

The equivalent potential of water molecules for the electronic structure of histidine

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2007 J. Phys.: Condens. Matter 19 116102

(<http://iopscience.iop.org/0953-8984/19/11/116102>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 129.252.86.83

The article was downloaded on 28/05/2010 at 16:35

Please note that [terms and conditions apply](#).

The equivalent potential of water molecules for the electronic structure of histidine

Chunjie Li, Haoping Zheng and Xuemei Wang

Pohl Institute of Solid State Physics, Tongji University, Shanghai 200092,
People's Republic of China

Received 10 November 2006, in final form 22 January 2007

Published 27 February 2007

Online at stacks.iop.org/JPhysCM/19/116102

Abstract

In order to get more reliable electronic structure for protein in aqueous solution, it is necessary to construct a simple, easy-to-use equivalent potential of water molecules for protein electronic structure calculations. First-principles, all-electron, *ab initio* calculations have been performed to construct the equivalent potential of water molecules for the electronic structure of histidine (His) in solution. The process consists of three steps. First, the geometric structure of the His + $n\text{H}_2\text{O}$ system is determined by searching for the lowest energy of the system using a free cluster calculation method. Then, on the basis of the geometric structure obtained, the electronic structure of His with the potential of water molecules is calculated using the self-consistent cluster-embedding method. Finally, after replacing water molecules with dipoles, the electronic structure of His with the potential of dipoles is calculated. The dipoles are adjusted so that the electronic structure of His with the potential of dipoles is close to that of water molecules. The calculations show that the major effect of water molecules on the electronic structure of His is as follows: (1) the energy gap broadens by 33%; (2) two states, containing the contribution of carbon 2p electrons from the atom C_α , remain approximately unchanging, while the other states rise by about 0.049 Ryd which causes the exchange of state positions. The effect of water molecules on the electronic structure of His can be simulated by the dipole potential.

1. Introduction

The structural taxonomy of proteins has been intensively studied using crystallography and nuclear magnetic resonance (NMR) spectroscopy for decades. More than 30 thousand protein structures are known. In contrast, the computational research into the electronic structure of protein by first-principles calculation has not yet been well performed because of the huge number of atoms a protein consists of, and the limitation of both the computational conditions and the traditional calculation methods. But knowledge of the electronic structure is essential

for understanding the properties and functions of a molecule. It is from this perspective that a feasible computation technique that can greatly reduce the computational effort is needed for solving such problems.

'Self-consistent cluster-embedding calculation' (SCCE), a first-principles, all-electron, *ab initio* calculation method, was developed on the basis of density functional theory (DFT) [1, 2]. Unlike the traditional calculation method, the one-electron wavefunctions obtained by the SCCE method are localized—each one-electron wavefunction is localized only in part of the region occupied by the system. This approach results in two advantages: (1) the SCCE method can be applied to any complex system, and the localized valence electrons in a material can be better described using the localized wavefunctions obtained in the SCCE calculation; (2) the computational effort is reduced greatly, from the scaling with N^3 of traditional method to scaling with N (N is the number of atoms of the system), while the calculation precision is kept. This makes the electronic structure calculation for protein a reality. The SCCE calculations have been successfully applied to several insulators, semiconductors, metals, crystals with defects and impurities, and surfaces [2–6]. For the first time, the first-principles, all-electron, *ab initio* calculation of electronic structure of a real protein was completed in 2000 [7, 8]. To date, the electronic structures of three proteins with four three-dimensional structures have been obtained [7–10] (see the Website www.esprotein.org.cn). However, the former calculations did not include solvent influence due to limited computational resources and the following reasons: water molecules are usually quivering, turning and fungible at high speed in action with protein. There is no fixed chemical bonding 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 structure 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.

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 it approaches the upper limit for a supercomputer. So it is impossible to add a large amount of water molecules to the calculation. It is necessary to construct an equivalent potential for water molecules, which must be simple, easy-to-use, effective and with little additional computational effort.

Some equivalent potentials had been constructed for the geometric structure calculation of proteins, say, protein folding and molecule dynamics calculation [11, 12]. For example, people include the solvation effect in energy parameters, or use the solvation model which considers the effect of solvent as a smooth solvation potential represented by an equivalent function [13, 14]. The solvent hydrogen bond influence (water, ethanol and resin) on the stability of the peptide has been calculated by Guo and Karplus by first-principles calculation, in 1994 [15]. In 1995, the solvent's effect on several small protein molecules such as the bovine pancreatic trypsin inhibitor (BPTI) has also been studied by Schaefer [16]. In recent years, some other models have been developed such as the Poisson–Boltzmann (PB) continuum model [17] and MM/PBSA model [18]. 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 at its primary stage, there is no suitable equivalent potential of water molecules 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—the building blocks of protein. Those for two amino

acids, cysteine (Cys) and lysine (Lys), have already been successfully calculated [19, 20]. In this paper, the equivalent potential of water molecules for the electronic structure of histidine (His) is constructed by first-principles calculation. His has 20 atoms with an alkaline lateral chain. It belongs to the polar amino acids, and is hydrophilic. In neutral solution, the histidine's N-terminal gets a proton to become H_3N^+ , the C-terminal loses a proton to become COO^- , and the tip of lateral chain is a benzene ring.

This work is based on two considerations. (1) There is no fixed chemical bond between water molecules and protein in solvent, so no fixed relative position of water molecules and protein. The water molecules are with the greatest 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 the solvent environment. (2) The electronic structure of protein is calculated using the SCCE method. The dipoles made up of point charges can be easily added to the SCCE calculation with little 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 His.

The work consists of three steps. First, the geometric structure of the His + $n\text{H}_2\text{O}$ system is determined using the free cluster calculation. Second, on the basis of the geometric structure obtained, the electronic structure of His with the potential of water molecules is calculated using the self-consistent cluster-embedding (SCCE) method. Third, after replacing water molecules with dipoles made up of point charges, the electronic structure of His with the potential of dipoles is calculated. The dipoles are adjusted so the electronic structure of His 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 His is established using the dipoles made up of point charges. The equivalent potential developed in this paper can be applied directly to the electronic structure calculation of protein using the SCCE method.

2. Theoretical model

The free cluster calculation and the self-consistent cluster-embedding calculation methods have been described in detail elsewhere (see [2, 21, 24] and Website www.esprotein.org.cn). Here we only give a brief overview for completeness.

According to the density functional theory (DFT) [25, 26], 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 $\text{Ryd} = e^2/2a_0 = 13.6 \text{ eV}$):

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

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

$$\rho(\mathbf{r}) = \rho^\uparrow(\mathbf{r}) + \rho^\downarrow(\mathbf{r}) = \sum_{\text{occupied} \dots l} |\Phi_l^\uparrow(\mathbf{r})|^2 + \sum_{\text{occupied} \dots m} |\Phi_m^\downarrow(\mathbf{r})|^2, \quad (2)$$

$$T_{\text{ni}}[\rho] = \sum_{\text{occupied} \dots l} \int \Phi_l^{\uparrow*}(\mathbf{r})(-\nabla^2)\Phi_l^{\uparrow}(\mathbf{r}) \, d\mathbf{r} + \sum_{\text{occupied} \dots m} \int \Phi_m^{\downarrow*}(\mathbf{r})(-\nabla^2)\Phi_m^{\downarrow}(\mathbf{r}) \, d\mathbf{r}, \quad (3)$$

where the superscripts \uparrow and \downarrow represent spin up and spin down, respectively. Using formulae (2) and (3), a stationary state single-electron Schrödinger equation, i.e., the well known Kohn–Sham equation [26], is obtained by the variation of functional (1) with respect to $\Phi_n^{\sigma*}(\mathbf{r})$ under the conservation rule $\int \rho(\mathbf{r}) \, d\mathbf{r} = N$:

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

where the exchange–correlation potential is

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

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

In practical calculations, however, the two conditions above cannot be satisfied. The following approximations are used. First, the approximation of $E_{\text{xc}}[\rho]$: the exchange–correlation potential $V_{\text{xc}}^{\sigma}(\mathbf{r})$ is averaged over all one-electron states with spin σ (such as LDA). Second, we give up the unconstrained condition: each $\Phi_n^{\sigma}(\mathbf{r})$ is constrained to satisfy a certain boundary condition. Please note that the first approximation means that all $\Phi_n^{\sigma}(\mathbf{r})$ in Kohn–Sham equation (4) now correspond to the same $V_{\text{xc}}^{\sigma}(\mathbf{r})$ (and single-electron Hamiltonian), which has the symmetry of $\rho^{\sigma}(\mathbf{r})$, and all $\Phi_n^{\sigma}(\mathbf{r})$ are now constrained by this added symmetry if considering no boundary condition. But this added constraint is not physically essential, and can be removed by a boundary condition. The second approximation means that all wavefunctions which do not satisfy the boundary condition are thrown away, although they may be the true solutions for the system. When one kind of $\{\Phi_n^{\sigma}(\mathbf{r})\}$ is chosen (a boundary condition is applied), this means that a kind of non-interacting electron 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 the variational principle, the calculated energy will be close to the true ground-state energy only if the trial one-electron wavefunctions $\Phi_n^{\sigma}(\mathbf{r})$ describe real electrons well. And such a set of $\{\Phi_n^{\sigma}(\mathbf{r})\}$ is the best approximation solution. For example, a set of Bloch functions can give a good description to quasi-free electrons, but cannot do it for localized electrons. The latter can be best described by a set of localized one-electron wavefunctions.

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

2.1. Extended non-interacting electron model

Each one-electron wavefunction $\Phi_n^\sigma(\mathbf{r})$ is constrained to spread over the whole region occupied by the system.

Under this model, equation (4), with a periodic boundary condition, can be used for a perfect periodic crystal. The single-electron Hamiltonian with a periodic boundary condition has the periodicity of $\rho(\mathbf{r})$, the Bloch theorem is valid and a band structure calculation is performed. For a free cluster, equation (4) is solved with the finite boundary condition $\Phi_n^\sigma(r) \xrightarrow{r \text{ goes away from cluster}} 0$, and the single-electron Hamiltonian has the point symmetry of the free cluster.

2.2. Localized non-interacting electron model

Each one-electron wavefunction $\Phi_n^\sigma(\mathbf{r})$ 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: the system is divided into k embedded clusters, and the $N\Phi_n^\sigma(\mathbf{r})$ are divided into k groups. The $\Phi_n^\sigma(\mathbf{r})$ 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; then the other $k - 1$ embedded clusters are regarded as the fixed environment, and their atoms are the same as the surrounding atoms. Each surrounding atom has a core region. The electronic density of the i th embedded cluster is represented by $\rho_1(\mathbf{r})$; the electronic density of the other $k - 1$ embedded clusters is $\rho_2(\mathbf{r})$ which has small overlap with $\rho_1(\mathbf{r})$. Because all $N\Phi_n^\sigma(\mathbf{r})$ are localized, we have ($N = N_1 + N_2$)

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

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

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

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

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

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

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

where the M_2 is the number of surrounding atoms. It is obvious that 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 finite boundary condition caused by the $V_{\text{or}}(\mathbf{r})$:

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

The physical reasons for boundary condition (7) were given in [1]. Consider the formulae (2'), (6) and (7); it is easy to see that $\int \rho_1(\mathbf{r}) V_{\text{or}}(\mathbf{r}) d\mathbf{r} = 0$. So the $V_{\text{or}}(\mathbf{r})$ in equation (4') does not change the energy functional (1). Actually what we have done is only transform the boundary condition (7) into an equivalent potential $V_{\text{or}}(\mathbf{r})$ in equation (4'), so 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:

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

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

In practical calculations, 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 the His + 7H₂O system

The geometric structure and electronic structure of the His + 7H₂O system are determined by free cluster calculation. The reasons for taking seven water molecules are as follows:

- (1) It is reasonable to assume that the His electronic states, whose energies are much lower than the Fermi energy, are not 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 actually use seven water molecules to simulate the water solvent effect on the His electronic states near the Fermi level. If we choose ten electronic states near the Fermi level, the degrees of freedom are few, and seven water molecules (containing 21 atoms, more than the 20 atoms of His) should be reasonable.
- (2) The valence electrons in the amino acid residue are localized electrons. The protein's electronic states near the Fermi level are mainly the localized electrons of the N-terminal H₃N⁺, 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 molecules are needed near the N-terminal H₃N⁺, one water molecule is needed near the C-terminal COO⁻, and three water molecules are needed near the neutral tip of the lateral chain. Thus, seven water molecules are needed for His.

Table 1. Atomic coordinates of the His cluster.

No.	Atom	X (Å)	Y (Å)	Z (Å)
1	C	1.295	1.036	0.483
2	C	0.709	-0.204	1.156
3	C	-0.546	-0.625	0.429
4	N	-1.837	-0.208	0.7
5	C	-2.668	-0.842	-0.206
6	N	-1.919	-1.618	-1.001
7	C	-0.63	-1.514	-0.649
8	N	1.625	0.721	-0.941
9	C	2.55	1.457	1.209
10	O	2.619	2.565	1.717
11	O	3.534	0.659	1.292
12	H	-3.74	-0.73	-0.263
13	H	0.199	-2.027	-1.115
14	H	-2.119	0.439	1.422
15	H	1.436	-1.013	1.123
16	H	0.468	0.025	2.192
17	H	0.568	1.845	0.516
18	H	2.022	1.561	-1.397
19	H	0.765	0.433	-1.439
20	H	2.317	-0.048	-0.973

- (3) Of course, the more water molecules that are used, the better approximation the results give. However, more water molecules would make the adjusting of geometric structure and the dipoles much more difficult. Limited by the computational capacity, we choose seven water molecules.

The software for the free cluster calculation was developed by the group of Professor Callaway in the Department of Physics and Astronomy, Louisiana State University (USA) [21–23]. Electronic structures of many molecules and clusters have been calculated using this software [27–31].

A linear combination of Gaussian orbitals is used as the basis function. The optimized Gaussian bases are the same as those used in the electronic structure calculation for proteins [7–10]: 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 is 757. There are 451 002 grid points used for numerical calculation of the exchange–correlation energy.

3.1. Initial geometric structure of the His + 7H₂O system

The coordinates of 20 atoms of His in solution are given in table 1; these were originally from a PDB structure file provided by the Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, the Rockefeller University (<http://prowl.rockefeller.edu/aainfo/struct.htm>). In neutral solvent (pH = 7), the His N-terminal gets a proton to become H₃N⁺, and the C-terminal loses a proton to become COO⁻. So polar water molecules mainly influence the His charged H₃N⁺, COO⁻, and the tip of the lateral chain. At the beginning, the seven water molecules are put randomly around the His: three water molecules near the H₃N⁺, one near the COO⁻, and three near the tip of the lateral chain.

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

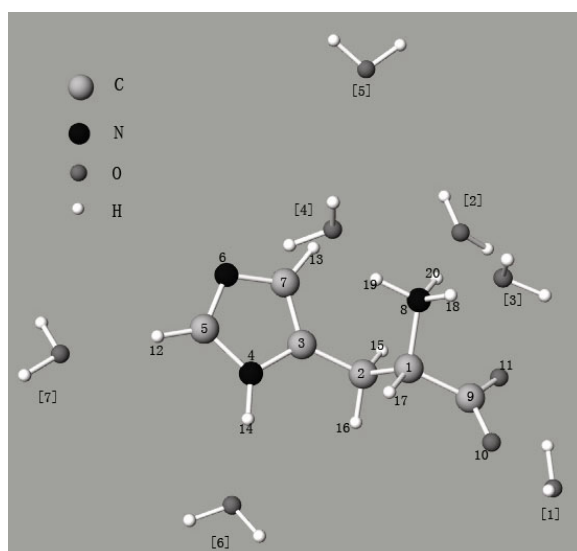


Figure 1. Final geometric structure of the His + 7H₂O system.

3.2. Adjusting of seven water molecules

For a real protein in solution at room temperature when the hydrogen bonds break easily, it is the free energy, which contains the entropy part, that determines the protein folding as well as the protein's final solution structure. For a certain solution structure at any moment, there is always one electron distribution (electronic structure) that makes the system's total energy minimum, because the movement of an electron is much faster than that of a nucleus. So the entropy affects firstly the protein's solution structure, which then determines the electronic structure of the protein. However, what we are interested in is neither the protein folding nor the solution structure, but the electronic structure of the protein with a known solution structure. In the His + 7H₂O system, the His has already been in its solution structure which is unchanging, and the structure of the water molecule is unchanging too. Only the relative position of the seven water molecules and His can be altered, around a certain structure having the minimum total energy. So the entropy difference caused by different positions of the seven water molecules is very small and can be reasonably neglected. Then the seven water molecules are most probably at the positions that make the total energy minimum, and the electronic structure of the protein with water molecules at these positions can be considered approximately as the electronic structure of protein in the solvent environment.

In the calculation, the water molecules are moved according to the applied forces, while the geometric structure of each water molecule is kept. The seven molecules are adjusted one by one. After hundreds of adjustments, the geometric structure of the His + 7H₂O system with the lowest total energy is obtained. The total energy of the final structure is -2157.1160 Ryd, 0.1070 Ryd lower than that of the original structure. The final atomic coordinates of the seven water molecules are given in table 2. The final geometric structure of the His + 7H₂O system is shown in figure 1, where the atoms and water molecules are numbered according to tables 1 and 2, respectively. There are three water molecules near H₃N⁺ (H₂O [2, 3] and [4]), one near COO⁻ (H₂O [1]), and three near the benzene ring of the lateral chain (H₂O [5, 6] and [7]).

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

No. of water molecule	Atom	X (Å)	Y (Å)	Z (Å)
[1]	O	4.5474	1.4705	3.8128
	H	4.6566	0.8189	3.1202
	H	4.1258	2.2130	3.3800
[2]	O	3.3725	-1.3911	-1.2786
	H	3.2887	-2.1873	-1.8033
	H	4.0913	-1.5761	-0.6741
[3]	O	2.6042	3.0051	-2.4025
	H	2.4447	3.6062	-3.1301
	H	3.4208	3.3140	-2.0101
[4]	O	-0.2697	0.5799	-2.8479
	H	-1.1512	0.3854	-2.5296
	H	-0.4083	1.0131	-3.6901
[5]	O	2.1147	-4.0535	-4.2382
	H	1.4878	-4.4162	-4.8640
	H	2.9717	-4.2319	-4.6256
[6]	O	-2.9817	1.9774	2.5211
	H	-3.9235	2.0130	2.6883
	H	-2.6436	2.8005	2.8739
[7]	O	-5.7963	-1.0721	-0.0245
	H	-5.9550	-1.9567	-0.3541
	H	-6.6688	-0.7234	0.1581

Of course, it is not absolutely certain that we have reached the best structure, i.e., the best relative positions of water molecules and His which has the lowest total energy. But we are sure that the total energy of our final structure is very close to the truth. So using the final structure, the calculated electronic structure of His in the potential of water molecules will be a good approximation to that in a real aqueous solution. The reasons are as follows. (i) We do not study the relative positions of the water molecules and His. (ii) There is no fixed chemical bond between the water molecules and the protein in solution, so there is no fixed relative position of the water molecules and the protein. The water molecules are most probably at the position which makes the total energy of the system minimum. (iii) In order to reduce the computational effort, charge density fitting is used in both free cluster calculation and band structure calculation: to calculate the electronic structure using a pseudo-charge density, which differs from the real one but can give a total energy very close to that calculated using the real charge density. It is deemed that the electronic structure calculated using the pseudo-charge density is a good approximation to the real electronic structure of the system.

4. The electronic structure of the His in the potential of the water molecules

On the basis of the geometric structure of the His + 7H₂O system determined in section 3, the system is now divided into eight clusters for the self-consistent cluster-embedding calculation (SCCE). The His molecule is the first cluster, and each water molecule is one cluster. The software for the SCCE calculation was developed by our group [2, 24], and has been applied to several insulators, semiconductors, metals, crystals with defects and impurities, surfaces, and proteins [2–12].

Table 3. Parts of eigenvalues and Mulliken populations of His with the potential of seven water molecules.

State	Energy (Ryd)	Mulliken populations							
		C		N		O		H	
		s	p	s	p	s	p	s	p
43	-0.0840	3.5340	0.5896	-1.1418	3.3921	-0.0032	-0.0010	-5.3757	0.0060
42	-0.1058	2.0244	-0.1382	4.9622	0.4838	-0.0348	0.0102	-6.3170	0.0095
The above are unoccupied states									
41	-0.3578	0.0102	0.0055	0.0178	0.0020	0.0005	0.9726	-0.0088	0.0001
40	-0.3635	-0.0585	0.0947	0.0321	0.0003	0.0452	0.9152	-0.0296	0.0006
39	-0.4114	-0.0033	0.6334	0.0347	0.0811	0.0143	0.2561	-0.0198	0.0035
38	-0.4380	0.1916	0.1589	0.0581	0.0369	-0.0005	0.5606	-0.0073	0.0018
37	-0.4880	-0.0644	0.1895	0.2351	0.6213	0.0012	0.0039	0.0111	0.0024
36	-0.5260	-0.0066	0.1324	0.1295	0.7434	-0.0001	0.0108	-0.0132	0.0039
35	-0.6217	-0.0072	0.4823	0.0087	0.0382	0.0062	0.4787	-0.0079	0.0011
34	-0.6569	-0.0339	0.3065	0.0207	0.0198	0.1117	0.4985	0.0745	0.0022

As proven in the section 2, for the SCCE calculation, the potential is not changed; only the one-electron wavefunctions are chosen for being localized: each one-electron wavefunction belongs to one cluster, and localizes in the region of the cluster. So the electronic structure of His is separated from that of the water molecules, i.e., the electronic structure of the His in the potential of the 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: $\rho_1(\mathbf{r})$ for the embedded cluster is self-consistently changed during the iterations, while the rest of the system serves as a fixed environment, $\rho_2(\mathbf{r})$. (ii) *Inter-cluster iteration*. The eight embedded clusters are synchronously calculated by eight CPUs. After the convergence of the intra-cluster iterations of all eight 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 a converged result. Table 3 gives the information on orbits 34–43.

5. The equivalent potential of water molecules with His 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 the two H atoms are replaced by a positive point charge located in the middle of a line connecting the two H atoms. Adjusting the point charges and positions of the dipoles, the electronic structure of His is recalculated using the SCCE calculation. There is now no inter-cluster iteration; however, only the intra-cluster iteration of the His cluster needs to be performed because the dipoles have no electron.

The electronic structure obtained in section 4 is considered as approximately the electronic structure of His in the 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.* $\overline{\Delta E^\sigma} = \frac{1}{N^\sigma} [\sum_{n=1}^{N^\sigma} (\varepsilon_n^\sigma - \varepsilon_{n0}^\sigma)^2]^{1/2}$, where the ε_n^σ and ε_{n0}^σ are the eigenvalues of the n th electronic state with the spin σ calculated in this section and in section 4, respectively. N^σ is the number of electrons with spin σ .

Table 4. Point charges and coordinates of seven dipoles.

No. of dipole	Charge, e	X (Å)	Y (Å)	Z (Å)
(1)	+0.8000	4.3278	1.5344	3.0215
	-0.8000	4.4839	1.4890	3.5842
(2)	-0.5000	3.0947	-0.9618	-1.3135
	+0.5000	3.4122	-1.8817	-1.2736
(3)	-0.7000	2.6042	3.0051	-2.4025
	+0.7000	2.9328	3.4601	-2.5701
(4)	-0.5750	-0.2697	0.5799	-2.8479
	+0.5750	-0.7798	0.6993	-3.1098
(5)	-0.8500	2.0257	-4.0049	-4.1905
	+0.8500	2.1407	-4.2754	-4.6972
(6)	-0.6550	-3.0855	2.1250	2.6105
	+0.6550	-3.3873	2.5544	2.8704
(7)	-0.6550	-5.9735	-1.1642	-0.0498
	+0.6550	-6.4892	-1.4322	-0.1233

- (2) *The equivalent mean square deviation of the charge density.* $\overline{\Delta C^\sigma} = \frac{1}{N} [\sum_{i=1}^N \sum_{j=1}^N (\sum_{n=1}^{N^\sigma} C_{ni}^{\sigma*} C_{nj}^\sigma - \sum_{n=1}^{N^\sigma} C_{ni0}^{\sigma*} C_{nj0}^\sigma)^2]^{1/2}$, where the C_{ni}^σ and C_{ni0}^σ are the expansion coefficients of the eigenfunctions of the n th electronic states with the spin σ calculated in this section and in section 4, respectively. N is the number of Gaussian bases used to expand the one-electron wavefunction. (Please note the formula (2): $\rho^\sigma(\mathbf{r}) = \sum_{n=1}^{N^\sigma} |\Phi_n^\sigma(\mathbf{r})|^2 = \sum_{n=1}^{N^\sigma} [\sum_{i=1}^N C_{ni}^{\sigma*} U_i^*(\vec{r})] [\sum_{j=1}^N C_{nj}^\sigma U_j(\vec{r})] = \sum_{i=1}^N \sum_{j=1}^N [\sum_{n=1}^{N^\sigma} C_{ni}^{\sigma*} C_{nj}^\sigma] U_i^*(\vec{r}) U_j(\vec{r})$, where the Gaussian bases $U_i(\vec{r})$ are the same in the two calculations.)

The initial charge of the dipoles is set to $0.5e$. The distance L between the positive charge and negative charge is 0.5857 Å, which remains unchanging. After the convergence, the values of the two criteria are calculated. Then the point charges and the positions of the seven dipoles are adjusted one by one, until the two criteria reach the minimum. Thus, the potential of seven dipoles simulates the potential of water molecules for the electronic structure of His. The final charges and coordinates of dipoles are in table 4, and the structure is shown in figure 2, where the atoms and dipoles are numbered according to tables 1 and 4, respectively. Table 5 gives the eigenvalues and Mulliken populations of the ten orbits near the Fermi level of His in the potential of the dipoles.

6. Discussion

In order to study the water solvent effect on the electronic structure of His, the electronic structure of isolated His is calculated by free cluster calculation. The total energy of isolated His is -1091.7324 Ryd. The results are given in table 6.

The eigenvalues of orbits 34–43 of His with three potentials are given in table 7. The last row gives the energy gap: $E_g = E_{42} - E_{41}$. Figure 3 shows the comparison between the three sets of eigenvalues of His for the potential of dipoles, that of water molecules and no potential, respectively.

The properties of His are mainly determined by the electronic states near the Fermi level. In tables 3, 5 and 6, the Mulliken population analysis shows that in all three cases, the electronic

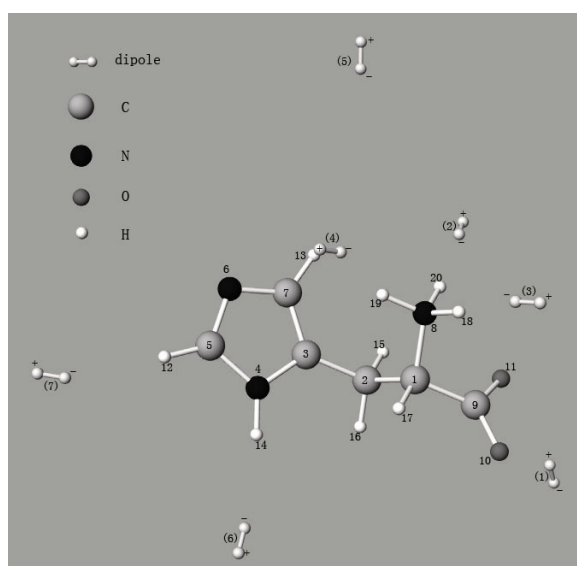


Figure 2. The final geometric structure of the His + 7dipoles system.

Table 5. Parts of the eigenvalues and Mulliken populations of His with the potential of seven dipoles.

State	Energy (Ryd)	Mulliken populations							
		C		N		O		H	
		s	p	s	p	s	p	s	p
43	-0.0585	0.4899	0.3923	0.5123	0.3231	0.0042	-0.0049	-0.7272	0.0102
42	-0.0763	0.2389	0.1865	1.6472	0.1652	0.0099	0.0058	-1.2629	0.0095
The above are unoccupied states									
41	-0.3423	-0.0505	0.1196	0.0015	0.0014	0.0040	0.9385	-0.0070	0.0004
40	-0.3566	0.0123	0.0035	0.0094	0.0041	0.0016	0.9762	-0.0041	0.0001
39	-0.4108	0.0512	0.3900	0.0268	0.0317	0.0111	0.4765	0.0103	0.0025
38	-0.4359	0.1368	0.4399	0.0397	0.0434	0.0047	0.3323	-0.0008	0.0039
37	-0.5117	-0.0644	0.1444	0.2504	0.6389	0.0014	0.0045	0.0224	0.0023
36	-0.5270	-0.0107	0.1957	0.0782	0.7418	0.0009	0.0064	-0.0151	0.0046
35	-0.6250	0.0219	0.4442	0.0172	0.0380	0.0183	0.4659	-0.0066	0.0012
34	-0.6395	0.0023	0.2575	0.0091	0.0153	0.1866	0.4735	0.0585	0.0019

states below the Fermi level are similar: (a) one hybridized state of an oxygen 2p electron (from COO^-) and a carbon 2p electron (from atom 1: C_α); (b) one state of an oxygen 2p electron from COO^- ; (c) one hybridized state of an oxygen 2p electron (from COO^-), and carbon 2p and 2s electrons (from atom 1: C_α); (d) one hybridized state of a carbon 2p electron (from the benzene ring) and an oxygen 2p electron (from COO^-); (e) two nitrogen 2p electrons from the benzene ring of the lateral chain (states 37 and 36); (f) one hybridized state of oxygen 2p and carbon 2p electrons from COO^- (state 35); (g) one hybridized state of oxygen 2p, 2s, and carbon 2p electrons from COO^- (state 34). But the sequences of states (a), (b), (c) and (d) are different in three cases (see table 7): (1) for isolated His, the sequence is 41, 40, 39 and 38; (2) for His with the potential of water molecules, it is 40, 41, 38 and 39; (3) for His with the

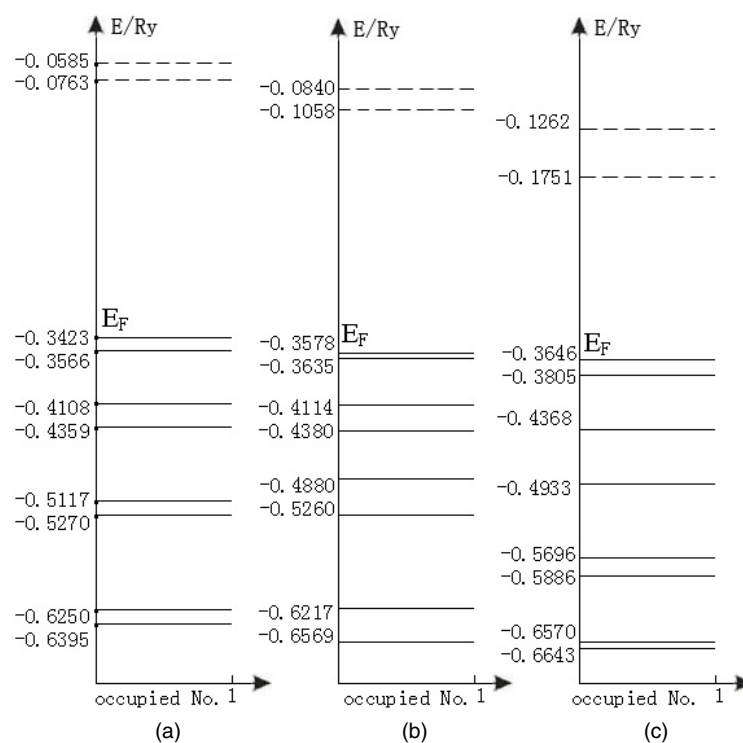


Figure 3. Comparison of three sets of eigenvalues: (a) dipoles; (b) water molecules; (c) none. (The full line represents an occupied state; the dashed line represents an unoccupied state.)

Table 6. Parts of eigenvalues and Mulliken populations for isolated His.

State	Energy (Ryd)	Mulliken populations							
		C		N		O		H	
		s	p	s	p	s	p	s	p
43	-0.1262	-0.0708	0.7151	-0.0575	0.3308	-0.0008	0.0038	0.0644	0.0149
42	-0.1751	-0.3798	0.0786	0.3735	0.0964	0.0139	0.0095	0.7939	0.0140
The above are unoccupied states									
41	-0.3646	-0.0561	0.1536	-0.0278	-0.0043	-0.0062	0.9217	0.0186	0.0006
40	-0.3805	0.0108	0.0081	0.0088	0.0037	-0.0004	0.9695	-0.0006	0.0001
39	-0.4368	0.1145	0.1615	-0.0291	-0.0026	-0.0008	0.7369	0.0192	0.0004
38	-0.4933	-0.0059	0.8229	0.0502	0.0841	0.0006	0.0476	-0.0055	0.0060
37	-0.5696	-0.0862	0.1402	0.2964	0.6124	0.0015	0.0053	0.0287	0.0017
36	-0.5886	0.0004	0.2252	0.0326	0.7521	-0.0005	0.0022	-0.0175	0.0055
35	-0.6570	0.0113	0.4492	0.0103	0.0276	0.0153	0.4808	0.0045	0.0009
34	-0.6643	-0.0032	0.2366	0.0044	0.0107	0.2251	0.4933	0.0319	0.0011

potential of dipoles, it is 41, 40, and the states 38 and 39 are now indistinguishable from the viewpoint of Mulliken populations.

By comparing the third and fourth columns of table 7, as well as figures 3(b) and (c), the main effects of water molecules are seen to be as follows: (1) the energy gap broadens by 33%; (2) two states (40 and 38) of the third column of table 7, which contain the contribution of the

Table 7. Three sets of eigenvalues of His.

Eigenvalues	Energy (Ryd)		
	Dipoles	Water molecules	No potential
43 (unoccupied)	-0.0585	-0.0840	-0.1262
42 (unoccupied)	-0.0763	-0.1058	-0.1751
41 (E_F)	-0.3423	-0.3578	-0.3646
40	-0.3566	-0.3635	-0.3805
39	-0.4108	-0.4114	-0.4368
38	-0.4359	-0.4380	-0.4933
37	-0.5117	-0.4880	-0.5696
36	-0.5270	-0.5260	-0.5886
35	-0.6250	-0.6217	-0.6570
34	-0.6395	-0.6569	-0.6643
$E_g = E_{42} - E_{41}$	0.2631	0.2520	0.1895

carbon 2p electrons of the atom C_α , remain approximately unchanged, while the other six states rise by about 0.049 Ryd which causes the exchanges of state positions between states 41 and 40, and between states 39 and 38.

The comparison of the second and third columns of table 7, as well as figures 3(a) and (b), reveals that below the Fermi level, the eigenvalues of His with the potential of dipoles are close to those of His with the potential of water molecules: five eigenvalues are almost unchanging, and three eigenvalues have deviations of about 0.02 Ryd. Above the Fermi level, the energy gap is broadened by 4.4%. This shows that the effect of water molecules on the electronic structure of His can be simulated by the dipole potential.

7. Conclusion

The geometric structure of the His + 7H₂O system with the lowest total energy is determined by free cluster calculation. On the basis of the geometric structure above, the electronic structure of His with the potential of water molecules is calculated using the self-consistent cluster-embedding calculation. Then the water molecules are replaced by adjustable dipoles. The dipoles are adjusted so that the electronic structure of His with the potential of dipoles is close to that with water molecules. The calculations show that the major effect of water molecules on the His electronic structure is as follows: (1) the energy gap broadens by 33%; (2) two states containing the contribution of carbon 2p electrons of the atom C_α remain approximately unchanging, while the other six states rise by about 0.049 Ryd which causes the exchanges of the state positions. The effect of water molecules on the electronic structure of His can be simulated by the potential of dipoles; the eigenvalues and Mulliken populations calculated using the two kinds of potentials are very close. So we established a simple, easy-to-use, with almost no additional computational effort, dipole potential which simulated the effect of water molecules on the His electronic structure.

More dipole potentials simulating the potentials of water molecules for other amino acids will be constructed. All results will be directly applied to electronic structure calculations for protein in aqueous solution.

Acknowledgments

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 Centre. The calculations were performed on the supercomputer DAWN 4000A of Shanghai Supercomputer Centre of China.

References

- [1] Zheng H P 1997 *Phys. Lett. A* **226** 223
Zheng H P 1997 *Phys. Lett. A* **231** 453
- [2] Zheng H P 1993 *Phys. Rev. B* **48** 14868
- [3] Zheng H P 1995 *Physica B* **212** 125
- [4] Zheng H and Rao B K 1997 *Phys. Rev. B* **55** 4174
- [5] Zheng H, Wang Y and Ma G 2002 *Eur. Phys. J. B* **29** 61
- [6] He J and Zheng H P 2002 *Acta Phys. Sin.* **51** 2580 (in Chinese)
- [7] Zheng H P 2000 *Phys. Rev. E* **62** 5500
- [8] Zheng H P 2000 *Prog. Phys.* **20** 291 (in Chinese)
- [9] Zheng H P 2002 *Mod. Phys. Lett. B* **16** 1151
- [10] Zheng H P 2003 *Phys. Rev. E* **68** 051908
- [11] Lazaridis T and Karplus M 1999 *J. Mol. Biol.* **288** 477
- [12] Lazaridis T and Karplus M 1999 *Proteins Struct. Funct. Genet.* **35** 133
- [13] Klamt A and Schüürmann G 1993 *Chem. Soc. Perkin Trans.* **2** 799
- [14] York D M and Karplus M 1999 *J. Phys. Chem. A* **103** 11060
- [15] Guo H and Karplus M 1994 *J. Phys. Chem.* **98** 7104
- [16] Schaefer M and Karplus M 1995 *J. Phys. Chem.* **100** 1578
- [17] Ren C L *et al* 2004 *J. Adhes.* **80** 831
- [18] Massova I and Kollman P A 2000 *Perspect. Drug Discovery Des.* **18** 113
- [19] Wang X M, Zheng H P and Li C J 2006 *Eur. Phys. J. B* **52** 255
- [20] Li C J, Zheng H P and Wang X M 2006 *Sci. China G* **36** 480 (in Chinese)
- [21] Chen H 1988 *Dissertation for the Doctoral Degree* Louisiana State University, America
- [22] Chen H, Callaway J and Misra P K 1988 *Phys. Rev. B* **38** 195
- [23] Chen H and Callaway J 1991 *Phys. Rev. B* **44** 2289
- [24] Zheng H 1993 *Dissertation for the Doctoral Degree* Louisiana State University, America
- [25] Hohenberg P and Kohn W 1964 *Phys. Rev. B* **136** 864
- [26] Kohn W and Sham L J 1965 *Phys. Rev. A* **140** 1133
- [27] Zheng H P and He J 2001 *J. Tongji Univ.* **29** 593 (in Chinese)
- [28] Xu W H and Zheng H P 2003 *J. Tongji Univ.* **31** 374 (in Chinese)
- [29] Lin S J and Zheng H P 2004 *J. Tongji Univ.* **32** 551 (in Chinese)
- [30] Hao J A and Zheng H P 2004 *Acta Phys. Sin.* **53** 1044 (in Chinese)
- [31] Zheng H P and Hao J A 2005 *Chin. Phys.* **14** 529