ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Ab initio studies of targets for pharmaceutical intervention

Thesis submitted for the degree of
Doctor Philosophiæ

Candidate:
Marialore Sulpizi

Supervisors:
Prof. Paolo Carloni
Prof. Antonino Cattaneo

October 2001

SISSA - SCUOLA INTERNAZIONALE SUPERIORE STUDI AVANZATI
TRIESTE
Via Beirut 2-4

TRIESTE

SISSA ISAS
SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI
INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Ab initio studies of targets for pharmaceutical intervention

Thesis submitted for the degree of
Doctor Philosophiæ

Candidate:
Marialore Sulpizi

Supervisors:
Prof. Paolo Carloni
Prof. Antonino Cattaneo

October 2001

Table of contents

Introduction

Ab initio Modelling of Drug/Target interactions

The design of novel molecular systems, conducted with forethought and planning, plays a fundamental role for drug discovery. The fundamental assumption of the 'rational' drug design approach is that the event that produces the beneficial effects of drugs is the molecular recognition and binding of ligands to the active site of specific targets, such as enzymes, receptors and nucleic acids. The effect of binding can be promotion or inhibition of signal transduction, of enzymatic activity or molecular transport.

Computer-aided design (CAD) is one of the most important approaches to rational drug design(Kubinyi et al., 1997; Leach, 1996; Doucet and Weber, 1996). The approach complements the structure-based drug design, in which the 3D structure is known by X-ray crystallography or NMR. A variety of computational chemistry methodologies are used in CAD.

Ligands are screened from a library and docked in the target by using *scoring functions*, which allows for a crude and fast estimation for protein-drug interactions. These functions are based on the statistical analysis of known structure of protein-ligand complexes(Koppensteiner and Sippl, 1998), physicochemical properties(Bohm and Klebe, 1996), molecular dynamics force field calculations with implicit or explicit solvent, free energy perturbation calculations(Gilson et al., 1997). The SMoG(DeWitte and Shakhnovich, 1996) pseudo-energy function is an example of scoring function based on a statistical analysis of high-resolution crystal structures of protein-ligand complexes. The simplest physicochemical scoring functions are based on number of contacts, within a specified distance, between ligand and receptor (i.e. HOOK(Eisen et al., 1994)). More complex empirical functions contain terms which account for lipophilicity,

electrostatics, hydrogen bonding, and the area of the contact region (i.e. LUDI(Bohm, 1992)).

For a quantitative estimation of ligand affinity, more sophisticated methods are required. In classical molecular dynamics simulations (MD), the interatomic interactions are modelled via *empirical* energy functions. MD has a long and successful record in the simulation of biological systems(Wang et al., 2001). Still, in spite of the very many successes there are many areas where the use of effective potentials might not be fully appropriate. This is due to the fact that there is a large class of phenomena that depend on the electronic structure in such an intricate way that they are difficult to be modelled via effective potentials. Examples include the treatment of polarization effects(Alber, 1998), many-body effects (such as those which are responsible for the planarity of π systems(Hutter et al., 1996a)) and, most importantly, the simulation of bond-forming/bond-breaking processes. These may be more reliably modelled by *ab initio* quantum-chemical methods. In this respect first principle density functional theory (DFT) is emerging as a very powerful technique to study biomolecular systems, as it can describe relatively large systems accurately (in particular, it includes correlation effects) at the computational cost usually significantly lower than post-Hartree Fock methods. In particular, use of gradient correction for the exchange-correlation functional has opened the route to an accurate description of H-bonding(Klein, 2001; Tuckerman et al., 1997; Molteni and Parrinello, 1998; Silvestrelli and Parrinello, 1999a; Bruge et al., 1999), a crucial interaction in biochemical systems(Diaz et al., 2000; Sagnella et al., 1996) (Carloni and Rothlisberger, 2000).

Motivation of the present work

In this thesis we further explore the capability of first principle methods to provide insights on drug/target interactions in different contexts.

In the first part of this work, we address the issue whether DFT methods can be used as a potential tool for drug-screening. First principle calculations are particularly interesting for screening the energetics of drug/target interactions, as they do not involve the painstaking procedure of developing each set of new parameters for each novel drug. In this context, we use *ab initio* methods as a novel tool to determine a scoring function in

a series of prodrug / target (herpes simplex type 1 thyimidine kinase) complexes for gene-therapy based anticancer approaches. This work, accompanied by experimental data provided by Prof. Folkers' Lab (ETH, Zurich) provides a new, very simple, *ab initio*-based approach to the construction of scoring functions for drug-screening.

In the second part of the thesis we investigate the capability of DFT to describe non trivial interactions which are encountered in several inhibitor/enzyme complexes of pharmaceutical interest. Clearly, the description of these non-trivial phenomena might require the use of electronic structure methods. Here we present an example of cation-π interaction found in the *human immunodeficiency virus* reverse transcriptase (HIV-1 RT), one of the major targets for anti-AIDS therapy(Furman et al., 2000)). Furthermore, we provide a description of the hydroxyl-π interactions in the active site of μ-glutathione S-transferase(Xiao et al., 1996) (μ-GST), whose differential expression has been implicated in the development of cancers as well as their resistance to chemotherapeutic drugs ((McCallum et al., 2000) and reference therein).

Finally we present a classic problem treated by quantum-chemical methods: the simulation of an enzymatic reaction. We focus on a class of cysteine proteases, the caspases. These enzymes are extremely important targets for pharmaceutical intervention in therapies against Alzheimer's and other neurodegenerative processes, yet very few inhibitors have been so far designed. Since an important class of inhibitors is the given by the transition state analogs, it is of importance to fully understand the enzymatic reaction, that is the hydrolysis of peptides. Because of the crucial importance of temperature and environment(Karplus, 2000; Glennon and Warshel, 1998; Varnai and Warshel, 2000; Villa et al., 2000) effects for enzymatic catalysis, we use here a hybrid Car-Parrinello Molecular dynamics / Molecular mechanics approach recently developed in the Lab of Prof. U. Roethlisberger (Laio et al., 2001). This technique allows to evaluate the intermolecular interactions at the active site from electronic structure calculations as the simulation proceeds(Car and Parrinello, 1985). Steric and electrostatic effects of the protein scaffold on the quantum region are included using classical MD approach on the rest of the system. The free energy of the process is calculated using a thermodynamic integration approach(Ciccotti et al., 1989; Carloni et al., 2000; Piana et al., 2001).

This thesis is organized as follows. The first chapter provides a description of the used computational techniques. The following chapter describes the systems investigated here and summarizes our findings. The subsequent three chapters are devoted to a detailed description of my thesis work. In a final chapter we draw some conclusions and provide a perspective for possible future applications, which could follow this work.

Chapter 1

Methods

Computer simulations are a powerful tool to predict molecular properties that are inaccessible to experiments, once the reliability of the molecular models, force fields and computational procedures has been established.

Here I will describe the computational chemistry techniques used in this thesis. I will first outline the basic formalism of density functional theory and Car-Parrinello molecular dynamics. Then I will give a brief introduction to classical MD simulations, with particular reference to GROMOS force field. Finally I will briefly describe the QM/MM approach used in the caspase reaction simulation.

1.1 Density Functional Theory

1.1.1 The Hohenberg-Khon theorems and the Kohn-Sham method

Density functional Theory is an exact theory of the ground state of a many particle system. The mathematical formulation was given by Hohenberg and Kohn(Hohenberg and Kohn, 1964) (HK) in the mid-1960s and it allows to replace the Scrödinger equation for the *N*-electron wave function with a computational scheme involving the electron density $\rho(\mathbf{r})$.

Hohenberg and Kohn (HK) theorems state the uniqueness of the ground-state density ρ associated with a given external potential v_{ext} (e.g. the one exerted by the nuclei on the electrons) and the variational character of the density energy functional.

The total energy $E = <\psi | H | \psi>$, in term of the density $\rho(\mathbf{r})$ reads:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ext}[\rho] = T[\rho] + V_{ee}[\rho] + \int d\mathbf{r} v_{ext}(\mathbf{r})\rho(\mathbf{r}),$$

where $T[\rho]$ is the kinetic energy, $V_{ee}[\rho]$ is a term representing the electron-electron potential energy. For a given trial density $\tilde{\rho}$, such as $\tilde{\rho} > 0$ and $\int \tilde{\rho}(\mathbf{r}) d\mathbf{r} = N$ the variational principle reads:

$$E_0 \leq E_v[\tilde{\rho}].$$

As consequence, the existence of a universal density functional, $F_{HK}[\rho] = T[\rho] + V_{ee}[\rho]$, independent from the external potential v_{ext} was established. However, the exact form of F_{HK} is not known.

Based on the variational character of the functional, the ground state density and energy can be calculated, in principle, by minimization procedures.

$$\frac{\delta}{\delta n}\left[E[\rho] - \mu\left(\int d\mathbf{r}\rho(\mathbf{r}) - N\right)\right] = 0 \qquad (1)$$

A practical scheme for the solution of equation (1) was given by Kohn and Sham(Kohn and Sham, 1965), in which one-electron orbitals were introduced to describe the electron density. They defined a non-interacting reference system of Hamiltonian H_s for which the ground state density is exactly ρ:

$$H_s = \sum_{i=1}^{N}\left(-\tfrac{1}{2}\nabla_i^2\right) + \sum_{i=1}^{N} v_s(\mathbf{r}_i)$$

For that system both the density and the kinetic energy are well defined in term of single particle wavefunction ψ_i:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \qquad (2)$$

$$T_s[\rho] = \sum_{i=1}^{N} < \psi_i \,| -\tfrac{1}{2}\nabla^2 \,| \psi_i >$$

The idea was that in such way the kinetic energy can be computed simply to good accuracy, leaving a small residual correction that is handled separately.

The HK functional can be written in terms of the non-interacting kinetic energy as follow:

$$F_{HK}[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho]$$

where

$$E_{xc}[\rho] = T_s[\rho] - T[\rho] + V_{ee} - J[\rho]$$

is the exchange and correlation energy containing the difference between T and T_s and the non-classical part of V_{ee}.

The Euler-Lagrange equations become:

$$\begin{cases} \mu = \dfrac{\delta E_{HK}[\rho]}{\delta\rho(\mathbf{r})} = v_{eff}(\mathbf{r}) + \dfrac{\delta T_s[\rho]}{\delta\rho(\mathbf{r})} \\ v_{eff}(\mathbf{r}) = v_{ext}(\mathbf{r}) + \int d\mathbf{r}' \dfrac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} + v_{xc}(\mathbf{r}) \end{cases} \qquad (3)$$

where $v_{eff} = \dfrac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})}$ is the exchange-correlation potential. By substituting ρ (eq. 2) in eq. 3 one obtains the N one-electron equations:

$$\left[-\frac{1}{2}\nabla^2 + v_{eff}(\mathbf{r})\right]\psi_i = \varepsilon_i \psi_i \qquad \text{i=1,N}$$

which can be solved self consistently to obtain ρ ed E.

1.1.2 Exchange-correlation energy: From LDA to GGA

To solve KS equation an explicit form for the $E_{xc}[\rho]$ is required. The search for an accurate form is still an open challenge in the DFT community. The simplest approximation for $E_{xc}[\rho]$ was proposed by Khon and Sham(Kohn and Sham, 1965) in 1965 and assume that the exchange-correlation energy per electron at a point **r** is equal

to the exchange-correlation energy per electron in a homogeneous electron gas that has the same electron density at the point **r**. It follows that

$$E_{xc}^{LDA}[\rho] = \int d\mathbf{r}\rho(\mathbf{r})\varepsilon_{xc}^{hom}[\rho(\mathbf{r})]$$

where ε_{xc}^{hom} is the exchange and correlation energy per particle of a uniform electron gas of density $\rho(\mathbf{r})$. The corresponding exchange and correlation potential becomes:

$$v_{xc}^{LDA}(\mathbf{r}) = \frac{\delta E_{xc}^{LDA}}{\delta\rho(\mathbf{r})} = \varepsilon_{xc}[\rho(\mathbf{r})] + \rho(\mathbf{r})\frac{\partial\varepsilon(\rho)}{\partial\rho}.$$

In LDA, corrections to the exchange-correlation energy due to the inhomogeneities in the electronic charge density around **r** are ignored. Considering this inexact nature of the approximation, it may at first seem somewhat surprising that such calculations are so successful. This can be partially attributed to the fact that LDA gives the correct sum rule to the exchange-correlation hole. That is, there is a total electronic charge of one electron excluded from the neighborhood of the electron at position **r**.

LDA has proved useful and reasonably accurate in many systems, such as metals and semiconductors. However it fails, for example, in hydrogen-bonded systems, where spatial variations in the electron density are too rapidly varying to be described adequately by LDA. Thus, the investigation of biomolecules requires approximations other than LDA, where E_{xc} is approximated by an expansion in higher order terms of the density gradient (general gradient approximations (GGA)). If we consider only the first order gradients, the exchange-correlation energy can be written:

$$E_{xc}[\rho] = \int \mathbf{d}^3\mathbf{r}\rho(\mathbf{r})\mathrm{f}(\rho,\nabla\rho).$$

In our calculations we used the gradient corrections approximations (GCA) of Becke(Becke, 1988):

$$E_x^B = E_x^{LDA} - \beta\int \rho^{4/3}\frac{x^2}{(1+6\beta\sinh^{-1}x)}d\mathbf{r}, \qquad \text{with}: x = \frac{|\nabla\rho|}{\rho^{4/3}}$$

The β parameter was obtained by fitting E_x values obtained by atomic Hartree-Fock calculations for noble gas atoms(Becke, 1988).

For the correlation energy we have used the parameterization introduced by Lee, Yang and Parr (LYP)(Lee et al., 1988; Colle and Salvetti, 1975)

$$E_c = -a\int \frac{1}{1+d\rho^{-1/3}}\left\{\rho + b\rho^{-2/3}\left[C_F\rho^{5/3} - 2t_W + (\tfrac{1}{9}t_W + \tfrac{1}{18}\nabla^2\rho)\right]e^{-c\rho^{-1/3})}\right\}d\mathbf{r}$$

where $C_F = \frac{3}{10}(3\pi^2)^{2/3}$, $t_W(\mathbf{r}) = \frac{1}{8}\frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} - \frac{1}{8}\nabla^2\rho$ and a = 0.049, b = 0.132, c = 0.2533, d = 0.349.

The BLYP approximation has been successfully used in a variety of H-bonded systems ((Bruge et al., 1999) and references therein).

1.1.3 Basis set for the electronic orbitals

When considering a periodic system a natural approach to solve the Kohn-Sham equations, is to expand the KS in terms of plane wave (PW) basis set.

Each Bloch function $\psi_i^{\mathbf{k}}$ of band i and momentum **k** results an infinite sum of the form:

$$\psi_i^{\mathbf{k}}(\mathbf{r}) = \sum_G c_i^{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}},$$

where **k** belongs to the first Brillouin zone of the crystal, and **G** is a reciprocal lattice vector.

This approach offers the advantages that results may be systematically converged with respect to basis set by variation of a single parameter, namely the cut-off energy E_{cut} for the kinetic energy T^{PW}.

$$T^{PW} = \frac{1}{2}|\mathbf{k}+\mathbf{G}|^2 \leq E_{cut}$$

Because the non-local nature of the plane waves, usually a large amount of basis function is necessary. On the other hand the use of plane waves permits the use of Fast Fourier Transform (FFT) techniques which are computationally efficient. Moreover PW are free of basis set superposition errors and allow computation of the total energy of different atomic arrangements with the same accuracy.

Use of the plane waves is not indicated to reproduce the rapid oscillations of the wave function in closeness of the atomic nuclei. In particular the core electrons, which are highly localized around the nuclei, need a very large set of plane waves for an accurate description of chemical bonds. However, the core levels are well separated in energy from the valence electrons, and they do not play a role in the chemical bonding properties of the system.

Thus, the core electron orbitals can be omitted from the KS equations and their effect on the valence electrons is described by using effective potentials (the pseudopotentials).

Pseudopotentials are usually derived from all electron (AE) atomic calculations, comparing the AE wave function with the pseudo wave function obtained by omitting the core electrons and including the pseudopotential term. Norm-conserving pseudopotentials needs to satisfy:

(i) The valence AE and pseudopotential eigenvalues must be the same. $\varepsilon_l^{PP} = \varepsilon_l^{AE}$.

(ii) The valence pseudo-wavefunction should contain no nodes. This requirement avoids the undesired spatial oscillation associate to the presence of nodes.

(iii) The atomic radial pseudo $R_l^{PP}(r)$ and AE radial $R_l^{AE}(r)$ atomic wave functions with angular momentum *l* must be identical beyond a chosen cutoff r_{cl}:

$$R_l^{PP}(r) = R_l^{AE}(r) \text{ for } r > r_{cl}$$

(iv) The charge enclosed within r_{cl} for the two wave functions must be the same:

$$\int_0^{r_{cl}} |R_l^{PP}(r)|^2 r^2 dr = \int_0^{r_{cl}} |R_l^{AE}(r)|^2 r^2 dr .$$

The last requirement guarantees the transferability of the pseudopotential, by means the independence from the chemical environment. Several pseudopotential recipes have been proposed(Kerker, 1980; Bachelet et al., 1982; Vanderbilt, 1985). In our calculation we follows the implementation of Troullier and Martins(Troullier and Martins, 1991). These pseudopotentials are accurate also for atoms for which the use of previous pseudopotential was restricted, like first row atoms, transition metals and rare earths.

In simulating molecular cluster a common procedure is to choose a large cell (a supercell) around the simulated molecules, in order to reduce the interactions between the system and its images. In our calculation we used a procedure that allows treating the system as isolated. In particular for the calculations of chapter 3 and 4 the procedure of Barnett and Landman(Barnett and Landman, 1993; Hockney, 1970), while for the calculation in chapter 5 the approach of Martyna(Martyna and Tuckerman, 1999) was used.

To reinforce the non- periodic boundary conditions the one electron wavefunctions are expanded in plane waves within the cell

$\psi_i = \frac{1}{\sqrt{V}} \sum_{\mathbf{g}} C_i(\mathbf{g}) e^{i\mathbf{gr}}$ for **r** inside the cell

$\psi_i(\mathbf{r}) = 0$ for **r** outside the cell

with the boundary condition that both $\psi_i(\mathbf{r})$ and $\nabla\psi_i(\mathbf{r})$ are negligibly small at the cell borders.

1.2 Car-Parrinello Molecular Dynamics

1.2.1 The Car-Parrinello Lagrangian and the equation of motion

In classical mechanics the Lagrangian can be written in terms of the dynamical variables which normally are atomic positions and velocities. The Car-Parrinello Lagrangian can similarly be written, but also includes a term for the electronic wavefunction:

$$\mathrm{L} = \sum_i \tfrac{1}{2}\mu \int |\dot{\psi}_i(\mathbf{r})|^2 d\mathbf{r} + \sum_{Ion} \tfrac{1}{2} M\dot{\mathbf{R}} - E[\{\psi_i\},\{\mathbf{R}\}] + \sum_{ij} \lambda_{ij} \left(\langle \psi_i(\mathbf{r}) | \psi_j(\mathbf{r}) \rangle - \delta_{ij} \right)$$

where μ is a fictitious mass which is associated this the expansion coefficients of the Kohn-Sham electronic wavefunctions and E is the Kohn-Sham energy functional. This is analogous to the usual form of the Lagrangian where the kinetic energy term is replaced with the fictitious dynamics of the wavefunctions and the Kohn-Sham energy functional replaces the potential energy.

λ_{ij} are the Lagrange multipliers associated with the constraint necessary to ensure orthonormality of the one electron wavefunctions during the dynamics

$\int \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}) d\mathbf{r} = \delta_{ij}$.

In terms of molecular dynamics, these can be thought of as additional forces on the wavefunctions, which maintain orthonormality throughout the calculation.

From L, it follows that the Lagrange equations of motion are:

$$\begin{cases} \dfrac{\delta}{\delta t}\dfrac{\delta L}{\delta \dot{\psi}_i^*} = \dfrac{\delta L}{\delta \psi_i^*} \\ \dfrac{\delta}{\delta t}\dfrac{\delta L}{\delta \dot{\mathbf{R}}} = \dfrac{\delta L}{\delta \mathbf{R}} \end{cases}$$

$$\begin{cases} \mu \ddot{\psi}_i(\mathbf{r}) = -H_{KS}\psi_i(\mathbf{r}) + \sum_j \lambda_{ij}\psi_j(\mathbf{r}) \\ M_{Ion}\ddot{\mathbf{R}} = -\dfrac{dE[\{\psi_i\},\{\mathbf{r}_i\}]}{d\mathbf{R}} \end{cases} \qquad (4)$$

In general the nuclear trajectories generated by equation (4) and those obtained from the classical lagrangian of the physical system do not coincide, unless $E[\{\psi_i\},\{\mathbf{r}_i\}]$ is at the instantaneous minimum. However, the parameter μ and the initial conditions $\{\psi_i\}_0, \{\mathbf{r}_i\}_0$ can be chosen in such a way that the time scale for the electronic degrees of freedom is much shorter than that of the nuclei. If in this case, and in the presence of particular conditions on the electronic spectrum of the H_{KS} such as a gap at the Fermi level, the nuclear trajectory, initially lying on the Born-Oppenheimer surface will deviate from it only after a very long time.

A schematic view of such an ionic trajectory is given in Fig. 1.1.

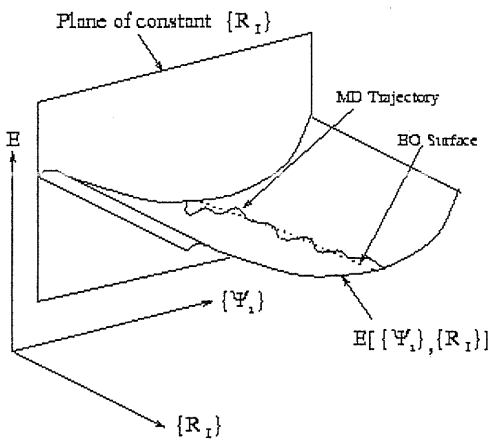

Figure 1.1 *A schematic view of ionic trajectory, according to CP dynamics, on the BO.*

In other words by appropriately choosing μ and $\{\psi_i\}_0, \{\mathbf{r}_i\}_0$ it is possible to put the coupled electronic and ionic systems into a metastable state, in which the system behave without significant energy transfer from the electronic to the ionic degrees of freedom, with the electrons following adiabatically the nuclei remaining close to the BO surface.

In such a metastable state meaningful temporal average on ionic degrees of freedom can be computed. It can be shown that by approximating H_{KS} with a linear operator the dynamics generated by (4) will consist of a superposition of harmonic modes with frequency equal to $\omega_{ij} = \sqrt{\frac{2(\varepsilon_i^* - \varepsilon_j)}{\mu}}$ where the lowest frequency to be compared to the ionic one is $\sqrt{\frac{2E_g}{\mu}}$ where E_g is the electronic energy gap. For the systems investigated in this thesis, E_g is of the order of some eV, resulting in well-separated energy scales for ions and electrons.

To integrate the equations of motion over discrete time intervals δt the Verlet(Swope et al., 1982) algorithm was used. With this algorithm the position at time $t+\delta t$ is calculated as:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\dot{\mathbf{r}}(t) + \tfrac{1}{2}\delta t^2\ddot{\mathbf{r}}(t)$$

and the velocities are calculated as:

$$\dot{\mathbf{r}}(t+\delta t) = \dot{\mathbf{r}}(t) + \tfrac{1}{2}\delta t[\ddot{\mathbf{r}}(t) + \ddot{\mathbf{r}}(t+\delta t)].$$

1.2.2 Constant temperature simulations: The Nosé thermostat.

Constant temperature MD can be obtained by coupling the ionic degrees of freedom to a Nosé thermal bath(Nose, 1984) which provides an additional term in the Lagrangian:

$$L' = L + K_{Nose} - V_{Nose}.$$

$K_{Nose} = \frac{1}{2}Q\dot{s}^2$ is the kinetic energy, where Q is a proper parameter and *s* represents the additional degree of freedom coming from the thermal bath. *s* is associated with an extra potential energy term that reads

$$V_{Nose} = (F+1)k_B T \ln s.$$

F is the total number of the degrees of freedom of the system. Equations of motion for the systems become

$$Q\ddot{s} = \sum_i m\dot{\mathbf{r}}_i^2 s - (F+1)k_B\frac{T}{s}$$

Q can be arbitrarily chosen and its value determines the strength of the coupling; high values of Q resulting in low coupling and vice versa.

1.3 Classical Molecular Dynamics Simulations

In classical molecular dynamics simulation methods electronic degrees of freedom are integrated out. The equations for the nuclei motion in (3) are obtained replacing the electronic potential with an empirical approximation, called force field.
Here I will provide a short description of molecular dynamics methods used in GROMOS96, which is the package simulation I mainly used in this thesis calculations.

1.3.1 Classical force field: GROMOS96

The GROMOS force field has the following form:

$$V = V_{bon} + V_{nonb}$$

with a bonded term and a non-bonded one.

$$V_{bon} = V_{bonds} + V_{angles} + V_{har} + V_{tig} =$$
$$= \sum_{n=1}^{N_b} \tfrac{1}{4} K_{b_n} \left[b_n^2 - b_{0_n}^2\right]^2 + \sum_{n=1}^{N_\theta} \tfrac{1}{2} K_{\theta_n} \left[\cos\theta_n - \cos\theta_{0_n}\right]^2 +$$
$$+ \sum_{n=1}^{N_\xi} \tfrac{1}{2} K_{\xi_n} \left[\xi_n - \xi_{0_n}\right]^2 + \sum_{n=1}^{N_\varphi} \tfrac{1}{2} K_{\varphi_n} \left[1 + \cos(\delta_n)\cos(m_n \varphi_n)\right]^2$$

The bonded term includes four different contributions representing bond stretching (V_{bonds}), bond-angle bending (V_{angles}), (harmonic) improper dihedral (V_{har}) and (trigonometric) dihedral-angle torsion (V_{tig}).

$$V_{nonb} = V_{LJ} + V_{CRF} =$$
$$\sum_{\substack{nonbonded \\ pairs(i,j)}} \left[\frac{C_{12}(i,j)}{\left(r_{ij}^{4D}\right)^6} - C_6(i,j) \right] \frac{1}{\left(r_{ij}^{4D}\right)^6} + \sum_{\substack{nonbonded \\ pairs(i,j)}} \frac{q_i q_j}{4\pi\varepsilon_0} \left[\frac{1}{r_{ij}^{4D}} - \frac{\tfrac{1}{2} C_{rf} \left(r_{ij}^{3D}\right)^2}{R_{rf}^3} - \frac{1 - \tfrac{1}{2} C_{rf}}{R_{rf}} \right]$$

The non-bonded term include a first contribution describing the Wan der Waals interactions and a second one describing the electrostatic interactions. For both of them the sum is extended to all couples ij with a serie of excluded pair including neighbor atom pairs.
The electrostatic term includes the Poisson-Boltzmann reaction field correction, which is included to approximate the interactions over the long-range cut-off. R_{rf} is the reaction

field cut-off, distance for which the correction is zero, while C_{rf} is the coefficient governing the size of the reaction field term.

The GROMOS force field uses of the *concept of charge group*. It means that the atoms that belong to a charge group are chosen such that their partial atomic charges add up to 0. Defining charge groups makes the electrostatic interaction, between two groups, a dipolar interaction which goes to 0 as $1/r^3$ and permit to choose a smaller cut-off radius for the electrostatics. Moreover GROMOS electrostatic calculations neglect the high frequency components of the long-range part of Coulomb interaction with a considerable reduction of computer time.

1.3.2 MD algorithm: Leap-frog

For the integration of the equations of motion, GROMOS uses the Leap-frog algorithm. In this algorithm, the velocities are first calculated at time *t+1/2dt*; these are used to calculate the positions, *r*, at time *t+dt*.

$$\begin{cases} \mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t+\frac{\delta t}{2}) \\ \mathbf{v}(t+\frac{\delta t}{2}) = \mathbf{v}(t-\frac{\delta t}{2}) + \mathbf{a}(t)\delta \end{cases}$$

In this way, the velocities *leap* over the positions, then the positions *leap* over the velocities. The advantage of this algorithm respect to Verlet one is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time *t* can be approximated by the relationship:

$$\mathbf{v}(t) = \frac{1}{2}\left[\mathbf{v}(t-\frac{\delta t}{2}) + \mathbf{v}(t+\frac{\delta t}{2})\right].$$

1.3.3 Constraints using the SHAKE method

The use of holonomic constraints for chemical bonds provides several advantages:

i) for high frequencies for which $h\nu >> K_BT$ a quantum behavior is expected and a constraint treatment can be better than a classical description in term of harmonic oscillators;

ii) application of constraints reduces the range of frequencies of a system alleviating the problem of slow energy relaxation between high and low frequency modes;

iii) application of constraints will generally save computing effort. In fact the length of the time step Δt in a MD simulation is limited by the highest frequency ν_{max} occurring in the system of interest ($\Delta t << \nu_{max}$). In particular hydrogen, are involved in the fastest bond stretching vibrations, which are often of little interest in classical simulations of biological systems. By constraining these degrees of freedom the time step can be therefore increased without prejudicing the accuracy of the simulation;

The SHAKE(Ryckaert et al., 1977) method can be used to impose distance constraints. For two given atoms k_1 and k_2 equations of motion have to be integrated using the distance constraint:

$$\sigma_k(\mathbf{r}) = r^2_{k_1k_{21}} - d^2_{k_1k_2} = 0$$

where the constraint distance is given by $d^2_{k_1k_2}$. Applying Lagrange's method of undetermined multipliers, equations of motion become

$$m_i\ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i}E - \nabla_{\mathbf{r}_i}\sum_{k=1}^{N_c}\lambda_k(t)\sigma_k(\mathbf{r}) \qquad (5)$$

The equation (5) is solved in GROMOS within the Leap-frog algorithm with an iterative procedure up to fixed degree of convergence.

1.3.4 Temperature and pressure scaling

Equilibration of the system at constant temperature and pressure before the QM/MM calculations can be achieved by constant pressure and temperature algorithms. The one implemented in GROMOS is the weak coupling method described in the paper by Berendsen *et al.*(Berendsen et al., 1984). The atomic equations of motion are modified such that the net result on the system is a first-order relaxation of the temperature T toward the preset reference value T_0.

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}(T_0 - T(t)) \qquad (6)$$

Since the temperature at time t is defined in terms of the ionic velocities according to

$$E_{kin}(t) = \sum_{i=1}^{N} \frac{1}{2} m_i v_i(t) = \frac{1}{2} N_{df} K_B T(t),$$

a change in the temperature of the system can be obtained by scaling the atomic velocities with a factor λ(t), which in term of the time step Δt, is

$$\lambda(t) = \left[1 + \frac{2c_v^{df}}{K_B}\frac{\Delta t}{\tau_T}\left[\frac{T_0}{T(t)} - 1\right]\right]^{1/2}.$$

The relaxation rate is controlled by the ration of heat capacity per degree of freedom c_v^{df} and the chosen *relaxation time* τ_T.

τ_T is an adjustable parameter and its value is chosen sufficiently small (strong coupling) to achieve the required temperature, but sufficiently large in order not to disturb the properties of the system by coupling to the temperature bath(Berendsen et al., 1984).

For pressure scaling methods based on the same principles as the constant temperature schemes are used. In this case role of temperature is played by the pressure and the role of the atomic velocities is played by atomic positions.

The weak coupling method leads to an equation for pressure analogous to (6)

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_0 - P(t)).$$

The hydrostatic pressure can be defined using the virial theorem and is:

$$P(t) = \frac{2}{3}\frac{[E_{kin} - W(t)]}{V(t)}$$

where $W(t)$ is the virial and $V(t)$ is the volume of the computational box.

A change in volume of the box and so in the pressure on the box can be obtained by scaling the atomic coordinates by a factor

$$\mu(t) = \left[1 - \kappa_T \frac{\Delta t}{\tau_P}[P_0 - P(t)]\right]^{1/3}.$$

In this case the relaxation rate is controlled by the ratio of isothermal compressibility κ_T and by the chosen *pressure relaxation time* τ_P. As for τ_T a proper choice should balance between efficient of the coupling and perturbation on the system.

Since the definition of the pressure depends on the kinetic energy, the pressure coupling should not be stronger then the temperature coupling ($\tau_P >> \tau_T$).

1.4 Mixed QM/MM simulations

Hybrid QM/MM approaches combines quantum chemical methods with classical force field environments(Warshel and Levitt, 1976; Singh and Kollman, 1986; Field et al., 1990; Sherwood, 2000).

The Hamiltonian for the hybrid system can be written as:

$$H = H_{QM} + H_{MM} + H_{QM/MM}$$

where H_{QM} is the Hamiltonian for the quantum system, H_{MM} is the Hamiltonian for the classical system and $H_{QM/MM}$ contains the interactions between the two parts.

Mixed QM/MM approaches have been developed applying both a semi-empirical(Warshel and Levitt, 1976; Singh and Kollman, 1986; Field et al., 1990; Bakowies and Thiel, 1996; Liu et al., 1996; Cummins and Gready, 1997; Luque et al., 2000) and *ab initio* (Deng et al., 1993; Wei and Salahub, 1994; Tunon et al., 1995; Stanton et al., 1996; Rothlisberger, 1998; Eichinger et al., 1999; Lyne et al., 1999; Pioda et al., 2000) level of theory.

For this thesis calculations the quantum part is treated at DFT level with BLYP gradient corrections, while the classical part is described by GROMOS96 force field.

The central question concerning hybrid methods is the definition of $H_{QM/MM}$, i.e. the way how the quantum and the classical subsystems are interfaced. In the calculations of this thesis a combination of Car-Parrinello molecular dynamics simulations and classical molecular simulation is applied. In particular the approach of Laio *et al.* for the electrostatic interactions (Laio et al., 2001) is briefly summarized below.

1.4.1 Bonded interactions

Saturating the valence of QM atoms. Bonded interactions involve QM atoms that are jointed to MM atoms leaving the QM atoms with unsaturated valence orbitals. The

solution we adopt is to cap the QM atoms with monovalent atoms described by suitable pseudopotantials. In particular U. Rothlisberger developed a pseudopotential for "monovalent" carbon, which accurately reproduces C-C bonds properties.
A classical van der Waals interaction is left for the boundary atoms in the quantum system (respecting the classical exclusion rules), since this boundary atom is typically a united atom in the GROMOS96 force filed, and has an effective size that cannot be accounted for otherwise within a simple pseudopotential scheme.

Keeping the local geometry. The remaining bonded interactions; in particular angular and dihedral terms involving MM and QM atoms are included in the classical force field. These terms permit to keep the geometry at the interface, however since classical terms have not been parametrized to describe chemical reactions, they might have limited accuracy in cases where the boundary is strongly distorted. Rearrangements of the QM system should not be hindered by these boundary terms that might, as in the case of dihedral interactions extend quite far into the quantum region.

The exclusion problem. A last remark involves the problem of the exclusion. Van der Waals exclusions are easy to handle, since this interaction is treated at a classical level. For the electrostatic, the effects of the nearby charges is significantly more difficult to exclude, since the classical exclusion, that are two body terms, cannot be univocally defined in the QM case where the interaction of the classical point charge is with the full QM density.
In the approach we use, the exclusions are introduced by an additional term

$$E_{excl} = -\sum_{excl} \frac{q_{RESP} q_{MM}}{r} . \qquad (7)$$

The charge q_{RESP} is determined by a variational RESP fit at every step(Laio et al., 2001), and since it depends explicitly on the density, the forces due to this term also act on the density in form of a compensating field. In the total energy, boundary bonds, angles, dihedrals and Van der Walls terms are taken into account by specifying them in the classical Hamiltonian, while the electrostatic exclusions are given by (7).

1.4.2 Non-bonded interactions

The non-bonded interactions include electrostatic and steric contributions between non-linked atoms. The Hamiltonian for that contribution can be written as:

$$H_{non\text{-}bonded} = \sum_{i \in MM} q_i \int dr \frac{\rho(r)}{|r - r_i|} + \sum_{\substack{i \in MM \\ j \in QM}} v_{vdw}(r_{ij}) \qquad (8)$$

where r_i is the position of the MM atom i, with charge q_i, $\rho(r)$ is the total (electronic plus ionic) charge of the quantum system and $v_{vdw}(r_{ij})$ is the van der Waals interaction between atom i and atom j.

Short range electrostatic interactions. Calculation of the electrostatic interactions poses two main problems the first is the electron spill-out problem, and the second one the very high computational cost for calculation of the electrostatic term in (8).
Spill-out problem comes from the possibility for the electron density to localize on the classical point charges of the MM atoms. This problem is particularly pronounced in a plane waves basis set approach.
A solution to the problem is found replacing the Coulomb potential in (8) with a suitable function $v_j(r)$ such that $v_j(r)$ goes to 1/r for large to a finite value for small r (see Fig. 1.2).

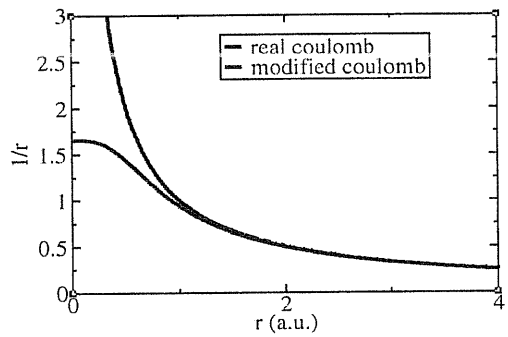

Figure 1.2. *In black* $v_j(r) = \frac{1}{r}$ *(Coulomb potential) and in blue* $v_j(r) = \frac{r_{cj}^n - r^n}{r_{cj}^{n+1} - r^{n+1}}$ *which goes as 1/r for large value of r and goes to a finite value for small r*

The electrostatic energy results:

$$H_{el} = \sum_{i \in MM} q_i \int dr \rho(r) v_j(|r - r_j|). \qquad (9)$$

An accurate description of hydrogens bond distances and energetics can be obtained with the following choice for v_j:

$$v_j(r) = \frac{r_{cj}^n - r^n}{r_{cj}^{n+1} - r^{n+1}}$$

where n=4 and r_{cj} is the covalent radius of the atom j (i.e. ~ 0.4 Å for hydrogen and ~ 0.7 Å for oxygen). The influence of the functional form of $v_j(r)$ on the bonding properties can be very dramatic. The previous choice of parameters has been tested to provide accurate results for the structural properties of a quantum water molecule in a box of classical water molecule without any *ad hoc* reparametrization of the force field(Laio et al., 2001).

Long range electrostatic interactions. As I mentioned before, the explicit calculation of (8) would be too expensive in a plane wave based approach. A straightforward computation requires calculation of order $N_r N_{MM}$ integrals, where N_r is the number of real space grid points (of order 100^3) and N_{MM} is the number of classical atoms (usually of order of 10000 or more in system of biochemical relevance).

A solution is found computing the exact calculation within a sphere of given radius and using a proper approximation for the region outside. The Coulomb interactions between the atoms outside this sphere and the QM part are simply described as the interaction between the MM charges and the electric multipole moments of the quantum region.

Up to quadrupolar order, the electrostatic field due to ρ at the position r_j of the MM atom j is given by

$$\int dr \frac{\rho(r)}{|r - r_j|} = C\frac{1}{|r - r_j|} + \sum_\alpha D^\alpha \frac{(r_j^\alpha - \bar{r}^\alpha)}{|r - r_j|^3} + \frac{1}{2}\sum_{\alpha\beta} Q^{\alpha\beta} \frac{(r_j^\alpha - \bar{r}^\alpha)(r_j^\beta - \bar{r}^\beta)}{|r - r_j|^5} + O\left(\frac{R_q}{|r_j - \bar{r}|}\right)^4 \qquad (10)$$

where $\bar{r}$ is the origin of the multipolar expansion (in our implementation the geometrical centre of the quantum system) and C, D^α and $Q^{\alpha\beta}$ are the total charge, the dipole and the quadrupole of the quantum charge distribution ρ.

$$C = \int dr \rho(r)$$

$$D^\alpha = \int dr \rho(r)\left(r^\alpha - \bar{r}^\alpha\right)$$

$$Q^{\alpha\beta} = \int dr \rho(r)\left[3\left(r^\alpha - \bar{r}^\alpha\right)\left(r^\beta - \bar{r}^\beta\right) - \delta^{\alpha\beta}|r - \bar{r}|^2\right]$$

Using (10) for the electrostatic interactions electrostatic Hamiltonian (9) can be expressed as:

$$H_{el} = \sum_{i \in NN} q_j \int dr \rho_e(r) v_j(|r - r_j|) + H_{lr} \quad (11)$$

where H_{lr} is given by

$$H_{lr} = C \sum_{j \notin NN} \frac{q_j}{\tau_j} + \sum_{\alpha} D^{\alpha} \sum_{j \notin NN} \frac{q_j}{\tau_j^3} \tau_j^{\alpha} + \frac{1}{2} \sum_{\alpha\beta} Q^{\alpha\beta} \sum_{j \notin NN} \frac{q_j}{\tau_j^5} \tau_j^{\alpha} \tau_j^{\beta}$$

with $\tau_j^{\alpha} = r_j^{\alpha} - \bar{r}^{\alpha}$.

The only free parameter of this approach is R_c, which is the radius of MM atoms explicitly included in electrostatic calculations. Tests have been performed(Laio et al., 2001) on the convergence respect to that parameter and a suitable radius resulted around 10 a.u. radius.

1.4.3 Potential and forces on the atoms

In the approach described, the non-bonded interactions are described by Hamiltonian (8). Energy conserving dynamics can be performed if the potential and the force are consistently calculated. In particular the potential is obtained by taking the functional derivative of (11) with respect to ρ.

$$V(r) = \frac{\delta H_{el}}{\delta \rho} = \sum_{j \in NN} q_j v_j(|r - r_j|) + V_{lr}(r)$$

where

$$V_{lr}(r) = \frac{\delta H_{lr}}{\delta \rho} = \sum_{j \notin NN} \frac{q_j}{\tau_j} + \sum_{\alpha} (r^{\alpha} - \bar{r}^{\alpha}) \sum_{j \notin NN} \frac{q_j}{\tau_j^3} \tau_j^{\alpha} + \frac{1}{2} \sum_{\alpha\beta} \left[3(r^{\alpha} - \bar{r}^{\alpha})(r^{\beta} - \bar{r}^{\beta}) - \delta^{\alpha\beta} |r - \bar{r}|^2 \right] \sum_{j \notin NN} \frac{q_j}{\tau_j^5} \tau_j^{\alpha} \tau_j^{\beta}$$

.

The forces on the atoms can be obtained by calculating the derivative of (11) with respect to atomic positions. In particular for $j \in NN$ set the forces are:

$$F_j^{\alpha} = q_j \int dr \rho(r) g_j(|r - r_j|) \frac{r^{\alpha} - r_j^{\alpha}}{|r - r_j|} \text{ with } g_j(r) = \frac{dv}{dr}.$$

while the forces on MM atoms not included in the NN set are:

$$F_j^{\gamma} = -q_j \left[\left(-\frac{C}{\tau_j^3} - 3\frac{1}{\tau_j^5} \sum_{\alpha} D^{\alpha} \tau_j^{\alpha} - \frac{5}{2} \frac{1}{\tau_j^7} \sum_{\alpha\beta} Q^{\alpha\beta} \tau_j^{\alpha} \tau_j^{\beta} \right) \tau_j^{\gamma} + \frac{D^{\gamma}}{\tau_j^3} + \frac{1}{\tau_j^5} \sum_{\alpha} Q^{\alpha\gamma} \tau_j^{\alpha} \right].$$

The correctness of the implementation can be directly verified by monitoring the conservation of energy for MD simulations. Some examples are given in ref.(Laio et al., 2001) and exhibit negligible shifts.

The relevant characteristics of the method I described can be summerized in the following points:

(i) the electrostatic coupling between the QM and MM parts is based on a fully Hamiltonian approach;

(ii) the occurrence of electron-spill out from the QM system onto neighboring classical point charges is impeded by the use of a suitable modification of the Coulomb interactions at short range;

(iii) boundary atoms involved in the chemical bonds between QM and MM part of the system are treated with empirical monovalent pseudopotentials;

(iv) the exclusion rules of standard (bio)molecular force field are incorporated in a consistent manner.

1.4.4 Partitioning the system

A central step in the modelling of any system within a QM/MM approach is the partitioning into QM subsystem and MM environment. A larger QM model increases the accuracy and the predictive power of the simulation. However, the computational cost of QM/MM simulation is almost completely determined by the size of the QM which, in a plane waves approach, scales as the third power of the system size.
An appropriate partitioning strongly depends on the system under study. One of the simplest examples for a possible choice of QM/MM partitioning is a typical solute/solvent system in which the former is treated quantum mechanically and the latter classical. In the specific case, the interactions between the two subsystems do not involve any chemical bonds. The simulation of enzymatic reactions on the other hand, involves almost invariably a quantum system that is a fraction of a larger, covalently

bonded molecule. So treatment of cuts between QM and MM atoms become a relevant point in a mixed approach simulation.
A good strategy in choosing QM/MM interface is to design the cut such that the QM and MM systems have, as computed based on the force field charges, an integer charge. Within the charge group concept used in GROMOS96, consistent cuts can be easily defined, since charge neutrality is guaranteed as long as complete charge groups are included in the QM system. In simulating proteins the best solution is, if possible, to cut non-polar bonds as C-C sidechain bonds. However accurate pseudopotential can be parameterized for specific situations in which a different bond has to be cutted. For example a boundary C_α can be parameterized to have the electronic properties that resemble more those of the backbone than those of a methyl group.

1.5 Calculated Properties

In this paragraph I will present some of the quantum mechanical properties which have been used as analysis instruments in this thesis.

1.5.1 Electron density shifts

A powerful instrument to investigate the chemical changes on a molecule for effect of interaction with other molecules or with the environment is the calculation of electron density shifts.

(i) The rearrangements of electronic density at the enzymatic active site ($\Delta\rho$) upon inclusion of the protein electric field was calculated as $\Delta\rho=\rho_{protein}-\rho_0$, where $\rho_{protein}$ is the electronic density for the quantum system in the all protein, while ρ_0 is the electronic density for the quantum system in vacuum.

(ii) The rearrangement of ρ upon formation of molecular complexes was calculated as $\Delta\rho=\rho_{1+2}-\rho_1-\rho_2$ where ρ_{1+2} is the electronic density for the complex and ρ_1 and ρ_2 are the electronic densities for single fragments.

(iii) Induced dipole moments on fragments can be calculated from integrating the induced electron density $\Delta\rho$ over the convolution of covalent radii around each group.

1.5.2 Maximally Localized Wannier functions

The Maximally localized Wannier functions provide a localized representation for the electronic states which is alternative to Block orbitals.

They are defined via a unitary trasformation of the Block function $\psi_n^{\mathbf{k}}(\mathbf{r})$, according to

$$w_n(\mathbf{r}-\mathbf{R}) = \frac{V}{(2\pi)^3}\int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{R}}\psi_n^{\mathbf{k}}(\mathbf{r})$$

where the labelling follows the lattice vector of the unit cell **R** instead that the wave vector **k**. Among all the possible sets of localized orbitals, Marzari and Vanderbilt(Marzari and Vanderbilt, 1997) basis set is the one which minimizes the total spread of the Wannier orbitals, given by the functional:

$$\Omega = \sum_n \left(\langle \mathbf{r}^2 \rangle_n - \langle \mathbf{r}_n \rangle^2 \right)$$

The centers $\langle \mathbf{r}_n \rangle$ of the maximally localized Wannier orbitals obtained with this procedure provide an immediate picture of electrons localization and of the chemical properties of a bond.

The geometrical analysis of centers of maximally localized Wannier functions(Silvestrelli and Parrinello, 1999b; Silvestrelli et al., 1998; Marzari and Vanderbilt, 1997) (WFC) is useful to investigate the polarity of a chemical bond, as shifts of WFC along the chemical bond correlate to differences of Pauling electronegativities(Hehre et al., 1986) ($\Delta\chi$) between the atoms forming the bond. In this thesis, the WFC analysis has been extended to the determination of changes in polarity of the chemical bonds upon formation of the intermolecular complexes. To illustrate our procedure, let us suppose that a chemical bond between atoms X and Y is in two different chemical environments, labeled 0 and 1. Then *an increase* of $\Delta\Delta\chi$(X-Y)= $\Delta\chi$(X-Y) $_1$-$\Delta\chi$(X-Y) $_0$ indicates that polarization toward atom X has *increased* on passing from 0 to 1.

1.5.3 Free energy calculations

In the fifth chapter of this thesis the activation free energy of an enzymatic reaction is calculated. In this paragraph I will briefly summarize the method I used.
The applied computational approach is based on a method developed for classical molecular dynamics studies of dissociation of an ion pair in polar solvent(Ciccotti et al., 1989).
In this scheme the reaction free energies are calculated as an integral of the average force f_s acting on the constraint along the reaction coordinate r.

$$\Delta F(r) = F(r) - F(r_0) = -\int_{r_0}^{r} dr' \, f_s(r')$$

The MD implementation is based on the relation between the mean force and the force on constraint required to keep the ion pair at a fixed distance during the simulation.
For the simple case of a distance constraint the force on constraint is equal to the corresponding Lagrange multiplier

$$\langle \lambda(r) \rangle = f_S(r) = -\left\langle \frac{\partial V}{\partial r_{AB}} - \frac{2K_B T}{r_{AB}} \right\rangle_{r_{AB}=r}$$

where the brackets denote an ensemble average over all configurations of solvent atoms and ions A and B with the condition that the ions be at a specified distance r_{AB}.
In practice, the value of r is increased in small increments using the final state of the previous simulation as the initial configuration for the next. The relative free energies for a set of discrete values of r are obtained as a discrete calculation of the integral:

$$\Delta F(r) = -\int_{r_0}^{r} dr' \langle \lambda(r') \rangle . \qquad (12)$$

The validity of equation (12) is not dependent on the nature of the interatomic forces between A and B, and it has been successfully extended to a variety of chemical systems calculations(Curioni et al., 1997; Meijer and Sprik, 1998a; Carloni et al., 2000).
In general in a condensed-phase system the true reaction coordinate can be a complicated quantity, a combination of a large set of atomic coordinates. However, in some cases, as in our specific case, it is reasonable to assume that there is one only reaction coordinate since the other degrees of freedom influence the system on much

more longer time-scales. In these cases the described approach can give also give quantitative information on the reaction free energy barrier.

Chapter 2

Outline

In this chapter I will give an overview on the systems investigated in this thesis and for each of them I will summerize our findings. The first paragraph will be dedicated to the application of *ab initio* calculations as a method for drug screening, in particular to characterize the HSV1-TK / substrates interactions and relate their molecular properties to catalysis. In a second paragraph a description of cation-π and OH-π interactions in two adducts of pharmacological interest is presented. Electrostatics and quantum effects are described and compared in the two cases. In a third paragraph, I will describe a study on the reaction mechanism of caspase-3, a target in the anti Alzheimer's therapy. Both energetics and structural aspects are described. Furthermore this study has allowed to construct a structural model of the intermediate/enzyme complex which provide a base for design of new inhibitors.

2.1 The rational of catalytic activity of Herpes Simplex Virus thymidine kinase: a combined biochemical and quantum chemical study

In chapter 3 of this thesis, DFT calculations are applied the rational screening of a series of Herpes Simplex Virus 1 thymidine Kinase (HSV1-TK) drugs and inhibitors.
TK has an important role in Gene therapy –based anticancer research since it selectively phosphorylates nucleoside analogs which can either inhibit viral DNA polymerase or cause toxic effects when incorporated into viral DNA stopping replication of the cancerous cells(Elion et al., 1977; Keller et al., 1981).
The motivation of our work was to relate structural properties of TK_{HSV1} ligands to their chemical reactivity and try to explain the different catalytic activity of a series of drugs presenting an apparently similar binding mode. Up to now substrates for HSV1-TK were the simple result of combinatorial chemistry and if they have been analyzed in terms of affinity, there was no explanation for their different catalytic behaviour.
This work was possible thanks to the large number of available crystal structures of several drug-enzyme adducts(Bennett et al., 1999; Brown et al., 1995; Champness et al., 1998; Prota et al., 2000; Wild et al., 1995; Wild et al., 1997).
Calculations focus on a set of ligands carrying a representative set of the large spectrum of sugar-mimicking moieties and for which structural information of the TK_{HSV1}-ligand complex are available (Fig.2.1 a,b). On the experimental side, Prof. Folker's group in ETH-Zurich provided us the catalytic constants of these compounds, measured under the same experimental conditions, using an UV-spectrophotometric assay. Calculations point to the crucial role of electric dipole moment of ligands and its interaction with the negatively charged residue Glu225. A striking correlation is found between the energetics associated to this interaction and the k_{cat} values measured under homogeneous conditions (Fig.2.1c).
This finding uncovers a fundamental aspect of the mechanism governing substrate diversity and catalytic turnover and thus represents a significant step forward to the rational design of novel and powerful prodrugs for antiviral and TK_{HSV1}-linked suicide gene therapies.
Thus first principle methods reveal themselves as potential tool for fast drug screening.

In fact a simple geometry optimization for a compound in the enzyme active site and its description in term of electronic properties would not require much more computational cost than building up an *ad hoc* force field.

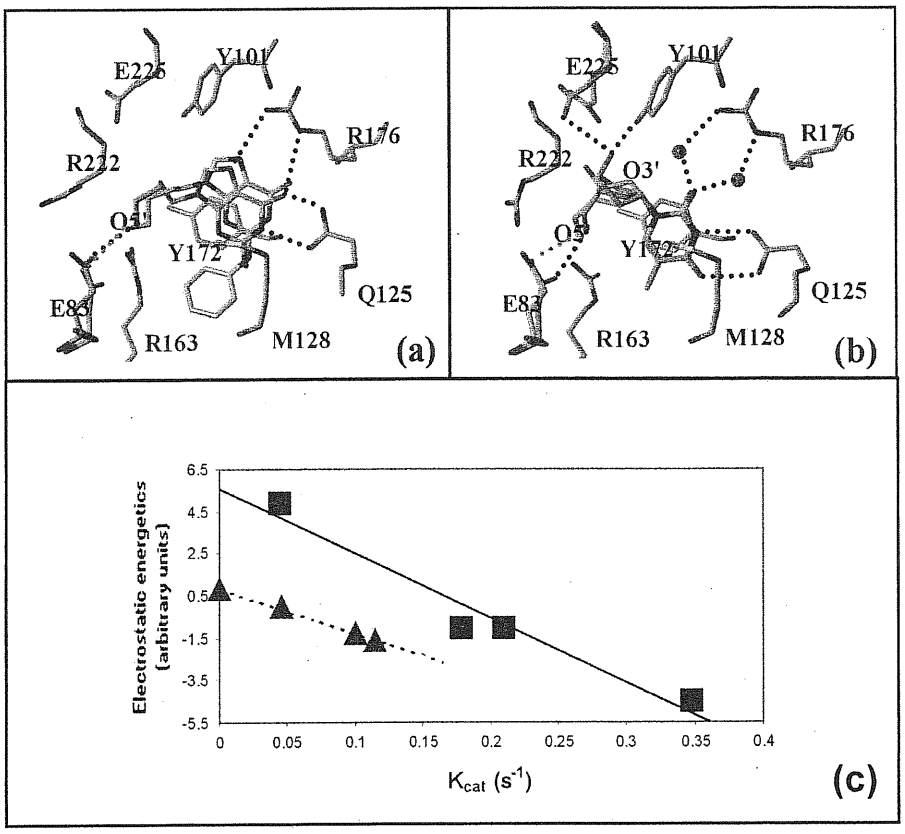

***Figure 2.1**. Active site superposition of representative ligands of the guanine (a) and thymine (b) series. Correlation between k_{cat} and the sugar-moiety-dipole-Glu225 electrostatic energy (electrostatic energetics). Thymine and guanine derivative are displayed as squares and triangles, respectively. Linear fits are also plotted (R^2 values are 0.954 and 0.994 of thymine and guanine, respectively).*

2.2 Cation-π versus OH-π Interactions in Proteins: examples from system of pharmacological interest.

In the next chapter, structure and bonding of a cation-π and a facet OH-π adducts are investigated using density functional theory with gradient-corrections for the exchange-correlation functional. Our calculations are carried out for two specific model complexes representing i) the thymine/Arg 72 adduct in the ternary complex of HIV-1 reverse transcriptase (RT) with a DNA template primer and a deoxynucleoside

triphosphate(Huang et al., 1998) ii) the Tyr6-Thr13 adduct in μ-gluthatione transferase (μ-GST)(Xiao et al., 1996) (Fig. 2.2).

Both investigated systems have pharmacological relevance. HIV-1 RT is one of the major tagets for anti-AIDS therapy(Furman et al., 2000); μ-GST-differential expression has been implicated in the development of cancers as well as their resistance to chemotherapeutic drugs(McCallum et al., 2000) (and reference therein).

The calculations point to the fundamental role of electrostatics. In HIV-1 RT, the π electronic density of thymine is essentially uneffected by the presence of the arginine guanidinium group; on the contrary, tyrosine is significantly polarized by the interaction with the hydroxyl group and other groups present in the μ-GST enzyme. The influence of Thr13 induced-polarization on Tyr6 pKa is compared with that of other interacting groups at the active site. Electrostatic interactions provide similar stabilization energies to the two π complexes.

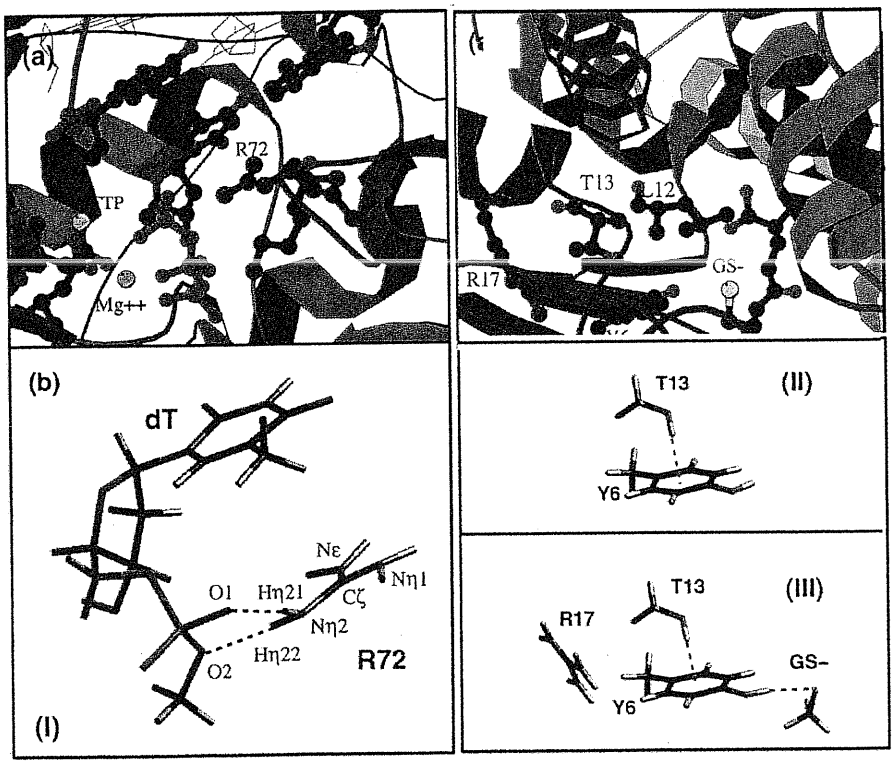

Figure 2.2 *(left) (a) Thymine binding site in HIV-1 RT(Huang et al., 1998) and (b) complex* ***I*** *used in the calculations. (right) (a) Active site of μ-GST – GS⁻ complex(Xiao et al., 1996) and (b) model complexes* ***II-III*** *used in the calculations.*

2.3 Reaction Mechanism of Caspases: Insights from QM/MM Car-Parrinello Simulations

In chapter 5, I will describe a study on the energetics and the structural features of the enzymatic reaction mechanism of caspase-3, a cysteine protease which is supposed to play a key role in the development of neurodegenerative disease.

Caspase-3 is in fact a key enzyme in cell apoptosis and it is has been proved to be directly involved in the neurodegeneration of the Alzheimer's disease(Fasulo et al., 2000). Up to now research of caspase-3 inhibitors has been based only on combinatorial chemistry but has not completely solved efficiency and selectivity problems.

The investigation of the enzymatic mechanism is of interest for pharmaceutical application, as the knowledge of the transition state may provide the starting point for the rational design of new inhibitors.

In the study of Caspase-3 reaction the requirement of quite large active site and the necessity to include protein geometric and electric effects has led us to use a mixed car-Parrinello/Classical Molecular Dynamics approach. We use the electrostatic treatment of the interface developed in Rothlisberger's group in Zurich(Laio et al., 2001).

We calculated the free energy barrier of the reaction with the thermodynamic integration method, which has been successfully used in a variety of chemical reactions(Curioni et al., 1997; Meijer and Sprik, 1998a; Carloni et al., 2000), among which that performed by another protease, in particular the aspartyl protease from HIV-1(Piana et al., 2001b).

These calculations show that the attack of the hydrolytic water molecule implies an activation free energy of ca. $\Delta F_A \sim 20$ kcal/mol and leads to a previously unrecognized gem-diol intermediate can readily ($\Delta F_A \sim 5$ kcal/mol) evolve to the enzyme products (Fig2.3). These findings help to elucidate the striking difference in catalytic activity between the caspase family and the cysteine protease of papain family, such as the well studied papain and the cathepsins.

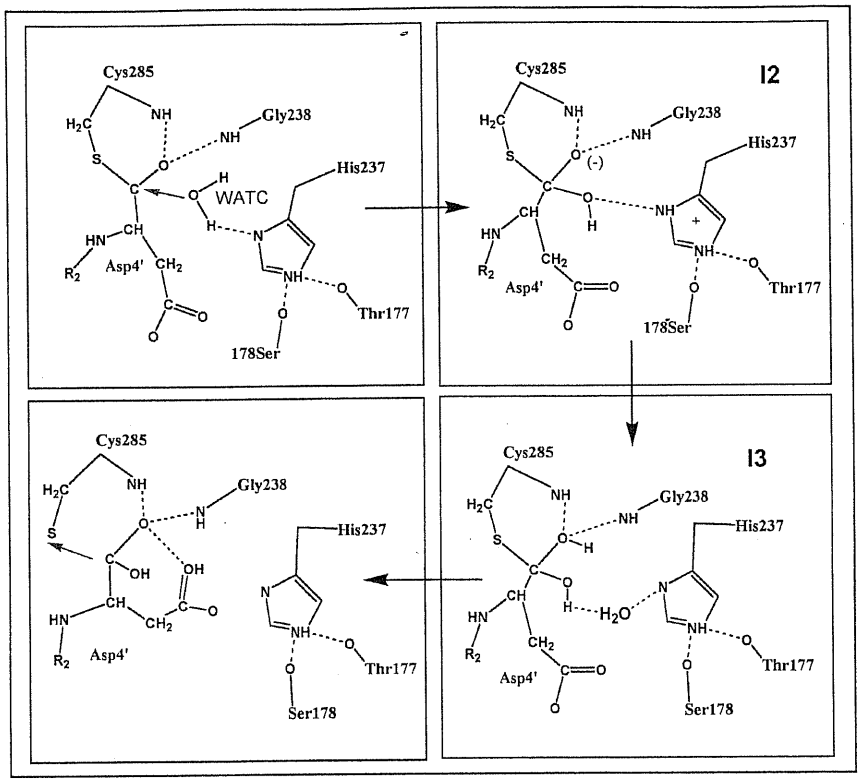

Figure 2.3 *Reaction mechanism for Caspase-3 as it emerges from QM/MM calculations. Deacylation step precedes throught the formation of a first anionic intermediate* ***I2*** *which then evolves to the neutral tetrahedric intermediate* ***I3****. Energy barrier associated with intermediate formation and dissociation are respectively aroound 20kcal/mol and 5 kcal/mol, respectively.*

Chapter 3

The rational of catalytic activity of Herpes Simplex Virus thymidine kinase: a combined biochemical and quantum chemical study

Most antiherpes therapies exploit the large substrate acceptance of herpes simplex type-1 thymidine kinase (TK_{HSV1}) relative to the human isoenzyme. The enzyme selectively phosphorylates nucleoside analogs which can either inhibit viral DNA polymerase or cause toxic effects when incorporated into viral DNA. To relate structural properties of TK_{HSV1} ligands to their chemical reactivity we have carried out *ab initio* quantum chemistry calculations within the density functional theory (DFT) framework in combination with biochemical studies. Calculations have focused on a set of ligands carrying a representative set of the large spectrum of sugar-mimicking moieties and for which structural information of the TK_{HSV1}-ligand complex are available. k_{cat} of these ligands have been measured under the same experimental conditions using an UV-spectrophotometric assay. The calculations point to the crucial role of electric dipole moment of ligands and its interaction with the negatively charged residue Glu225. A striking correlation is found between the energetics associated to this interaction and the k_{cat} values measured under homogeneous conditions. This finding uncovers a fundamental aspect of the mechanism governing substrate diversity and catalytic turnover and thus represents a significant step forward to the rational design of novel and powerful prodrugs for antiviral and TK_{HSV1}-linked suicide gene therapies.

3.1 Introduction

3.1.1 Gene Therapy for Anticancer Research

The thymidine kinase from *Herpes simplex virus type* 1 (TK_{HSV1}) salvages (2'-deoxy)thymine (dT) into the virus' metabolism by converting it to thymidine monophosphate (Chen and Prusoff, 1978). In contrast to the human isoenzyme, TK_{HSV1} acts as phosphorylating agent towards a large variety of nucleoside analogs such as (North)-methanocarba-thymidine (n-MCT), aciclovir (ACV), ganciclovir (GCV), penciclovir (PCV). These analogs exhibit chemical diversity for the nucleobase as well as for the sugar-like chain moiety (Elion et al., 1977; Keller et al., 1981) (Fig. 3.3). Substrate diversity of TK_{HSV1} provides the molecular basis for effective and selective treatment of virus infections. It is also exploited for gene-therapy approaches involving TK_{HSV1} as suicide gene by anticancer intervention (Culver et al., 1992) or the control of graft-versus-host-disease by allogeneic bone marrow transplantation (Bonini et al., 1997).

In anticancer gene therapy, viral tk genes are introduced in the tumoral cells, and then treatment with nucleoside analogs is provided. When the phosphorylated analogs are incorporated in the cell DNA, the elongation of the replicating DNA strand is stopped and cell death follows (Fig.3.1). This also cause a bystander effect, so that not every cell within a tumor has to be transduced by the *tk* gene in order to wipe out the tumor. . This strategy has been evaluated for possible treatment of localized brain tumors, liver metastases, peritoneal-based metastases, and mesotheliomas. Although the bystander effect is a positive phenomenon associated with this approach, its unpredictability, along with difficulties in transduction itself by vectors, has kept the cure rates low.

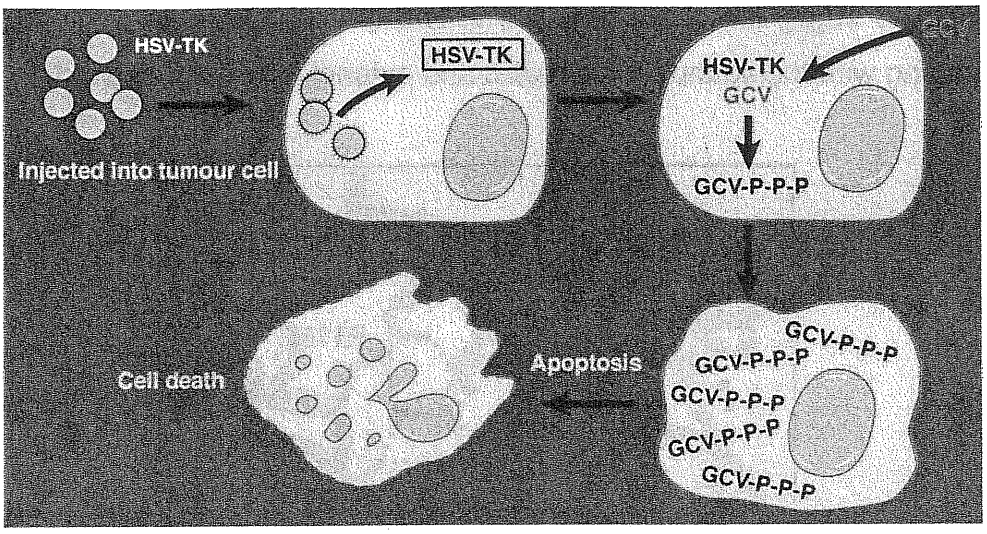

Figure 3.1. Anticancer Gene Therapy. First HSV gene is introduced in the tumor cell, where it express viral TK. Then the cell is treated with nucleoside anlogs (e.g. ganciclovir). Phosphorylated analogs block cell DNA replication leading to cell death.

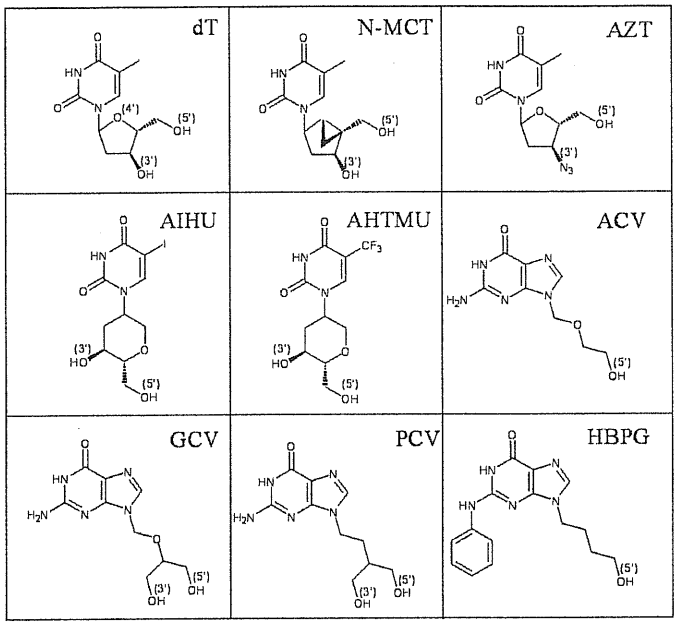

Figure 3.2. Chemical formulas of selected (fraudulent) substrates and inhibitors of TK_{HSV1}. dT is the natural substrate; n-MCT, ACV, PCV, GCV, AHIU, AHTMU and AZT are prodrugs; and HBPG is an inhibitors. The 5'OH and 3'OH groups belonging to dT and their mimics are labeled.

3.1.2 HSVTK: the structure

The recent determination of the X-ray structure of a large spectrum of ligand/enzyme complexes (Bennett et al., 1999; Brown et al., 1995; Champness et al., 1998; Prota et al., 2000; Wild et al., 1995; Wild et al., 1997) has opened a new avenue for the understanding structure/function relationships as well as of the functional role of aminoacids involved in the binding. It has emerged that the enzyme accommodates the

nucleobase and sugar-like chain moieties in two different pockets (P1 and P2 in Fig. 3.3), interacting with the ligand by a specific and extensive H-bond network.

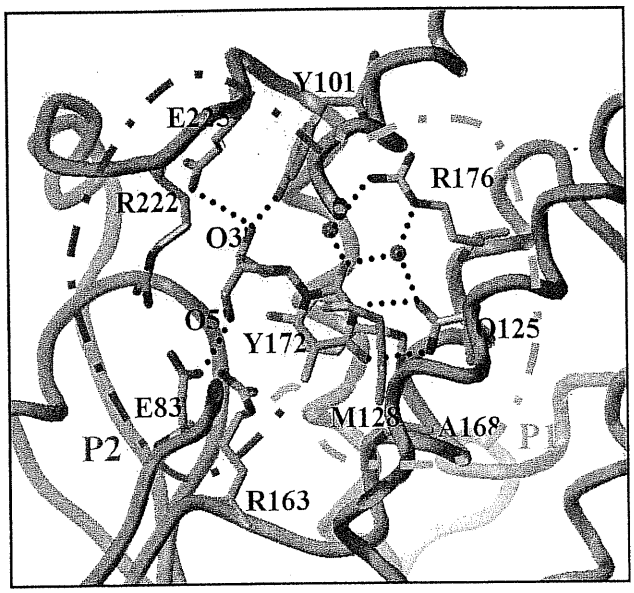

***Figure 3.3**. X-ray structure for the thymidine kinase active site (Wild et al., 1997). P1 and P2 are the active site portions which accommodate the nucleobase and the sugar like moiety, respectively. Water molecules are shown as red spheres and hydrogen bonds are depicted as dotted line. dT and the amino acids are displayed as capped sticks and are color coded (C: orange; O: red; N: blue; S: yellow). The rest of the protein is shown as ribbon model. The picture was generated with the program SYBYL (Tripos Inc., St. Louis)*

In the pocket P1, the nucleobase moiety (either a thymine or a guanine ring) is stabilized by direct H-bonds with the highly conserved Gln125 and with Arg176 by means of two ordered water molecules. The pyrimidine ring is further fixed by a peculiar interaction with Tyr172 and Met128 (Wild et al., 1997) in a sandwich-like orientation (Alber et al., 1998) (Kussmann-Gerber et al., 1998) (Fig. 3.3).

In P2, the sugar-like chains interacts with the protein via its hydroxyl groups. The 3'OH (and its mimics) forms specific H-bonds with Tyr101 and/or Glu225 (Figs. 3.3, 3.4) while 5'OH forms a direct H-bond or water-mediated interactions with Arg163, Glu83 and Arg222 (Fig. 3.3, 3.4). In contrast, the polar C1'-O4'-C4' function belonging to dT and the correspondent ether groups of ACV and GCV interact neither with polar or charged groups of the protein nor with the solvent; instead, they point towards a hydrophobic region made up of the Trp88, Ile97, Met128 side chains (Fig. 3.5).

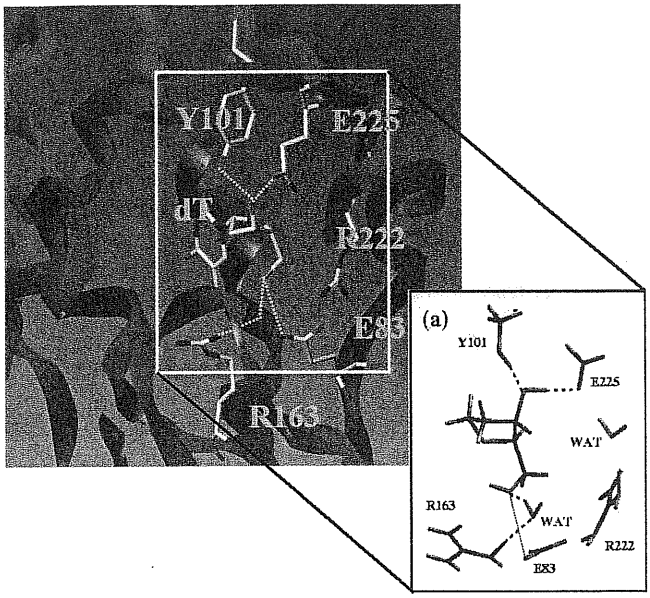

***Figure 3.4**. Quantum mechanical model for dT/HSV-1 TK complex. Water molecules are represented as red spheres and hydrogen bonds are shown as dotted lines.*

3.1.3 Motivation of present work

Key aspects of nucleotide binding have been recently elucidated by theoretical (Alber et al., 1998) and experimental (Bennett et al., 1999; Brown et al., 1995; Champness et al., 1998; Prota et al., 2000; Wild et al., 1995; Wild et al., 1997) approaches. In contrast, fundamental questions regarding the nature of the interactions between the ribose-like moiety and the enzyme are still opened. First, it is rather intriguing that the accessibility of the 5'OH mimic to the ATP cofactor is roughly the same (Fig. 3.6), yet the HBPG molecule, which shares the same binding mode as structurally related prodrug aciclovir (Bennett et al., 1999), *inhibits* the enzyme. Furthermore, it is not known why k_{cat} values of prodrugs are much smaller than that of the natural substrate (Table 3.1) although the protein/sugar mimics H-bonding interactions are very similar (Fig. 3.6). Finally, as discussed above, it is rather surprising that the environment accommodating inner sugar ring C1'-O4'-C4' group of natural substrate is totally hydrophobic (Fig. 3.5).

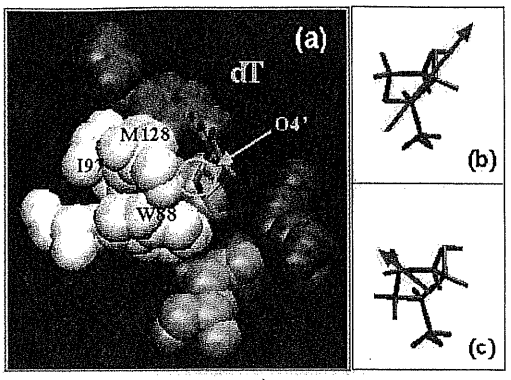

Figure 3.5. *(a) dT-TK_{HSV1} complex. Orientation of O4' in its hydrophobic pocket; sugar (b) and methylene derivative (c) electric dipoles.*

In this paper we address these fundamental issues by performing a combined quantum chemical and biochemical investigation of sugar- and sugar-like moieties of substrates and inhibitors of TK_{HSV1}. The theoretical methodology is based on gradient-corrected density functional theory. This approach allows an accurate description of electrostatic and hydrogen bonded complexes (Piana and Carloni, 2000; De Santis and Carloni, 1999; Roethlisberger et al., 2000; Roethlisberger and Carloni, 1999; Alber et al., 1998), such as those investigated here. The biochemical study involves the measurements of the catalytic constant (k_{cat}) of dT- and prodrugs using a modified UV-spectrophotometric assay based on pyruvate kinase and lactate dehydrogenase(Keller et al., 1981). This assay does not require the use of radioactively labeled compounds. This provided a set of homogenous data, which is necessary for a proper comparison with the quantum chemical calculation.

As we will show, the interaction between the *electric dipole* of the sugar mimicking moiety and the *negative charge* of the conserved residue Glu225 correlates with k_{cat} measurements and thus, is a critical factor for the phosphorylation reaction.

Compound	k_{cat} (s^{-1})[a]	k_{cat} in (s^{-1}) from literature	k_{cat} as rate % of dT[b]
dT	0.348 ± 0.004	0.35[c] – 0.45[d]	100%
ACV	0.115 ± 0.025	0.015[d] - 0.101[e]	27%
GCV	0.100 ± 0.017	0.102[d]	n.d.
PCV	0.045 ± 0.019	n.d	9%
AHTMU	0.210 ± 0.048	n.d	n.d.
n-MCT	0.178 ± 0.010	n.d.	n.d.
AZT	0.044 ± 0.018	0.056[c]	n.d.
HBPG	0	0[f]	n.d.

Table 3.1. *Values of catalytic activity (k_{cat}) of compounds listed in Fig. 3.2. The values are the results of at least five independent experiments.*

[a] This work; [b] The rate in % were determined using the substrates at 250 μM (VereHodge, 1993); [c] in (Pilger et al., 1999); [d] in (Kokoris et al., 1999); [e] in (Kussmann-Gerber et al., 1999); [f] HBPG is an inhibitor (Xu et al., 1995); n.d.: not determined

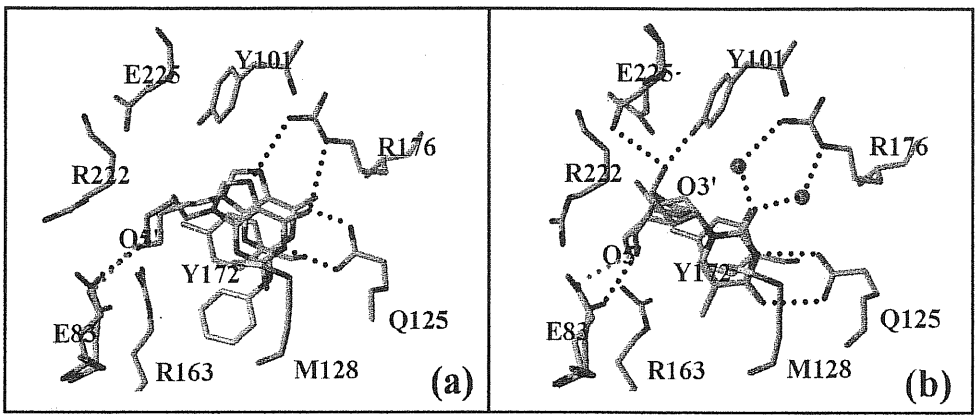

***Figure 3.6**. Active site superposition of representative ligands of the guanine (a) and thymine (b) series. a) The structure of the substrate ACV in complex with TK_{HSV1} (Bennett et al., 1999) is superimposed with that of the inhibitor HBPG (Bennett et al., 1999). It clearly shows that the inhibitor in orange shares the same binding mode as the substrate ACV in cyan. In both cases the 5'OH mimicking group are at hydrogen bond distance from E83 and points towards the phosphate moiety of ATP. For seek of clarity the superimposed GCV and PCV sharing the same binding mode of ACV (Bennett et al., 1999) are not shown. H-bond network of ACV is shown as dotted line as a representative for all guanine derivatives b) The structure of the natural substrate dT (cyan) ligated to TK_{HSV1} (Champness et al., 1998) is superimposed with this of TK_{HSV1}:n-MCT (Prota et al., 2000) (orange). Both natural substrate and the prodrug share the same binding mode with the 5'OH mimicking group pointing towards E83. AHTMU and AZT (Bennett et al., 1999) have the same binding mode and they are not shown for clarity.*
H-bond network of dT is shown as dotted line as a representative for all thymine derivatives in respect to the base and the 5'OH position. In contrast, the interaction of 3'OH mimicking group with E225 and Y101 varies depending on the ligand. All residues of the TK_{HSV1}:HBPG and TK_{HSV1}:n-MCT which take the same conformation as in TK_{HSV1}:ACV and TK_{HSV1}:dT, respectively were omitted for seek of clarity.

3.2 Methods

3.2.1 Biochemical essays

Materials. PreScission protease was purchased from Pharmacia. AZT and reagents for enzyme assays were obtained from Sigma. Strain BL21 (Pharmacia) served as expression host. n-MCT and PCV were kindly provided by Dr. Marquez (NIH, Bethesda, USA) and Dr. Johannsen (Forschungszentrum Rossendorf Institute of Bioinorganic and radiopharmaceutical Chemistry research center, Rossendorf, D) respectively. 5-Trifluoromethyl anhydrohexitoluridine (AHTMU), an AHIU analogues, and HBPG were gift of Dr. De Clercq (Rega Institute Katholike Universiteit Leuven, Be) and Dr. Wright (University of Massachussets, USA), respectively. ACV and GCV were purchased by

Glaxo-Wellcome and Roche, respectively. dT was obtained from Fluka.

Expression and purification of TK_{HSV1}. TK_{HSV1} was expressed as gluthathion-S-transferase-fusion protein in competent *E. coli* BL21 using the vector pGEX-6P-2-TK (Prota et al., 2000). The protein was purified by gluthathion-affinity chromatography followed by on-column PreScission protease cleavage using a previously described protocol (Prota et al., 2000). Purification was monitored by SDS-PAGE and led to a > 90 % pure TK_{HSV1} , which was directly used for k_{cat} determination. Total protein concentration was measured using the Bio-Rad Protein Assay (Bio-Rad).

Spectrophotometric assay for k_{cat} determination. An UV-spectrophotometric test based on a lactate dehydrogenase - pyruvate kinase coupled assay (Keller et al., 1981) was employed to monitor ADP formation during substrate phosphorylation. The concentration of ATP and the ligands were 5 mM and 1 mM, respectively. These concentrations are at least five time higher than the binding affinities of the studied compounds allowing the measurement of V_{max} and thus of k_{cat} ($k_{cat} = V_{max}/[E]$). The change in Absorbance at 340 nm was recorded over time and correlate with the k_{cat} of the analyzed substrates for which values were known from the literature and allows the determination of k_{cat} for compounds that are not available in the radiolabelled form (Table 3.1).

3.2.2 Quantum chemical calculations

Our structural models for quantum chemical calculation on the sugar moiety of dT-, ACV-, GCV-, PCV-, n-MCT- and HBPG- TK_{HSV1} were constructed from the correspondent crystal structures (Pdb-entries: 1kim, 2ki5, 1ki2, 1ki3, 1e2k, 1qhi) (Bennett et al., 1999; Champness et al., 1998; Prota et al., 2000). The resolution of the structures ranges from 1.7 to 2.4 Å.

The structure of AHTMU-TK_{HSV1} has not been solved yet, while the structure of the complex AHIU-TK_{HSV1} is known (Pdb entry: 1ki6) (8). AHIU is chemically and structurally extremely similar to AHTMU (Fig.3.2): the two analogs differ only for the group at position 5 of the nucleobase ring (a iodine in AHIU, a CF_3 in AHTMU). All available structural data show that this substitution does not affect the binding orientation of the

sugar mimicking moiety (Champness et al., 1998). The initial configuration of AHTMU is therefore built from this X-ray structure.

The X-ray structure of the AZT-TK_{HSV1} complex is not available and therefore one has to resort to theoretical structural models. Here we used molecular dynamics based models reported earlier (Christians et al., 1999).
The complexes included the sugar-like moieties of the ligands and part of side chains of all the groups directly interacting with it or forming water-mediated hydrogen bonds (Tyr101, Arg163, Glu83 and Glu225, Arg222). Trp88, Ile97, Met128 side chains, were not included as i) they would have dramatically increased the size of our model complexes; ii) they do not contribute significantly to the electrostatic interactions (the focus of the present work) as they form only weak hydrophobic contacts with the sugar; iii) the position of these groups is essentially the same in all the complexes investigated here (Bennett et al., 1999; Champness et al., 1998; Prota et al., 2000) and therefore their contribution is expected to be rather constant. Hydrogen atoms were added assuming standard bond lengths and bond angles. In Fig. 3.4 the quantum-mechanical model for dT, the natural substrate, is shown in details. The overall charge of our complexes is 0.

Calculations were performed within the framework of density functional theory (DFT) (Hohenberg and Kohn, 1964) in its Kohn and Sham (Kohn and Sham, 1965) formulation. In this approach the use of gradient corrections is crucial for correctly describe hydrogen bond interactions. Here we used the prescription of Becke, Lee, Yang and Parr as it has been shown to provide an accurate description of water and hydrogen bonding in biological systems (Piana and Carloni, 2000; De Santis and Carloni, 1999; Molteni et al., 1999; Alber et al., 1999). The basis set consisted of a plane wave basis set up to a cutoff of 70 Rydberg, the interactions between valence electrons and ionic cores being described by pseudopotentials of the Martins-Troullier type (Troullier and Martins, 1991). Only the gamma point was used. The dT-TK complex was inserted in orthorombic box of edges 14x15x16 $Å^3$. Similar box sizes were used for all the other complexes. In all circumstances the separation between periodic images was at least 6 Å. However the Coulombic interaction between images was screened using the approach of Barnett et al. (Barnett and Landman, 1993). Geometry optimization was performed using the direct inversion iterative subspace (DIIS) (Pulay ,

1980). Cα atoms were constrained to crystallographic position to take into account for the reduced mobility of residues due to protein environment. A similar procedure has been reported elsewhere. (Piana and Carloni, 2000)

The electronic structure was described with the geometrical analysis of centers of maximally localized Wannier functions (WFC) (Silvestrelli et al., 1998; Silvestrelli and Parrinello, 1999; Marzari and Vanderbilt, 1997). This analysis is useful to investigate the polarity of chemical bonds, as WFC shifts are related quantitatively to the differences $\Delta\chi$ of Pauling electronegativities (Hehre et al., 1986) with respect to the non interacting compounds(Alber et al., 1999): an increase of $\Delta\chi$ corresponds to an increase of polarization of the electronic density in the X-Y bond toward the most electronegative atom X. Effects of protein electric field were estimated through a comparison of Wannier-centers shifts for the system in vacuum or in the presence of the protein electric field. The procedure was identical to that of ref.(Piana and Carloni, 2000) Interaction energies between the substrate and the active site are calculated within the central dipole approximation (Leach, 1996) where the dipole is the result of the quantum calculation. Interaction energy was calculated as $E_{charge\text{-}dipole} = (\xi\, \mu \cdot \mathbf{R})/(4\pi\varepsilon_0 R^3)$, where **R** is the distance vector between the centers of charges of the sugar moiety and the residue of charge ξ; μ is the dipole moment for the sugar moiety.
Even if the expansion is limited to large distances between two interacting groups, this model can provide quantitative results for a comparison between systems which have similar geometries. We have reported the energetics in arbitrary units.

3.3 Results

3.3.1 Catalytic Activity of TK_{HSV1} .

The catalytic activity of TK_{HSV1} with the compounds listed in Fig. 3.2 was measured under the same conditions. Table 3.1 shows that the measured k_{cat} of dT agrees with those reported in literature (Kokoris et al., 1999; Pilger et al., 1999). Also the k_{cat} of ACV corresponds to that measured in two independent laboratories (Kussmann-Gerber et al., 1999; VereHodge, 1993). In the literature a value of 0.015 (s^{-1}) is also reported for ACV (Kokoris et al., 1999), but in consideration to the new reported values this seems to be

an outliner. The k_{cat} values of GCV and PCV are also congruent with the literature values (Kokoris et al., 1999; VereHodge, 1993). The same agreement has been found for AZT whose k_{cat} value is 0.044 (s^{-1}). The k_{cat} values of the new compounds AHTMU and n-MCT have been reported here for the first time.

3.3.2 Quantum Chemistry Calculations

Overall, the geometry of the complexes are in very good agreement with the experimental structural data as shown by the low root mean squared deviation values between experimental and optimized structures (Table 3.2). In particular Table 3.3 shows that the structural parameters of the geometry optimized systems are in good agreement with experimental data. Thus, our first principle approach appears to reproduce very reliably the structural properties of the adducts. Geometry optimization permits also us to establish the H-bond network between the substrate and the enzyme which is crucial in determining the electrostatic properties of the adducts and which is only indirectly deducible from the crystal structure where H are missing.

Compound	RMS (Å)	k_{cat} (s^{-1})	Dipole(D)	Dipole-GLU225 energy	Dipole-ARG163 energy
dT	0.19	0.348 ± 0.004	1.74	-4.5	-3.1
ACV	0.26	0.115 ± 0.025	1.96	-1.6	-2.2
GCV	0.27	0.100 ± 0.017	1.74	-1.3	-3.4
PCV	0.35	0.045 ± 0.019	1.31	0.0	-0.4
AHTMU	0.26	0.210 ± 0.048	0.40	-1.0	-0.4
n-MCT	0.17	0.178 ± 0.010	2.83	-1.0	-4.9
AZT	0.24	0.044 ± 0.018	3.26	4.9	4.7
HBPG	0.15	0	1.74	0.8	-2.0

Table 3.2. *Structure, catalytic activity and electrostatic interactions in ligand-TK_{HSV1} complexes. RMS deviations between X-ray and optimized structure are reported in Å; dipole moments in Debye. Glu225- and Arg163- sugar-like chain interactions are reported in arbitrary units.*

Enzyme-sugar interactions. One of the most important physical properties of polar molecules such as ribose and its analogs is given by their electric dipole moment. Fig. 3.7 offers a visual representation of the calculated dipoles projected to the geometry-optimized structure. The dipole of the sugar moiety of the natural substrate aligns strikingly to the negatively charged Glu225 group. Interestingly, the dipoles of most the

prodrugs investigated here also point to this residues. The calculated stabilization energy resulting from the electrostatic interaction is maximum for the natural substrate (Table 3.3). The dipoles of AZT and the inhibitor HBPG do not point towards the negative residue and as a consequence the interaction energy is repulsive.

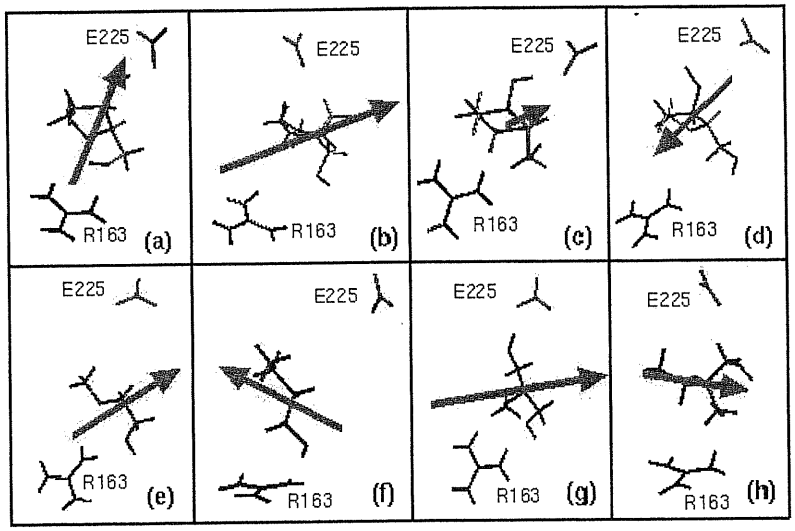

Figure 3.7. *Electric dipoles of HSV-1 TK sugar like chains. dT (a), n-MCT (b), AHTMU (c), AZT (d), ACV (e), HBPG (f), GCV (g), PCV(h).*

	Bond length (Å)		Angle (degree)		Torsional angle (degree)
O5-C5	1.48 (0.01)	O5-C5-C4	110 (-2)	O5-C4-C3-C2	56 (6)
C5-C4	1.52 (0.01)	C5-C4-C3	116 (5)	C5-C4-C3-C2	-105 (-14)
C4-C3	1.57 (0.06)	C4-C3-C2	103 (0)	C4-C3-C2-C1	-36 (10)
O3-C3	1.44 (0.04)	O3-C3-C4	110 (6)	C3-C2-C1-O4	46 (9)
C3-C2	1.54 (0.02)	C3-C2-C1	100 (-4)	O3-C3-C2-C1	81 (17)
C2-C1	1.52 (0.04)	C2-C1-O4	104 (8)	C2-C1-O4-C3	-39 (-20)
C1-O4	1.47 (0.01)	C1-O4-C4	106 (-13)	C1-O4-C4-C3	15 (23)
O4-C4	1.48 (0.05)	O4-C4-C3	106 (8)	O4-C4-C3-C2	14 (-17)

Table 3.3 *dT/TK optimized structure: bond lengths, angles and torsional angles of the sugar moiety. Deviations from the crystallographic structure are shown in brackets.*

In an attempt at correlating electrostatic interactions with kinetics data, we first subdivide the compounds in two major classes. The first includes the compounds bearing a thymine- or a thymine substituted base (namely dT, AHTMU, n-MCT and AZT); the second those bearing guanine (ACV, HBPG, GCV, PCV).

The subdivision is necessary as the computational model did not include the base and consequently the electrostatic calculations do not consider the contributions of the

nucleobases (Alber et al., 1998), whose dipole moments are different but additive [1]. Fig. 3.8 points out the correlation between the interaction energy dipole-Glu225 and the catalytic constant of the studied ligands appears clearly. The slope of the linear fit depends from the type of sugar mimicking moiety of the ligand. The additivity of the sugar like chains and nucleobase dipoles is confirmed by a rigid shift between the linear fit for the guanine and thymine derivatives (Fig. 3.8).

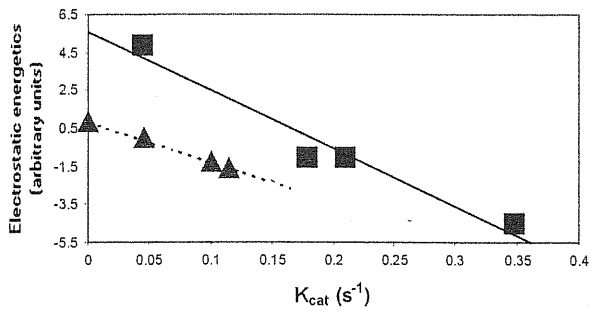

Figure 3.8. *Correlation between k_{cat} and the sugar-moiety-dipole-Glu225 electrostatic energy (electrostatic energetics). Thymine and guanine derivative are displayed as squares and triangles, respectively. Linear fits are also plotted (R^2 values are 0.954 and 0.994 of thymine and guanine, respectively).*

The role of Arg222, Arg163 and Glu83 interacting groups. Although there is no direct correlation between the k_{cat} and the interaction Arg163-Dipole, the calculated interaction between Arg163 and the sugar is important, consistently with previously published mutagenesis experiments (Black and Loeb, 1993). From the crystallographic and mutagenesis studies, Arg222 and Glu83 are essential elements of the kinase machinery, Arg222 forms the anion hole with Arg220 to make the phosphate atom more electrophilic and Glu83 behaves as base on the O5' atom causing an increase in nucleophily. Thus, it is not surprising that Arg222- and Glu83-Dipole interactions appears not to have any direct correlation with k_{cat} (data not shown) being their influence based on a totally different mechanisms.

Sugar/TK$_{HSV1}$ hydrophobic interactions. Our calculations have pointed to the crucial role of the electric dipole moment of the sugar moieties. We now address to the following question: In the natural substrate, is the polar C1'-O4'-C4' (which faces a hydrophobic pocket, Fig 3.5a) important for the correct orientation of the dipole? To

[1] The contribution of the nucleobase of the nucleotides to the dipole interaction results additive, even if different for thymine and guanine. Since the chemical bond between the nucleobase and the sugar moiety is non-polar we have verified the additivity of the

answer this question, we calculate the change of electric dipole associated to replacement of O4' with the CH_2 apolar group (Fig.3.5b). Our calculation show that the resulting dipole is both smaller and different in orientation relative to the sugar moiety (Fig 3.5b,c). We conclude that the polar function is essential for a correct alignment of the dipole to E225 charge.

Protein environment effects. The field of the protein may be very important for the chemistry of the active site of this and other enzymes (Warshel, 1998). Here we estimate the effect of the protein frame by comparing the electronic structure of the complexes in vacuum with those in the presence of the protein. A convenient representation of the electronic structure is given by the Wannier functions, whose centers (WFC) represent chemical concepts such as lone pairs and chemical bonds. Fig. 3.9 shows that there is no appreciable displacement of the WFC; so that it is possible to conclude that main contribution to the interaction is included in the model we have chosen for quantum mechanics calculations.

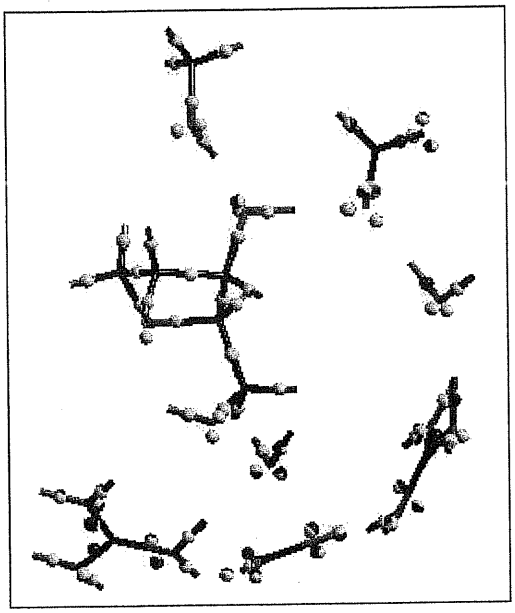

Figure 3.9. *Electronic structure of dT/HSV-1 TK active site: the Wannier centers (Silvestrelli et al., 1998; Silvestrelli and Parrinello, 1999; Marzari and Vanderbilt, 1997) in the presence and in the absence of the protein electrostatic potential are represented as yellow and green spheres, respectively.*

nucleobase and sugar moiety dipoles.

3.4 Discussion and Conclusion

Deciphering the binding mode of prodrugs to the enzyme and their catalytic turnover may help to rationally design more and potent prodrugs for antiviral therapy to overcome the problems of resistance or to be used in enzyme-prodrug gene therapy as well as improve variants of TK_{HSV1} for gene therapeutic approaches (Encell et al., 1999).

The mechanisms ruling binding affinities toward TK_{HSV1} have been previously presented (Champness et al., 1998; Prota et al., 2000; Alber et al., 1998; Perozzo et al., 2000). However, (Prota et al., 2000)the intriguing question posed by the recently published crystal structure of the inhibitor HBPG (Bennett et al., 1999) as to what is the structural basis for the different properties between inhibitor and substrate sharing the same binding mode remained open. This issue prompted us to perform a combined biochemical and quantum-chemical study. Two key parameters have been considered, on the one hand binding affinity and on the other hand catalytic turnover.
Because not all k_{cat} values were available and some conflicting kinetics results had been reported in the literature (Table 3.1), the catalytic activity of TK_{HSV1} towards all ligands have been measured under the same experimental conditions. In this way, a homogeneous set of measured k_{cat} values have been obtained and compared with the calculations.
The calculated sugar dipole points to Glu225 and the resulting stabilization energy correlates strikingly with the catalytic activity (Fig. 3.8). These results are consistent with a previously published study on E225L TK_{HSV1}. Indeed this mutant exhibits a loss in binding affinity proportional to the loss of one hydrogen bond and more interesting a dramatic drop in catalytic activity (Pilger et al., 1999).

Glu225 contributes to the binding affinity of the prodrug by H-bond interactions with the 3'OH group (Brown et al., 1995; Champness et al., 1998; Prota et al., 2000; Wild et al., 1995; Wild et al., 1997). Our calculations point to its important role in catalysis derived from its interaction with the dipole. In contrast no correlation have been found between the calculated terms and binding affinity expressed as K_i or K_M.
This type of calculations are very straight forward and can be extended in the future to new compounds.
Being electrostatic the dominating component of the sugar-like chain/protein

interactions and correlations, one might think about using molecular mechanics calculation for performing similar studies. We performed such calculation on some of the complexes and indeed the results are very promising (data not shown). These calculations however required the development of the force filed parameters of the prodrugs, which in turn required quantum chemical calculation. Therefore, the calculation effort using force field methodologies is not reduced compared to the quantum chemical calculation. On the other hand, the latter, being a parameter-free approach, allows an automatic procedure to study prodrug/enzyme interactions.

Our study provides also a rationale for the presence of polar group in the hydrophobic pocket of the enzyme (Fig. 3.5). Indeed, replacement of the O4' with a CH_2 group ether provides a dramatic reduction and change of orientation of the dipole (Fig 3.5b,c), which in turn could decrease the stabilizing dipole/Glu225 interactions.

In conclusion we indicated a substrate-enzyme molecular interaction which shows to be relevant to catalysis. We have shown a nice correlation between dipole-charge (substrate-Glu225) interaction and catalytic activity which candidates this electrostatic contribution as stabilizing mechanism for catalysis. Up to now all prodrugs have been the result of nucleobase or sugar moiety substitution with the only rational guide of mimicking the sugar moiety assuming the importance of elements such as the O4' but not knowing the exact role. With our work we suggest a guide element in the rational design of novel nucleosides analogs having an increased phosphorylating activity.

Chapter 4

Cation-π versus OH-π Interactions in Proteins

In this first chapter, structure and bonding of a cation-π and a facet OH-π adducts are investigated using density functional theory with gradient-corrections for the exchange-correlation functional. Our calculations are carried out for two specific model complexes representing i) the thymine/Arg 72 adduct in the ternary complex of HIV-1 reverse transcriptase (RT) with a DNA template primer and a deoxynucleoside triphosphate(Huang et al., 1998) ii) the Tyr6-Thr13 adduct in μ-gluthatione transferase (μ-GST)(Xiao et al., 1996). We find that electrostatic interactions play an important role and provide similar stabilization energies to the two π complexes. In HIV-1 RT, the π electronic density of thymine is essentially uneffected by the presence of the arginine guanidium group; on the contrary, tyrosine is significantly polarized by the interaction with the hydroxyl group and other groups present in the μ-GST enzyme. The influence of Thr13 induced-polarization on Tyr6 pKa is compared with that of other interacting groups at the active site.

4.1 Introduction

The role of on-face interactions of π-rings for molecular recognition processes and protein folding is now widely recognised(Burley and Petsko, 1986; Scrutton and Raine, 1996; Shoemaker et al., 1990; Dougherty, 1996; Ma and Dougherty, 1997; Stauffer et al., 1990; McCurdy et al., 1992; Shi et al., 1994; Poralla et al., 1994; Abe and Prestwich, 1995; Poralla, 1994; Barak et al., 1997; Ermer and Eling, 1994; Allen et al., 1996; Suzuki et al., 1992; Rodham et al., 1993; Kashino et al., 1988; Waksman et al., 1992; Atwood et al., 1991; Jedrzejas et al., 1995; Ji et al., 1992). Aromatic groups interact with charged side-chains in cation-π complexes(Burley and Petsko, 1986; Scrutton and Raine, 1996; Shoemaker et al., 1990; Dougherty, 1996; Ma and Dougherty, 1997; Stauffer et al., 1990; McCurdy et al., 1992; Shi et al., 1994; Poralla et al., 1994; Abe and Prestwich, 1995; Poralla, 1994; Barak et al., 1997; Poralla, 1994) and, more rarely(Barak et al., 1997), with NH(Ermer and Eling, 1994; Allen et al., 1996; Suzuki et al., 1992; Rodham et al., 1993; Kashino et al., 1988; Waksman et al., 1992; Atwood et al., 1991) or OH(Jedrzejas et al., 1995; Ji et al., 1992) H-bond donors in facial H-bonded complexes.

Ab initio quantum chemical calculations at the Hartree-Fock (HF)(Mecozzi et al., 1996a; Mecozzi et al., 1996b), HF-MP2(Minoux and Chipot, 1999; Duan et al., 2000; Tarakeshwar et al., 1999; Liu et al., 1993) and B3LYP(Caldwell and Kollman, 1995) levels of theory have provided useful insights on the nature of the binding. Stabilization of the complexes is believed to arise from electrostatics interactions(Mecozzi et al., 1996b; Mecozzi et al., 1996a; Duan et al., 2000; Caldwell and Kollman, 1995), involving the π-ring quadrupole and either the positive charge (in cation-π complexes)(Caldwell and Kollman, 1995) or the NH or OH dipoles (in facial H-bonds). Additional non-electrostatic terms, from charge polarization and donation(Minoux and Chipot, 1999; Caldwell and Kollman, 1995; Schneider, 2001; Schneider et al., 1992) (in cation-π complexes) to covalent interactions, exchange forces(Tarakeshwar et al., 1999) (in facial H-bonds) may also play a significant role.

Here we attempt at dissecting key facets of cation-π and facial H-bond interaction properties by performing first principles quantum-chemical calculations on two representative examples of cation- and OH-π complexes present in enzymatic active sites. The first is the Arg17–thymine complex (Fig. 4.1a) in human immunodeficiency virus reverse transcriptase (HIV-1 RT)(Huang et al., 1998). The second is the Thr13-Tyr6 complex in μ-glutathione S-transferase (μ-GST)(Ji et al., 1992; Xiao et al., 1996) (Fig. 4.2b), a class of enzymes catalyzing the nucleophilic attack of glutathione anion sulfur (GS-) to a variety of aromatic compounds.

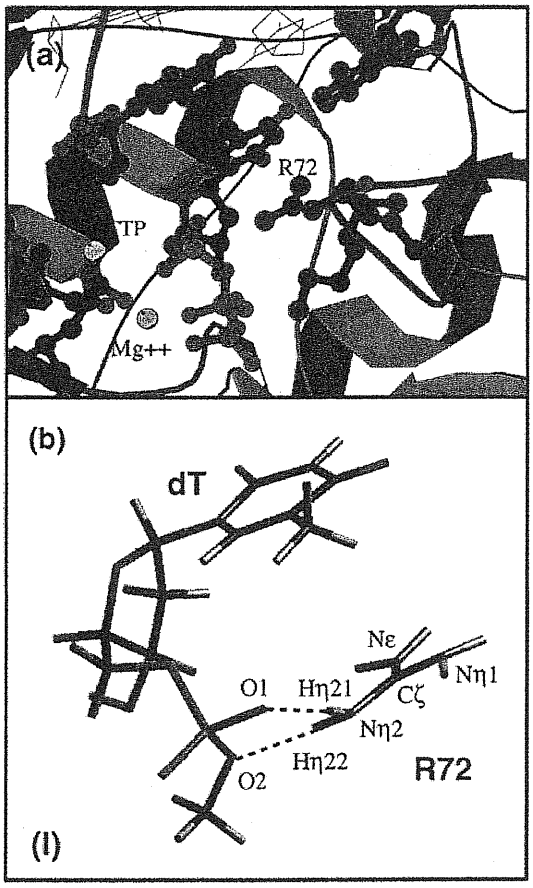

*Figure 4.1. Thymine binding site in HIV-1 RT(Huang et al., 1998) (Figure made with the Molscript program(Kraulis, 1991; Kraulis, 1997) (a)) and (b) complex **I** used in the calculations.*

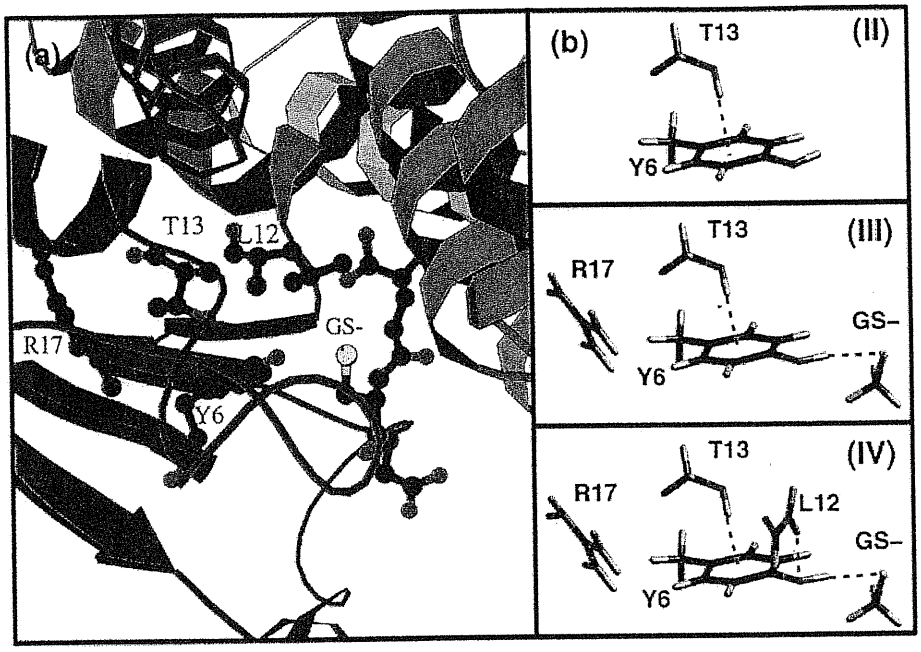

Figure 4.2. *Active site of μ-GST – GS⁻ complex(Xiao et al., 1996) (Figure made with the Molscript program(Kraulis, 1991; Kraulis, 1997) (a)) and (b) model complexes* ***II-IV*** *used in the calculations.*

Calculations are carried out at the *same level of theory* so as to allow a meaningful comparison of structural and electronic properties of the two systems. We use gradient-corrected density functional theory (DFT) with planewaves basis-set, which has already revealed itself as a powerful tool to describe structure and bonding of rather large complexes representing substrate/enzymes adducts(Carloni et al., 1995; De Santis and Carloni, 1999; Rothlisberger et al., 2000; Rothlisberger and Carloni, 1999; Segall et al., 1998a; Segall et al., 1998b; Segall et al., 1998c; Segall et al., 1999; Alber, 1998; Pantano et al., 2000; Molteni et al., 1999; Armstrong, 1998).

4.2 Calculation details

4.2.1. Model Complexes.

Complex **I** (Fig. 4.1 (b)) was built from the ternary catalytic complex of HIV-1 with a substrate (dTTP) and the DNA-primer and template(Huang et al., 1998) (entry 1RTD in the Brookhaven Database(Bernstein et al., 1999)) (Fig. 4.1). It included the Arg72 guanidinum group and deoxy-thymidine triphosphate, modeled as deoxy-thymidine

monophosphate.

Complexes **II-IV** (Fig. 4.2 (b)) were built starting from the X-ray structure of rat μ-GST (EC 2.5.1.18) complexed with GS^- at 2.20 Å resolution(Xiao et al., 1996) (entry 6GST in the Brookhaven database(Bernstein et al., 1999)) (Fig. 4.2). The complexes included the following groups: Tyr6, Thr13 (**II**), Tyr6, Thr13, Arg17, GS^- (**III**), Tyr6, Thr13, Arg17, GS^-, Leu12, (**IV**). The latter groups were modeled as cresol, methanol, methylthiolate, guanidinium and formamide, respectively[2]. Tyr6 hydroxyl group and glutathione were assumed protonated and deprotonated, respectively, as suggested by kinetic solvent deuterium isotope(Huskey et al., 1991) and by MP2 *ab initio* calculations(Zheng and Ornstein, 1997).

4.2.2 Computational Details.

The quantum problem was solved within the framework of DFT. Becke, Lee, Yang and Parr (BLYP)(Becke, 1988; Lee et al., 1988) gradient corrections were used for the exchange-correlation functional. The basis set for the valence electrons consisted of plane-waves expanded up to a cutoff of 70 Ry. The interaction between valence electrons and ionic cores were described by norm-conserving pseudopotentials of the Martins-Troullier type(Troullier and Martins, 1991). Our models were treated as isolated systems as in ref. (Barnett and Landman, 1993). Complexes **I**, **II**, **III**, **IV** were inserted in orthorombic simulation cells of lattice parameters 11.0 x 10.0 x 11 Å, 11.5 x 12.0 x 10 Å , 14.0 x 16.0 x 13.0 Å, 15.0 x 18.2 x 12.8 Å, respectively. All non-hydrogen atoms were kept fixed to the crystallographic coordinates as in Ref (Liu et al., 1993). Geometry optimization was performed using the direct inversion iterative subspace (DIIS) method(Pulay , 1980) for both electronic and ionic degrees of freedom.

Some calculations included the external electric potential Φ of the μ-GST– GS^- / water system. Φ was added up to the external potential due to the nuclei-electron interactions. In atomic units, $\Phi(r) = \Sigma_i q_i /(r_i-r)$, where q_i are the RESP(Bayly et al., 1993; Cornell et al., 1995) atomic point charges at the atomic position r_i.

All calculations were performed with a parallel version- of the Car-Parrinello(Car and

[2] Arg17 guanidinium group was also included in II-IV to ensure electroneutrality to these systems.

Parrinello, 1985) code developed by J. Hutter et al.(Hutter et al., 1996b).

4.2.3 Calculated Properties.

(i) ***Polarization effects.*** Shifts of electron density throughout the entire regions of the complexes ($\Delta\rho$) were calculated as the differences between the densities of the complexes and those of the fragments forming the complexes(Gu and Scheiner, 1999). The resulting integrated charges on single groups (Δq) were calculated by integrating positive and negative contributions of the charge on spheres centered on nonhydrogen atoms with a 2 a.u. radii.

(ii) ***Polarity of chemical bonds.*** The geometrical analysis of centers of maximally localized Wannier functions (WFC)(Silvestrelli et al., 1998; Silvestrelli and Parrinello, 1999b; Marzari and Vanderbilt, 1997) is useful to investigate the polarity of a chemical bond, as shifts of WFC along the chemical bond correlate to differences of Pauling electronegativities(Hehre et al., 1986) ($\Delta\chi$) between the atoms forming the bond(Alber et al., 1999). In this paper, the WFC analysis has been extended to the determination of changes in polarity of the chemical bonds upon formation of the intermolecular complexes. To illustrate our procedure, let us suppose that a chemical bond between atoms X and Y is in two different chemical environments, labeled 0 and 1. Then *an increase* of $\Delta\Delta\chi(X\text{-}Y) = \Delta\chi(X\text{-}Y)_1 - \Delta\chi(X\text{-}Y)_0$ indicates that polarization toward atom X has *increased* on passing from 0 to 1.

(iii) ***Electrostatic energy.*** Interaction energies between two moieties 1 and 2 (thymine and arginine guadininium in **I** and threonine and tyrosine in **II-IV**) were calculated as a multipole expansion up to the sixth term:

$$E = E_{charge(1)\text{-}charge(2)} + E_{charge(1)\text{-}dipole(2)} + E_{dipole(1)\text{-}charge(2)} + E_{dipole(1)\text{-}dipole(2)} + E_{charge(1)\text{-}quadrupole(2)} + E_{quadrupole(1)\text{-}charge(2)} =$$

$$\xi_1\xi_2/(4\pi\varepsilon_0 R)+(\xi_1\,\mu_2\cdot\mathbf{R})/(4\pi\varepsilon_0 R^3)+(\xi_2\,\mu_1\cdot\mathbf{R})/(4\pi\varepsilon_0 R^3)+((\mu_1\cdot\mu_2)/R^3-3(\mu_1\cdot\mathbf{R})(\mu_2\cdot\mathbf{R})/R^5)/(4\pi\varepsilon_0)+\xi_1 Q_2(3\cos^2\theta_1-1)/(8\pi\varepsilon_0 R^3)+\xi_2 Q_1\,(3\cos^2\theta_2-1)/(8\pi\varepsilon_0 R^3).$$

Where: **R** is the distance vector between the centers of charges of the two moieties; ξ_1 and $\xi_2$are the charges of moieties 1 and 2; θ_1 and θ_2 are the angles defined by **R** and the quadrupole eigenvectors; μ_1 and μ_2 are dipole momenta for the moiety 1 and 2. They are calculated from the electron density. $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are quadrupole 3x3 tensors (calculated as: $Q_{xx}=\Sigma_j q_j x_j x_j$; $Q_{xy}=\Sigma_j q_j x_j y_j$; $Q_{xz}=\Sigma_j q_j x_j z_j$; $Q_{yy}=\Sigma_j q_j y_j y_j$; $Q_{xz}=\Sigma_j q_j x_j z_j$; $Q_{zz}=\Sigma_j q_j z_j z_j$) where q_J are the RESP charges(Bayly et al., 1993; Cornell et al., 1995) of atom **j** located at position $\mathbf{r}_j=(x_j,y_j,z_j)$[3] (see ref.(Leach, 1996)).

As the expansion is limited to large distances between two interacting groups, this model cannot provide quantitative results. Indeed, it is used here exclusively for a comparison between the two systems (RT and μ-GST active sites) which have similar geometries.

It is well established that in cation-apolar aromatic complexes only the charge-quadrupole term plays a role(Ma and Dougherty, 1997). Even though thymine has a non-negligible permanent dipole, the charge-quadrupole term is expected to provide by far the leading contribution in complex I: simple symmetry considerations suggest that the dipole-charge contribution must be small (Fig. 4.3). Consistently, explicit calculations of $E_{dipole\text{-}charge}$ shows that this is about one order of magnitude smaller than $E_{charge\text{-}quadrupole}$. In complex II-IV, $E_{charge\text{-}charge}= E_{dipole\text{-}charge} =E_{quadrupole\text{-}charge} =0$. Thus, the largest contribution is that due to dipole-dipole interactions.

[3] Qundrupole moments of apolar aromatic rings may be rather underestimated in DFT(Meijer and Sprik, 1998a). Test calculations of the quadrupole moment at the 6-31G* Hartree-Fock – MP2 level (using the Gaussian98 program)(Frisch, 1995) provided similar risults and the correspondent quadrupole/charge terms calculated with DFT and MP2 agree within 20 %.

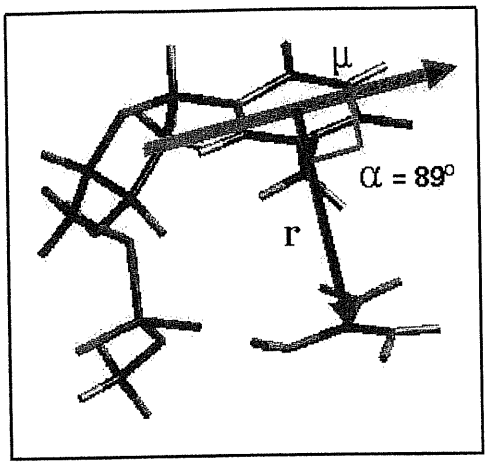

Figure 4.3 *Charge-dipole interactions in complex I: thymine ring dipole (μ) is practically orthogonal to the position vector* ***r*** *connecting thymine and Arg 72 centers of charges.*

4.3 Results

In this section, we investigate cation- and OH- π interactions in two specific systems, the Arg72-thymine cation-π complex in HIV-1 RT (Fig. 4.1) and the Thr13-Tyr6 OH-π complex in μ-GST (Fig. 4.2). Comparison is also made, within the density functional theory framework, between facial and standard H-bonds.

4.3.1 The Arg72-thymine complex in HIV-1 RT

Structural Properties. Geometry optimization of complex **I** was performed keeping all non-hydrogen atoms fixed to the crystallographic coordinates. As a test calculation, complex **I** was also *fully* geometry-optimized (maximum gradient on the nuclei $5 \cdot 10^{-5}$a.u.). The results, in terms of chemical bonding, charge polarization and electrostatics (i.e. the quantities calculated in the text), exhibit negligible differences with those of the *partially* optimized structure. For istance, the total integrated polarization charge Δq (see Calculated Properties section below) are 0.61 and 0.65 electrons for the partially and fully optimized structures, respectively. Fig. 4.4 points out the structural similarities between fully- and partially- optimized structures.

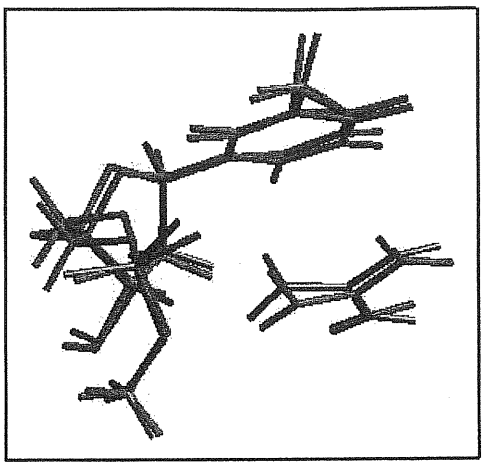

*Fig. 4.4. Comparison between fully (blue) and partially (red) geometry-optimized structures of complex **I**. In the partially optimized structure, only the hydrogens atoms have been energy-minimized.*

The arginine residue lies flat in a stacking conformation with a thymine ring in a rather common pattern of arginine - π adducts(Ma and Dougherty, 1997) (Fig. 4.1a) (minimum Arg/thymine distance ~ 4 Å). Arg72 guanidinium forms also two standard H-bonds to the phosphate group (d(Hη21-O1)=1.65 Å, d(Hη22-O2)=1.90 Å). Hydrogen atoms of the guanidinium group lose their planarity so as to optimize the interaction (Fig. 4.1b) (maximum deviation of arginine from planarity 29°). Similar structural features have been found in other contextes(Hutter et al., 1996a).

Binding. A simple electrostatic model indicates that quadrupole/charge interactions stabilize the complex by ~1.5 kcal/mol. In this system, even if thymine ring dipole (μ) is relevant, it is practically orthogonal to the position vector **r** connecting thymine and Arg 72 centers of charges and so the dipole-charge interaction is negligible (see Fig. 4.6).

Quantitative estimation of electrostatic interactions is not possible as the domain of validity of the model is limited to large distances between the two moieties. Similar contributions may be expected also for Phe- and Tyr-Arg complexes characterized by similar binding mode and Arg-π ring distance.

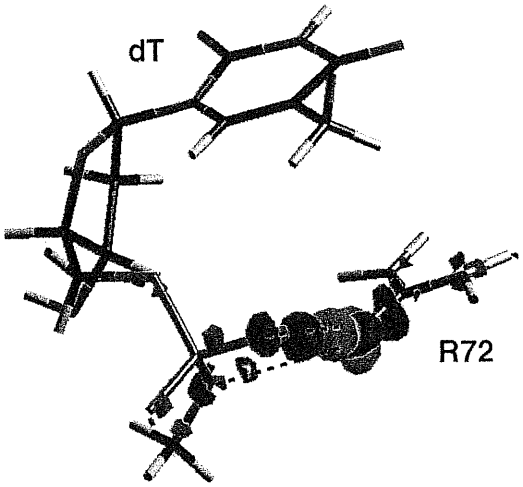

Figure 4.5. *The cation-π complex: Electronic density difference map for complex* ***I.*** *Green regions represent loss of electronic density as a result of formation of the complex, relative to the isolated moieties. Magenta regions refer to increased density. The contour shown is 0.005 e^-/au^3.*

The thymine ring is not polarized. Fig. 4.5, which plots the density difference upon formation of the complex, shows that polarization effects are sizeable only for the Arg72-phosphate moiety.

The phosphate group affects dramatically the polarity of arginine NH bonds involved in H-bonding: the calculated electronegativity difference ($\Delta\chi$) on passing from isolated fragments to the complex are as large as $\Delta\chi$(Nη2-Hη22)=0.17 and $\Delta\chi$(Nη2-Hη21)= 0.67, that is, on the same order of magnitude of the change of polarity between the CH in methane to the N-H bond in ammonia(Alber et al., 1999).

While electrostatics interactions play a fundamental role, covalent contributions appear to be negligible, as evinced by an analysis of the orbital characters.

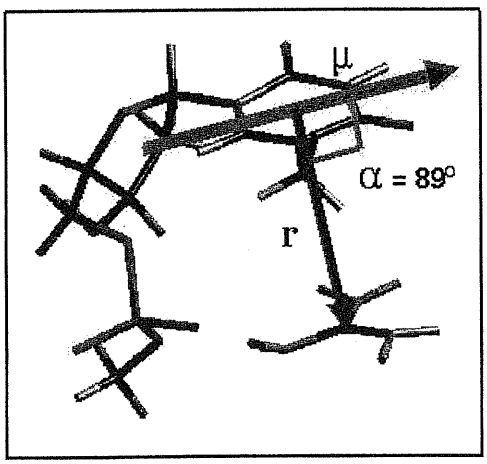

Fig. 4.6. *Charge-dipole interactions in complex I: thymine ring dipole (μ) is practically orthogonal to the position vector* ***r*** *connecting thymine and Arg 72 centers of charges.*

4.3.2 The Thr13-Tyr6 complex in μ-GST

Structural properties. In the active site of the enzyme-substrate complex (Fig. 4.2a), Thr13 hydroxyl group is believed to form a facial H-bond with the Tyr6 ring(Liu et al., 1993), but the precise orientation of the OH group is not known. In our structure with optimized hydrogens positions, the hydroxyl group forms an angle of 59° with the ring plane and it points toward the inner region of the tyrosine ring close to the carbons in *ortho* position with respect of the OH group (Fig. 4.7a). Its distance to the Tyr6 ring is short (2.15 Å).
Tyr6 interacts also with other groups in the active site, by forming conventional H-bonds with the substrate (GS-) and with Leu-12 (Fig. 4.2).

Binding. The largest contribution to electrostatics is provided by the interaction between the Try6 and Thr13 dipoles, which turns out to be oriented in an almost completely antiparallel fashion (Fig. 4.7a). Because of the orientation of the dipoles and of the proximity of the OH group to Tyr6, the dipole/dipole stabilization energy turns out to be as large as ~ 1kcal/mol, that is of the same order of (albeit smaller than) quadrupole-charge interactions in cation-π complexes. In contrast, the Tyr6 quadrupole-Thr13 dipole term is very small and does not contribute significantly to the overall interaction. Indeed, this term is a higher order in the multipolar expansion and decreases with a higher power of distance **r** than the dipole/dipole term (see computational section).

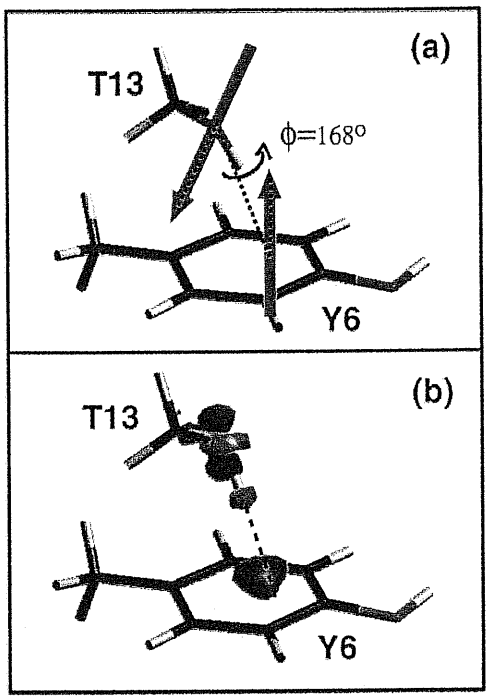

Figure 4.7. *The OH-π complex: (a) Electrostatic interactions: orientation of Thr13 and Tyr6 dipoles (red arrows, calculated values 1.7 and 1.3 D, respectively). (b) Electronic density difference map for complex* ***II****. Green regions represent loss of electronic density as a result of formation of the complex, relative to the isolated moieties. Magenta regions refer to increased density. The contour shown is 0.0015 e^-/au^3 .*

Thus, we conclude that the complex is stabilized by favorable dipole-dipole interactions, comparable in strength with quadrupole/charge interactions in cation-π complexes. These findings may provide a rationale for previous MP2 calculations, which indicated that energetics depend dramatically on the orientation of the threonine hydroxyl group(Liu et al., 1993).

While no charge transfer is observed, the aromatic ring experiences significant polarization effects, its integrated charges being ~ 0.01 e^-; this may be caused by the short distance between the threonine hydroxyl group and the tyrosine ring. Inclusion of groups interacting with Tyr6 (the GS- substrate anion and Leu12 backbone unit (complex **IV**) increases these polarization effects (Tyr6 charge ~0.08 e^-).

To further investigate polarization on the π-ring induced by Thr13 and the other interacting groups, we calculate the changes in polarity of the Tyr6 aromatic CH bonds upon formation of the active site complex. Polarity changes are calculated based on the maximally localized Wannier Functions (see computational section), which by definition are not suited to describe delocalized bonds. Thus, our analysis cannot be extended to tyrosine ring CC CO and OH bonds, which are partially or totally delocalized.

Model Complex	Δχ(CH)
II in vacuo	0.0026
III in vacuo	0.0030
IV in vacuo	0.0037
IV with electric field of residues within 8 Å from Tyr6 center of charge	0.0075
IV with electric field of the entire protein	0.0064

***Table 4.1**. The OH-π complex: Shifts of Pauling electronegativity(Hehre et al., 1986) Δχ(CH) as a result of formation of the OH -π complex, relative to isolated Tyr6.*

Table 4.1 shows that on binding to Thr13, these are more polarized toward the nonhydrogen atom relative to the gas phase. Additional polarization is achieved by including also the H-bonds with GS- and the Leu12 backbone unit (Table 4.1).
The Table shows that also other residues at the active site play a significant role, whereas long-range effects are small: the polarization of the residues within a sphere of 8 Å radius from Tyr6 center of mass are even larger than the total polarization. An important contribution of the strong polarization is likely to be due to Glu29, which is the only charged group within the sphere radius.
We conclude that the OH facial H-bond does polarize the ring, and that interacting groups at the active site dramatically increase polarization.

Also in this case, analysis of the orbital character suggests that covalent contributions do not play a significant role.

4.4 Discussion and Conclusions

In this study we have used first principles calculations to characterize the electronic properties of the specific OH- and cation- π complexes.
In the OH-π complex, our multipole expansion electrostatic model points to the previously unrecognized crucial role of Thr13-dipole/Tyr6-dipole interactions. The dipole-dipole stabilization energy term turns out to be on the same order of magnitude of the most important term in the cation-π complex, that is the charge/quadrupole contribution (~1kcal/mole). This is consistent with the fact that, in the multipole expansion, the two terms are expected to be on the same order of magnitude. As DFT fails in reproducing dispersion attractions, our study cannot provide information on this

contribution(Krystian and Pulay, 1994)
The charge/quadrupole contribution appears considerably larger than the dipole/dipole interaction, yet, because of the limitations of our model (see computational section) energies can only be approximations of the complex set of interactions that take place within a protein molecule. For this reason, only qualitative trends can be sought.

Novel information is also obtained on polarization effects. While these do not play a significant role for cation π interactions, as they are relatively large upon formation of the OH--π adduct, possibly as a consequence of the close proximity of the OH group to the aromatic ring. In contrast, charge donation is instead not observed. Standard H-bonding (Fig. 4.6) involves even larger polarization effects than the facial H-bond. Electron density difference maps of CH---O(Gu and Scheiner, 1999) have shown that also this type of H-bonding involves rearrangement of the charge.

Our calculations provide further support to the proposal that Thr13 acts as electron-withdrawing group toward the aromatic system(Liu et al., 1993), as the OH dipole polarizes the aromatic CH bonds toward the carbon atoms (Table 4.1). Polarization is increased if the substrate and other groups interacting with Tyr6 are included (Table 4.1). Thus, the presence of Thr13 is expected to *stabilize* the tyrosinate anion relative to its conjugate base, consistent with previous MP2 calculations, which point to a significant increase of GS^- proton affinity upon Thr13 binding to Tyr6(Liu et al., 1993). Thus, these and the previous calculations provide independently a rationale for the decrease of GS^- pKa in mutants on the 13 position where the OH group has been eliminated (T13 μ-GST A, T13 μ-GST V, T13 μ-GST S)(Liu et al., 1993). Furthermore, our calculations suggest that several other groups present at the active site are also very important for lowering Tyr6 pKa, from GS^- (as previously proposed(Zheng and Ornstein, 1997; Liu et al., 1993)) to Leu12 amide group and Glu29. We conclude that the protein appears to be engineered so as to enhance Tyr6 acidity.

In summary, several conclusions can be drawn by the present study. First, electrostatic contributions are comparable in the two complexes. Second, the π electronic density is significantly polarized (though less than a standard H-bond) by the interaction with the OH group but it is not by interacting with the σ charged group. Finally, in the OH-π complex, protein environment is crucial for charge polarization and Tyr6 acidity.

Chapter 5

Reaction Mechanism of Caspases: Insights from QM/MM Car-Parrinello Simulations

Caspases are fundamental targets for pharmaceutical interventions in a variety of diseases involving disregulated apotosis. Here, we present a QM/MM Car-Parrinello study of key steps of the enzymatic reaction for a representative member of this family, caspase-3. The hydrolysis of the acyl-enzyme complex is described at the density functional (BLYP) level of theory while the protein frame and solvent is treated using the GROMOS96 force field. These calculations show that the attack of the hydrolytic water molecule implies an activation free energy of ca. $\Delta F_A \sim 20$ kcal/mol in good agreement with experimental data, and leads to a previously unrecognized gem-diol intermediate that can readily ($\Delta F_A \sim 5$ kcal/mol) evolve to the enzyme products. Our findings assist in elucidating the striking difference in catalytic activity between caspases and other structurally well-characterized cysteine proteases (papains and cathepsins) and may help in the design of novel transition-state analog inhibitors.

5.1 Introduction

Several strategies for therapies against diseases involving disregulated apoptosis (such as Alzheimer's(Shimohama, 2000), Parkinson's(Jordan et al., 2000), and cancer(Kaufmann and Gores, 2000)) are based on inhibition of caspases(Talanian et al., 2000), a family of cysteine proteases involved in all apoptosis pathways(Salvesen and Dixit, 1997). Twelve different caspases have been characterized so far(Talanian et al., 2000). Although the activity and specificity patterns of these enzymes are clearly distinct(Ventimiglia et al., 2001), their overall reaction mechanism is expected to be highly analogous(Wilson et al., 1994). In fact, all caspases recognize specific four-residue sequences and they cleave with essentially no exception peptide bonds located after an Asp group (Fig. 5.1). In addition, the 3D structures of caspases determined so far (caspases 1(Rano et al., 1997; Okamoto et al., 1999; Wei et al., 2000; Huang et al., 2001),3(Rotonda et al., 1996; Mittl et al., 1997; Lee et al., 2000; Riedl et al., 2001), 7(Wei et al., 2000; Huang et al., 2001), and 8(Watt et al., 1999; Blanchard et al., 1999; Xu et al., 2001)) are highly homologous and structurally alike(Walker et al., 1994) (Fig. 5.2a). Finally, the conformational properties of the two catalytically relevant residues (a Cys and a His residue(Fersht, 1997)) are very similar (Fig. 5.2a).

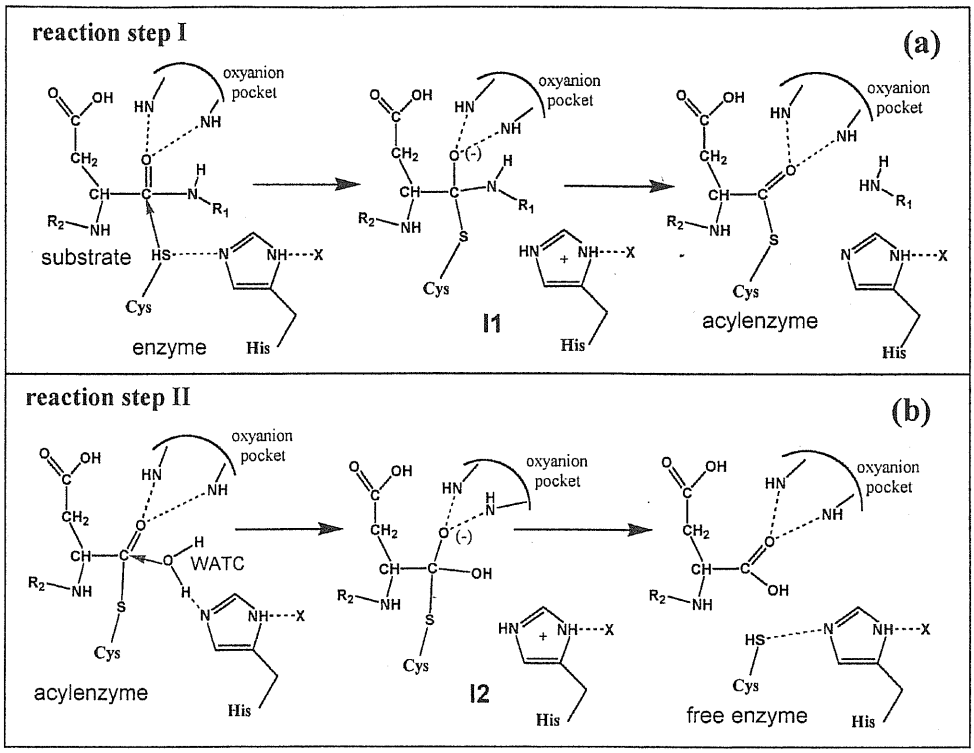

Figure 5.1. The reaction mechanism of peptide hydrolysis catalyzed by caspases.(Wilson et al., 1994) ***(a)*** *Reaction step I, nucleophilic attack of the catalytic Cys to the substrate backbone leading to the formation of the acyl-enzyme complex. The peptide bond which is cleaved follows an Asp group.* ***(b)*** *Reaction step II, nucleophilic attack of WATC to the acyl-enzyme leading to the products.* ***I1*** *and* ***I2*** *are the reaction step intermediates for the two steps of the reaction. In caspase-3, the oxyanion pocket is constituted by the amid-protons of Gly238 and Cys285; the catalytically relevant residues are Cys285 and His237; the H-bond acceptor group X is the Thr177 backbone(Rotonda et al., 1996).*

Since it is a successful strategy to design new selective inhibitors based on transition state analogs(Liu et al., 1996), it is of importance to fully understand the details of the enzymatic reaction and in particular, to characterize the structural and electronic properties of the activated complexes. So far, the mechanistic aspects of the enzymatic reaction of caspases have solely been inferred from indirect evidence drawn from structural information on enzyme/inhibitor complexes. The postulated mechanism is a straightforward adaptation of the generic protease picture(Fersht, 1997) (Fig. 5.1). In the first step, the thiol group belonging to the catalytically active Cys residue (Fig. 5.1a) is deprotonated by a histidine acting as a general acid. The resulting thiolate species attacks the carbonyl carbon of the aspartic peptide bond forming a first tetrahedral intermediate (**I1** in Fig. 5.1a). The oxyanion group of **I1** is believed to be stabilized through additional hydrogen-bond interactions with two backbone amide groups, constituting the so-called oxyanion pocket. Protonation of the α-amide nitrogen by the catalytic histidine leads to the cleavage of the peptide bond, release of the C-terminal

product fragment and the formation of an acyl-enzyme complex. In the second step of the reaction, a water molecule from the solvent forms an H-bond with the catalytic His (WATC in Fig. 5.1b). WATC performs a nucleophilic attack on the carbonyl carbon of the acyl-enyme complex forming a second tetrahedral intermediate (**I2**). The reaction proceeds by protonation of the covalently linked cysteine and subsequent release of the free sulfhydryl group and the final carboxy product (Fig. 5.1b). During the reaction, the acidity of the catalytically relevant histidine may be enhanced through interactions with H-bond donor(s) (X in Fig. 5.1), potential third members of the catalytic triad(Fersht, 1997).

To provide a detailed description of key steps of the enzymatic reaction, we have undertaken a theoretical study on one exemplary member of the caspase family, the downstream caspase-3. This enzyme is especially amenable for such an investigation since it acts as one of the central death executioner and as such is involved in virtually every model of apoptosis.(Porter and Janicke, 1999) It is also implicated in a large number of human diseases for which either excessive (such as in the case of ischemic damage and neurodegenerative disorders) or insufficient (e.g. for cancer and autoimmune diseases) apoptosis occurs(Talanian et al., 2000). Caspase-3 exhibits a large substrate diversity, as it cleaves a variety of proteins involved in cell maintenance and/or repair, among them poly(ADP-ribose) polymerase, p21-activated kinase 2, gelsolin, and DNA-dependent protein kinase.(Cohen, 1997; Liu et al., 1997; Enari et al., 1998) Finally, it is one of the best biochemically and structurally characterized isoenzymes. In particular, the crystal structure of the human isoenzyme complexed with the aldehyde tetrapeptide Ac-DEVD-CHO (Ac = acetyl group) is available(Rotonda et al., 1996). The structure is similar to the acyl-enzyme intermediate of Fig. 5.1b. Thus, it offers a particularly suitable starting point for an investigation of the second reaction step, namely the nucleophilic attack of WATC to the acyl-enzyme complex. In fact, this structure contains the natural substrate recognition sequence (DEVD) covalently bound to Cys285 with only minor differences with respect to the acyl form of the enzyme.
The theoretical tool used in this investigation is a mixed quantum mechanical/molecular mechanical (QM/MM) Car-Parrinello(Car and Parrinello, 1985) approach, in which the essential groups participating in the reaction are treated using an electronic structure method. The rest of the protein is described within the framework of a classical force field description (Laio et al., 2001). In this way, the forming and breaking of chemical

bonds during the reaction can be simulated directly within a fully dynamical description that takes explicit account of the steric and electrostatic effects of the protein environment. The free energy of the process is calculated using a thermodynamic integration technique(Ciccotti et al., 1989). The same method has previously been applied successfully for a variety of enzymatic reactions(Curioni et al., 1997; Meijer and Sprik, 1998b; Carloni et al., 2000; Piana et al., 2001).

Our simulations show that the attack of the hydrolytic water molecules involves an activation barrier of ca. 20 kcal/mol that is compatible with experimental measurements(Stennicke and Salvesen, 1997). However, at variance, with the postulated textbook mechanism, the hydrolysis proceeds via a gem-diol intermediate that readily forms the final products. Thus, our calculations provide structural models of the involved transition states and intermediates, which in turn might serve as possible templates for the design of new specific inhibitors for caspase-3. Furthermore, our investigation helps to elucidate different mechanistic aspects between caspases and other cysteine proteases of the papain family, such as papain(Drenth et al., 1968; Kamphuis et al., 1984) itself, which has been the first to be crystallized and other cysteine proteases of interest as pharmacological target, such as cathepsins B(Musil et al., 1991; Jia et al., 1995; Turk et al., 1995; Yamamoto et al., 2000), L(Coulombe et al., 1996; Guncar et al., 1999), K(Zhao et al., 1997; Yamashita et al., 1997; Thompson et al., 1997), H(Guncar et al., 1998), and S(Fengler and Brandt, 1998); Trypanosoma Cruzi enzyme cruzain(Gillmor et al., 1997; Brinen et al., 2000); bleomycin hydrolase and deubiquitinating enzyme(Johnston et al., 1997; Johnston et al., 1999).

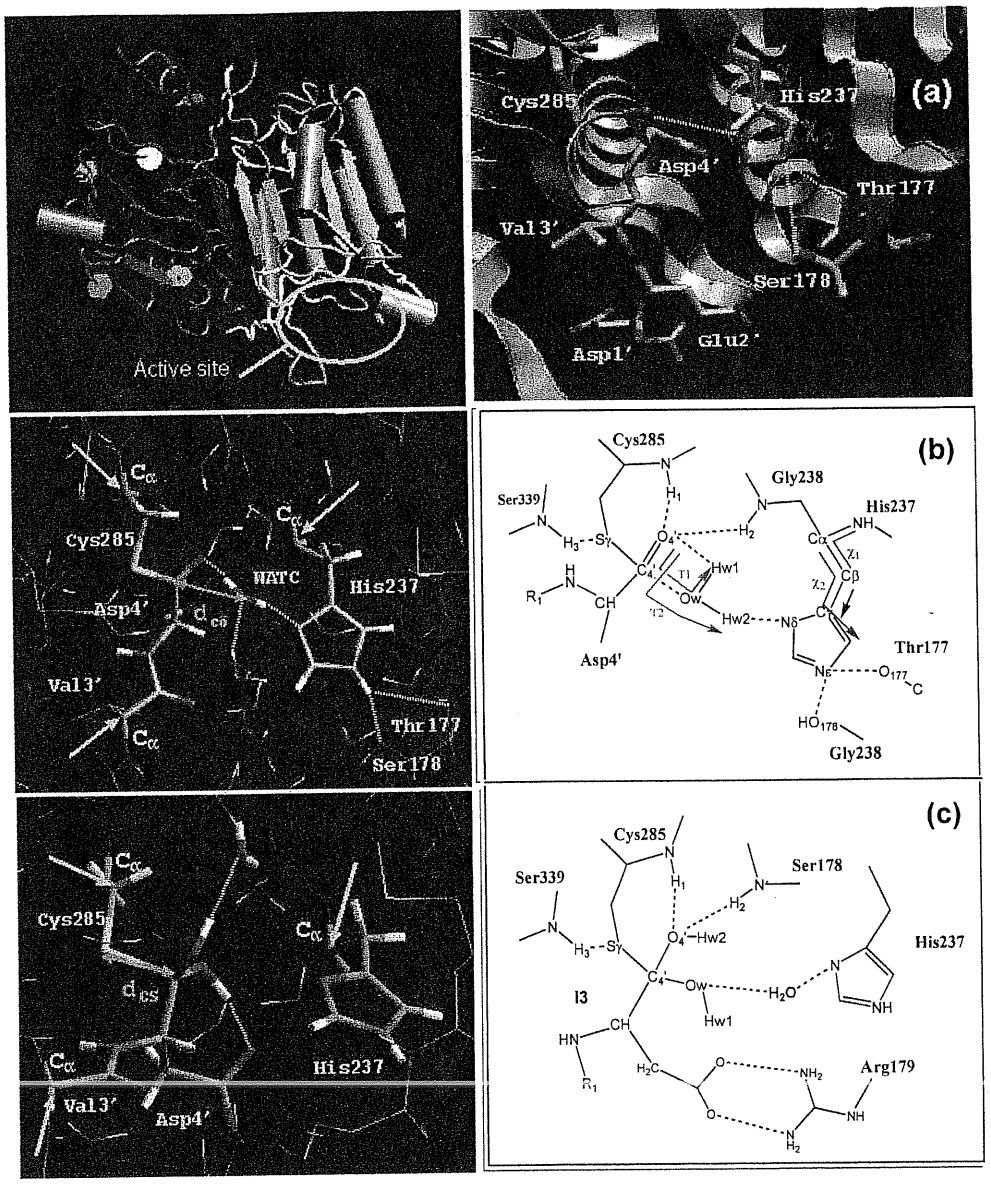

Figure 5.2. ***(a)*** *Caspase 3/Ace-DEVD-COH inhibitor 3D structure(Rotonda et al., 1996). Left: The global folding, which is identical to that of all the other caspases whose structure has been determined (caspases 1(Rano et al., 1997; Okamoto et al., 1999; Wei et al., 2000; Huang et al., 2001),3(Rotonda et al., 1996; Mittl et al., 1997; Lee et al., 2000; Riedl et al., 2001), 7(Wei et al., 2000; Huang et al., 2001), and 8(Watt et al., 1999; Blanchard et al., 1999; Xu et al., 2001)). The enzyme consists of two identical subunits composed of two chains, which are folded in a compact cylinder. The fold is dominated by a central core of a six–stranded β-sheet flanked by two helices on one side and three on the other. The active site region is indicated with a yellow circle. Right: Active site, the groups relevant to catalysis (Cys285, His237, Ser178 and Thr177, substrate) are reported as cylinders. The χ_1 and χ_2 conformational torsion angles of His237 are also displayed. In all the inhibitor/caspase complex structures, d(Cys285-S_γ-His237-N_δ) ranges from 5.6 Å(Rotonda et al., 1996) to 6.3 Å(Mittl et al., 1997) while the Ramachandran angles of His237 remain around χ_1~ -80° and χ_2 ~110° throughout.* ***(b)*** *Model of the acyl-enzyme/substrate complex. Left: 3D Structure and chosen reaction coordinate for the hydrolysis reaction. Right:: Atom labeling, H-bond interactions and T1 and T2 torsion angles. The last provide WATC orientation respect to the acyl-enzyme carbonyl.* ***(c)*** *Model of intermediate* ***I3*** *as obtained from QM/MM simulations. Left: 3D structure and constraint coordinate for the decomposition of the intermediate; Right: labels and H-bond interactions.*

In ***(b)*** *and* ***(c),*** *QM and MM atoms are drawn as cylinders and sticks, respectively. The C_α (Cys285), C_α (His237) and C_α (Val3') are boundary atoms (described with monovalent pseudopotentials) at the QM/MM interface indicated with an arrow.*

5.2 Computational details

5.2.1 Structural Model for Reaction Step 2.

The structure of caspase-3/Ac-Asp1'Glu2'Val3'Asp4'-CHO(Rotonda et al., 1996) presents the fold typical for caspases (Fig. 5.2a). It also contains the preferred recognition motif of caspase-3 (DEVD). It differs from the acyl complex of Fig. 5.1b only by two features, the replacement of carbonyl group of Asp4' with a CHO group and the absence of the catalytic water (WATC in Fig. 5.1b). Thus, to construct our initial model of the acyl-enzyme (Fig. 5.2b), the hybridization of the Asp4'-C is changed into *sp2* so that the Asp4'-CO bond points towards the oxyanion pocket, which is formed by the amide nitrogen atoms of Gly238 and Cys285. In this way, the orientation of Asp4'-CO is assumed to be in the most likely arrangement for the deacylation reaction to occur(Wilson et al., 1994) (Fig. 5.2b). Furthermore, the water molecule WATC is added in such a way as to form H-bond interactions with Asp4'- and His237.
Hydrogen atoms, invisible to the X-ray diffraction experiment, are added assuming standard bond lengths and angles. The system is immersed in a box of edges 86.6 x 73.3 x 76.4 Å containing 12,858 water molecules[4]. The overall negative charge of the protein-substrate complex is neutralized by adding 6 sodium ions located close to residues Glu191 A, Glu213 A, Asp313 A and the corresponding analogues in the other symmetric unit of the homodimer (numbering is according to the pdb release 1PAU(Rotonda et al., 1996)). All of these residues are located far from the active site.

5.2.2 QM/MM Car-Parrinello calculations.

The system is partitioned into two regions, one treated at the quantum-mechanical level (QM hereafter) and the other classically (MM hereafter). For the first part of the reaction step II, i.e. the initial nucleophilic attack of the water molecule, the QM region includes WATC and the side chains of Cys285, His237 and Asp4' (Fig. 5.2b). For the subsequent step (the decomposition of the gem-diol intermediate formed in the first part of the hydrolysis), the QM portion is based on intermediate **I3** as obtained from the previous calculations (Fig. 5.2c). It includes the side chains of Cys285, His237 and Asp4', the tetrahedral intermediate and a water molecule (H2O in Fig. 5.2c), which is in

hydrogen bonding contact to the reactive complex. This additional water molecule is treated quantum mechanically to allow for its possible participation as a proton shuttle during the decomposition of the intermediate (See Chart 1, which describes the reaction mechanism, as obtained by the QM/MM calculations). All the other atoms have been considered as belonging to the MM region.

Initially, the system has been equilibrated performing purely classical MD calculations based on GROMOS96(van Gunsteren et al., 1996) in combination with a P3M implementation to treat the long-range electrostatic interactions.(Hunenberger, 2000) The mesh used for the P3M calculation was 64x64x64. A leap-frog integration algorithm with a time step of 1.5 fs was used throughout with all bond lengths kept fixed applying a SHAKE algortihm(Ryckaert et al., 1977). Constant NVT simulations were performed by coupling the system to a Berendsen thermostat(Berendsen et al., 1984) with a relaxation time of 0.1 ps. This protocol was used to equilibrate hydrogen atoms and solvent molecules. For the force field based calculations within the QM/MM scheme, the same setup was used except that a velocity Verlet propagation algorithm was applied with a time step of 0.145 fs and no Berendsen thermostat was imposed.

Density-functional theory (DFT) calculations of the QM region are carried out using the BLYP recipe for the exchange and correlation part.(Becke, 1988; Lee et al., 1988) The basis set for the valence electrons consists of plane-waves expanded up to a cutoff of 70 Ry. The interactions between valence electrons and ionic cores are described by norm-conserving pseudopotentials of the Martins-Troullier type(Troullier and Martins, 1991).

The C_α - C_β (Cys285), C_α - C_β (His237), and C_α-C (Val4') bonds cross the QM/MM interface. To saturate the chemical bond with Cys285-C_β , His237-C_β and Val504-C_α., parameterized monovalent pseudopotentials of Cys285-C_α , His237-C_α and Val504-C are used(Laio et al., 2001).

Our models are treated as isolated systems as in ref.(Martyna and Tuckerman, 1999) DFT-based Car-Parrinello molecular dynamics is carried out with a time step of 6 a.u. at an average temperature of T = 300K.

The electrostatic interactions between quantum and classical regions of the system are described within a fully Hamiltonian coupling scheme(Laio et al., 2001). In this approach, the short-range electrostatic interactions between QM and MM part are taken explicitly into account within a sphere of radius 5.3 Å around every QM atom using an

[4] The density of the solvent in our system is 1g/cm^3.

appropriately modified Coulomb potential that ensures that no electron spill-out occurs. The electrostatic interactions with the more distant MM atoms are treated using a multipole expansion for the QM region. Bonded and van der Waals interactions between QM and MM parts were modeled with the standard GROMOS96 force field(van Gunsteren et al., 1996). Further details of the mixed QM/MM Car-Parrinello method used in this work are given in ref.(Laio et al., 2001).

The overall procedure of our calculations involved an initial equilibration of the solvent molecules (except WATC) and hydrogen atoms by 150 ps of classical MD at room temperature followed by a relaxation of the active site by 1.5 ps of unconstrained QM/MM calculations. Using the final snapshot of the system resulting from this preequilibration studies, the investigation of the first part of the hydrolysis reaction, i.e. the nucleophilic attack of WATC to the tetrahedral intermediate **I1** was started. To this end, QM/MM simulations were performed in which the reactive system was constrained to follow a suitable reaction coordinate, in this case the distance between the carbon atom of the acyl group and the oxygen atom of the nucleophilic water molecule $d_{CO}=d(C_{sub}-O_{wat})$ (Fig. 5.2b (left)). The average constraint force was calculated for d_{CO} = 2.9, 2.6, 2.3, 2.1, 1.9 and 1.8 Å. Each point was simulated for 1.5 ps of which the first 0.5 ps were used for equilibration(Carloni et al., 2000) resulting in a total of 9 ps of mixed QM/MM simulations. At a constraint distance of d_{CO} =1.7 Å, a gem-diol intermediate (**I3** in Chart 1) was formed for which 3 ps of unconstrained dynamics were performed. Starting from the last configuration, the second part of the reaction (the decomposition of the gem-diol intermediate) was investigated in an analogous way, i.e. by constraining the system along the C-S (d_{CS}) distance as suitable reaction coordinate. d_{CS} was increased from 1.9 Å to 2.4 Å in increments of 0.1 Å while all the other degrees of freedom were free to relax. Also in this case, each point was simulated for 1.5 ps of which the first 0.5 ps were used for equilibration. For both reaction steps, the transition states were located by the change of sign of the constraint force and the activation free energies (ΔF) were calculated within the thermodynamic integration technique as an integral of the MD-averaged constraint forces along the reaction coordinate Q_i : $\Delta F=\int dQ_i\ f_s(Q_i)$. (Ciccotti et al., 1989) The calculated value of ΔF was compared to experimental value of kcat using transition state theory(Fersht, 1997): Under the reasonable assumption that we work in excess of substrate, $kcat/k_M)\approx kcat = (k_BT/h)\exp(-\Delta F/RT)$.

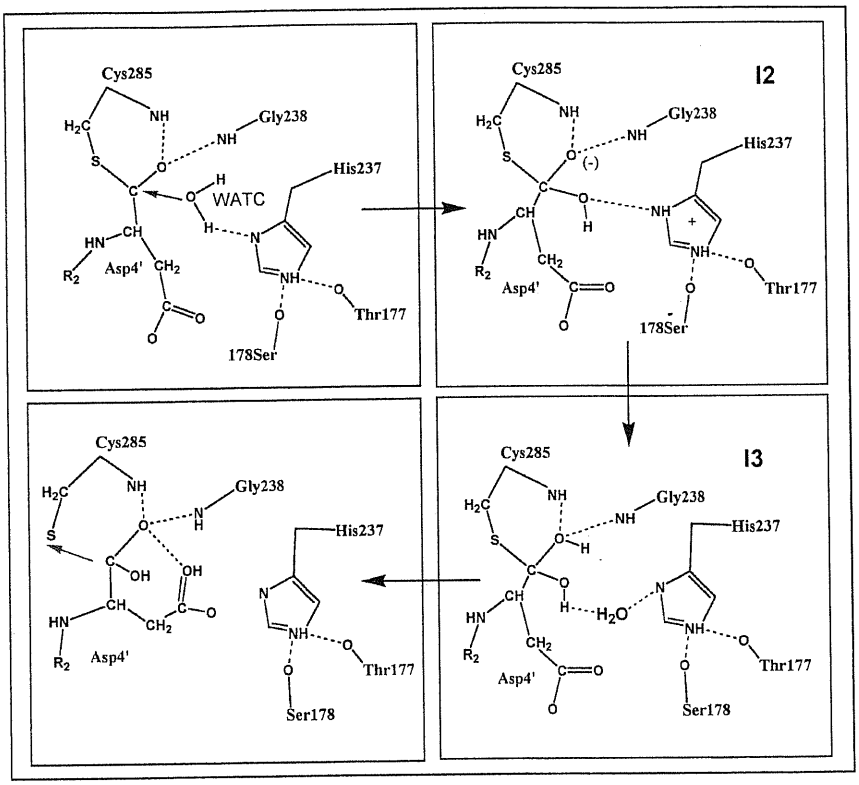

Chart 1

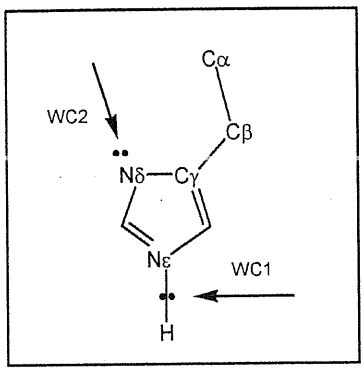

Chart 2

5.2.3 Calculated Properties.

1) The polarization of the electronic density ρ at the active site due the electric field of the surrounding protein was determined as $\Delta\rho_{ext}=\rho_{qm}-\rho_{qm,0}$, where ρ_{qm} and $\rho_{qm,0}$ are the electronic densities of the quantum system in the presence of the protein external field and in vacuum, respectively. 2) The rearrangement of ρ upon formation of the catalytic complex was calculated as $\Delta\rho_{form}= \rho_{qm}-\rho_{as}-\rho_{WATC}$ where ρ_{as} is the electronic density of the quantum system without WATC and ρ_{WATC} is the one of WATC alone; both ρ_{as} and

ρ_{WATC} refer to densities that are calculated in the presence of the external protein field. Starting from $\Delta\rho_{ext}$ we calculated the total polarization integrated over all the quantum box (Δn_{ext}), the polarization integrated over WATC (Δn_{WATC}) and over the substrate carbonyl (Δn_{CO}). The definition of the volumes of WATC and CO was based on the covalent radii(Shriver et al., 2001). 3) Induced dipole moments on Asp4'-CO and WATC groups were obtained by integrating the induced electron density $\Delta\rho$ over spheres of covalent radius(Shriver et al., 2001) r_c around each of the atom ($d = \int_0^{r_c} \Delta\rho_{form}(r) r dr$). 4) The analysis of the geometric centers of maximally localized Wannier functions (WFC)(Silvestrelli et al., 1997; Silvestrelli et al., 1998; Marzari and Vanderbilt, 1997c), calculated as previously reported(Alber et al., 1999), were used to investigate the polarity of chemical bonds, as shifts of WFC correlate to differences of Pauling electronegativities between the involved atoms(Alber et al., 1999).

5.3 Results

Here we describe the complete second part of the enzymatic reaction, i.e. the nucleophilic attack of the catalytic water (WATC) to the carbonyl carbon of the acyl-enzyme complex followed by decomposition of the formed intermediate. Chart 1 summarizes the obtained reaction pathway schematically. In our QM/MM model, the side chains of the active site residues Cys285 and His237, the backbone of Asp4' and WATC are treated at the DFT level, whereas the rest of the protein is described classically (Fig.2b, left panel). We follow the first part of the reaction by performing constrained QM/MM simulations in which the WATC-carbonyl distance (d_{CO} in Fig. 5.2b, left panel) is successively shortened. The subsequent decomposition of the formed reaction intermediate **I3** (shown in Fig. 5.2c) is simulated by constrained QM/MM calculations in which the distance between Cys285-S_γ-Asp4'-C (d_{CS} in Fig. 5.2c, left panel) is stretched in increments of 0.1 Å.

The structure of the acyl-enzyme complex, built from 150 ps of classical and 1.5 ps of QM/MM Car-Parrinello MD, is presented in Fig. 5.2b. The carbonylic oxygen of the substrate is stabilized via hydrogen bonds to Cys285-NH and Gly238-NH forming the oxyanion pocket. WATC also interacts with O4' and in addition with His237-N_δ .

5.3.1 Nucleophilic Attack of WATC to the Acyl-Enzyme.

In this Section, we give a description of the structural determinants and energetics associated with the first step of the deacylation process. Upon shortening of d_{CO} from 2.9 to 2.3 Å, the H-bond between Gly238-NH_2 and Asp4'-O is lost and the one between Cys285-NH_1 and Asp4'-O is strengthened (Tab.1). The orientation of the catalytic histidine, His237, is maintained throughout the reaction, as evinced by the torsional angles χ_1 and χ_2, which do not change significantly during the dynamics (Tab.1). This pronounced conformational stability can likely be attributed to the extended H-bond network of His237 which not only interacts with WATC but also forms and breaks H-bonds to the Thr177 backbone carbonyl and the side chain of Ser178 (see Tab. 5.1). The active site polarization of His237 can be analyzed in terms of the environment-induced shift of the Wannier centers(Silvestrelli et al., 1997; Silvestrelli et al., 1998; Marzari and Vanderbilt, 1997c) that represent the Nε-H bond and the Nδ lone pair (Chart 2). Tab. 5.2 shows that the lone-pair of Nδ is transferred towards the nitrogen atom in passing from vacuum to the protein surrounding; an inverse effect occurs for the Nε-H bond.

d_{CO}	C_4-$O_{4'}$	S_γ-$C_{4'}$	H1-$O_{4'}$	H2-$O_{4'}$	H3-S_γ	Hw1-$O_{4'}$	Hw2-N_δ	N_ε-O_{177}	N_ε-O_{178}	S_γ-C_4-$O_{4'}$	T1	T2	χ_1	χ_2
2.9	1.25±0.02	1.84±0.06	2.5±0.2	3.2±0.2	2.2±0.1	1.8±0.1	2.0±0.2	2.3±0.6	2.4±0.5	119±3	10±19	-76±17	-78±12	110±17
2.7	1.26±0.03	1.83±0.05	1.9±0.2	2.3±0.4	2.3±0.2	1.9±0.2	2.0±0.2	2.7±0.5	1.9±0.2	119±3	14±17	-70±16	-82±9	107±15
2.6	1.25±0.01	1.84±0.04	1.9±0.2	2.0±0.3	2.2±0.1	1.9±0.1	1.9±0.1	2.8±0.2	1.8±0.2	119±3	24±9	-64±5	-83±7	111±7
2.3	1.26±0.01	1.87±0.09	1.9±0.1	2.0±0.2	2.3±0.2	2.0±0.1	1.9±0.1	3.0±0.3	1.8±0.1	119±3	22±12	-71±9	-79±8	114±16
2.1	1.25±0.02	1.93±0.09	1.9±0.1	2.0±0.2	2.2±0.2	2.6±0.3	2.3±0.5	2.8±0.4	2.1±0.4	117±3	52±24	-51±21	-79±10	108±14
1.9	1.26±0.01	1.97±0.06	1.9±0.1	2.0±0.2	2.3±0.1	2.7±0.1	2.0±0.3	3.4±0.2	1.8±0.2	115±3	58±9	-47±8	-80±11	106±17
1.8	1.27±0.02	2.01±0.10	1.8±0.2	2.0±0.2	2.4±0.2	2.6±0.1	1.8±0.2	3.2±0.2	1.7±0.1	114±3	53±12	-53±10	-81±10	111±12
1.7(I)	1.29±0.02	2.05±0.11	1.8±0.1	2.0±0.2	2.3±0.1	2.6±0.1	1.6±0.2	3.4±0.3	1.7±0.1	113±3	51±8	-59±8	-80±10	116±9
1.7(II)	1.29±0.02	2.17±0.13	1.9±0.1	1.9±0.2	2.4±0.1	2.6±0.1	1.10±0.09	2.8±0.5	2.0±0.3	111±4	55±15	-53±8	-78±8	112±11
1.7(III)	1.47±0.02	1.93±0.10	2.7±0.2	2.5±0.2	2.4±0.2	2.4±0.1	1.8±0.2	1.6±0.5	2.3±0.5	107±3	12±20	-27±4	-81±9	117±10

***Table 5.1**. Nucleophilic attack of WATC: Structural properties. Selected MD-averaged distances (Å) and angles (°), averaged over 1 ps, at different values for the constraint distance d_{CO}. Standard deviations are also reported. Atom labeling as in Fig 5.2.*

The orientation of the catalytic water molecule WATC is also fully maintained during the dynamics, as shown by the almost constant values of the torsional angles T_1 and T_2 as defined in Fig. 5.2b (Tab. 5.1). A calculation of the rearrangement of the electron density Δn_{form} upon the WATC/thiohemiacetal interaction, respect to separate fragments, reveals that the polarization is almost equally distributed on WATC and on Asp4'-CO and increases with decreasing d_{CO} (see Tab. 5.3 and Fig. 5.3).

Redistributions of the electron density on the two interacting groups lead to the formation of induced dipole moments on both Asp4'-CO and WATC. These induced dipole moments, calculated from the change in electron density $\Delta\rho_{form}$ as defined in the Method Section, are the main contribution from the total rearrangement of the electron density.

Our QM/MM approach also allows for a direct determination of the environment effects during the course of the reaction. The electric field of the protein leads to a sizeable polarization of the reactants (around 0.6-0.7e, see Tab. 5.2). A simplified calculation of the active site in vacuum would thus not be able to account for the specific features of the electronic structure. Analysis of the single contributions to the electric field indicate that a shell of radius 8 Å around carbonyl carbon of the substrate provides more than 80% of the total field on the reactants. Thus, in this specific case long-range effects of the protein electric field are of less importance for the reactant state. In particular, there are no secondary structure elements, which affect the energetics of the reaction in a significant way.

d_{CO} (Å)	Δn_{ext} (e)	Δn_{WATC} (e)	Δn_{CO} (e)	μ_{WATC} (D)	μ_{CO}(D)	WS1 (Å)	WS2(Å)
2.9	0.71	0.02	0.03	0.12	0.16	0.021	-0.018
2.3	0.66	0.03	0.03	0.08	0.12	0.018	-0.014
1.8	0.63	0.03	0.04	0.14	0.23	0.016	-0.007

Table 5.2. *Nucleophilic attack of WATC: Changes in electronic properties upon inclusion of the protein electric field. Δn_{ext} (total polarization), Δn_{WATC} (polarization on WATC), Δn_{CO} (polarization on substrate carbonyl group), μ_{WATC} (induced dipole on WATC), μ_{CO} (induced dipole on substrate carbonyl), WS1 and WS2 (shifts of the Wannier centers representing the Nε-H bond and Nδ lone pair, respectively) are calculated for selected snapshot at various d_{CO} distances. Polarizations are expressed in number of electrons (e), dipoles in Debyes (D) and shifts in the Wannier centers in Å.*

d_{CO} (Å)	Δn_{form} (e)	Δn_{WATC} (e)	Δn_{CO} (e)	μ_{WATC} (D)	μ_{CO}(D)
2.9	0.31	0.04	0.02	0.15	0.11
2.7	0.41	0.06	0.05	0.18	0.31
2.6	0.42	0.06	0.04	0.16	0.28
2.3	0.39	0.06	0.04	0.23	0.23
2.1	0.39	0.05	0.05	0.20	0.36
1.9	0.69	0.09	0.09	0.22	0.59
1.8	0.78	0.12	0.10	0.22	0.60

Table 5.3. *Nucleophilic attack: of WATC: Changes in electronic properties upon formation of the acyl/water complex. Δn_{form} (total polarization), Δn_{WATC} (polarization on WATC), Δn_{CO} (polarization on substrate carbonyl group), μ_{WATC} (induced dipole on WATC, in Debye), μ_{CO} (induced dipole on substrate carbonyl) are calculated for selected snapshots at various at different WATC-substrate distances d_{CO}. Polarizations are expressed in number of electrons (e^-), dipoles in Debyes (D).*

Further decrease of d_{CO}, up to 1.8 Å causes an increase of the polarization of the QM part (Tab. 5.3). Similarly, the electronegativity of WATC and its induced dipole moment increase (Tab. 5.3 and Fig. 5.3c). Furthermore, the S_γ-$C_{4'}$ bond weakens while the $C_{4'}$-$O_{4'}$ bond appears to stiffen as indicated by the smaller fluctuations of the $C_{4'}$-$O_{4'}$ bond length with respect to the ones of the S_γ-$C_{4'}$ bond. (Tab. 5.1). The carbonyl carbon of the substrate approaches *sp3* hybridization (as indicated by the S_γ-$C_{4'}$-$O_{4'}$ angle, which is reduced by about 5 degrees from 119° to 114°, Tab. 5.1) leading the system towards the tetrahedral conformation of the intermediate **I1**. WATC maintains its Hw1-bond with Nδ(His237) while it breaks its Hw2-bond with Asp4'-C to interact with a water from the solvent. This results in an increased value of the average distance of Hw1-$O_{4'}$ (Tab.1). As a consequence, the orientation of WATC with respect to Asp4'-CO changes in such a way as to provide a stabilizing interaction of the induced dipole moments on WATC and Asp4'-CO, which exhibit an antiparallel orientation (Tab. 5.1, and Fig. 5.3c).

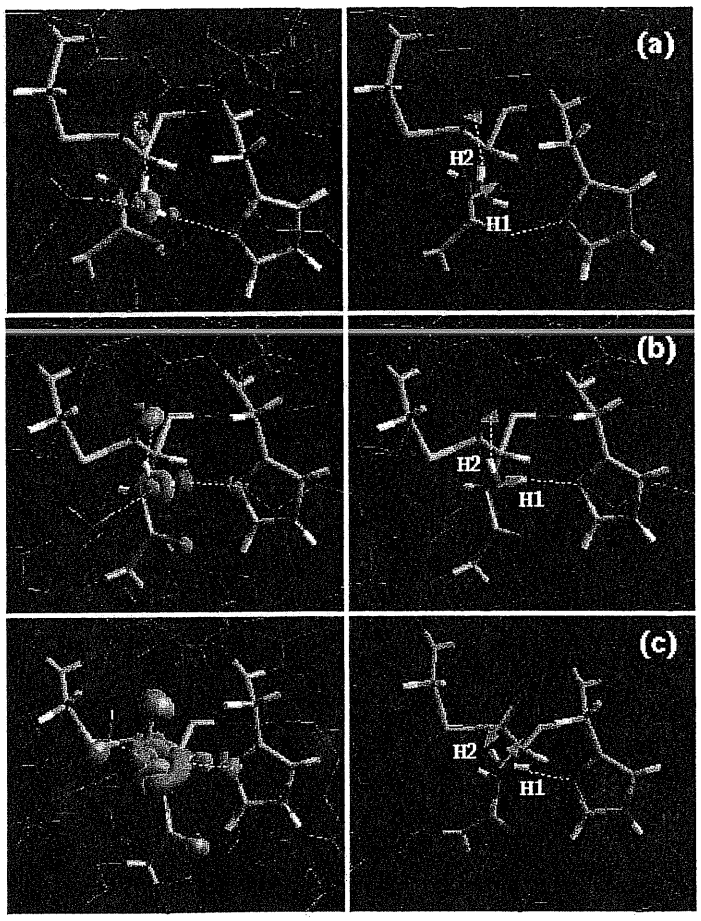

***Figure 5.3**. Water nucleophilic attack to the acyl-enzyme. Left: Electronic density difference map for d = 2.9 Å **(a)**, d= 2.3 Å **(b)** and d = 1.8 Å **(c)**. Light blue isocontours represent loss of electronic density as a result of complex formation relative to the isolated moieties. Orange regions refer to increased densities.*

*The contour shown is 0.005 e^-/au^3. Right: Induced dipole moments as result of complex formation are also displayed for d_{CO}=2.9 Å **(a)**, d_{CO}=2.3 Å **(b)** and d_{CO}=1.8 Å **(c)**.*

After about 1 ps of constraint dynamics at d_{CO} = 1.7 Å, WATC transfers a proton to His237-Nε (Fig. 5.4a). At this point, the constraint force changes sign indicating that the transition state of the reaction has been passed. The free energy barrier of the process can be calculated from the average force on the constraint (Tab. 5.4). The calculated value, ΔF ≈ 20 kcal/mol (Fig. 5.5), is slightly higher than the value calculated (Fersht, 1997) (ΔF ≈ 17.7 kcal/mol) from the measured experimental rate of ≈ $1s^{-1}$ (Stennicke and Salvesen, 1997).

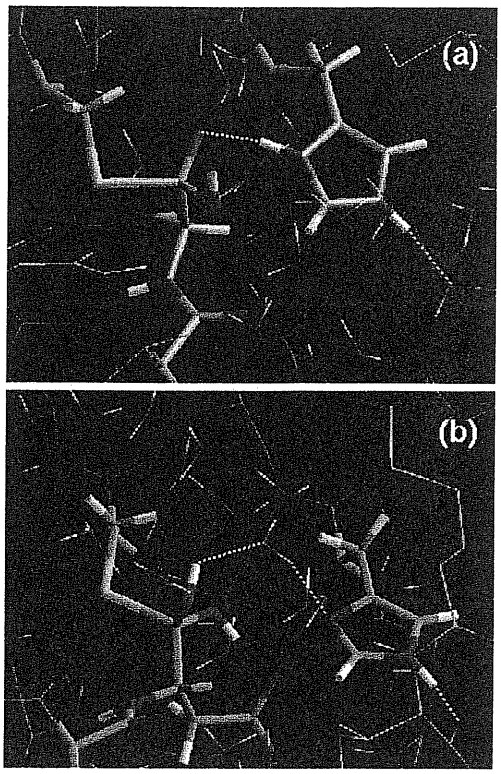

Figure 5.4. *Water nucleophilic attack to the acyl-enzyme. : Structure of transition state **(a)** and of the intermediate (**I3** in Chart 1) **(b)**. The H-bond pattern of His237 is shown with yellow dotted lines.*

d_{CO} (Å)	Force on constraint (a.u.)
2.9	0.010±5E-04
2.7	0.016±4E-04
.6	0.022±6E-04
2.3	0.036±2E-04
2.1	0.039±2E-04
1.9	0.045±4E-04
1.8.	0.048±5E-04
1.8(rev)	0.0040±2E-04
1.9(rev)	0.0043±2E-04
d_{CS} (Å)	**Force on constraint (a.u.)**
1.9	0.0000±1E-03
2.1	-0.0049±4E-04
2.2	-0.0100±3E-04
2.3	-0.0042±3E-04
2.4	-0.0015±1E-03

Table 5.4. *Constraint forces associated to several d_{CO} and d_{CS} distances from constraint QM/MM simulations of the two reaction steps. (rev) stands for backward simulation.*

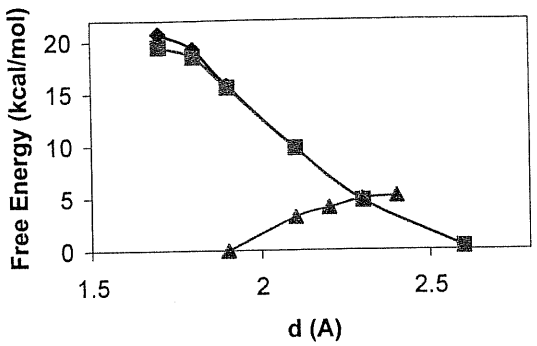

***Figure 5.5**. Activation free energy ΔF plotted as function of the reaction constraints d_{CO} (blue and magenta for forward and backward reaction) and d_{CS} (green).*

As the observation of a proton transfer is limited to a single event during our QM/MM calculations, we can clearly not sample this process in a statistically meaningful way. However, we have endeavored to quantify the sensitivity of the calculated free energy with respect to this event by estimating the error on the constraint force. To this end, we have also followed this part of the pathway in the inverse direction by increasing d_{CO} starting from the product side. The resulting free energy barrier is slightly lower (Tab. 5.4, and Fig. 5.5) and the estimated corresponding error due to hysteresis is only ≈ 1 kcal/mol.

After ca.1.8 ps upon release of the constraint, His237-N_δ transfers spontaneously its N_δ proton to the carbonylic oxygen. Although this event contrasts with the assumed standard reaction mechanism, which assumes that the proton is transferred to the sulfur atom of the thioether link(Wilson et al., 1994), the $C_{4'}$-$O_{4'}$ bond is much more polar (and thus a better proton acceptor) than the C-S bond (Tab. 5.1), suggesting that the proton affinity of $O_{4'}$ is substantially larger than the one of S.

This event leads to a neutral intermediate **I3** (Chart I, and Fig. 5.4b) that is stable during the additional time-scale explored (≈1ps). The interactions with the oxyanion pocket are well maintained after formation of the activation of the complex in Fig. 5.4a; in contrast, they slightly weaken due to the formation of the intermediate **I3**, as evidenced by the H1-$O_{4'}$ and H2-$O_{4'}$ distances (Tab. 5.1). The conformation of His237 remains fairly well maintained all through the simulation. After **I3** formation, His237 H-bonds to the solvent (Fig. 5.4b), and as a result the torsional angle T_2 increases (Tab. 5.1). Furthermore, when Hw1 is transferred to $O_{4'}$ the S_γ-$C_{4'}$ bond becomes stronger as indicated by the reduction in average bond length, while the $C_{4'}$-$O_{4'}$ bond length enlarges significantly in accordance with increasing single bond character. At the same time the S_γ-$C_{4'}$-$O_{4'}$ angle approaches 107°, a value typical for a *sp3* configuration (Tab.1).

5.3.2 Dissociation of the Gem-Diol Intermediate I3.

The final step of the deacylation process is the dissociation of the C-S bond of the tetrahedral intermediate, leading to the products and the free enzyme. This second step of the reaction has been followed starting with the last snapshot of the previous simulation (**I3** in Chart I). Constrained QM/MM simulations were performed in which the Cys285-S_γ-$C_{4'}$ bond has been gradually elongated starting from its equilibrium length (Fig. 5.2c).

As d_{CS} is increased, several modifications concerning the tetrahedral intermediate take place. In particular $C_{4'}$ evolves towards a planar hybridization, as is apparent from a decrease of all the bond angles involving $O_{4'}$-$C_{4'}$. The slightly higher value of S_γ-$C_{4'}$-$O_{4'}$ is due to the involvement in H-bonds to Cys285-N_{H1} which increases its bending. At d_{C-S}=2.1-2.2 A, several proton hopping events occur from the $O_{4'}$ to the Asp4' side-chain (Fig. 5.6). As d_{CS} is further increased, the proton definitely localizes on the side chain of Asp4' (Fig.6b). A measure of proton delocalization between the backbone carbonyl and the side chain of Asp4' can be deduced from its average distance from $O_{4'}$ and $O_{\delta1}$ shown in Tab. 5.5. For d_{CS} ~2,1-2.2 Å, the proton is delocalized on both oxygens. As the d_{CS} increases, the proton is located on Asp4' $O_{\delta1}$ and the $C_{4'}$-$O_{4'}$ bond length is reduced, indicating an increased double bond character. During the entire simulation, the active site conformation remains essentially non-altered. In particular, the $O_{4'}$ pocket and the S_γ coordination are well-maintained (see Tab. 5.5).

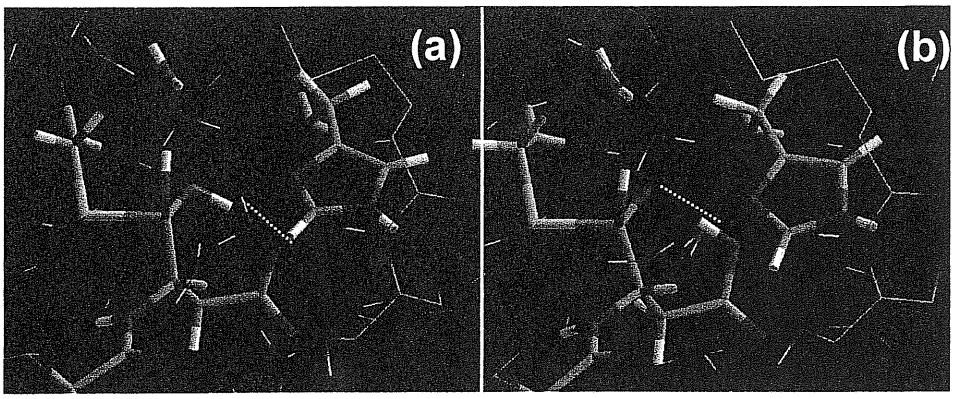

Figure 5.6. *Dissociation of the products. Several proton hopping events occurs between O_4* ***(a)*** *and $O_{\delta1}$* ***(b)*** *during the reaction. Upon shortening of the C-S distance, the proton finally localizes on $O_{\delta1}$ (intermediate* ***I3****).*

d_{CS} (Å)	S_γ-$C_{4'}$-$O_{4'}$	S_γ-$C_{4'}$-O_w	S_γ-$C_{4'}$-C	$C_{4'}$-$O_{4'}$	H1-$O_{4'}$	H2-$O_{4'}$	H_3-S_γ	H_{w2}-$O_{4'}$	H_{w2}-$O_{\delta1}$
2.1	110±5	105±7	107±2	1.36±0.01	1.78±0.05	2.8±0.9	2.1±0.1	1.2±0.09	1.3±0.1
2.2	110±2	102±2	106±2	1.34±0.05	1.85±0.05	2.90±0.03	2.2±0.2	1.3±0.2	1.2±0.2
2.3	109±3	101±1	104±1	1.31±0.07	1.9±0.1	2.4±0.2	2.3±0.1	1.51±0.03	1.07±0.05
2.4	106±2	101±2	101±1	1.29±0.03	1.88±0.04	2.14±0.07	2.3±0.2	1.57±0.05	1.04±0.01

***Table 5.5.**: Selected structural parameters at different values of the constraint distance d_{CS}. The atom labeling is shown in Fig 2c. Average values and standard deviations are reported, distances are expressed in Å, angles in degrees.*

The calculated free energy barrier turns out to be only ≈ 5 kcal/mol (see Fig. 5.5), i.e. much lower than that of the previous reactions step corresponding to the nucleophilic attack of WATC. Thus, the rate-limiting barrier of the deacylation reaction is clearly associated with the formation of the tetrahedral intermediate rather than its dissociation into the products.

5.4 Discussion and conclusions

Surprisingly in view of the great importance of the caspase family, no theoretical studies have yet been performed on these systems. In this work, we have presented a description of the second step of the enzymatic reaction, the hydrolysis of the acyl-enzyme complex. Our mixed QM/MM Car-Parrinello approach, which is based on the caspase-3/Ac-DEVD-CHO structure(Rotonda et al., 1996), has provided a detailed quantum mechanical treatment of this fundamental step of the reaction including electrostatic and dynamic effects of the protein frame. Our calculations show that the reaction proceeds via nucleophilic attack of WATC to the acyl-enzyme complex via a tetrahedral transition state (Fig. 5.4a). Thus, the postulated second intermediate (**I2,** Chart 1) actually corresponds to a transition state. Subsequently, a previously unrecognized gem-diol intermediate is formed (**I3**, Chart I), which in turn easily decomposes to the products. This new mechanism could be general for the caspase family.

Our simulation of the nucleophilic attack of a water molecule suggests that the protein uses the oxyanion pocket (the amido hydrogens of the backbone of Cys285 and Gly238 Fig.2 b) and His237 to impose a highly ordered supramolecular structure for the

hydrolysis process (Fig. 5.2b) thus reducing the activation free energy of the reaction(Piana et al., 2001). His237 acts as multiple proton shuttle: first, it accepts a proton from WATC thereby enhancing the nucleophilic power of WATC; then, it protonates intermediate **I3** (Chart I), thus weakening the C-S bond. The catalytic histidine residue is activated by an H-bond with the side chain of Ser178 (Fig. 5.2b). This interaction does not emerge from the X-ray structure. The formation of the reaction intermediate **I3** is accompanied by a strong rearrangement of the electronic structure (as can be monitored by the corresponding changes in the electron density and in the shifts of the Wannier centers, Tab. 5.2), and the formation of induced dipole moments on both WATC and Asp4'-CO (Tab.3 and Fig. 5.3), whose interaction partly counter balances the repulsive energy due to the constrained relative positions of the two moieties. The resulting reaction intermediate (**I3** in Chart 1) is the gem-diol tetrahedral complex in Fig. 5.4b, in which the carbon atom of the carbonyl group is covalently linked to the sulfur atom of Cys285 and the C_α (Asp4'). The calculated activation free energy of the deacylation step is due to the hydrolysis of the acyl-enzyme and is fairly compatible with the experimentally free energy value (≈ 18 kcal/mol)[1,] (Stennicke and Salvesen, 1997).

The dissociation of the tetrahedral intermediate **I3** is investigated by constrained QM/MM simulations in which the $C_{4'}$-S_γ bond is elongated successively. It occurs through a proton hopping from the hydroxy group bound to the oxyanion pocket of the Asp4'-sidechain (Fig. 5.6). As a consequence of this proton transfer, the H-bond between the sidechains Asp4' and Arg179 becomes weaker permitting an easier release of the substrate from the active site pocket. Thus, the calculations suggest that the substrate itself plays an active role in catalysis contributing to the deacylation process as proton acceptor. The associated energy barrier is much lower than that of the previous reaction step (≈5 kcal/mol).

In conclusion, the deacylation process passes through the formation of an anionic tetrahedral transition state, which rapidly evolves towards the neutral gem-diol intermediate **I3**. The latter, which has not emerged from the previous mechanistic hypothesis of the mechanism(Wilson et al., 1994) can be used in future studies as a template for the design of intermediate analogs[5]. The subsequent dissociation of the C-S thioether bond is concerted with a proton transfer to the Asp sidechain of the substrate (Fig. 5.6). This mechanism is new alternative to the direct proton hopping from

[5]The structure of the **I3**/caspase-3 complex is available on the web at http://www.sissa.it/cm/bc.

the charged His to the sulfur atom of the anionic intermediate, which has been recently proposed for the correspondent reaction in aqueous solution(Strajbl and Warshel, 2001). Here, we propose that the presence of a proton acceptor such as the Asp group, missing in the simulation in water, may dramatically affect the reaction pathway.
Our calculations also provide additional clues about the differences in catalytic activity of caspases relative to other cysteine proteases, such as papains. Caspase-3 has a lower catalytic efficiency (k_{cat} about $1s^{-1}$) than that measured for papain(Menard et al., 1991). Both enzymes appear to function using a chemically similar catalytic triad: Cys285, His237 and Ser178 for caspase-3, and Cys25, His159 and Asn175 for papain(Brocklehurst et al., 1998). Our calculations suggest that the decrease in catalytic efficiency on passing from papain to caspase-3 may be ascribed to both *conformational* and *electrostatic* and properties. Indeed, inspection of the corresponding crystal structures shows in fact that Cys285 and His237 have a markedly different location in the active site. Their side-chains are closer in papain (by ~(Lee et al., 2000)2.0 – 2.5 Å) (Kamphuis et al., 1984; Kamphuis et al., 1985; Yamamoto et al., 1992; LaLonde et al., 1998) than those in caspase-3 (see Fig. 5.2a), and the conformational parameters χ_1 and χ_2 of His238 are totally different[6]. In addition, our calculations suggest that secondary-structure elements belonging to the protein frame are not relevant to the induced field in the active site. In particular, the electric field experienced by the catalytic water molecule is of the same order of the electric field of bulk water. This is in contrast to most enzymes as well as to papain and cathepsins, where an α-helix has been suggested as responsible for a relevant contribution to the electric field of the active(Kortemme and Creighton, 1995). Thus, the active site of caspases instead seems to act merely as a geometrical constraint.

Finally, the simulations presented here can usefully compared with other calculations with a similar setup on hydrolases such as HIV-1 protease and GTP hydrolase. In particular it can help to clarify role of catalytic water polarization in relationship with enzymatic efficiency. All these enzymes must stabilize the highly nucleophilic agent OH- by deprotonation of the catalytic water(Fersht, 1997). In GTP hydrolase enzyme, the proton acceptor is the amino group of Gln61, which is a very weak base comparable to bulk water. Here, the catalytic water is strongly polarized towards the oxygen(Cavalli

[6] χ_1~-80° and χ_2~ 110° for caspases(Rotonda et al., 1996; Mittl et al., 1997; Lee et al., 2000) (Fig. 2a) , χ_1~-175° and χ_2~-120° for papain(Kamphuis et al., 1984)

and Carloni, 2001) so as to increase its acidity. In contrast, in HIV-1 protease(Piana et al., 2001), water, which is polarized as in the bulk, can easily donate a proton to the fairly good base Asp25 (pKa ≈ 4.4). Finally, here in caspase-3, water, which is *less* polarized than in the bulk, easily transfers a proton to the excellent base His237 (pK ≈ 6.5).

Conclusions and perspectives

In the work presented here we apply first principle methods for the study of the enzyme/substrate interactions in systems of pharmaceutical interest.

In the first part, a rational drug screening approach based on DFT calculations is presented. We performed a quantum-chemical study on a representative set of the large spectrum of sugar-mimicking moieties for which structural information of the TK_{HSV1}-ligand complexes is available. If the mechanisms ruling binding affinities toward TK_{HSV1} have been previously presented (Champness et al., 1998; Prota et al., 2000; Alber et al., 1998; Perozzo et al., 2000), no rational explanation was available for the different catalytic behaviors. Moreover the question posed by the recently published crystal structure of the inhibitor HBPG (Bennett et al., 1999) as to what is the structural basis for the different properties between inhibitor and substrate sharing the same binding mode remained open.

With our calculations we have addressed these issues by carrying out DFT calculations on models of substrate/enzyme adducts. The calculations suggest a role of the electric dipole moment of ligands and its interaction with the negatively charged residue Glu225 emerges. A striking correlation is found between the energetics associated to this interaction and the k_{cat} values measured under homogeneous conditions by Prof. Folkers' lab. This finding uncovers a fundamental aspect of the mechanism governing substrate diversity and catalytic turnover and thus represents a significant step toward the rational design of novel and powerful prodrugs for antiviral and TK_{HSV1}-linked suicide gene therapies.

Our study provides also a rationale for the presence of polar group in the hydrophobic pocket of the enzyme. Indeed, replacement of the O4' with a CH_2 group ether provides a dramatic reduction and change of orientation of the dipole, which in turn could decrease the stabilizing dipole/Glu225 interactions.

This type of calculations are very straight forward and can be extended in the future to the design of new and more potent prodrugs for antiviral therapy to overcome the problems of resistance or to be used in enzyme-prodrug gene therapy as well as improve variants of TK_{HSV1} for gene therapeutic approaches (Encell et al., 1999).

In the second part of my thesis, DFT methods have also shown to provide an accurate description of interactions in ligands/protein complexes involving aromatic molecules. In particular they have been applied to a cation-π interaction occurring in HIV RT and to hydroxyl-π interaction occurring in active site μ-Gluthathione S-transferase. Calculations confirm the essential role of electrostatic and provide a detailed analysis of quantum effects.

In the following part of this thesis one of the first application of QM/MM methods to the description of an enzymatic reaction is presented. In particular we focused on Caspase-3 reaction mechanism. Surprisingly, in view of the great importance of the caspase family, no theoretical studies have been performed so far.
In this work I presented a description of the deacylation step, that is the nucleophilic attack of catalytic water to acylenzyme, followed by the decomposition of the reaction intermediate. Our mixed QM/MM approach has provided a detailed quantum mechanical treatment of this fundamental step of the enzymatic reaction including electrostatic and dynamical effects of the protein frame.
Our simulation suggests that the deacylation process pass through the formation of the anionic intermediate, which rapidly evolves towards the neutral gem-diol intermediate. The consequent C-S bond dissociation is concerted with a proton transfer to substrate Asp sidechain. Such mechanism is an alternative mechanism to direct proton hopping from charged His to anionic intermediate S, which has been recently proposed for the correspondent reaction in aqueous solution in presence of His group(Strajbl and Warshel, 2001). We propose here that the presence of a proton acceptor such as the Asp group, missing in the simulation in water, may dramatically affect the mechanism.

The above mechanism could be general for caspases. Indeed, different classes of caspases possess distinct activity patterns, yet they are expected to share the same enzymatic mechanism because of their high structural similarity.

Our calculations may provide some hints on the difference in catalytic activity of caspases relative to other cysteine proteases, such as papains. We suggest that the decrease in catalytic efficiency on passing from papain to caspase-3 may be ascribed to both conformational and electrostatic properties. Indeed, our calculations suggest that secondary-structure elements belonging to the protein frame are not relevant to active site electric field. In particular, the electric field experienced by the catalytic water is of the same order of the electric filed of bulk water. Finally the calculations have provided a comparison with other hydrolases such as HIV-1 protease and GTP hydrolase, in particular helping to clarify role of catalytic water polarization in relationship with enzymatic efficiency.

The calculated structure of the transition state provides novel information, which was not currently available from experiments and can be used as starting point in Prof. Cattaneo's lab for a rational drug design based on Pauling transition state theory (structure is now available on http://www.sissa.it/cm/bc).

Bibliography

Abe,I. and Prestwich,G.D. (1995). Molecular cloning, characterization, and functional expression of rat oxidosqualene cyclase cDNA. Proc. Natl. Acad. Sci. U. S. A *92*, 9274-9278.

Alber,F. (1998). Density functional studies on herpes simplex virus type 1 thymidine kinase-substrate interactions: the role of Tyr-172 and Met-128 in thymine fixation. Proteins *31 (4)*, 453-459.

Alber,F., Folkers,G., and Carloni,P. (1999). Dimethyl Phosphate: Stereoelectronic versus Enviroments Effects. J. Phys. Chem. B *103*, 6121-6126.

Alber,F., Kuonen,O., Scapozza,L., Folkers,G., and Carloni,P. (1998). Density functional studies on herpes simplex virus type 1 thymidine kinase-substrate interactions: the role of Tyr-172 and Met-128 in thymine fixation. Proteins *31*, 453-459.

Allen,F.H., Hoy,V.J., Howard,J.A.K., Thalladi,V.R., Desiraju,G.R., and Wilson,C.C. (1996). J. Am. Chem. Soc. *119*, 3477.

Armstrong,R.N. (1998). Mechanistic imperatives for the evolution of glutathione transferases. Curr Opin Chem Biol *5*, 618-623.

Atwood,J.L., Hamada,F., Robinson,K.D., Orr,G.W., and Vincent,R.L. (1991). Nature *349*, 683.

Bachelet,G.B., Hamann,D.R., and Schlüter,M. (1982). Phys. Rev. B *26*, 4199.

Bakowies,D. and Thiel,W.J. (1996). J. Phys. Chem. *100*, 10580.

Barak,D., Ordentlich,A., Segall,Y., Velan,B., Benschop,H., De Jong,L., and Shafferman,A. (1997). Carbocation-Mediated Processes in Biocatalysts. Contribution of Aromatic Moieties. J. Am. Chem Soc. *119*, 3157-3158.

Barnett,R.N. and Landman,U. (1993). Born-Oppeneimer molecular dynamics simulations of finte systems: Structure and dynamics of (H2O)2. Phys. Rev. B *48*, 2081-2097.

Bayly,C.I., Cieplack,P., Cornell,W.D., and Kollman,P.A. (1993). J. Phys. Chem. *97*, 10269-10280.

Becke,A.D. (1988). Density-functional exchange-energy approximation with corrct asymptotic behavior. Phys. Rev. A *38*, 3098-3100.

Bennett,M.S., Wien,F., Champness,J.N., Batuwangala,T., Rutherford,T., Summers,W.C., Sun,H., Wright,G., and Sanderson,M.R. (1999). Structure to 1.9 A resolution of a complex with herpes simplex virus type-1 thymidine kinase of a novel, non-substrate inhibitor: X-ray crystallographic comparison with binding of aciclovir. FEBS Lett. *443*, 121-125.

Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., Di Nola,A., and Haak,J.R. (1984). J. Chem. Phys. *81*, 3684-3690.

Bernstein,F.C.I., Koetzle,T.F., and Williams,G.J.B. (1999). The protein data bank: A computer-based archival file for macromolecular structures. J. Mol. Biol. *112*, 535-542.

Black,M.E. and Loeb,L.A. (1993). Identification of important residues within the putative nucleoside binding site of HSV-1 thymidine kinase by random sequence selection: analysis of selected mutants in vitro. Biochemistry *32*, 11618-11626.

Blanchard,H., Kodandapani,L., Mittl,P.R., Marco,S.D., Krebs,J.F., Wu,J.C., Tomaselli,K.J., and Grutter,M.G. (1999). The three-dimensional structure of caspase-8: an initiator enzyme in apoptosis. Structure. Fold. Des *7*, 1125-1133.

Bohm,H.J. (1992). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J Comput Aided Mol. Des *6*, 61-78.

Bohm,H. and Klebe,G. (1996). What can we learn from molecular recognition in protein-ligand complexes for the design of new drug. Angew. Chem. Int. Ed. Engl *35*, 2588-2614.

Bonini,C., Ferrari,G., Verzeletti,S., Servida,P., Zappone,E., Ruggieri,L., Ponzoni,M., Rossini,S., Mavilio,F., Traversari,C., and Bordignon,C. (1997). HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia [see comments]. Science *276*, 1719-1724.

Brinen,L.S., Hansell,E., Cheng,J., Roush,W.R., McKerrow,J.H., and Fletterick,R.J. (2000). A target within the target: probing cruzain's P1' site to define structural determinants for the Chagas' disease protease. Structure Fold. Des *8*, 831-840.

Brocklehurst,K., Watts,A., Patel,M., and Thomas,E. (1998). Comprehensive biological catalysis. Academic Press).

Brown,D.G., Visse,R., Sandhu,G., Davies,A., Rizkallah,P.J., Melitz,C., Summers,W.C., and Sanderson,M.R. (1995). Crystal structures of the thymidine kinase from herpes simplex virus type-1 in complex with deoxythymidine and ganciclovir. Nat. Struct. Biol. *2*, 876-881.

Bruge,F., Bernasconi,M., and Parrinello,M. (1999). Density-functional study of hydration of ammonium in water clusters. J. Chem. Phys. *110*, 4734-4736.

Burley,S.K. and Petsko,G.A. (1986). Amino-aromatic interactions in proteins. FEBS Lett. *203*, 139-143.

Caldwell,J.W. and Kollman,P.A. (1995). J. Am. Chem. Soc. *117*, 4177-4178.

Car,R. and Parrinello,M. (1985). Unified Approach for molecular dynamics and density-functional theory. Phys. Rev. Lett. *55*, 2471-2474.

Carloni,P., Bloch,P., and Parrinello,M. (1995). J. Phys. Chem. *99*, 1338.

Carloni,P. and Rothlisberger,U. (2000). Simulation of enzymatic systems: perspectives from Car-Parrinello molecular dynamics simulations. In Theoretical Biochemistry - Processes and properties of biological systems, Ericksson, ed. Elsevier Science).

Carloni,P., Sprik,M., and Andreoni,W. (2000). Key Steps of the cis-Platin-DNA Interaction: Density Functional Theory-Based Molecular Dynamics Simulation. J. Phys. Chem. B *104*, 823-835.

Cavalli,A. and Carloni,P. (2001). Enzymatic GTP Hydrolysis: Insights from an *Ab initio* Molecula Dynamics Study. submitted to.

Champness,J.N., Bennett,M.S., Wien,F., Visse,R., Summers,W.C., Herdewijn,P., de Clerq,E., Ostrowski,T., Jarvest,R.L., and Sanderson,M.R. (1998). Exploring the active site of herpes simplex virus type-1 thymidine kinase by X-ray crystallography of complexes with aciclovir and other ligands. Proteins *32*, 350-361.

Chen,M.S. and Prusoff,W.H. (1978). Association of thymidylate kinase activity with pyrimidine deoxyribonucleoside kinase induced by herpes simplex virus. J. Biol. Chem. *253*, 1325-1327.

Christians,F.C., Scapozza,L., Crameri,A., Folkers,G., and Stemmer,W.P. (1999). Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling. Nat. Biotechnol. *17*, 259-264.

Ciccotti,G., Ferrario,M., Hynes,J.T., and Kapral,R. (1989). Constrained Molecular Dynamics and the Mean Potential for an Ion Pair in a Polar Solvent. Chem. Phys. *129*, 241-251.

Cohen,G.M. (1997). Caspases: the executioners of apoptosis. Biochem. J *326 (Pt 1)*, 1-16.

Colle,R. and Salvetti,D. (1975). Theor. Chim. Acta *37*, 329.

Cornell ,W.D., Cieplack ,P., Bayly ,C.I., Gould ,I.R., Merz ,K.M., Ferguson ,D.M., Spellmeyer ,D.C., and Fox ,T. (1995). J. Am. Chem. Soc. *117*, 5179-5197.

Coulombe,R., Grochulski,P., Sivaraman,J., Menard,R., Mort,J.S., and Cygler,M. (1996). Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. EMBO J *15*, 5492-5503.

Culver,K.W., Ram,Z., Wallbridge,S., Ishii,H., Oldfield,E.H., and Blaese,R.M. (1992). In vivo gene transfer with retroviral vector-producer cells for treatment of experimental brain tumors [see comments]. Science *256*, 1550-1552.

Cummins,P. and Gready,J. (1997). J. Comp. Chem. *18*, 1496.

Curioni,A., Sprik,M., Andreoni,W., Schiffer,H., Hutter,J., and Parrinello,M. (1997). Density Functional Theory-Based Molecular Dynamics Simulation of Acid-Catalyzed Chemical Reactions in Liquid Trioxane. J. Am. Chem. Soc. *119*, 7218-7229.

De Santis,L. and Carloni,P. (1999). Serine Proteases: An *Ab initio* Molecular Dynamics Study. Proteins: Struc. Func Gen. *37*, 611-618.

Deng,Z., Martyna,G.J., and Klein,M.L. (1993). Electronic states in metal-ammonia solutions. PHYSICAL REVIEW LETTERS *71*, 267-270.

DeWitte,R. and Shakhnovich,E. (1996). J. Am. Chem Soc. *118*, 11733-11744.

Diaz,N., Suarez,D., and Merz,K. (2000). Zinc Metallo-beta-Lactamases from Bacterioides fragilis: A Quantum Chemical Study on Model Systems of the Active site. J. Am. Chem Soc. *122*, 4197-4208.

Doucet,J. and Weber,J. (1996). Computer-Aided Molecular Design. Academic Press).

Dougherty,D.A. (1996). Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp. Science *271*, 163-168.

Drenth,J., Jansonius,J.N., Koekoek,R., Swen,H.M., and Wolthers,B.G. (1968). Structure of papain. Nature *218*, 929-932.

Duan,G., Smith,V.H., and Weaver,D.F. (2000). J. Phys. Chem. A *104*, 4521.

Eichinger,M., Parrinello,M., Hutter,J., and Parrinello,M. (1999). J. Chem. Phys. *21*, 10452.

Eisen,M.B., Wiley,D.C., Karplus,M., and Hubbard,R.E. (1994). HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins *19*, 199-221.

Elion,G.B., Furman,P.A., Fyfe,J.A., de Miranda,P., Beauchamp,L., and Schaeffer,H.J. (1977). Selectivity of action of an antiherpetic agent, 9-(2- hydroxyethoxymethyl) guanine. Proc. Natl. Acad. Sci. U. S. A *74*, 5716-5720.

Enari,M., Sakahira,H., Yokoyama,H., Okawa,K., Iwamatsu,A., and Nagata,S. (1998). A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature *391*, 43-50.

Encell,L.P., Landis,D.M., and Loeb,L.A. (1999). Improving enzymes for cancer gene therapy. Nat. Biotechnol. *17*, 143-147.

Ermer,O. and Eling,A. (1994). J. Am. Chem. Soc. , Perkin Trans. 2, 925.

Fasulo,L., Ugolini,G., Visintin,M., Bradbury,A., Brancolini,C., Verzillo,V., Novak,M., and Cattaneo,A. (2000). The neuronal microtubule-associated protein tau is a substrate for caspase-3 and an effector of apoptosis. J. Neurochem. *75*, 624-633.

Fengler,A. and Brandt,W. (1998). Three-dimensional structures of the cysteine proteases cathepsins K and S deduced by knowledge-based modelling and active site characteristics. Protein Eng *11*, 1007-1013.

Fersht,A. (1997). Enzyme structure and mechanism. (New York).

Field,M.J., Bash,P.A., and Karplus,M. (1990). J. Comp. Chem. *11*, 700.

Frisch, M. J. Gaussian 98. *Gaussian, Inc.* 1995. *Pittsburgh PA.*
Ref Type: Computer Program

Furman,P.A., Painter,G.R., and Anderson,K.S. (2000). An analysis of the catalytic cycle of HIV-1 reverse transcriptase: opportunities for chemotherapeutic intervention based on enzyme inhibition. Curr. Pharm. Des *6*, 547-567.

Gillmor,S.A., Craik,C.S., and Fletterick,R.J. (1997). Structural determinants of specificity in the cysteine protease cruzain. Protein Sci. *6*, 1603-1611.

Gilson,M.K., Given,J.A., Bush,B.L., and McCammon,J.A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys. J *72*, 1047-1069.

Glennon,T. and Warshel,A. (1998). Energetics of the catalytic reaction of ribonuclease A: A computational study of alternative mechanisms. J. Am. Chem Soc. *120*, 10234-10247.

Gu,Y. and Scheiner,S. (1999). J. Am. Chem. Soc. *121*, 9411-9422.

Guncar,G., Podobnik,M., Pungercar,J., Strukelj,B., Turk,V., and Turk,D. (1998). Crystal structure of porcine cathepsin H determined at 2.1 A resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. Structure *6*, 51-61.

Guncar,G., Pungercic,G., Klemencic,I., Turk,V., and Turk,D. (1999). Crystal structure of MHC class II-associated p41 Ii fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J *18*, 793-803.

Hehre,W.J., Radom,L., Schleyer,P.V.R., and Polple,J. (1986). *Ab initio* molecular orbital theory., p. 140.

Hockney,R.W. (1970). Methods Comput Phys *9*.

Hohenberg,P. and Kohn,W. (1964). Inhomogeneous electron gas. Phys. Rev. *136*, B864-B871.

Huang,H., Chopra,R., Verdine,G.L., and Harrison,S.C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science *282*, 1669-1675.

Huang,Y., Park,Y.C., Rich,R.L., Segal,D., Myszka,D.G., and Wu,H. (2001). Structural basis of caspase inhibition by XIAP: differential roles of the linker versus the BIR domain. Cell *104*, 781-790.

Hunenberger,P. (2000). Optimal charge-shaping functions for the particle-particle-particle-mesh (P3M) method for computing electrostatic interactions in molecular simulations. J. Chem. Phys. *113*, 10464-10476.

Huskey,S.W., Huskey,W.P., and Lu,A.Y.H. (1991). J. Am. Chem. Soc. *113*, 2283-2290.

Hutter,J., Carloni,P., and Parrinello,M. (1996a). J. Am. Chem. Soc. *118*, 8710.

Hutter, J., Ballone, P., Bernasconi, N., Focher, P., Fois, E., Godecker, S., Parrinello, M., and Tuckerman, M. CPMD. MPI fur Festkorperforschung and IBM Zurich Research Laboratory[3.0h]. 1996.
Ref Type: Computer Program

Jedrzejas,M.J., Singh,S., Brouillette,W.J., Laver,W.G., Air,G.M., and Luo,M. (1995). Structures of aromatic inhibitors of influenza virus neuraminidase. Biochemistry *34*, 3144-3151.

Ji,X., Zhang,P., Armstrong,R.N., and Gilliland,G.L. (1992). The three-dimensional structure of a glutathione S-transferase from the mu gene class. Structural analysis of the binary complex of isoenzyme 3- 3 and glutathione at 2.2-A resolution. Biochemistry *31*, 10169-10184.

Jia,Z., Hasnain,S., Hirama,T., Lee,X., Mort,J.S., To,R., and Huber,C.P. (1995). Crystal structures of recombinant rat cathepsin B and a cathepsin B- inhibitor complex. Implications for structure-based inhibitor design. J Biol. Chem *270*, 5527-5533.

Johnston,S.C., Larsen,C.N., Cook,W.J., Wilkinson,K.D., and Hill,C.P. (1997). Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 A resolution. EMBO J *16*, 3787-3796.

Johnston,S.C., Riddle,S.M., Cohen,R.E., and Hill,C.P. (1999). Structural basis for the specificity of ubiquitin C-terminal hydrolases. EMBO J *18*, 3877-3887.

Jordan,J., Galindo,M.F., Cena,V., and Gonzalez-Garcia,C. (2000). [Cysteine proteinase and neurodegeneration]. Rev Neurol. *31*, 333-340.

Kamphuis,I.G., Drenth,J., and Baker,E.N. (1985). Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. J Mol. Biol. *182*, 317-329.

Kamphuis,I.G., Kalk,K.H., Swarte,M.B., and Drenth,J. (1984). Structure of papain refined at 1.65 A resolution. J Mol. Biol. *179*, 233-256.

Karplus,M. (2000). Aspects of protein reaction dynamics: Deviations from simple behavior. J. Phys. Chem. B *104*, 11-27.

Kashino,S., Tomita,M., and Haisa,M. (1988). Acta Cryst C *44*, 780.

Kaufmann,S.H. and Gores,G.J. (2000). Apoptosis in cancer: cause and cure. Bioessays *22*, 1007-1017.

Keller,P.M., Fyfe,J.A., Beauchamp,L., Lubbers,C.M., Furman,P.A., Schaeffer,H.J., and Elion,G.B. (1981). Enzymatic phosphorylation of acyclic nucleoside analogs and correlations with antiherpetic activities. Biochem. Pharmacol. *30*, 3071-3077.

Kerker,G.P. (1980). J. Phys. C *13*, L189.

Klein,M.L. (2001). Chemistry. Water on the move. Science *291*, 2106-2107.

Kohn,W. and Sham,L.J. (1965). Self-consistent equations including exchange and correlation effects. Phys. Rev. *140*, A1133-A1138.

Kokoris,M.S., Sabo,P., Adman,E.T., and Black,M.E. (1999). Enhancement of tumor ablation by a selected HSV-1 thymidine kinase mutant. Gene Ther. *6*, 1415-1426.

Koppensteiner,W.A. and Sippl,M.J. (1998). Knowledge-based potentials--back to the roots. Biochemistry (Mosc.) *63*, 247-252.

Kortemme,T. and Creighton,T. (1995). Ionisation of cysteine residues at the termini of model alpha-helical peptides. Relevance of unusual thiol pKavalues in proteins of the thoredoxin family. J. Mol. Biol. *253*, 799-812.

Kraulis,P.J. (1991). J. Appl. Cryst. *24*, 946-950.

Kraulis, P. J. MolScript. [2.1]. 1997.
Ref Type: Computer Program

Krystian,S. and Pulay,P. (1994). Chem. Phys. Lett. *229*, 175-180.

Kubinyi,H., Martin,Y., and Folkers,G. (1997). 3D Qsar in Drug Design. Kluwer Academic Publishers).

Kussmann-Gerber,S., Kuonen,O., Folkers,G., Pilger,B.D., and Scapozza,L. (1998). Drug resistance of herpes simplex virus type 1--structural considerations at the molecular level of the thymidine kinase. Eur. J. Biochem. *255*, 472-481.

Kussmann-Gerber,S., Wurth,C., Scapozza,L., Pilger,B.D., Pliska,V., and Folkers,G. (1999). Interaction of the recombinant herpes simplex virus type 1 thymidine kinase with thymidine and aciclovir: a kinetic study. Nucleosides Nucleotides *18*, 311-330.

Laio,A., VandeVondele ,J., and Rothlisberger,U. (2001). A Hamiltonian Electrostatic coupling Scheme for Hybrid Car-Parrinello Molecula Dynamics Simulations. submitted to JCP.

LaLonde,J.M., Zhao,B., Smith,W.W., Janson,C.A., DesJarlais,R.L., Tomaszek,T.A., Carr,T.J., Thompson,S.K., Oh,H.J., Yamashita,D.S., Veber,D.F., and Abdel-Meguid,S.S. (1998). Use of papain as a model for the structure-based design of cathepsin K inhibitors: crystal structures of two papain-inhibitor complexes demonstrate binding to S'-subsites. J Med. Chem *41*, 4567-4576.

Leach,A.R. (1996). Molecular modelling. Principles and applications. (Singapore: Addison Wesley).

Lee,C., Yang,W., and Parr,R.G. (1988). Development of the Colle-Salvetti correlation-energy formulainto a functional of the electron density. Phys. Rev. B *37*, 785-789.

Lee,D., Long,S.A., Adams,J.L., Chan,G., Vaidya,K.S., Francis,T.A., Kikly,K., Winkler,J.D., Sung,C.M., Debouck,C., Richardson,S., Levy,M.A., DeWolf,W.E., Jr., Keller,P.M., Tomaszek,T., Head,M.S., Ryan,M.D., Haltiwanger,R.C., Liang,P.H., Janson,C.A., McDevitt,P.J., Johanson,K., Concha,N.O., Chan,W., Abdel-Meguid,S.S.,

Badger,A.M., Lark,M.W., Nadeau,D.P., Suva,L.J., Gowen,M., and Nuttall,M.E. (2000). Potent and selective nonpeptide inhibitors of caspases 3 and 7 inhibit apoptosis and maintain cell functionality. J. Biol. Chem. *275*, 16007-16014.

Liu,H., Muller-plathe,F., and van Gustern,W. (1996). J. Mol. Biol. *261*, 454.

Liu,S., Ji,X., Gilliland,G.L., Stevens,W.J., and Armstrong,R.N. (1993). Second-Sphere Electrostatic Effects in the Active Site of Glutathione S-Transferase. Observation of an On-Face Hydrogen Bond between the Side Chain of Threonine 13 and the pi-Cloud of Tyrosine 6 and Its Influence on Catalysis. J. Am. Chem. Soc. *115*, 7910-7911.

Liu,X., Zou,H., Slaughter,C., and Wang,X. (1997). DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. Cell *89*, 175-184.

Luque,F.J., Reuter,N., Cartier,A., and Ruiz-Lopez,M.F. (2000). J. Phys. Chem. *104*, 10923.

Lyne,P., Hodosceck,M., and Karplus,M. (1999). J. Phys. Chem. *103*, 3462.

Ma,J.C. and Dougherty,D.A. (1997). Chem. Rev. *97*, 1303-1324.

Martyna,G. and Tuckerman,M. (1999). A reciprocal space based method for treating long range interactions in *ab initio* and force-field-based calculations in clusters. J. Chem. Phys. *110*, 2810-2821.

Marzari,N. and Vanderbilt,D. (1997). Maximally-localized Wannier functions for composite energy bands. Phys. Rev. B *56*, 12847-12865.

McCallum,S.A., Hitchens,T.K., Torborg,C., and Rule,G.S. (2000). Ligand-induced changes in the structure and dynamics of a human class Mu glutathione S-transferase. Biochemistry *39*, 7343-7356.

McCurdy,A., Jimenez,L., Stauffer,D.A., and Dougherty,D.A. (1992). J. Am. Chem. Soc. *114*, 10314-10321.

Mecozzi,S., West,A.P., Jr., and Dougherty,D.A. (1996a). J. Am. Chem. Soc. *118*, 2307-2308.

Mecozzi,S., West,A.P., Jr., and Dougherty,D.A. (1996b). Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. Proc. Natl. Acad. Sci. U. S. A *93*, 10566-10571.

Meijer,E. and Sprik,M. (1998a). *Ab initio* Molecular Dynamics Study of the Reaction of Water with Formaldehyde in Sulfuric Acid Solution. J. Am. Chem. Soc. *120*, 6345-6355.

Meijer,E. and Sprik,M. (1998b). *Ab initio* Molecular Dynamics Study of the Reaction of Water with Formaldehyde in Sulfuric Acid Solution. J. Am. Chem. Soc. *120*, 6345-6355.

Menard,R., Khouri,H.E., Plouffe,C., Laflamme,P., Dupras,R., Vernet,T., Tessier,D.C., Thomas,D.Y., and Storer,A.C. (1991). Importance of hydrogen-bonding interactions involving the side chain of Asp158 in the catalytic mechanism of papain. Biochemistry *30*, 5531-5538.

Minoux,H. and Chipot,C. (1999). J. Am. Chem. Soc. *121*, 10366-10372.

Mittl,P.R., Di Marco,S., Krebs,J.F., Bai,X., Karanewsky,D.S., Priestle,J.P., Tomaselli,K.J., and Grutter,M.G. (1997). Structure of recombinant human CPP32 in complex with the tetrapeptide acetyl-Asp-Val-Ala-Asp fluoromethyl ketone. J. Biol. Chem. *272*, 6539-6547.

Molteni,C., Frank,I., and Parrinello,M. (1999). An Excited State Density Functional Theory Study of the Rhodopsin Chromophore. J. Am. Chem. Soc. *121*, 12177-12183.

Molteni,C. and Parrinello,M. (1998). Glucose in aqueous solution by first principles molecular dynamics. J. Am. Chem Soc. *120*, 2168-2171.

Musil,D., Zucic,D., Turk,D., Engh,R.A., Mayr,I., Huber,R., Popovic,T., Turk,V., Towatari,T., Katunuma,N., and . (1991). The refined 2.15 A X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J *10*, 2321-2330.

Nose,S. (1984). J Chem Phys *81*, 511.

Okamoto,Y., Anan,H., Nakai,E., Morihira,K., Yonetoku,Y., Kurihara,H., Sakashita,H., Terai,Y., Takeuchi,M., Shibanuma,T., and Isomura,Y. (1999). Peptide based interleukin-1 beta converting enzyme (ICE) inhibitors: synthesis, structure activity relationships and crystallographic study of the ICE-inhibitor complex. Chem Pharm. Bull. (Tokyo) *47*, 11-21.

Pantano,S., Alber,F., and Carloni,P. (2000). Proton Dynamics in an Enzyme Substrate: an *Ab initio* Molecular Dynamics Study. J. Mol. Struc. (Theochem).

Perozzo,R., Jelesarov,I., Bosshard,H.R., Folkers,G., and Scapozza,L. (2000). Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. A calorimetric study. J. Biol. Chem. *275*, 16139-16145.

Piana,S. and Carloni,P. (2000). Conformational flexibility of the catalytic asp dyad in HIV-1 protease: An *ab initio* study on the free enzyme [In Process Citation]. Proteins *39*, 26-36.

Piana,S., Carloni,P., and Parrinello,M. (2001). Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. submitted.

Pilger,B.D., Perozzo,R., Alber,F., Wurth,C., Folkers,G., and Scapozza,L. (1999). Substrate diversity of herpes simplex virus thymidine kinase. Impact Of the kinematics of the enzyme. J. Biol. Chem. *274*, 31967-31973.

Pioda,G., Rothlisberger,U., Rothlisberger,U., and Togni,A. (2000). Organometallics *19*, 2144.

Poralla,K. (1994). Bioorg. Med. Chem. Lett *4*, 285-290.

Poralla,K., Hewelt,A., Prestwich,G.D., Abe,I., Reipen,I., and Sprenger,G. (1994). A specific amino acid repeat in squalene and oxidosqualene cyclases. Trends Biochem. Sci. *19*, 157-158.

Porter,A.G. and Janicke,R.U. (1999). Emerging roles of caspase-3 in apoptosis. Cell Death. Differ. *6*, 99-104.

Prota,A., Vogt,J., Perozzo,R., Pilger,B.D., Wurth,C., Marquez,V., Russ,P., Schulz,G., Folkers,G., and Scapozza,L. (2000). Kinetics and high resolution structure of HSV 1 TK and the engeneered Y101F mutant in complex with an antiviral drug with conformationallyrestricted sugar ring pucker. Biochemistry.

Pulay ,P. (1980). Convergence acceleration of iterative sequences. The case of SCF iteration. Chem. Phys. Lett. *73*, 393-398.

Rano,T.A., Timkey,T., Peterson,E.P., Rotonda,J., Nicholson,D.W., Becker,J.W., Chapman,K.T., and Thornberry,N.A. (1997). A combinatorial approach for determining protease specificities: application to interleukin-1beta converting enzyme (ICE). Chem. Biol. *4*, 149-155.

Riedl,S.J., Renatus,M., Schwarzenbacher,R., Zhou,Q., Sun,C., Fesik,S.W., Liddington,R.C., and Salvesen,G.S. (2001). Structural basis for the inhibition of caspase-3 by XIAP. Cell *104*, 791-800.

Rodham,D.A., Suzuki,S., Suenram,R.D., Lovas,R.J., Dasgupta,S., Goddard III,W.A., and Blake,G.A. (1993). Nature *362*, 735.

Rothlisberger, U. Combined Qauntum Mechanical and Molecular Mechanical Methods. J.Gao and M.A.Thompson. ACS Symposium Series 712. 1998. Washington, DC, Am. Chem. Soc.
Ref Type: Serial (Book,Monograph)

Rothlisberger,U. and Carloni,P. (1999). *Ab initio* Molecular Dynamics Studies of a Synthetic Biomimetic Model of Galactose Oxidase. Intl. J. Quant. Chem. *73*, 209-218.

Rothlisberger,U., Carloni,P., Doclo,K., and Parrinello,M. (2000). A comparative study of galactose oxidase and active site analogs based on QM/MM Car Parrinello simulations. J. Biol. Inorg. Chem. *5*, 236-250.

Rotonda,J., Nicholson,D.W., Fazil,K.M., Gallant,M., Gareau,Y., Labelle,M., Peterson,E.P., Rasper,D.M., Ruel,R., Vaillancourt,J.P., Thornberry,N.A., and Becker,J.W. (1996). The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis. Nat. Struct. Biol. *3*, 619-625.

Ryckaert,J.P., Ciccotti,G., and Berendsen,H.J.C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular Dynamics of n-alkanes. J. Comp. Phys. *23*, 327-341.

Sagnella,D.E., Laasonen,K., and Klein,M.L. (1996). *Ab initio* molecular dynamics study of proton transfer in a polyglycine analog of the ion channel gramicidin A. Biophys. J *71*, 1172-1178.

Salvesen,G.S. and Dixit,V.M. (1997). Caspases: intracellular signaling by proteolysis. Cell *91*, 443-446.

Schneider,H.J. (2001). Angew. Chem. Int. Ed. Engl *30*, 1417-1436.

Schneider,H.J., Schiestel,T., and Zimmermann,P. (1992). J. Am. Chem. Soc. *114*, 7698-7703.

Scrutton,N.S. and Raine,A.R. (1996). Cation-pi bonding and amino-aromatic interactions in the biomolecular recognition of substituted ammonium ligands. Biochem. J *319 (Pt 1)*, 1-8.

Segall,M., Payne,M., Ellis,S.W., Tucker,G., and Boyes,R. (1998b). Phys Rev E *57*, 4618.

Segall,M., Payne,M., Ellis,S.W., Tucker,G., and Boyes,R. (1998a). Xenobiotica *28*, 15.

Segall,M., Payne,M., Ellis,S., Tucker,G., and Boyes,R. (1998c). N Chem Res Toxicol *11*, 962.

Segall,M., Payne,M., Ellis,S., Tucker,G., and Eddershaw,P. (1999). Xenobiotica *29*, 561.

Sherwood,P. (2000). Modern Methods and Algorithm of Quantum Chemistry., J.Grotendorst, ed. (Julich: John von Neumann Institute for Computing), p. 257.

Shi,Z., Buntel,C.J., and Griffin,J.H. (1994). Isolation and characterization of the gene encoding 2,3-oxidosqualene- lanosterol cyclase from Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A *91*, 7370-7374.

Shimohama,S. (2000). Apoptosis in Alzheimer's disease--an update. Apoptosis. *5*, 9-16.

Shoemaker,K.R., Fairman,R., Schultz,D.A., Robertson,A.D., York,E.J., Stewart,J.M., and Baldwin,R.L. (1990). Side-chain interactions in the C-peptide helix: Phe 8 ... His 12+. Biopolymers *29*, 1-11.

Shriver,D.F., Atkins,P., and Langford,C.H. (2001). *Inorganic Chemistry*. (New York: Freeman), p. 58.

Silvestrelli,P.L., Marzari,N., Vanderbilt,D., and Parrinello,M. (1998). Maximally-localized Wannier functions for disordered systems: application to amorphous silicon. Solid State Commun *107*, 7-11.

Silvestrelli,P.L. and Parrinello,M. (1999a). Structural, electronic, and bonding properties of liquid water from first principles. J. Chem Phys. *111*, 3572-3580.

Silvestrelli,P.L., Bernasconi,M., and Parrinello,M. (1997). *Ab initio* infrared spectrum of liquid water. Chem. Phys. Lett. *277*, 478-482.

Silvestrelli,P.L. and Parrinello,M. (1999b). Phys. Rev. Lett. *82*, 3308-3311.

Singh,U.C. and Kollman,P.A. (1986). J. Comp. Chem. *7*, 718.

Stanton,R.V., Little,L.R., and Merz,K.M. (1996). J. Phys. Chem. *99*, 11266.

Stauffer,D.A., Barrans,J.R.E., and Dougherty,D.A. (1990). Angew. Chem. Int. Ed. Engl *29*, 915-918.

Stennicke,H.R. and Salvesen,G.S. (1997). Biochemical characteristics of caspases-3, -6, -7, and -8. J. Biol. Chem. *272*, 25719-25723.

Strajbl,M. and Warshel,A. (2001). *Ab initio* evaluation of the Free energy Surfaces for general Base/Acid Catalyzed Thiolysis of Formammide and the Hydrolysis of Methyl Thiolformate: A Reference Solution Reaction for studies of Cysteine Proteases. J. Phys. Chem. B *105*, 4471-4484.

Suzuki,S., Green,P.G., Bumgarner,R.E., Dasgupta,S., Goddard III,W., and Blake,G.A. (1992). Science *257*, 942.

Swope,W.C., Andersen,H.C., Berens,P.H., and Wilson,K.R. (1982). J Chem Phys *76*, 637.

Talanian,R.V., Brady,K.D., and Cryns,V.L. (2000). Caspases as targets for anti-inflammatory and anti-apoptotic drug discovery. J Med. Chem *43*, 3351-3371.

Tarakeshwar,P., Choi,H., Lee,S., Lee,J., Kim,K., Ha,T., Jang,J., Lee,J., and Lee,H. (1999). A theoretical investigation of the nature of the pi-H interaction in ethene-H2O, benzene-H2O, and benzene-(H2O). J. Chem Phys. *111*, 5838-5850.

Thompson,S.K., Halbert,S.M., Bossard,M.J., Tomaszek,T.A., Levy,M.A., Zhao,B., Smith,W.W., Abdel-Meguid,S.S., Janson,C.A., D'Alessio,K.J., McQueney,M.S., Amegadzie,B.Y., Hanning,C.R., DesJarlais,R.L., Briand,J., Sarkar,S.K., Huddleston,M.J., Ijames,C.F., Carr,S.A., Garnes,K.T., Shu,A., Heys,J.R., Bradbeer,J., Zembryki,D., Veber,D.F., and . (1997). Design of potent and selective human cathepsin K inhibitors that span the active site. Proc. Natl. Acad. Sci. U. S. A *94*, 14249-14254.

Troullier,N. and Martins,J.L. (1991). Efficient pseudopotentials for plane-wave calculation. Phys. Rev. B *43*, 1943-2006.

Tuckerman,M.E., Marx,D., Klein,M.L., and Parrinello,M. (1997). On the Quantum Nature of the Shared Proton in Hydrogen Bonds. Science *275* , 817-820.

Tunon,I., Martins-Costa,M.T.C., Millot,C., and Ruiz-Lopez,M.F. (1995). J. Mol. Model. *1*, 196.

Turk,D., Podobnik,M., Popovic,T., Katunuma,N., Bode,W., Huber,R., and Turk,V. (1995). Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. Biochemistry *34*, 4791-4797.

van Gunsteren,W.F., Billeter,S.R., Eising,A.A., Hünenberger,P.H., Krüger,P.K.H.C., Mark,A.E., Scott,W.R.P., and Tironi,I.G. (1996). Biomolecular Simulation: The GROMOS96 Manual and User Guide . (Zürich: vdf Hochschulverlag AG).

Vanderbilt,D. (1985). Optimally smooth norm-conserving pseudopotentials. Phys. Rev. B *32*, 8412.

Varnai,P. and Warshel,A. (2000). Computer simulation studies of the catalytic mechanism of human aldose reductase. J. Am. Chem Soc. *122*, 3849-3860.

Ventimiglia,R., Lau,L., Kinloch,R., Hopkins,A., Karran,E., Petalidis,L., and Ward,R. (2001). Role of caspases in neuronal apoptosis. DRUG DEVELOPMENT RESEARCH *52*, 525-533.

VereHodge,R.A. (1993). Famciclovir and Penciclovir . The mode of action of famciclovir including its conversion to penciclovir. Antiviral Chemistry and Chemotherapy *4*, 67-84.

Villa,J., Strajbl,M., Glennon,T., Sham,Y., Chu,Z., and Warshel,A. (2000). How important are entropic contributions to enzyme catalysis? Proc. Natl. Acad. Sci. U. S. A *97*, 11899-11904.

Waksman,G., Kominos,D., Robertson,S.C., Pant,N., Baltimore,D., Birge,R.B., Cowborn,D., Hanafusa,H., Mayer,B.J., Overduin,M., Resh,M.D., Rios,C.B., Silverman,L., and Kurijan,J. (1992). Nature *358*, 646.

Walker,N.P., Talanian,R.V., Brady,K.D., Dang,L.C., Bump,N.J., Ferenz,C.R., Franklin,S., Ghayur,T., Hackett,M.C., Hammill,L.D., and . (1994). Crystal structure of the cysteine protease interleukin-1 beta- converting enzyme: a (p20/p10)2 homodimer. Cell *78*, 343-352.

Wang,W., Donini,O., Reyes,C., and Kollman,P. (2001). Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. ANNUAL REVIEW OF BIOPHYSICS AND BIOMOLECULAR STRUCTURE *30*, 211-243.

Warshel,A. (1998). Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. J. Biol. Chem. *273*, 27035-27038.

Warshel,A. and Levitt,M. (1976). Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol. Biol. *103*, 227-249.

Watt,W., Koeplinger,K.A., Mildner,A.M., Heinrikson,R.L., Tomasselli,A.G., and Watenpaugh,K.D. (1999). The atomic-resolution structure of human caspase-8, a key activator of apoptosis. Structure. Fold. Des *7*, 1135-1143.

Wei,D. and Salahub,D. (1994). Chem. Phys. Lett. *224*, 291.

Wei,Y., Fox,T., Chambers,S.P., Sintchak,J., Coll,J.T., Golec,J.M., Swenson,L., Wilson,K.P., and Charifson,P.S. (2000). The structures of caspases-1, -3, -7 and -8 reveal the basis for substrate and inhibitor selectivity. Chem Biol. *7*, 423-432.

Wild,K., Bohner,T., Aubry,A., Folkers,G., and Schulz,G.E. (1995). The three-dimensional structure of thymidine kinase from herpes simplex virus type 1. FEBS Lett. *368*, 289-292.

Wild,K., Bohner,T., Folkers,G., and Schulz,G.E. (1997). The structures of thymidine kinase from herpes simplex virus type 1 in complex with substrates and a substrate analogue. Protein Sci. *6*, 2097-2106.

Wilson,K.P., Black,J.A., Thomson,J.A., Kim,E.E., Griffith,J.P., Navia,M.A., Murcko,M.A., Chambers,S.P., Aldape,R.A., Raybuck,S.A., and . (1994). Structure and mechanism of interleukin-1 beta converting enzyme. Nature *370*, 270-275.

Xiao,G., Liu,S., Ji,X., Johnson,W.W., Chen,J., Parsons,J.F., Stevens,W.J., Gilliland,G.L., and Armstrong,R.N. (1996). First-sphere and second-sphere electrostatic effects in the active site of a class mu gluthathione transferase. Biochemistry *35*, 4753-4765.

Xu,G., Cirilli,M., Huang,Y., Rich,R.L., Myszka,D.G., and Wu,H. (2001). Covalent inhibition revealed by the crystal structure of the caspase- 8/p35 complex. Nature *410*, 494-497.

Xu,H., Maga,G., Focher,F., Smith,E.R., Spadari,S., Gambino,J., and Wright,G.E. (1995). Synthesis, properties, and pharmacokinetic studies of N2-phenylguanine derivatives as inhibitors of herpes simplex virus thymidine kinases. J. Med. Chem. *38*, 49-57.

Yamamoto,A., Tomoo,K., Doi,M., Ohishi,H., Inoue,M., Ishida,T., Yamamoto,D., Tsuboi,S., Okamoto,H., and Okada,Y. (1992). Crystal structure of papain-succinyl-Gln-Val-Val-Ala-Ala-p-nitroanilide complex at 1.7-A resolution: noncovalent binding mode of a common sequence of endogenous thiol protease inhibitors. Biochemistry *31*, 11305-11309.

Yamamoto,A., Tomoo,K., Hara,T., Murata,M., Kitamura,K., and Ishida,T. (2000). Substrate specificity of bovine cathepsin B and its inhibition by CA074, based on crystal structure refinement of the complex. J Biochem. (Tokyo) *127*, 635-643.

Yamashita,D.S., Smith,W.W., Zhao,B.G., Janson,C.A., Tomaszek,T.A., Bossard,M.J., Levy,M.A., Oh,H.J., Carr,T.J., Thompson,S.K., Ijames,C.F., Carr,S.A., McQueney,M., DAlessio,K.J., Amegadzie,B.Y., Hanning,C.R., AbdelMeguid,S., DesJarlais,R.L., Gleason,J.G., and Veber,D.F. (1997). Structure and design of potent and selective cathepsin K inhibitors. J Am Chem Soc *119*, 11351.

Zhao,B., Janson,C.A., Amegadzie,B.Y., D'Alessio,K., Griffin,C., Hanning,C.R., Jones,C., Kurdyla,J., McQueney,M., Qiu,X., Smith,W.W., and Abdel-Meguid,S.S. (1997). Crystal structure of human osteoclast cathepsin K complex with E-64. Nat. Struct. Biol. *4*, 109-111.

Zheng,Y.J. and Ornstein,R.L. (1997). J. Am. Chem. Soc. *119*, 1523-1528.