

Charles University in Prague
Faculty of Science
Department of Physical and Macromolecular Chemistry



Non-covalent interactions in fundamental biological
processes

Doctoral Thesis Abstract

Mgr. Vojtěch Klusák

Supervisor:
Mgr. Lubomír Rulíšek, CSc.

Institute of Organic Chemistry and Biochemistry AS CR
Center for Biomolecules and Complex Molecular Systems

Praha 2010

Universita Karlova v Praze
Přírodovědecká fakulta
Katedra fyzikální a makromolekulární chemie



Nekovalentní interakce v základních biologických
procesech

Souhrn disertační práce

Mgr. Vojtěch Klusák

Školitel:

Mgr. Lubomír Rulišek, CSc.

Ústav organické chemie a biochemie AV ČR
Centrum biomolekul a komplexních molekulárních systémů

Praha 2010

Introduction

Understanding inter- and intra-molecular interactions is the key for our insight into the properties of the biomolecular systems which, in turn, maintain and govern virtually all the processes in biology. Attempts to draw the structure-function relationship at the atomistic or electronic level bring us quite often beyond experimental resolution and capabilities. Modern tools of computational chemistry enable us to focus with satisfactory degree of reliability on the details of the studied process, gather additional (often complementary) information and ascribe particular structural features to measurable quantities.

Under normal conditions, in the processes related to biomolecular reorganization or assembly (folding or unfolding of protein or nucleic acid complexes or membrane formation), the covalent bonds do not break. Such processes are controlled by non-covalent interactions. Despite individual non-covalent interactions are relatively weak, they are numerous. Large number of small non-covalent forces governs equilibria between folding or unfolding of biomolecules, ligand binding or release, quaternary protein structures assembly or denaturation, and so on. Most of these equilibria are subtle and bio-systems are held in delicate balance between large number of countervailing forces and it is the small difference between these large numbers that determines direction of the process.

The presented thesis is focused on accurate description of the molecular properties, especially non-covalent interactions and interpretation of their role in various biological processes.

The first part of the thesis is focused on the stabilization of biomolecular (protein) structures, exemplified in a receptor-ligand complex and a network of interactions in the core of the protein. This includes the discussion of the general phenomena resulting from the studied and analyzed characteristics, as well as of some pitfalls related to their theoretical treatment.

In the second part of the thesis the key structures (reaction intermediate and transition states) of a prominent enzymatic reaction

(peptide hydrolysis) in a specific enzyme and the interactions important for their stabilization have been calculated and analyzed. It allowed us to suggest and discuss mutation experiments carried out in our collaborator's laboratory and relate the calculated data to the reactions of other hydrolytic enzymes. Finally, the theoretical aspects of the quantum chemical treatment are discussed as well.

Methods

In the studies collected in this thesis the methodology was focused mainly on (i) selection and preparation of an appropriate model and (ii) selection of an appropriate computational method.

An example of receptor-ligand interaction the complex of *bombyx mori* pheromone binding protein (PBP) with the native pheromone bombykol ((10E,12Z)-hexadeca-10,12-dien-1-ol) is presented. The stabilization of the protein core has been studied on the small protein rubredoxin. Theoretical aspects of the noncovalent interactions, especially the intramolecular ones, are demonstrated on the [*n*]helicenes, molecules consisting of *all-ortho* annulated benzene rings.

The reaction mechanism of peptide bond hydrolysis has been studied for the dizinc human glutamate carboxypeptidase II (GCPII), and the findings were complemented by calculations of plain formamide hydrolysis, water (as a general acid/base) assisted formamide hydrolysis, AlaAla dipeptide hydrolysis and hydrolysis of formamide in the active site of monozinc peptidase thermolysin.

Quantum chemical calculations were performed using several programs. Most of the calculations were performed using Turbomole program. In some cases Gaussian program has been used for the geometry optimization. Molpro has been used for the CCSD(T) calculations. The molecular mechanics calculations included in the QM/MM procedure were performed by the AMBER program package.

Localization of the enzyme catalysis in the protein active site makes it possible to construct the appropriate quantum chemical

model which is possible to use in quantum chemical calculations and which can reflect all chemically important aspects of the real process. For modeling the effects of the active site environment, the hybrid quantum and molecular mechanical (QM/MM) scheme implemented in the program ComQum has been used.

Results

Using simple model fragments as representatives of the amino acid residues, the interaction energies of their complexes with pheromone were calculated *ab initio* (at the level RI-MP2/aug-SVP). It has been shown that the pheromone molecule is not just expelled into the binding cavity from the outer environment (polar sensillar lymph) due to its hydrophobicity. On the contrary, the pheromone is attracted by several aromatic residues in the cavity (via X-H... π and π ... π interactions) that interact with practically the whole hydrocarbon unsaturated chain of the pheromone.

Next, we have evaluated the stabilization energies provided by the principal interactions of two phenylalanine side chains inside a small protein rubredoxin. We have shown that stabilization inside the hydrophobic core of a small protein, rubredoxin, determined by means of high-level correlated *ab initio* calculations (complete basis set limit estimate of MP2 stabilization energy + CCSD(T) correction term), is surprisingly strong. This attraction originates in London dispersion energy between aromatic rings or between an aromatic ring and an aliphatic chain, and is comparable to classical H-bonding. Moreover, residues of the aromatic nature can participate in several strong interactions at once, which may be crucial for the role of key residues in establishing networks inside a protein.

We have demonstrated that the intramolecular basis set superposition error (BSSE) can dramatically influence the calculated energies of the folded molecular systems and cause predictions of erroneous geometries as well. Namely, in the case of [*n*]helicenes, it has been shown that the intramolecular BSSE can be so large that the calculations predict clearly absurd results, such as the higher stability of [*n*]helicene compared with [*n*]phenacene for $n > 6$ (using MP2 and

medium to large basis sets, such as TZVP or aug-cc-pVDZ). The (RI-)DFT-D method, on the other hand, has been shown much less susceptible to BSSE than MP2 and CCSD(T) and for the energy optimized structures properly accounts for the dispersion energy.

Using the QM/MM approach we have identified one reaction intermediate and two corresponding transition states in the glutamate carboxypeptidase reaction cycle. One transition state connects the intermediate with the structure of Michaelis complex (enzyme – substrate complex), and the other transition state leads to the peptide bond cleavage. The agreement with the experimental value of reaction barrier was satisfactory. In simple terms, we proposed to describe the *N*-Ac-Asp-Glu hydrolysis by GCPII as the reaction between a deactivated hydroxide with an activated peptide bond assisted by glutamate as a general acid/base. The fascinating and unique role of di-zinc center with regard to the reaction mechanism described is in making the nucleophile, the OH⁻ anion, available, but at the same time enabling control of its nucleophilicity. The proceeding of the reaction is depending on proper orientation of the substrate and of the proton shuttle (Glu424) provided by the interactions with the enzyme active site.

The accuracy of several quantum chemical methods on the model systems representing important intermediates and transition states in several amide hydrolysis reactions was critically addressed. It was shown that the (RI-)MP2 and SCS-MP2 methods perform reasonably well for the closed shell systems including zinc ions. Non-hybrid DFT functionals (PBE, TPSS) underestimate the barriers whereas the B3LYP exhibits an overestimation of the activation barrier. For the transition state structures the empirical term in case of (RI-)DFT-D does not perform such a reliable and systematic correction of the difference between the results of DFT methods and CCSD(T) benchmark values.

Conclusions

We demonstrated the importance of dispersion energy for an accurate description of non-covalent interactions of aromatic residues.

It has been shown that interaction energies of such interactions are competitive and sometimes can prevail over the hydrogen bonding.

It has been shown that an extreme care must be taken when modeling larger molecular systems. One of the phenomena that is not present in the small molecular systems but can dramatically influence the description of larger molecules is the intramolecular basis set superposition error (BSSE).

The (RI-)DFT-D method has shown to provide very good results, thus being a practical solution to the intramolecular BSSE problem and at the same time covers the dispersion interaction for the energy optimized structures.

However, it was shown that for the description of transition states DFT-D method is not reliable, not even in the case of closed shell systems. On the other hand, the standard DFT methods, lacking proper dispersion description, performed reasonably well in that case.

In conclusion, it has been shown that an accurate modeling of biomolecules in biological processes is a complex problem and it is not always possible to *a priori* neglect the role of more distant or weakly bound components (residues). However, once all the physical aspects of the (bio)chemical processes are well described, computational modeling provides us with the invaluable information about the details of the studied biological phenomena.