, B.K. Rao, S.N. Khanna, P. Jena, Phys. Rev. B **55**, 4174 (1997)
5. H. Zheng, Y. Wang, G. Ma, Eur. Phys. J. B **29**, 61 (2002)
6. J. He, H. Zheng, Acta Phys. Sin. (in Chinese) **51**, 2580 (2002)
7. Shaojie Lin, Haoping Zheng, Acta Phys. Sin. (in Chinese) **54**, 4680 (2005)
8. Haoping Zheng, Shaojie Lin, J. Phys. Conf. Ser. **29**, 129 (2006)
9. Haoping Zheng, Phys. Rev. E **62**, 5500 (2000)
10. Haoping Zheng, Progress in Physics (in Chinese) **20**, 291 (2000)
11. Haoping Zheng, Modern Phys. Lett. B **16**, 1151 (2002)
12. Haoping Zheng, Phys. Rev. E **68**, 051908 (2003)
13. Themis Lazaridis, Martin Karplus Proteins: Structure, Function, and Genetics **35**, 133 (1999)
14. Themis Lazaridis, Martin Karplus, J. Mol. Boil. **288**, 477 (1998)
15. L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936)
16. A. Klamt, G. Schuurmann, J. Chem. Soc. Perkin. Trans. **2**, 799 (1993)
17. Hong Guo, Martin Karplus, J. Phys. Chem. **98**, 7104 (1994)
18. Michael Schaefer, Martin Karplus, J. Phys. Chem. **100**, 1578 (1996)
19. F. Eckert, A. Klamt, AIChE J. **48**, 369 (2002)
20. J.B. Foresman, T.A. Keith, K.B. Wiberg, J. Snoonian, M.J. Frisch, J. Phys. Chem. **100**, 16098 (1996)
21. P. Hohenberg, W.J. Kohn, Phys. Rev. B **136**, 864 (1964)
22. W. Kohn, L.J. Sham, Phys. Rev. A **140**, 1133 (1965)
23. H. Chen, Ph.D. thesis, Department of Physics and Astronomy, Louisiana State University 1988 (unpublished)
24. H. Chen, J. Callaway et al., Phys. Rev. B **38**, 195 (1988)
25. H. Chen, J. Callaway et al., Phys. Rev. B **44**, 2289 (1991)
26. H. Zheng, Ph.D. thesis, Department of Physics and Astronomy, Louisiana State University 1993 (unpublished)
27. Haoping Zheng, Jun He, J. Tongji Univ. (in Chinese) **29**, 593 (2001)
28. Wanhong Xu, Haoping Zheng, J. Tongji Univ. (in Chinese) **31**, 374 (2003)
29. Shaojie Lin, Haoping Zheng, J. Tongji Univ. (in Chinese) **32**, 551 (2004)
30. Jingan Hao, Haoping Zheng, Acta Phys. Sin. (in Chinese) **53**, 1044 (2004)
31. Haoping Zheng, Jingan Hao, Chinese Physics **14**, 529 (2005)