Wenge Han
Dr. sc. hum.

Ab initio and Hybrid Quantum Mechanical/Molecular Mechanical (QM/MM) Studies of Biological Systems

Born on 23, 05, 1967 in Nei Mongol, P.R. China.
Doctor degree study in Biophysics, from SS 1997 to WS 1999.
Graduate study in Solid State Physics, 09/1989 – 03/1992, Tianjin University, P.R. China.
Undergraduate study in Applied Physics, 08/1985 – 07/1989, Tianjin University, P.R. China.

Doktorvater: Prof. Dr. Sándor Suhai
Institut: Deutsches Krebsforschungszentrum (DKFZ)

The purpose of this thesis is to study the conformational properties of the biomolecules in aqueous solution, and the role of the electrostatic interactions within a biomolecule. We used different methods to investigate the systems with different sizes. The work can be divided into four parts:

1. Density functional studies on N-Methylacetamide- water complexes

For the understanding of the solvation effect on the structural properties and biological functions of proteins, it is important to know how proteins interact with an aqueous environment through hydrogen bonding. N-Methylacetamide (NMA), one of the simplest models for the main chain of proteins, has been extensively studied in this respect. So far, all theoretical investigations concerned with the structural, energetic and vibrational aspects of hydrogen bonding in NMA complexes either used the conventional *ab initio* molecular orbital theory or molecular mechanics procedures. *Density functional theory* (DFT) methods may provide, on the other hand, an interesting and powerful methodological alternative in this field. Many evidences show that the results of DFT calculations are fairly consistent with experimental data for a number of molecular properties and are well comparable to post-Hartree-Fock (HF) calculations such as the second order Møller-Plesset perturbation method (MP2) at a relatively lower cost. We, therefore, would like to see how well the DFT methods in studying the H-bonding properties. Twelve *trans*- and *cis*-NMA–water complexes have been studied here using DFT functionals and compared with some other *ab initio* calculations.

We found both the DFT functionals B3LYP and BLYP predicted the geometries and binding energies which are in agreement with the MP2 and experimental results. The H-bonding cooperative effects were found in two of the *trans*-NMA+2H_2O complexes, in which the two water molecules attach to the N–H and C=O groups of *trans*-NMA, respectively. As a result, a *trans*-NMA complex with one H-bond on either N–H or C=O group favors a second H-bond formed at the other group. This cooperative effect may enhance the hydrogen bonding between peptide bonds and make the formation of secondary structures more favorable. No substantial H-bonding cooperativity was found for *cis*-NMA-water complexes. This suggests that *cis*-NMA in an aqueous solution is still unstable and this is why the concentration of *cis*-NMA in water is only a few percent.

2. A study of aqueous N-Acetyl-L-alanine N´-Methylamide: structures and Raman, VCD, and ROA spectra

There are two main methods to predict the solvation effect on properties of the solutes. The first one is to explicitly include the solvent molecules. The second one is to use the reaction field continuum models where explicit solvent molecules are not considered. In the case of the H_2O solution, the explicit solvation model can reveal detailed information about the H-bonding between the solute and water molecules. For *ab initio* calculations, however, it is not possible to treat the bulk solvent molecules explicitly. On the other hand, the reaction field solvent continuum models may predict the effect of the bulk water but cannot give detailed information about H-bonding.

N-Acetyl-L-alanine N´-Methylamide (AAMA) is another primary model of the protein backbone. Two major degrees of freedom, ϕ and ψ, determine its conformation. Here we decided to compare different solvation models to study the effect of hydration on the AAMA geometries, relative energies, and vibrational properties, and through comparing the calculated vibrational spectra with the experimental data, to predict what kind(s) of AAMA conformation(s) will occur in aqueous solution. We applied the solvent continuum model, added four explicit water molecules, and finally combined the two approaches on eight AAMA conformers. For the four lowest energy (on the B3LYP/6-31G* potential energy surface) AAMA+4H_2O complexes, we calculated the B3LYP/6-31G* Hessians and atomic polar tensors, HF/6-31G* atomic axial tensors, HF/6-311+G** electric dipole-electric dipole polarizability derivatives, HF/6-31G* electric dipole-magnetic dipole polarizabilities, and electric dipole-electric quadrupole polarizabilities, which gave us the required quantities to simulate the vibrational absorption (VA), Raman, vibrational circular dichroism (VCD), and Raman optical activity (ROA) spectra.

We found that the explicit water molecules stabilize two AAMA structures (P_{II} and α_R), which are not stable in the isolated state. The influence of the explicit water molecules on the AAMA vibrational spectra was also studied. The solvent continuum model applied to the AAMA+4H_2O complexes further modified the orientations of the water molecules and influenced the vibrational modes and intensities. By comparing the calculated and the observed Raman, VCD, and ROA spectra, we suggest that the P_{II} structure of AAMA (ϕ~–93°, ψ~128°) is the dominant one in aqueous solution.

3. QM/MM study of the active site of free papain and of the NMA-papain complex

As is well known, the quantum mechanical (QM) methods are capable of providing the detailed information of chemical structure and reactivity. With the developing of the computer facilities and computing methods, larger and larger molecular systems can be studied by accurate *ab initio* methods, DFT or semiempirical methods. Nevertheless, there are many problems of chemical interest that can not yet be handled because of the computational expense. These include the studies of molecular properties and chemical reactivities in the presence of solvents, and the investigations of enzymatic reactions in proteins. It is not possible to use QM methods for explicitly describing the solvent or the bulk protein environment. Empirical molecular mechanical (MM) force fields, on the other hand, have been successfully employed in many molecular dynamics and Monte Carlo simulations for large systems. They are, however, unable to give an adequate description of charge reorganizations and the forming or breaking of chemical bonds.

Hybrid QM/MM models have been developed to solve this dilemma and to take the merits of both the quantum and classical approaches. The system to be investigated can be divided into two parts, the QM part and the MM part. The QM methods applied to the QM part are able to describe the bond forming and breaking, the electron transfer processes, and the transition state properties. The MM methods applied to the MM part, on the other hand, give the description of the solvent or the remaining part of the system. If one molecule is partitioned into two parts, a linking atom, a

localized bond orbital, or a pseudobond will be added to fill the open valence of the atom in the cutting point of the QM part.

Using the hybrid QM/MM method by combining the Gaussian 94 program package with the AMBER 4.1 force field, we studied the active site of free papain and the NMA–papain complex using the HF and B3LYP methods for the QM regions and the Amber 4.1 force field for the bulk enzyme environment.

We found that a covalent tetrahedral intermediate structure of NMA–papain complex could be obtained only when the amide N atom of the NMA was protonated through a proton transfer from His-159 in the catalytic site. Our results support the previous assumption that a proton transfer from His-159 to the amide N atom of the substrate occurs prior to or concerted with the nucleophilic attack of the Cys-25 sulfur atom to the carbonyl group of the substrate. We also found that the electrostatic field produced by the protein environment is very important for the properties of papain active site. The thiolate-imidazolium ion pair in the catalytic site, which is crucial for the catalytic activity of the enzyme, can be stabilized only in the presence of the environment produced by the protein. The groups that form this ion pair, the imidazole ring of His-159 and the sulfur atom of Cys-25, are coplanar in the free form of papain. However, binding of the substrate to the active site disturbs this coplanarity.

4. SCC-DFTB/MM studies of H-bonded systems and of N-Acetyl-(L-Ala)$_n$ N'-Methylamide helices in water solution

The *ab initio* methods, like HF, MP2 or DFT functional B3LYP, normally give accurate results. However, with the increasing size of the QM region, the application of these methods may be impractical.

Recently, Elstner *et al.* have presented the development of a self-consistent-charge density-functional tight binding scheme (SCC-DFTB), which has been shown to be able to give a reliable description of reaction energies, geometries and vibrational properties of small organic molecules and biomolecules. We therefore combined the SCC-DFTB scheme with the AMBER 4.1 force field, so that we can apply the hybrid SCC-DFTB/MM method to larger biological systems which *ab initio* QM/MM can not handle. As the application of this hybrid method, we studied the water solvation effect on the helical structures (α and 3_{10}) of L-alanine (L-Ala) homopolypeptides. Firstly, some small systems which contain hydrogen bonds with water molecules were tested. Then treating the water molecules with the AMBER force field (TIP3P model), and the L-Ala homopolypeptides with the SCC-DFTB method, we studied the stability of the L-Ala α and 3_{10} helices with different number of residues solvated with water molecules using layers of variable thickness.

We found that, in gas-phase, the α-helices of (L-Ala)$_n$ (n = 4, 5, 8, 11) are less stable than the corresponding 3_{10}-helices. In water solution, however, the α-helices are stabilized and, compared with 3_{10}-helices, the α-helices have stronger charge-charge interactions with the surrounding water molecules. This may be explained by the larger dipole moment of α-helices in aqueous solution, which will influence and organize the orientations of the surrounding water molecules. Comparing the SCC-DFTB energies between the α and 3_{10} helices in water solution, we found that the 3_{10}-helix is intrinsically more stable than the corresponding α-helix structure by 8-13 kcal/mol for (L-Ala)$_n$ (n = 4, 5, 8, 11). Whether the helix occurs in α or 3_{10} conformation then highly depends on the environment. The α-helical structure of (L-Ala)$_n$ favors a fully solvated environment where the solvation effect may overcome the intrinsic energy difference between the α and 3_{10} structures.