SISSA  **ISAS**

SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI -
INTERNATIONAL SCHOOL FOR ADVANCED STUDIES
Via Beirut 4, 34014 Trieste, Italy

Application of novel molecular simulation approaches for the investigation of structure and dynamics of biomolecules

Thesis submitted for the degree of *Doctor Philosophiae*

Candidate: Pietro Vidossich

Supervisor: Prof. Paolo Carloni

December 2005

INDEX

1. ROLE OF BIOMOLECULAR SIMULATIONS FOR STRUCTURAL GENOMICS

The sequencing of the genome of human and other organisms and the growing impact of structural biology on biomedical research have prompted scientific groups from several countries to undertake projects in the emerging field of structural genomics. Computational biology tools play a key role in this framework by expanding the knowledge of the role of proteins both in normal biological processes and in disease. *Biology-based computations*, such as those of structural bioinformatics, by using evolutionary concepts are able to extend the domain of structural biology to many more proteins whose structure has been solved by X-ray crystallography or by NMR. *Physics-based computations,* such as biomolecular simulation, simulate biomolecular motion according to the laws of physics, providing quantitative information on biomolecular dynamics and energetics as well as helping interpret biophysical data (e.g. spectroscopic data). This information is more and more recognized as the key to understand biological function(Karplus and McCammon, 2002;Allen et al., 2004). With the progress of computational techniques, combined with the power of parallel computing, the domain of application of molecular simulations has been constantly increased for the last two decades. New approaches are the focus of this thesis.

Biomolecular simulations span a broad range of computational techniques. In all cases, the accuracy depends dramatically on the potential energy function used to describe the molecular interactions in the system.

Biomolecular effective potentials (or force fields) targeting proteins have become more and more accurate in the last years, and are well suited to describe large systems at equilibrium(Mackerell, 2004). Empirical force fields-based molecular dynamics simulation are routinely used to refine X-ray and NMR structures, to follow the dynamics of proteins

and their adducts with other proteins or nucleic acids[1], to calculate ligand-receptor binding free energies and search conformational space.

Processes depending on the quantum nature of the electronic degrees of freedom (e.g. enzymatic reactions or the calculation of molecular electronic properties such as polarizabilty and NMR parameters) cannot be described by this relatively simple approach. One has to use quantum chemistry methods, which clearly are much more demanding computationally. Mixed first-principles/classical molecular dynamics simulations(Laio A. et al., 2002), which combine the first-principles molecular dynamics method of Car and Parrinello(Car and Parrinello, 1985) with empirical force-fields, allow to perform molecular dynamics calculations in which one part of the protein (the QM region, typically chosen as the chemically active site) is treated with higher accuracy, usually at the Density Functional level(Hohenberg and Kohn, 1964), while the rest of the protein (the so-called MM region) is treated by molecular mechanics. The QM/MM Hamiltonian include Van der Waals, electrostatics and polarization effects of the environment on the electronic properties of the active site.

In this thesis I have used recent developments of computational techniques to gain insight into biomolecular structure and function.

The first application regards the simulations of rare events. Among the recent methods that have been proposed to calculate free energy changes of slow processes along a predetermined reaction coordinate(Rodinger and Pomes, 2005), is Multiple Steering Molecular Dynamics (MSMD)(Isralewitz et al., 2001). In this approach, the system is pulled along the reaction coordinate by an external potential. The method is based on Jarzynski's identity(Jarzynski, 1997), relating the free energy change to the average of the

[1]currently, MD simulations cover tipically few tens ns.

accumulated work over an ensemble of transformations. Here we have applied such an approach to calculate the free-energy profile of water permeation through the aquaporin-1 channel. Classical MD simulations were used to investigate structural properties of water inside the channel. The calculations, which are based on the recently determined x-ray structure at 2.2 Å resolution(Sui et al., 2001), were carried out on one monomeric subunit immersed in a water-membrane mimic bilayer.

The second application focuses on the quantum chemical calculations of NMR parameters (Helgaker et al., 1999). Many empirical rules relate chemical shifts to structural properties(Xu and Case, 2001), yet very interesting systems display unusual resonance lines for which such parametrization cannot be used. In this case, ab initio, parameter free, calculation of molecular electronic properties can help in assigning experimental features and explaining their microscopic origin. Here we have performed hybrid QM/MM Car-Parrinello molecular dynamics simulations to investigate the hydrogen bonding pattern at the binding site of the complexes of Aspartic Proteases with a class of potent inhibitors, that of phosphinate and phosphonate derivatives. These inhibitors form H-bonds (usually referred to as Low Barrier Hydrogen Bond (LBHB)) in which the interoxygen distance between the catalytic Asp and the PO_2 moiety is as small as ~2.5 Å. 1H-NMR is a powerful probe for detecting LBHB's, as the chemical shifts are characterized by highly downfield signals. Our calculations are in fair agreement with the 1H-NMR data available for endothiapepsin, thus validating the computational setup used, and provide a picture of the H-bond pattern at the binding site.

We have then extended our investigation to DNA and RNA. Recent NMR experiments(Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b) on DNA and RNA duplexes have revealed trans-hydrogen bond Deuterium Isotope Effects (DIE's) on the chemical shifts of carbon atoms in A:T/U base pairs. These experiments suggested

that the hydrogen bonds of A:U base pairs of RNA are stronger than those of A:T base pairs of DNA and that different base pair conformations may modulate the transmission of the isotope effect across the hydrogen bond. To investigate these findings, classical Molecular Dynamics (MD) has been used to sample conformational preferences in selected DNA and RNA duplexes, while ab-initio calculations has been used to investigate the intrinsic interaction energy of A:T and A:U base pairs and to calculated the DIE on chemical shifts. Moreover, QM/MM and large scale DFT calculations, comprising a triplet of base pairs, were performed to assess the effect of the duplex environment.

The hybrid QM/MM calculations could be carried out thanks to an existing collaboration between SISSA and the EPFL lab directed by U. Rothlisberger. Thus, I wish to thank her and her group for continuos support in the use of the Code.

The following chapter briefly describes the methods used in the calculations. Chapters 3, 4 and 5 each describe one of the above mentioned studies. Each chapter include an introduction to the problem studied, focussing on its biological relevance. Finally, Chapter 6 summarizes the main results.

1.1 REFERENCES

Allen,T.W., O.S.Andersen, and B.Roux. 2004. On the importance of atomic fluctuations, protein flexibility, and solvent in ion permeation. *Journal of General Physiology* 124:679-690.

Car,R. and M.Parrinello. 1985. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letter.* 55:2471-2474.

Helgaker,T., M.Jaszunski, and K.Ruud. 1999. Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants. *Chemical Reviews* 99:293-352.

Hohenberg,P and W.Kohn. 1964. Inhomogeneous Electron Gas. *Phys. Rev. B* 136:864-871.

Isralewitz,B., M.Gao, and K.Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* 11:224-230.

Jarzynski,C. 1997. Nonequilibrium equality for free energy differences. *Physical Review Letters* 78:2690-2693.

Karplus,M. and J.A.McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* 9:646-652.

Laio A., VandeVondele J., and Roethlisberger U. 2002. A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116:6941.

Mackerell,A.D. 2004. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* 25:1584-1604.

Rodinger,T. and R.Pomes. 2005. Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Current Opinion in Structural Biology* 15:164-170.

Sui,H., B.G.Han, J.K.Lee, P.Walian, and B.K.Jap. 2001. Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414:872-878.

Vakonakis,I. and A.C.LiWang. 2004a. N1...N3 hydrogen bonds of A:U base pairs of RNA are stronger than those of A:T base pairs of DNA. *J. Am. Chem. Soc.* 126:5688-5689.

Vakonakis,I. and A.C.LiWang. 2004b. Trans-hydrogen bond deuterium isotope effects of A:T base pairs in DNA. *J. Biomol. NMR* 29:65-72.

Xu,X.P. and D.A.Case. 2001. Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13 ' chemical shifts in proteins using a density functional database. *Journal of Biomolecular Nmr* 21:321-333.

2. COMPUTATIONAL METHODS

Computer simulations at different level of accuracy are able to predict several properties of biomolecules, otherwise inaccessible by experiments. The theory and the implemented algorithms used throughout this thesis are described below. The approximations and the limitations of these methods when applied to the biomolecular systems are also discussed. In the following, the energy functions used in this thesis will be presented. Then, the principles of Molecular Dynamic simulations for the sampling of configurational space will be outlined.

2.1. ENERGY FUNCTIONS

Knowledge of the energy of the system at a given nuclear configuration is of fundamental importance in molecular modeling. This may be obtained through the solution of the time independent Schroedinger equation. The degree of accuracy at which this may be solved depends on the number of atoms involved. Density Functional Theory (DFT) is a useful approximation to treat systems up to hundreds of atoms. When bigger systems are of interest, empirical potentials are the only feasible alternative.

2.1.1. The Schroedinger equation

In the non relativistic quantum theory, the time independent Schroedinger equation gives the energy of a stationary state constituted by M atoms and N electrons(McWeeny and Sutcliffe, 2005):

$$H\psi = E\psi$$

where the Hamiltonian operator is:

$$H = T + V_{ext} + V_{ee}$$

T is the kinetic operator and the external potential V_{ext} is determined from the atomic configuration :

$$V_{ext} = -\sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_A}{|\mathbf{R}_I - \mathbf{r}_i|}$$

Vee is the electron-electron repulsion operator

$$V_{ext} = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{|\mathbf{r}_j - \mathbf{r}_i|}.$$

Within the Born-Oppenheimer approximation, nuclear and electronic degrees of freedom are separated and the problem is reduced to the calculation of the energy of the system for a given nuclear configuration. The electronic wavefunction

$$\psi = \psi(\mathbf{r_1}, \mathbf{r_2}, ...\mathbf{r_N})$$

is in principle dependent also on spin coordinates. As in this work we will always deal with closed-shell atomic systems, we will assume double occupancy of any electron energy level and will neglet the spin dependency in the Hamiltonian.

2.1.2. Density functional theory

Density functional theory (DFT) is a rigorous theory of the ground state of a many particle system. Hohenberg and Kohn in the mid-1960s gave a elegant mathematical formulation that was able to replace the Schrödinger equation for the N-electron wavefunction with a theoretical scheme involving the electron density $\rho(\mathbf{r})$ alone:

$$\rho(\mathbf{r}_1) = N \int \left|\psi(\mathbf{r}_1, \mathbf{r}_2 ... \mathbf{r}_N)\right|^2 d\mathbf{r}_2 d\mathbf{r}_3 ..... d\mathbf{r}_N$$

.

2.1.2.1. The Hohenberg-Kohn theorems

The first Hohenberg and Kohn theorem(hohenberg p., 1964) states the uniqueness of the ground-state density associated with a given external potential V_{ext} (within an additive constant). As V_{ext} univocally determines the Hamiltonian of the system, it follows that the ground state energy and all the observables are functionals of the electron density $\rho(\mathbf{r})$ alone.

The second theorem(hohenberg p., 1964) states that in the ground state for any trial density $\tilde{\rho}$, for which $\tilde{\rho} > 0$ and $\int \tilde{\rho}(\mathbf{r}) d\mathbf{r} = N$

$$E_0 = \min_{\tilde{\rho}} E_v\left[\rho\right]$$

with

$$E_v\left[\rho\right] = T\left[\rho\right] + E_{ee}\left[\rho\right] + E_{ne}\left[\rho\right] = \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{HK}\left[\rho\right]$$

where $v(r)$ is the external potential and $F_{HK}\left[\rho\right] = \left\langle \psi\left[\rho\right] \middle| H_e \middle| \psi\left[\rho\right] \right\rangle$ is the Hohenberg and Kohn functional. $H_e = \sum_{i=1}^{N} \left(-\frac{\nabla_i^2}{2} \right) + J\left[\rho\right] + Q\left[\rho\right]$ besides the kinetics energy operator, includes the classical and non-classical electron-electron repulsion terms. Thus the variational principle can be used to find the ground state density for any given nuclear configuration: $\rho(\mathbf{r})$ has to satisfy the stationary principle

$$\delta\left\{ E_v\left[\rho\right] - \mu\left[\int \rho(\mathbf{r}) d\mathbf{r} - N \right] \right\} = 0$$

leading to the following Euler-Lagrange equation

$$\mu = v(\mathbf{r}) + \frac{\partial F_{HK}[\rho]}{\partial \rho(\mathbf{r})}$$

F_{HK} does not depend from the external potential and can be assumed as an universal functional of the electronic density. Although the DFT is a rigorous theory, F_{HK} is not known analytically.

2.1.2.2. The Kohn-Sham equations

A practical and accurate approach to the determination of the ground state energy was introduced in 1965 by Kohn and Sham(kohn w. and sham l.j., 1965). They indirectly solved the problem mapping the system of interest in a non-interacting system with the same density and Hamiltonian

$$H_s = \sum_i^N \left(-\frac{\nabla_i^2}{2} \right) + \sum_i^N v_s(\mathbf{r}_i)$$

for which the density and the kinetics energy are well defined in terms of the KS orbitals:

$$\rho(\mathbf{r}) = \sum_i^N \left| \psi_i^{KS}(\mathbf{r}) \right|^2,$$

$$T_s[\rho] = \sum_i^N \left\langle \psi_i^{KS} \left| -\frac{1}{2}\nabla^2 \right| \psi_i^{KS} \right\rangle$$

The residual correction for kinetic energy of the interacting system is handled separately in F_{HK}, in fact:

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

where

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

is the exchange-correlation energy that includes the correction for the kinetic energy and the non classical part of the electron repulsion. The Euler-Lagrange equation turns out to be:

$$\mu = v_{eff}(\mathbf{r}) + \frac{\partial T_s[\rho]}{\partial \rho(\mathbf{r})}$$

with

$$v_{eff}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})}$$

The density of the new system of non-interacting electrons in a external potential v_{eff} corresponds to the initial system density. The corresponding total energy is

$$E = \sum_i^N \varepsilon_i - \frac{1}{2} \int \frac{\rho(\mathbf{r}')\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho] - \int \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} d\mathbf{r}$$

where ε_i are the KS orbital energies of the non-interacting system. According to the variational principle, the ground state can be obtained solving the Kohn-Sham equations

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

that are non linear and have to be solved self-consistently, as v_{eff} depends on the electronic density. The KS equations are in principle exact but the exchange-correlation functional is not known and one has to resort to approximated forms.

2.1.2.3. Exchange-correlation functionals

A successful approximation for the exchange-correlation energy is based on the exchange-correlation energy of the uniform electron gas. Within the so called local density approximation (LDA) (kohn w. and sham l.j., 1965)

$$E_{xc}^{LDA} = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{\frac{1}{3}}\int \rho^{4/3}(\mathbf{r})d\mathbf{r}$$

LDA calculations, although successful in the description of many systems, exhibit some deficiencies when studying hydrogen bonded systems like those of biological interest. To improve the accuracy, it is necessary to go beyond LDA and explicitly include the dependency upon the gradient of the density in E_{xc} functional. These corrections are called general gradient approximations (GGA).

2.1.2.4. Becke exchange functional

This correction(Becke, 1988) was introduced to reproduce the exact asymptotic behavior of the exchange energy and reads

$$E_x[\rho] = E_{xc}^{LDA}[\rho] - \beta \int \rho^{4/3}(\mathbf{r}) \frac{x^2}{1 + 6\beta \sinh^{-1} x} d\mathbf{r}$$

with

$$x = \frac{|\nabla \rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}$$

the parameter β is determined by a fit on the exact HF data and was fixed by Becke to be 0.0042 au.

2.1.2.5. Lee Yang Parr correlation functional

The LYP functional(Lee et al., 1988) for the correlation energy was derived from the Colle-Salvetti formula to calculate the correlation energies from HF second order density matrix, and it reads

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

where $C_F = \frac{3}{10}\left(3\pi^2\right)^{2/3}$, $t_W(\mathbf{r}) = \frac{1}{8}\frac{|\nabla \rho(\mathbf{r})|^2}{\rho(\mathbf{r})} - \frac{1}{8}\nabla^2 \rho(\mathbf{r})$, and a=0.04918, b=0.132, c=2533 and d=0.349 are parameters derived by fitting the HF results for the helium atom.

The BLYP recipe for the exchange-correlation functional are used here for all the DFT calculations, as its has been successfully used in a variety of H-bonded biological systems.

2.1.2.6. Plane waves basis set

The use of a plane waves basis set is convenient for solving the Kohn-Sham equations in a periodic system(Pickett, 1989). Each KS orbital can be written as

$$\psi_i^{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{g}} c_i^{\mathbf{k}}(\mathbf{g}) e^{i\mathbf{g}\cdot\mathbf{r}}$$

where V is the volume of the cell, **k** is a Bloch vector in the first Brillouin zone, **g** is a reciprocal lattice vector, c_i are the expansion coefficients and the summation is extended to infinite lattice vectors. 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}, that determines the accuracy of the calculations:

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

In the treatment of isolated clusters such as biological systems, with a low degree of symmetry, the Γ point approximation (**k**=0), offers a good accuracy with a reduction of the computational cost (the Fourier components are real).

Because of the non-local nature of the plane waves, usually a large amount of basis functions is necessary. On the other hand, the use of plane waves permits the use of Fast Fourier Transform (FFT) techniques, which are computationally very efficient.

2.1.2.7. Pseudopotentials

Plane waves are highly inefficient to describe the core electron wave functions. Indeed, they exhibit sharp spatial oscillations in closeness of the atomic nuclei, and a very large set of plane waves is required for an accurate chemical description. On the other hand, 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 frozen in the KS equations and only the valence electrons are described explicitly. Their interactions with nuclei and core electrons are described by using effective potentials, the pseudopotentials(Pickett, 1989).

Pseudopotentials are usually derived from all electron (AE) atomic calculations, and several types have been proposed to date. Here, we used norm-conserving pseudopotentials derived with the Martins-Troullier method(Troullier and Martins, 1991). These pseudopotentials are of the *smooth norm-conserving* type:

i) they contain no nodes;

ii) beyond a chosen cutoff radius r_{cl}, the normalized atomic radial pseudo-wave function R_l^{PP} is equal to the normalized radial all electron wave function R_l^{AE};

iii) the charge enclosed within r_{cl} is equal to the all electron charge

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

iv) the valence pseudopotential eigenvalues are equal to the valence all electron eigenvalues $\varepsilon_l^{PP} = \varepsilon_l^{AE}$;

v) at $r = r_{cl}$ the pseudo wavefunction and its first four derivatives should be continuous;

vi) the pseudopotential should have zero curvature at the origin.

The pseudo-wavefunction $\phi^{PP}=R_l^{PP}(r)Y_{l,m}(\Omega)$ is generated so that the radial part has the general form

$$R_l^{PP}(r) = \begin{cases} R_l^{AE}(r); r > r_{cl} \\ r^l e^{p(r)}; r \le r_{cl} \end{cases}$$

$$p(r) = c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}$$

where the coefficients of the polynomial *p(r)* are obtained by imposing the above conditions. The pseudopotential is calculated by an inversion of the KS equation and is unscreened by subtracting the Hartree and exchange-correlation contribution due to the valence electrons. Non linear core corrections may be used to account for the non linarity of the exchange correlation energy.

The functional form of the pseudopotential is

$$V^{PP}(r, r') = V_{loc}^{PP}(r) + \sum_{lm} V_l^{PP}(r) \mid Y_{lm} >< Y_{lm} \mid$$

where Y_{lm} are spherical harmonics. The semilocality of this functional form (local in the radial coordinate, non local in the angular ones) implies an increase in the computational cost. This difficulty is overcome by the method of Kleinman and Bylander(Kleinman and Bylander, 1982), which leads to a new functional form where the local and non local parts can be completely separated.

2.1.2.8. Electrostatic energy

Periodic boundary conditions are often used to minimize boundary effect and to mimic the presence of the bulk. In this scheme, a particle interacts with all other N-1 particles into the box and with all the images in an infinite array of periodic cells(Allen and Tildsley, 1987).

To simplify the calculations first is useful to split the energy function in two parts: a short-range term (involving atoms within a sphere of radius r_c: $r < r_c < L/2$) and a long-range term (for atoms that are more distant than r_c). The intramolecular and Lennard-Jones interactions can be treated as short-range interactions, since its energy functions decay rapidly. Even if these potential functions are not rigorously zero for $r \geq r_c$, truncation in the calculation at r_c will result in a systematic error which may be correct adding a tail contribution. For Coulomb interactions, instead, the tail correction diverges since the electrostatic energy function goes to zero only for r->∞.

The electrostatic energy of a system of nuclear charges Z_I at R_I and electronic charge density $\rho(\mathbf{r})$ is:

$$E_{electrostatic} = \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \sum_I \int d\mathbf{r} V_{core}^I(\mathbf{r})\rho(\mathbf{r}) + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|R_I - R_J|}$$

The Ewald method is an efficient way to treat long range interactions in periodic systems. The idea of the Ewald method is to add and subtract a diffuse charge distribution, chosen to be a Gaussian charge istribution, at each nuclear site so that the electrostatic energy may be written as the sum of two rapidly converging terms:

$$E_{electrostatic} = \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}' \frac{\rho_{tot}(\mathbf{r})\rho_{tot}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{1}{2}\sum_{I,J,\mathbf{L}} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J - \mathbf{L}|} erfc[\alpha|\mathbf{R}_I - \mathbf{R}_J - \mathbf{L}|] - \sum_I \frac{\alpha Z^2{}_I}{\sqrt{2\pi}}$$

where $\rho_{tot}(r) = \rho(r) + \sum_I \frac{Z_I \alpha^3}{\sqrt{\pi^3}} \exp[-(\alpha(\mathbf{r}-\mathbf{R}_I))^2]$, α is the width of the Gaussian charge distribution and erfc(x) the complementary error function. The first term is is calculated

from the solution of the Poisson's equation in reciprocal space, while the second term is evaluated in real space.

2.1.2.9. Electrostatic decoupling in clusters

The simulation of isolated clusters within a periodic boundary condition scheme needs some care, as self-interaction among replicas has to be cancelled. In our calculations, we have used the procedure developed by Martyna and Tuckerman (Martyna and Tuckerman, 1999) that allows treating the system as isolated.

The average electrostatic energy:

$$< \phi >= \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}') \rho(\mathbf{r}) \rho(\mathbf{r}')$$

can be represented in reciprocal space as:

$$< \phi >= \frac{1}{2V} \sum_{g} | \overline{\rho}(\mathbf{g}) |^2 \overline{\phi}(-g)]$$

where $\overline{\phi}(\mathbf{g})$ and $\overline{\rho}$ are the Fourier series expansion of the potential and the density, respectively. Writing the electrostatic potential as the sum of two arbitrary functions:

$$\phi(\mathbf{r}) = \phi^{long}(\mathbf{r}) + \phi^{short}(\mathbf{r})$$

such that $\phi^{short}(\mathbf{r})$ vanishes exponentially quickly at large distances from the system, while $\phi^{long}(\mathbf{r})$ contains all the long-range components of the physical potential, the electrostatic

energy for a cluster can be rewritten:

$$< \phi >= \frac{1}{2V} \sum_{g} | \bar{\rho}(\mathbf{g}) |^2 \, [\phi(-\mathbf{g}) + \phi^{screen}(-\mathbf{g})]$$

The screen function ϕ^{screen}, defined as:

$$\phi^{screen}(\mathbf{g}) = \bar{\phi}^{long}(\mathbf{g}) - \phi^{long}(\mathbf{g})$$

is the difference between the Fourier series $\bar{\phi}^{long}(\mathbf{g})$ and the Fourier transform $\phi^{long}(\mathbf{g})$ of the long-range potential. The screen function has the meaning of “screening” the interaction of the system with an infinite array of periodic images, thus canceling the self-interaction among replicas of the system.

2.1.3. Empirical force fields

Biomolecular force fields are generally based on the assumption that the energy may be expressed as a sum two body terms(Mackerell, 2004). This terms arise from different contributions and are generally atom centered. The parameters are obtained by fitting the potential with experimental data and high level quantum chemical calculations performed on small functional groups. In this work, we have used the AMBER force field(Cornell et al., 1995), where the additive analytical energy potential has the following form:

$$E = E_{bonds} + E_{angles} + E_{dihedrals} + E_{VdW} + E_{electrostatic} =$$

$$= \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals,n} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] +$$
$$+ \sum_{i<j} \left[\frac{A_{ij}}{(r_i - r_j)^{12}} - \frac{B_{ij}}{(r_i - r_j)^6} + \frac{q_i q_j}{4\pi\varepsilon_0 \mid r_i - r_j \mid} \right]$$

where

$E_{bonds} + E_{angles}$ represent bond stretching and angle bending as harmonic energy terms. The equilibrium bond and angle values r_{eq}, θ_{eq}, as well as the spring strengths K_r, K_θ are adjusted to reproduce structural parameters and normal mode frequencies.

$E_{dihedrals}$ is the Fourier decomposition of the torsional energy. Dihedral parameters are calibrated on small model compounds, comparing the energies with those obtained by ab initio quantum chemical calculations. Improper dihedral angles are inserted between non linked bonds to preserve planarity in aromatic rings.

E_{VdW} describes the Van der Waals interactions through a Lennard-Jones potential. Parameters are obtained so as to reproduce chemical-physical properties in simple organic liquids (densities, enthalpies of vaporization, free energies of solvation, etc).

$E_{electrostatic}$ is the electrostatic energy evaluated by assuming that the dielectric constant is equal to 1, and using the Restrained Electrostatic Potential (RESP) partial atomic charges(Bayly et al., 1993).

The RESP charges reproduce on a finite atoms distribution the electrostatic potential generated by the electronic density. Ab initio quantum chemical calculations at HF/6-31G* level of theory are performed on already optimized small molecules, and the electrostatic potential $\{V_j\}$ is calculated on a discrete grid. The point charges $\{q_i\}$ are fitted minimizing the following:

$$\chi^2_{\text{esp}} = \sum_{j=1}^{M} (V_j - \overline{V}_j)^2 + \sum_{i=1}^{N} \left[(q_i^2 + a^2)^{1/2} - b \right]$$

where $\{\overline{V}_j\}$ is the electrostatic potential generated by the point charges, *a* and *b* are appropriate parameters. The first term of χ^2_{esp} fits the potential generated by the ab initio calculation, while the hyperbolic terms tend to minimize the absolute charges, reducing the effect of poorly determined charges on buried atoms (as in the ESP charges).

The electrostatic and van der Waals interaction energies are calculated between atoms belonging to different molecules or for atoms in the same molecule separated by at least three bonds. Long range electrostatric interactions are calculated through the Particle Mesh Ewald Method, an efficient algorithm for the calculation of Ewald sums, which interpolates the electrostatic potential on a grid using smooth functions(Essman et al., 1995).

2.1.4. QM/MM potentials

Hybrid quantum mechanics-molecular mechanics methods permit quantum chemical calculations within the reactive center of a system in a classical force field environment (Singh and Kollman, 1986). The Hamiltonian for a 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 semiempirical and *ab initio* levels of theory.

In this thesis the quantum part is treated at the DFT level with BLYP(Becke, 1988;Lee et al., 1988) gradient corrections, whereas the classical part is described by the AMBER force field(Cornell et al., 1995) previously described. The crucial issue regarding hybrid methods is the definition of $H_{QM/MM}$, i.e. the treatment of the interactions between the classical and the quantum parts of the system. Below, the description of the interface approach is briefly reported, according to the treatment developed by Laio *et al*.(Laio A. et al., 2002b).

2.1.4.1. Bonded interactions

The bonded interactions at the interface are accounted by the classical force field. Bond, angular and dihedral terms involving MM and QM atoms are treated as in the classical MM Hamiltonian. These terms permit to keep stable the geometry at the interface. However, since they have not been parameterized to deal with chemical reactions, they might have limited accuracy in cases where the boundary is strongly distorted upon chemical transformations. When the partitioning among QM and MM regions occurs at a chemical bond, dangling bonds are saturated by hydrogen atoms. Spurious electrostatic interactions between the capping hydrogens and close-by classical atoms are excluded from the QM/MM Hamiltonian according to(Laio et al., 2004).

2.1.4.2. Non-bonded interactions

The remaining electrostatic and steric contributions between non-bonded atoms are taken into account as follow:

$$H_{QM/MM}^{non-bonded} = \sum_{i \in MM} q_i \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_i|} d\mathbf{r} + \sum_{\substack{i \in MM \\ j \in QM}} v_{VdW}(\mathbf{r}_{ij})$$

where r_i is the position of the MM atom i, with charge q_i, while $\rho(r)$ is the total (electronic plus ionic) charge of the quantum system. The first term describes the electrostatic interaction between the total density with the atomic partial charges of the classical system. The second term describes instead the van der Waals interaction between atoms i and j.

Calculation of the electrostatic interactions in this scheme has two major problems: (i) the electron spill-out phenomenon, and (ii) the very high computational cost. To solve these inconveniences different treatments are adopted for short and long range interactions.

Short range electrostatic interactions. The spill-out problem derives from the anomalous rearrangement of the electron density that tends to localize on the classical point charges of the MM atoms nearby the interface region. This problem is particularly pronounced in a delocalized plane waves basis set approach. A simple solution is found by modifying the Coulomb potential in the core region. Whereas the 1/r behavior is maintained for large r, for values of r shorter than the covalent radius of atoms the Coulomb potential goes to a finite value. The electrostatic term becomes:

$$H_{QM/MM}^{electrostatic} = \sum_{i \in MM} q_i \int \rho(\mathbf{r}) v_i(|\mathbf{r} - \mathbf{r}_i|) d\mathbf{r} = \sum_{i \in MM} q_i \int \rho(\mathbf{r}) \frac{r_{ci}^n - r^n}{r_{ci}^{n+1} - r^{n+1}} d\mathbf{r}$$

where n is fixed equal 4, and r_{ci} is the covalent radius of the atom i (i.e. ~ 0.4 Å for hydrogen and ~ 0.7 Å for oxygen). This 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 molecules without any *ad hoc* reparametrization of the force field(Laio A. et al., 2002a).

Long range electrostatic interactions. The explicit calculation of the electrostatic term is too expensive in a plane wave based approach. In fact, the calculation of about $N_r N_{MM}$ operations is required, 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). To reduce the computational cost, electrostatic interactions are taken into account within a multilayer approach. The explicit calculation is done within a sphere of given radius ($r<r_{c1}$). Outside this sphere, two approximations are adopted: (i) within a second larger radius ($r_{c1}<r<r_{c2}$) the Coulomb interactions are calculated between the MM atomic partial charges and the D-RESP charges of the QM system (see below for D-RESP charges derivation)

$$H_{QM/MM}^{electrostatic} = \sum_{\substack{i\in MM \\ j\in QM}}^{r_{c1}\le r\le r_{c2}} \frac{q_i Q_j^{RESP}}{\left|r_i - r_j\right|}$$

(ii) outside the outer layer ($r>r_{c2}$), the QM electronic density is simply described as its electric multipole expansion (up to quadrupolar order). Hence, the Hamiltonian for this longest range region becomes:

$$H_{QM/MM}^{electrostatic} = C\sum_{i\in MM}^{r\ge r_{c2}} q_i \frac{1}{\left|r - r_i\right|} + \sum_{\alpha} D^{\alpha} \sum_{i\in MM}^{r\ge r_{c2}} q_i \frac{(r_i^{\alpha} - \overline{r}^{\alpha})}{\left|r - r_i\right|^3} + \frac{1}{2}\sum_{\alpha\beta} Q^{\alpha\beta} \sum_{i\in MM}^{r\ge r_{c2}} q_i \frac{(r_i^{\alpha} - \overline{r}^{\alpha})(r_i^{\beta} - \overline{r}^{\beta})}{\left|r - r_i\right|^5} + O\left(\frac{q r_{c2}}{\left|r_i - \overline{r}\right|}\right)^4$$

where $\bar{r}$ is the origin of the multipolar expansion (the geometrical center of the quantum system) and *C* , *D* and *Q* are the total charge, the dipole and the quadrupole of the quantum charge distribution, respectively.

The only free parameters of this approach are r_{c1} and r_{c2}, which may be optimized for the system under investigation in order to speed up the calculation without affecting its accuracy.

2.2. MOLECULAR DYNAMICS SIMULATIONS

2.2.1. Car-Parrinello molecular dynamics

Ab initio molecular dynamics simulations are performed following the Car-Parrinello scheme(Car and Parrinello, 1985). In this approach the electronic degrees of freedom are treated as classical particles with a fictitious mass. As the forces on the nuclei are obtained by a derivative of the Lagrangian with respect the nuclear positions, a functional derivative with respect to the KS orbitals gives the 'forces' on the electrons. The Lagrangian proposed by Car and Parrinello for this motion is the following:

$$L^{CP} = \frac{1}{2}\sum_{i}^{N}\int \mu_i \left|\dot{\psi}_i(\mathbf{r})\right|^2 d\mathbf{r} + \frac{1}{2}\sum_{I} M_I \dot{\mathbf{R}}_I^2 - E\left[\{\psi_i\},\{\mathbf{R}_I\}\right] + \sum_{ij}\lambda_{ij}\left(\left\langle\psi_i(\mathbf{r})\middle|\psi_j(\mathbf{r})\right\rangle - \delta_{ij}\right)$$

where the Lagrange multipliers λ_{ij} are associated to the constrain necessary to ensure the orthornormality for the KS orbitals. The equations of motion derived from this Lagrangian are:

$$\begin{cases} \mu\ddot{\psi}_i(\mathbf{r}) = -H_{KS}\psi_i(\mathbf{r}) + \sum_j \lambda_{ij}\psi_j(\mathbf{r}) \\ M_I\ddot{\mathbf{R}}_I = -\dfrac{\partial E\left[\{\psi_i\},\{\mathbf{r}_i\}\right]}{\partial \mathbf{R}_I} \end{cases}$$

The nuclear trajectories generated by the latter equation and those of classical Lagrangian of the physical system do not coincide, unless E is at the instantaneous minimum. Actually, μ and the initial conditions can be chosen in such a way that the characterizing frequencies of the electronic degrees of freedom are much shorter than those of the nuclei, thus avoiding a significant energy transfer between them. In these conditions, the electrons will follow adiabatically the nuclei remaining close to the BO surface during the entire dynamics.

For insulators, such as the systems treated in this thesis, in which the energy gap is of some eV it turns out that the electronic and ionic frequencies are well separated and the Car-Parrinello dynamics can productively be performed.

2.2.2. Classical molecular dynamics

In classical MD simulations(A.R.Leach, 2001), the potential energy *E* acting on the atomic nuclei is described by simple functions of the ionic positions $E = E(\{\mathbf{r}_i\})$ (Mackerell, 2004). The motion of the system is generated by integrating the Newton's equation of motion over the molecular degrees of freedom. For a system of *N* particles of masses $\{m_i\}$ and positions $\{\mathbf{r}_i\}$, the equations of motion are:

$$F_i = -\nabla_{\mathbf{r}_i} E = m_i \ddot{\mathbf{r}}_i \qquad \text{with} \qquad i = 1, \ldots, N$$

where the potential energy *E* acting on the atomic nuclei may be described, neglecting the electronic interactions, by simple functions of the ionic positions $E = E(\{\mathbf{r}_i\})$.(Mackerell, 2004)

2.2.3. Integration of the equation of motion

In classical mechanics, the trajectory of the system is determined by its initial conditions, namely the initial positions and velocities, by integration of the Newton's equations of motions. As for complex systems these equations cannot be solved analytically, one has to resort to numerical techniques(A.R.Leach, 2001;Allen and Tildsley, 1987).

The integration algorithm can be derived from a Taylor series expansion of the atoms coordinates around an instant *t:*

$$r(t+\delta t) = r(t) + \dot{r}(t)\delta t + \frac{1}{2}\ddot{r}(t)\delta t^2 + \frac{1}{3}\dddot{r}\delta t^3 + \mathrm{O}(\delta t^4)$$

where *r*, *v* and *a* are the position, velocity and acceleration vectors, respectively, and δt is the time step. The length of this time interval is chosen so to ensure that variation in the potential energy function acting on the atoms within each time step is small. Summing the expansions of $\mathrm{r}(\mathrm{t}+\delta\mathrm{t})$ and $\mathrm{r}(\mathrm{t}-\delta\mathrm{t})$, one obtains the *Verlet algorithm*:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \frac{1}{2}a(t)\delta t^2 + O(\delta t^4)$$

which uses only positions and accelerations at time t and t-δt to calculate new positions at time t+δt with an error of order $O(\delta t^4)$. From the knowledge of the trajectory one can derive the velocity as follows:

$$v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} + O(\delta t^2)$$

with only $O(\delta t^2)$ order of accuracy. The advantages of the Verlet algorithm are its straightforwardness and modest storage requirements; the disadvantage is its moderate precision in the velocity calculation.

It is possible to obtain more accurate estimates of the velocity implementing other equivalent algorithms, differing for accuracy and computational speed. The simplest among these is the so-called *leap-frog algorithm*:

$$r(t+\delta t) \approx r(t) + v(t+\frac{1}{2}\delta t)\delta t$$

$$v(t+\frac{1}{2}\delta t) \approx v(t-\frac{1}{2}\delta t) + a(t)\delta t$$

which is currently used in the AMBER program. In this algorithm, first the velocities are calculated at half integer time steps t+1/2δt, then these are used to calculate the positions, at time t+δt. In this way, the velocities *leap* over the positions and the positions *leap* over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, the disadvantage is that they are not calculated at the same time as the

positions. As a consequence, in order to evaluate the kinetic and potential energy at a given *t*, the following approximation for v(t) is used:

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

2.2.4. Constraints

The time step of a simulation is severely limited by the highest frequency motion (i.e. bond vibrations) of the system. In particular hydrogen atoms are involved in the fastest bond stretching vibrations, that are of little interest in classical simulations of biological systems. Fixing these bond lengths during the simulation, the time step can be increased without affecting the accuracy of the simulation.

The set of N_c molecular constraints σ_k, one for each constrained bond, can be written as:

$$\sigma_k(\mathbf{r}_1, \ldots, \mathbf{r}_N) = 0 \qquad k = 1, \ldots, N_c$$

Satisfaction of the N_c constraints can be accomplished by applying the Lagrange's multiplier method. An extra term is added to the potential energy function, and the equations of motions become:

$$m_i \ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i} E - \nabla_{\mathbf{r}_i} \sum_{k=1}^{N_c} \lambda_k(t) \sigma_k(\mathbf{r}_1, \ldots, \mathbf{r}_N) \qquad i = 1, \ldots, N$$

After the time step δt, the system coordinate $\mathbf{r}_i'(t + \delta t)$ will differ by a quantity $\delta \mathbf{r}_i$ with respect to that of the unconstrained system $\mathbf{r}_i(t + \delta t)$: $\mathbf{r}_i'(t + \delta t) = \mathbf{r}_i(t + \delta t) + \delta \mathbf{r}_i$. The SHAKE algorithm

allows satisfying the set of k holonomic constraints by using an iterative procedure up to a fixed threshold of convergence(Ryckaert et al., 1977).

2.2.5. Temperature and pressure baths

In standard MD simulations, the total energy E is a constant of motion and the calculated time averages are equivalent to the averages done on a microcanonical ensemble (NVE). Often it is useful and/or necessary to perform simulations in other ensembles (i.e. NVT and NPT). Several algorithms have been implemented to achieve this issue.

In Car-Parrinello MD simulations, constant temperature dynamics can be obtained coupling the ionic degrees of freedom to a Nose' thermal bath(Nose', 1984;Hoover W.G., 1985), with a kinetic energy to be added in the CP Lagrangian:

$$T_{Nose'} = \frac{1}{2} Q \dot{s}^2$$

where Q is a parameter and s is the additional degree of freedom of the thermal bath. An extra potential energy term is associated with the bath and inserted in CP Lagrangian:

$$V_{Nose'} = (F+1) k_B T \ln s$$

where F is the total number of the degrees of freedom of the system. The equations of motion for the system become:

$$\ddot{\mathbf{r}}_i = \frac{1}{m_i s^2} \mathbf{f}_i - 2 \frac{\dot{s} \dot{\mathbf{r}}_i}{s}$$

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

where Q can be arbitrarily chosen to determine the strength of the coupling: high values result in a low coupling and vice versa.

The method proposed by Berendsen et al. (Berendsen H.J.C et al., 1984) has been used during the equilibration phase of classical MD simulations.

2.3. CALCULATED PROPERTIES

Once the trajectory and the velocities produced in a MD run have been stored, numerous structural properties can be calculated and eventually compared with experiments.

The ensemble average of an observable A of a system characterized by the Hamiltonian *H*, in the classical limit of statistical mechanics ($\hbar \to 0$), in most of the cases can be considered equivalent to its time average(A.R.Leach, 2001):

$$< A >= \frac{\int dp^N dr^N A\left(p^N, r^N\right) e^{-\beta H\left(p^N, r^N\right)}}{\int dp^N dr^N e^{-\beta H\left(p^N, r^N\right)}} \cong \lim_{t \to \infty} \frac{1}{t} \int dt' A\left[p^N\left(t'\right), r^N\left(t'\right)\right]$$

where $\beta = 1/k_B T$, and N is the number of degrees of freedom of the system. This equivalence is know as the ergodic hypothesis, and for many systems, such as for those of interest in this thesis, the validity of the hypothesis has been assessed computationally.

2.3.1. Structural properties

2.3.1.1. Root mean square displacement

The root mean square displacement (RMSD) of a set of N atoms at time *t* with respect to a reference structure, reads:

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \Delta r_i(\mathrm{t})^2}$$

where $\Delta \mathrm{r_i}(\mathrm{t}) = \mathrm{r_i}(\mathrm{t}) - \mathrm{r_i}(\mathrm{t_0})$ is the displacement of atom i during time *t*.

The RMSD per residue is the RMSD averaged over the atoms of the residue. Its time average is called root mean square fluctuations, RMSF:

$$RMSF_{res}^2 = \left\langle \Delta r_{res}{}^2 \right\rangle = \left\langle (r - \langle r \rangle)_{res}^2 \right\rangle$$

where the brackets ⟨⟩ indicate temporal average. This quantity can be compared with the crystallographic B-factor by the Debye-Waller relation:

$$B_{res} = \frac{8\pi^2}{3} RMSF_{res}^2$$

2.3.1.2. Gyration radius

The radius of gyration (G_R) is calculated as:

$$\mathrm{G_R} = \frac{1}{\mathrm{N_R}} \sqrt{\frac{1}{\mathrm{M}} \sum_{1=1}^{\mathrm{M}} (\mathrm{r_i} - \mathrm{r_G})^2}$$

where N_R is the number of residues, M is the number of steps for which averages are calculated, r_i is the position vector of atom i, and r_G is the position vector of the center of mass.

2.3.2. Electronic properties

Along this thesis some relevant electronic properties are calculated to characterize the system. Here, we have reported the principal means of analysis adopted.

2.3.2.1. D-RESP charges

Within the QM/MM scheme exists an inexpensive way to calculate at each time step restrained electrostatic potential derived charges (RESP) associated with the QM atoms(Laio A. et al., 2002c). RESP charges are fitted in order to reproduce the potential due to QM electronic density on the MM atoms that belong to the first sphere surrounding QM region, where electrostatic interactions between QM and MM region are treated explicitely. Since this contribution has to be computed anyway at every time step, RESP charges can be computed 'on the fly' with no additional computational cost. For this reason they are called dynamically generated RESP (D-RESP).

2.3.2.2. NMR chemical shifts

Nuclear Magnetic Resonance experiments measure the resonance frequencies of the nuclear spins of a sample in a uniform magnetic field $\vec{B}_{ext}$. The electrons of the sample react to the applied external field by creating a current distribution $\vec{J}_{ind}(\vec{r})$ which itself

induces an additional magnetic field. This induced field in inhomogeneous in space and depends very sensitively to the electronic structure of the system. Thus, nuclei in different chemical environments will resonate at different frequencies.

The resonance frequency depends upon the strength of the local magnetic field $\vec{B}_{loc}(\vec{r})$ experienced by each nucleus. Within the linear regime, $\vec{B}_{loc}(\vec{r})$ may be expressed as

$$\vec{B}_{loc}(\vec{R}) = \vec{B}_{ext} + \vec{B}_{ind}(\vec{R}) = B_{ext}\left[1 - \overline{\sigma}(\vec{R})\right]$$

where $\vec{B}_{ind}(\vec{R}) = \vec{B}_{ext}\vec{\sigma}(\vec{R})$ is the field induced by the applied uniform magnetic field $\vec{B}_{ext}$ and $\vec{\sigma}(\vec{R})$ is the shielding tensor. In the isotropic case $\sigma = \frac{1}{3}Tr\left[\vec{\sigma}(\vec{R})\right]$. σ is dimensionless, usually very small ≈ 10^{-6}-10^{-4} and measured in part per million (ppm).

$\vec{B}_{ind}(\vec{r})$ is generated by $\vec{J}_{ind}(\vec{r})$ and can be calculated through the Biot-Savart law:

$$\vec{B}_{ind}(\vec{r}) = \frac{1}{c}\int d\vec{r}\vec{J}_{ind}(\vec{r}') \wedge \frac{(\vec{r} - \vec{r}')}{\left|\vec{r} - \vec{r}'\right|^3}$$

The induced current density $\vec{J}^{ind}(\vec{r})$ is defined as:

$$\vec{J}^{ind}(\vec{r}) = \frac{e\hbar}{2mi}\left(\Psi^*(\vec{r})\nabla\Psi(\vec{r}) - \Psi(\vec{r})\nabla\Psi^*(\vec{r})\right) - \frac{e^2}{m}\vec{A}(\vec{r})\Psi^*(\vec{r})\Psi(\vec{r})$$

and may be calculated once the effect of the external magnetic field on the wavefunction is known. This perturbation is calculated in the framework of density functional perturbation theory using the variational approach developed by(Putrino A. et al., 2000).

The interaction between the electrons and a magnetic field in a non-magnetic insulator is described by the non-relativistic time-independent Hamiltonian(McWeeny and Sutcliffe, 2005):

$$H = \frac{1}{2m}\left[\vec{p} + \frac{e}{c}\vec{A}(\vec{r})\right]^2 + V(\vec{r})$$

where $\vec{p} = -i\hbar\nabla$ and $\vec{A}(\vec{r})$ is the vector potential associated to the magnetic field $\vec{B}(\vec{r})$ and satisfying $\vec{B}(\vec{r}) = \nabla \wedge \vec{A}(\vec{r})$. In a uniform magnetic field it is particularly convenient to choose for the vector potential $\vec{A}(\vec{r})$ the form:

$$\vec{A}(\vec{r}) = -\frac{1}{2}(\vec{r} - \vec{d}) \wedge B$$

where $\vec{d}$ is the gauge origin. For low magnetic field the first order linear correction to the ground state wavefunction is

$$\left|\Psi\right\rangle = \left|\Psi^{(0)}\right\rangle + B\left|\Psi^{(1)}\right\rangle$$

The first order correction to the ground state wavefunction $\left|\Psi^{(1)}\right\rangle$ resulting from the effect of the first order perturbation $H^{(1)} = \frac{e}{2}\vec{p}\cdot\vec{A}$ is evaluated by minimizing the second order energy functional which is variational in $\left|\Psi^{(1)}\right\rangle$ (Putrino A. et al., 2000):

$$E^{(2)} = \sum_i \left\langle \psi_i^{(1)} \left| H^{(0)} - \varepsilon_i^{(0)} \right| \psi_i^{(1)} \right\rangle + \sum_i \left[\left\langle \psi_i^{(1)} \left| H^{(1)} \right| \psi_i^{(0)} \right\rangle + \left\langle \psi_i^{(0)} \left| H^{(1)} \right| \psi_i^{(1)} \right\rangle \right] +$$
$$+ \frac{1}{2}\int d\vec{r}\, d\vec{r}\,' \frac{\delta^2 E_{Hxc}\left[\rho^{(0)}\right]}{\delta\rho(\vec{r})\delta\rho(\vec{r}\,')}\rho^{(1)}(\vec{r})\rho^{(1)}(\vec{r}\,')$$

Where $E_{Hxc}\left[\rho^{(0)}\right]$ is the Hartree, exchange and correlation functional and $\varepsilon_i^{(0)}$ are the eigenvalues of the unperturbed Hamiltonian.

As the perturbation is a magnetic field, the first order correction to the ground state wavefunction is purely imaginary. Thus $\rho^{(1)}(\vec{r}) = 0$ and the last term of the second order energy functional vanishes analytically leaving(Sebastiani and Parrinello, 2001b)

$$E^{(2)} = \sum_i \left\langle \psi_i^{(1)} \middle| H^{(0)} - \varepsilon_i^{(0)} \middle| \psi_i^{(1)} \right\rangle + \sum_i \left[\left\langle \psi_i^{(1)} \middle| H^{(1)} \middle| \psi_i^{(0)} \right\rangle + \left\langle \psi_i^{(0)} \middle| H^{(1)} \middle| \psi_i^{(1)} \right\rangle \right]$$

The choice of the gauge origin $\vec{d}$ is arbitrary and the total current is independent of the gauge origin chosen. However, for finite basis sets the numerical calculations are not exactly gauge invariant. To overcome this problem several methods, that employ different gauge origins, were proposed. In our calculations the "*d(r)=r*" version of the Continuous Set of Gauge Transformations (CSGT) method(Keith TA and Bader RFW, 1993) was used. In this method a gauge transformation with $\vec{d} = \vec{r}$ is performed for each point where $\vec{J}^{ind}(\vec{r})$ is evaluated. Within this gauge transformation $\vec{J}^{ind}(\vec{r})$ reads:

$$\vec{J}_i^{ind}(\vec{r}) = -\tfrac{e}{m} \sum_l \left\langle \psi_i^{(0)} \middle| \left[\vec{p} \middle| \vec{r} \right\rangle \left\langle \vec{r} \right| + \left| \vec{r} \right\rangle \left\langle \vec{r} \middle| \vec{p} \right] \frac{\left| \psi_l^{(0)} \right\rangle \left\langle \psi_l^{(0)} \right|}{\varepsilon_i - \varepsilon_l} \left\{ \left[\vec{r} \times \vec{p} \right] B \middle| \psi_i^{(0)} \right\rangle - \vec{p} \middle| \psi_i^{(0)} \right\rangle \left[\vec{B} \times \vec{r} \right] \right\}$$

In our calculations the electronic problem was solved within DFT with a plane waves basis set and pseudopotentials to represent the core contributions to the all electron wave function. Thus only the contributions of the valence electrons to the induced first order quantum electric current were taken into account.

Nevertheless it has been shown for several atom types that the core contribution determine an additive constant to the calculated chemical shift tensors. This constant turns out to be independent of the chemical environment.

In principle, the dependence of the exchange-correlation functional upon the electronic current should be explicitly taken into account as in Current Density Functional Theory (CDFT). In practice, the corrections to the chemical shift due to the use of CDFT functionals are negligible; therefore this dependence is omitted in the chemical shift calculations.

2.3.2.2.1. Isotopic shift on chemical shifts

The isotopic substitution effect on the chemical shift (DIE hereafter) is a small displacement (usually of the order of less than 1 ppm) of the nuclear magnetic resonance frequency of some nuclei that occurs when an atom of a chemical compound is substituted with one of its isotopes. Such a substitution leaves the ground state electronic wavefunction unchanged, but it affects the dynamical properties of the system. In particular, the vibrational levels of the molecule change.

Quantum chemical calculations on small molecular systems have shown that DIEs are dominated by bond stretching(Jameson, 1996). Thus, we made here a reasonable and simplifying assumption that the potential energy surface (PES) of nuclear motion is reduced to the stretching of the bond at which the isotopic substitution occurs. The PES was calculated by fixing the bond length at increasing values. Shielding constants were calculated through the variational approach of density functional perturbation developed by Sebastiani et al.(Sebastiani and Parrinello, 2001a) for each point along the PES. The calculated energy and shielding values were fitted to a polynomial of 8^{th} degree and of 6^{th}

degree, respectively, which provided a good fit. The 1D nuclear Shroedinger equation for the hydrogen motion:

$$\left[\frac{-h^2}{4\pi^2 m}\frac{d^2}{dq^2}+V(q)\right]\psi(q)=E\psi(q)$$

where the mass of the nuclei was set to 1 and 2 a.u. for Hydrogen and Deuterium, respectively, was solved variationally. Harmonic oscillator basis set functions:

$$\phi_n(q)=\frac{1}{(2^n n!\sqrt{\pi})^{1/2}}\exp(\frac{-q^2}{2})P_n(q)$$

where $P_n(q)$ is the Hermite polynomial of order n, were used to describe the vibrational ground state wavefunction(Wilson et al., 1955). One hundred terms were enough to assure that the calculated DIE's were converged. The DIE's were calculated as the difference between the shielding expectation values over the proton and deuterium probability distributions(Jameson, 1996):

$$DIE=\langle\psi_H|\sigma|\psi_H\rangle-\langle\psi_D|\sigma|\psi_D\rangle$$

2.3.3. Electrostatic calculations

Electrostatic interactions play an important role in many biological processes(Honig and Nicholls, 1995). In a continuum approximation of the solute/solvent system the electrostatic potential can be determined by solving the Poisson-Bolzmann (PB) equation:

$$\nabla \cdot \left[\varepsilon(r)\nabla\Phi(r)\right] - \varepsilon(r)\kappa(r)^2 \sinh[\Phi(r)] + 4\pi\rho^f(r)/KT = 0$$

where $\rho^f(r)$ is the fix charge density of the molecule, $\varepsilon(r)$ the distance dependent dielectric constant and $\kappa(r) = 1/\lambda^2$ with $\lambda = \sqrt{ekT/8\pi q^2}$ the Debye length.

An analytical solution of the equation ($\Phi(r) = \Phi(r;\varepsilon,\rho,\kappa^2)$) can be obtained only for very simple geometries and charge distributions, e.g. spheres. For proteins, the equation can be solved numerically by finite difference methods (Gilson and Honig, 1988).

2.3.4. Free energy calculations.

The Potential of Mean Force (PMF) associated to water permeation the Aquaporin-1 channel was calculated using the multiple steering MD approach. This approach, recently proposed by Jarzynski(C.Jarzynski, 1997b;C.Jarzynski, 1997a), allows to reconstruct the free-energy profiles by inducing an ensemble of finite-time transformations, collecting the instantaneous work done in each of them, and then performing a proper ensemble average(C.Jarzynski, 1997b;C.Jarzynski, 1997a):

$$e^{-\beta\Delta G(t)} = \left\langle e^{-\beta\Delta W(t)} \right\rangle$$

where ΔG(t) denotes the free-energy difference for the system at time t and 0, ΔW(t) the irreversible work done at time t, and the brackets represent an average over the set of transformations. In particular, if the transformations undergo by coupling a coordinate of

the system to a time-dependent harmonic external potential along a coordinate *z* of the form:

$$V(z,t) = \frac{k}{2}\left[z(t) - z_0 - vt\right]^2 ,$$

then, the equality can be written as(Hummer and Szabo, 2001):

$$G(z) = -k_B T \ln \left\langle \delta\left(z(t) - z\right) e^{-\beta W(t)} \right\rangle$$

In the previous equations *z(t)* represents the position of the system coordinate at time t, z_0 is the initial position of the minimum of the harmonic potential, *v* is the velocity at which the external potential is moved, *k* is the force constant, k_B is the Boltzmann constant and *T* is the reference temperature. To get a better convergence behaviour, the expression can be approximated by a cumulant expansion(Hummer and Szabo, 2001;Jensen et al., 2002):

$$G(z) = -k_B T \ln\left\langle \delta\left(z(t) - z\right)\right\rangle + \left\langle W(t)\right\rangle_z - \frac{1}{2k_B T}\sigma^2 \left\langle W(t)\right\rangle_z$$

In this case, $< >_z$ denotes an average restricted to those trajectories satisfying *z(t)=z* and σ^2 is the variance of the average work done(Jensen et al., 2002).

2.4. REFERENCES

Allen,M.P. and D.J.Tildsley. 1987. Computer simulation of liquids. Oxford.

Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P A. Electrostatic potential based method using charge restraints for determing atom-centered charges: the RESP model. *J.Phys.Chem.* 97, 10269. 1993.

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

Berendsen H.J.C, Postma J.P.M., van Gunsteren W.F., DiNola A., and Haak J.R. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684-3690.

Car,R. and M.Parrinello. 1985. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* 55:2471-2474.

Cornell,W.D., P.Cieplak, C.I.Bayly, I.R.Gould, K.M.Merz, Jr., D.M.Ferguson, D.C.Spellmeyer, T.Fox, J.W.Caldwell, and P.A.Kollman. 1995. A second-generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117:5179-5197.

Essman,U., L.Perera, M.L.Berkowitz, T.Darden, H.Lee, and L.G.Pedersen. 1995. A smooth particle mesh Ewald Method. *J. Chem. Phys.* 103:8577-8593.

Gilson,M.K. and B.Honig. 1988. Calculation of the Total Electrostatic Energy of A Macromolecular System - Solvation Energies, Binding-Energies, and Conformational-Analysis. *Proteins-Structure Function and Genetics* 4:7-18.

Hohenberg,P. and W. Kohn. 1964. *Phys. Rev. B* 136:864-871.

Honig,B. and A.Nicholls. 1995. Classical Electrostatics in Biology and Chemistry. *Science* 268:1144-1149.

Hoover W.G. 1985. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31:1695-1697.

Hummer,G. and A.Szabo. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A* 98:3658-3661.

Jameson,C.J. 1996. Understanding NMR chemical shifts. *Annual Review of Physical Chemistry* 47:135-169.

Jarzynski,C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. Phys.Rev.E 56, 5018-5035. 1997a.

Jarzynski,C. Nonequilibrium Equality for Free Energy Differences. Phys.Rev.Lett. 78, 2690-2693. 1997b.

Jensen,M.O., S.Park, E.Tajkhorshid, and K.Schulten. 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A* 99:6731-6736.

Keith,T.A. and Bader,R.F.W. 1993. Calculation of magnetic response properties using a continuous set of gauge transformations. C*hem. Phys. Lett.* 210:223-231.

Kleinman,L. and D.M.Bylander. 1982. Efficacious Form for Model Pseudopotentials. *Phys. Rev. Lett.* 48:1425-1428.

Kohn,W. and Sham,L.J. 1965. *Phys. Rev. A* 140:1133-1138.

Laio A., VandeVondele J., and Roethlisberger U. 2002a. A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116:6941-6947.

Laio A., VandeVondele J., and Roethlisberger U. 2002b. D-RESP: Dynamically Generated Electrostatic Potential Derived Charges from Quantum Mechanics/Molecular Mechanics Simulation. *J. Phys. Chem. B* 106:7300-7307.

Laio,A., F.L.Gervasio, J.VandeVondele, M.Sulpizi, and U.Rothlisberger. 2004. A variational definition of electrostatic potential derived charges. *J. Phys. Chem. B* 108:7963-7968.

Leach,A.R. Molecular Modelling. 2001. Pearson Education Limited. Harlow, Essex.

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

Mackerell,A.D. 2004. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* 25:1584-1604.

Martyna,G.J. and M.E.Tuckerman. 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.

McWeeny,R. and B.T.Sutcliffe. 2005. Methods of molecular quantum mechanics. London Academic Press.

Nose',S.J. 1984. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81:511-519.

Pickett,W.E. 1989. Pseudopotential Methods in Condensed Matter Applications. *Computer Physics Reports* 9:115-197.

Putrino A., Sebastiani D., Hutter J., and Parrinello M. 2000. Generalized variational density functional perturbation theory. *J. Chem. Phys.* 113:7102-7109.

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

Sebastiani,D. and M.Parrinello. 2001. A new ab-initio approach for NMR chemical shifts in periodic systems. *J. Phys. Chem. A* 105:1951-1958.

Singh,U.C. and P.A.Kollman. 1986. A Combined Abinitio Quantum-Mechanical and Molecular Mechanical Method for Carrying Out Simulations on Complex Molecular-Systems - Applications to the Ch3Cl + Cl- Exchange-Reaction and Gas-Phase Protonation of Polyethers. *Journal of Computational Chemistry* 7:718-730.

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

Wilson,E.B., J.C.Decius, and P.C.Cross. 1955. Molecular vibrations: the theory of infrared and Raman vibrational spectra. McGraw-Hill, New York.

3. DYNAMICS and ENERGETICS of WATER PERMEATION THROUGH the AQUAPORIN CHANNEL

3.1. INTRODUCTION

Sections 3.1.1. and 3.1.2. are introductory to the structure of the plasma membrane(Alberts et al. 2002).

3.1.1. Plasma Membrane and Membrane Proteins. Cells are enclosed by a membrane, the plasma membrane, constituted by a lipid bilayer in which membrane proteins are embedded (Fig. 1). The protein content is about 50% of the membrane mass.

The lipid bilayer provides an impermeable barrier that isolates the cytosol from the extracellular environment. The bilayer is formed by amphipatic molecules, which self assemble in acqueous environment.

Membrane proteins enable the cell to function and interact with the environment, performing most of the specific functions of the membrane. The protein may be bound to the membrane by noncovalent interactions or it may be covalently attached to it. Moreover, it may span the entire bilayer or reside only on one side of it (Fig. 2). Transmembrane proteins have a unique orientation in the membrane. To span the membrane, the primary sequence organizes into α helices or β sheets in order to satisfy hydrogen-bonding requirements. These secondary structure elements generally arrange into barrels. The vast majority of multipass transmembrane proteins in eukaryotic cells and in bacterial plasma membranes are constructed from trasmembrane α helices.

3.1.2. Membrane Transport. The lipid bilayer is a barrier for polar and charged solutes (Fig. 3). This barrier allows the cell to maintain concentrations of solutes in the cytosol that are different from those in the extracellular environment. However, the cell needs to recruit nutrients and to execrete waste metabolic products. For these purposes, specific transmembrane proteins have been developed, each of which transports a particular class of chemically related molecules, often only a specific molecule. Membrane transport proteins that mediate solute transfer are classified into carrier proteins and channel proteins (Fig. 4). The former bind the solute on one side of the membrane and undergo a series of conformational changes upon which the solute is released on the other side of the membrane. The latter form aqueous pores spanning the membrane through which solutes may pass. The transport may follow the electrochemical gradient (passive transport) or against it (active transport), in this case requiring a source of energy.

3.1.3. Aquaporins. The lipid bilayer has a finite permeability for water. However, certain cell types are exceedingly permeable to water(Agre, 2004). Based on this observation, the existence of water selective channels was postulated since the 50's, but it was only at the end of the 80's that the first functionally defined water channel protein (AQP-1) was isolated from red cell membranes(Agre, 2004). Nowadays it is recognized that aquaporins constitute a channel family evolved for rapid transport of water across biological membranes(Fujiyoshi et al., 2002). These channels are selective for water and repel hydronium ions (Borgnia et al., 1999;Saparov et al., 2001;Zeidel et al., 1992;Zeidel et al., 1994). Some members of the family are also permeated by small molecules, such as glycerol or urea (aquaglyceroporins) (Agre, 2004).

Several recent studies have revealed the structures of human aquaporin-1 (AQP-1) at 3.8(Murata et al., 2000) and at 3.7 Å(Ren et al., 2001) resolution by cryoelectron

microscopy, of aquaglyceroporin Glpf (the *E.coli* glycerol facilitator) at 2.2 Å by X-ray diffraction(Fu et al., 2000) and, more recently, bovine AQP-1 at 2.2 Å by X-ray diffraction(Sui et al., 2001). The protein is a tetramer composed by identical subunits, each of which contains six tilted transmembrane α-helices and two short membrane-inserted non-membrane-spanning helices, forming a right-handed bundle with N and C terminus located on the citoplasmatic side of the membrane (Fig. 5). Each channel subunit can be divided in three main functional regions (see Fig. 6): (i) the extracellular vestibule; (ii) the inner region, which is a long pore of ~ 2 Å radius responsible for the selectivity of the channel against cations and other small molecules (Fujiyoshi et al., 2002). In this region, four ordered water molecules have been detected in the x-ray structure(Sui et al., 2001); (iii) the cytoplasmic vestibule.

3.1.4. Motivation and outline of present work. Understanding the structures of both AQP-1 and Glpf has allowed molecular dynamics (MD) calculations to provide deep insights on the behaviour of glycerol and water transport as well as on the physical basis of selectivity against hydronium and other cations(de Groot et al., 2003;de Groot and Grubmuller, 2001;Tajkhorshid et al., 2002). Furthermore, multiple steering molecular dynamics (MSMD) simulations have been used to calculate the energetics of glycerol permeation in the Glpf channel(Jensen et al., 2002). It was found that the minima of the reconstructed free energy profile along the pore correspond to the sites where major H-bonds interactions can take place between glycerol and protein residues, and that the highest barrier against glycerol passage occurs at the narrowest part of the pore.

We present here the first free energy calculation based on the recently solved, high resolution X-ray structure of bovine AQP-1(Sui et al., 2001). To this aim, we use MSMD calculations on the channel, embedded in a water-membrane mimic bilayer (Fig. 7).

Because of the relatively large size of the system, we focus on one of the subunits, which has been shown to work independently from the other subunits(Shi et al., 1994). Furthemore, very recent MD simulations on the same channel structure, based on the GROMOS force field(Cornell et al., 1995), have been carried out for both on the monomeric and the tetrameric forms. These calculations show that the structural and dynamical properties of the two forms are rather similar, suggesting that the use of the monomeric form may be adequate to investigate this system (Mark Sansom, personal communication).

3.2. MODEL SYSTEMS AND METHODS

Structural models were based on the x-ray structure of bovine aquaporin-1 (AQP-1)(Sui et al., 2001). Only one monomer was considered. The water molecules detected in the x-ray structures were included in the models.

Residues 250-271 have not been detected in the structure. The influence of these residues on the electric field in the channel was estimated using the Poisson-Boltzmann equation (see below). Several structures, which differed for the conformation of residues 250-271, were generated using simulated annealing protocols in vacuum(kirkpatrick et al., 1983). The electrostatic potential in the channel pore, as calculated from the Poisson-Boltzmann equation, turned out not to differ largerly from that obtained without including these groups (Fig. 8). Thus, we decided not to include these groups in the calculations.

Three protonation states of His182 were considered. In model **A**, His182 is protonated in δ; in **B,** in ε; in **C**, His182 is diprotonated. The other histidines residues were protonated so as to provide the most plausible hydrogen bonding pattern in the structure; specifically,

His71 and His76 were protonated in the ε position, His36, His206 and His211 in the δ position.

The three structures were embedded in a water/*n*-octane/water bilayer, following the procedure reported in ref(Guidoni et al., 1999;Zhong et al., 1998) (Fig. 7). The thickness of the organic liquid was set to 31 Å based on the presence of such a layer of hydrophobic residues in the structure. Within this choice, the aromatic residues Trp11, Trp212, Trp247, Phe10, Phe208, Phe231 were located at the water/*n*-octane interface (Fig. 7). Solvent molecules were discarded whenever the distance between oxygen or hydrogen and any protein atom was smaller than the sum of the van der Waals radii. Chlorine anions were added so as to neutralize the charge of the models (+3,+3,+4 for **A**, **B**, **C**, respectively).

The computational protocol for the three models was the following: (i) equilibration of the solvent by 0.25 ns of MD at room conditions (298 K temperature and 1 atm pressure), while keeping the protein fixed; (ii) energy minimization of the entire system through three cycles of 3,000 steepest descent steps and 2,000 conjugate gradient steps, progressively allowing hydrogen atoms, side chain atoms and backbone atoms to move; (iii) thermalization of the entire systems by MD calculations, imposing a harmonic restraint of 10 kcal/mol on the four crystallographic water molecules inside the channel (see Fig. 6): 30 ps from 0 to 10 K, 20 ps from 10 to 50 K, 20 ps for each 50 K increase of temperature from 50 K up to 298 K; iv) 2.85 ns, 1.0 ns, 1.4 ns MD simulations at room conditions for **A**, **B**, **C**, respectively. The last 1.8 ns of the simulation of **A** were collected for analysis.

The following properties were calculated: i) root mean square displacement (rmsd) relative to the x-ray structure(A.R.Leach, 2001); ii) B factors (see, e.g.,(Costa and Carloni, 2003)); iii) water residence time along the channel; iv) diffusion coefficient D_α (α = x,y,z), from Einstein's relation: $2D_\alpha t = \langle |r(t)-r(0)|^2 \rangle$, where brakets <> denote trajectory average(A.R.Leach, 2001); v) water dipole moment (M) autocorrelation functions (ACF), calculated as $ACF(t) = \langle M_\alpha(t)M_\alpha(0) \rangle$ (α = x,y,z), where brakets <> denote trajectory

average(A.R.Leach, 2001); vi) persistency of hydrogen bonds interactions(Settanni, 2003); vii) the volume of the channel, as calculated with the HOLE program(Smart et al., 1993); viii) Ramachandran plots(Laskowski R A, 1993); ix) gyration radius $R_g = (\Sigma_i m_i r_i^2 / \Sigma_i m_i)^{1/2}$, where m_i and r_i are the mass and the position (with respect to the center of mass of the protein) of atom i, and the sum runs over all protein atoms.

3.2.1. Free energy calculations. The free energy associated to water permeation in model **A** was calculated using the multiple steering MD approach. The overall MSMD procedure and the parameters in the calculation were similar to those used by Schulten and co-workers for their investigation on glycerol permeation in the Glpf channel(Jensen et al., 2002). The structure of the channel after 1.2 ns of MD was divided into six sections of 5.0 Å length, in which MSMD calculations were carried out independently. In each section, 100 ps MD were carried out. During the dynamics, the water molecule closest to the section's boundaries (WATC hereafter) was subjected to a harmonic potential: a force constant *k*=2.0 kcal mol^{-1} $Å^{-2}$ was applied to the position of WATC oxygen atom. The MD snapshots at 70, 80, 90 and 100 ps were selected as starting configurations for the MSMD calculations. In each of these runs, the minimum of the harmonic potential was moved along the longitudinal axis of the channel at constant velocity, computing at each time the instantaneous work done to move WATC. The MSMD run parameters were: *k*=2.0 kcal mol^{-1} $Å^{-2}$, *v*=0.1 Å ps^{-1}. *z* was the longitudinal axis of the channel (see Fig. 6). $\langle W(t) \rangle_z$ was calculated as an average between the four MSMD simulations. The same free-energy value, and set to zero, was imposed at both channel outlets, where the water is fully hydrated by other water molecules. The free-energy profiles obtained in each section were matched at the boundaries, an then, the final free-energy profile was obtained by fitting the computed values with a combination of sine functions as in ref(Jensen et al., 2002).

In ref(de Groot and Grubmuller, 2001) the free energy profile was reconstructed from the MD simulation. It was calculated as G(z) = -RTln(n(z)), where n(z) is the frequency with which water molecules are observed along the pore during the free dynamics. Calculations using this formula, based on our 1.8 ns MD simulation, reproduced qualitatively the features obtained whith the steering dynamics method. However, as expected, the calculation suffers of poor statistical sampling. The same calculation is much better converged in the case of the MD simulation of ref(de Groot and Grubmuller, 2001), which was carried out on four AQP-1 subunits over 9 ns of dynamics.

3.2.2. Electrostatics calculations. The electrostatic potential in the channel was calculated by solving the Poisson-Boltzmann equation (PB). Calculations were carried out on a grid with 2 points/Å, with a temperature of 298 K and an ionic strength of 0.15 mM, using the program DELPHI(Rocchia et al., 2001). The dielectric constants of the protein was set to 2, while for water different values have been tested: 80 (bulk water), 50(Breed et al., 1996), 5(Partenskii et al., 1994).

The electrostatic field in the channel has been monitored during the dynamics using the simple coulomb expression of the AMBER force field(Cornell et al., 1995) with a dielectric constants equal to 1.

3.3. RESULTS

In this section, we first describe the structural and dynamical properties of our model of bovine AQP-1 (Fig. 7). Subsequently, the energetics of the permeation process is presented.

3.3.1. His182 protonation state. Water permeation is expected to depend crucially on the protonation state of His182, which faces the narrowest region of the water channel. Here, we have investigated the structural properties of three protomers of His182 by MD simulations: protomer **A**, in which His182 is protonated in δ; **B**, in which His182 is protonated in ε, and **C**, in which His182 is diprotonated. In **A**, His182 forms a three center hydrogen bond, which involves its backbone NH, its δ hydrogen and Val178 carbonyl oxygen. This interaction is very stable and it is maintained during the dynamics (Fig. 9). In **B**, His182 side chain, by flipping after $\approx$ 1 ns of dynamics, orients similarly to **A.** This conformational change may be due to the unfavourable interaction between the $N_{\varepsilon}H$ dipole and the positively charged Arg197 residue, which is located ~ 5 Å distant facing the channel. The double protonation of the histidine residue (model **C**) causes an unphysical enlargement of the channel. Indeed, the channel increases its size by as much as 0.7 Å in correspondence of His182, after 1.4 ns of MD (Fig. 10). This is possibly caused by the electrostatic repulsion between His182 and Arg197.

Thus, our calculations suggest that **A** is the most stable protomer of AQP-1, and all subsequent calculations were carried out on this model.

3.3.2. Structural Properties. The rmsd of the backbone atoms from the crystal structure fluctuates around an average value of 2.5 $\pm$ 0.2 Å in the last 1.8 ns of dynamics (Fig. 11). This value is slightly higher than that of the previous simulations of both human and bovine

AQP1 in its tetrameric form(de Groot et al., 2003;de Groot and Grubmuller, 2001). This might be expected as this simulation includes only the monomeric form. The helix bundle structure is fairly maintained during the dynamics, as it can be seen by a comparison of the initial and final MD structures (Fig. 11) as well as by a plot of the orientation and distances between the helices as a function of simulated time (Figs. 12 and 13). The gyration radius of the protein is conserved (19.0 $\pm$ 0.1 Å) and the dimensions of the channel (as calculated by the program HOLE(Smart et al., 1993), are also conserved during the dynamics (Fig. 14).

The B-factors of the Cα atoms are in qualitative agreement with the experimentally determined ones (Fig. 11). The largest discrepancy are observed for the extracellular and cytoplasmatic loops, that interact with neighbouring subunits in the x-ray structure and with the solvent in our simulation.

3.3.3. Water inside the channel. The number of water molecules present in the channel is 35 ($\pm$2). These water molecules exhibit one-dimensional diffusion in the channel, along the path identified by the HOLE algorithm(Smart et al., 1993). An estimation of the average water diffusion coefficient along this path, calculated from the average translational motion, is 0.71($\pm$ 0.04) x 10^{-5} cm^2 / s (Fig. 15). The estimation is approximate because the influence of the potential of mean force acting on the water molecules is here not taken into account. However, as expected, it is lower than the value in the bulk phase (5.19x10^{-5} cm^2 / s for the same water model), and it is larger than that calculated with the same water model and with the same formula, for the Glpf channel (Tajkhorshid et al., 2002). Four water molecules diffuse from the extracellular to the cytoplasmatic side within the last 1.8 ns of dynamics. This is consistent with the mesaured permeability of the channel (~2 molecules / (ns x channel))(Heymann J.B. and Engel A., 1999).

The permeability can also be estimated from calculated diffusion coefficient and the average pore radius(Heymann J.B. and Engel A., 1999), although in this case it is expected to be rather approximate for the argument given above. However, assuming an average pore radius of ~ 2 Å, this quantity turns out to be ~ 1 molecule / (ns x channel), that is of the same order of magnitude of the experiment.

In the inner region, the electric field $\vec{E}$ along the channel causes an orientation of the water dipoles. $\vec{E}$ is directed towards the extracellular vestibule for positive z values (see Fig. 6) and towards the intracellular vestibule for negative z values. It inverts its sign at the level of Asn78 and Asn194, two fully conserved residues(Heymann and Engel, 2000). As already suggested(de Groot and Grubmuller, 2001), $\vec{E}$ results mostly from the dipoles of two short membrane-inserted non-membrane-spanning helices (M3 and M7 in Fig. 6) facing each other with their N-terminus in the middle of the channel as well as from the polar groups lining the channel and forming H-bonds with the diffusing water molecules (Fig. 16).

As a consequence of the protein field, water molecules diffusing in the inner region show opposite orientation of the dipole moment with respect to the NPA motifs (Fig. 6 and Fig. 17): waters located in the extracellular half ($0 < z < 6$ Å) have dipoles pointing towards the extracellular vestibule, whereas in the cytoplasmatic part ($-6 < z < 0$ Å) display the opposite orientation. The dipoles reorient at the Asn78 and Asn194 side chains (Fig. 17). These residues are the only hydrogen bond donors present in the inner region, and help the translocation from one half of the channel to the other by coordinating the water oxygen.

Consistently with the structure of these water molecules, the orientational relaxation is much slower in the channel than in bulk phase, as shown by a comparison between the autocorrelation function of the water dipole moments in the channel and that in bulk phase (Fig. 18).

The location of crystallographic water oxygens correlates fairly with the sites where water molecules are more frequently observed (Fig. 19). In these locations, both x-ray and MD water molecules can H-bond to Asn78 and Asn194 side chains and with Gly192 and His76 backbone units.

We now turn our attention to water structure. For z > 14 Å and for z < -12 Å water molecules interact with progressely more water molecules, until for z=17 or –14 Å they resemble bulk waters (Fig. 20).

In the regions for 8 < z < 16 Å and for –12 < z < -8 Å (Fig. 6) water assume a funnel-like structure and in the inner region, the entire pore is filled with a single file of five ordered water molecules (Fig. 6), that is one more than what has been suggested based on experimental data(Sui et al., 2001).

The water H-bond pattern connecting water from one vestibule to the other along the single file is not (always) contiguous. At the constriction site (6 < z < 8 Å), the narrowest part of the pore, water molecules interact mainly with the protein (Fig. 21). The guanidinium group of Arg197, the only charged residue in the pore, acts as H-bond donor, while the Nε of the imidazole ring of His182 and the main chain carbonyl of Cys191 act as acceptors (Fig. 6). At the intracellular end of the inner region there is no equivalent of the constriction site found on the extracellular side, and the radius of the pore remains almost constant (~ 2 Å, Fig. 6) untill the intracellular vestibule is reached. Here the carbonyls of Ala75 and His76 (z ~ -6/-8) can form H-bonds with water molecules. His76 side chain does not extend into the channel and there are no other polar groups present. Unlike in the constriction site, water-water H-bonding is not broken. The main features of the H-bond pattern are very similar to those observed in previous simulations of both the human AQP-1 and the bacterial Glpf(de Groot and Grubmuller, 2001) channels.

3.3.4. Energetics. The free energy of permeation of water molecules through the channel (Fig. 6) has been calculated via MSMD simulations. The free energy profile has converged within four MSMD simulations, as the maximal variation in the free energy profiles obtained averaging three and four simulations was of 0.1 kcal / mol. This is less than the uncertainty of our free energy plot (0.2 kcal / mol). The steered water molecules have followed the same pathways as those observed in the MD simulations.

We follow here water permeation from the extracellular region to the cytoplasmatic region.

In the extracellular vestibule (8 < z < 16 Å, Fig. 6), the permeating water molecule starts to form H-bond interactions with the protein and the free energy barrier is never significantly larger than k_BT.

At the constriction region (6 < z < 8 Å, Fig. 6), water penetration may be assisted by formation of H-bonding interactions with His182 and Arg197 side-chains as well as Cys191 backbone unit. The barrier also in this case is not larger than k_BT.

Below this site (4 < z < 6 Å, Fig. 6), the permeating water may form H-bonds with Gly192 carbonyl group, with one or two other water molecules, and with the side chain of His182. The resulting free energy exhibit a minimum; consistenly, an ordered water molecule has been detected in the X-ray structure(Sui et al., 2001).

At the NPA motifs (-1.5 < z < 2 Å, Fig. 6), the permeating water may form H-bonds with two water molecules (which act as acceptors), as well as with Asn78 and Asn194 side chains, which act as donors. Here, the free energy reaches the absolute minimum. The latter is not symmetric around the NPA motif, possibly because of the presence of polar and charged groups (namely His182 and Arg197, 6 < z < 10 Å, Fig. 6) only towards the extracellular region. Also in this region an ordered water molecule has been detected(Sui et al., 2001).

In the region below, the permeating water molecule translocates towards the cytoplasmatic vestibule (-5 < z < -1.5 Å, Fig. 6). It passes through a increasingly hydrophobic region and

progressively decreases the number of H-bonds interactions with the protein frame, increasing its free energy. Specifically, it forms only one H-bond with Asn78, in a region where an ordered water molecule has been found. Then, it interacts with only water molecules.

Below this region, the permeating water forms again H-bond interactions with the protein frame. First, it forms an H-bonds to His76 (-7 < z < -5 Å, Fig. 6), and the free energy reaches a local minimum. In this region, an ordered water molecule has also been found. Then, after crossing a free energy maximum (z ~ -7.5 Å, Fig. 1), not much larger than k_BT, it H-bonds Ala75 and Gly74. Finally, it progressively recoveres the properties of bulk water (Fig. 20) in the cytoplasmatic region.

Inspection of the overall free energy profile indicates that the barrier to the permeation process correspond to the translocation from the NPA motifs towards either the cytoplasmatic or the extracellular region. The calculated barrier is 2.5 ± 0.2 kcal/mol, consistenly with the fact that the free energy of the process should be smaller than 5 kcal/mol(Heymann and Engel, 2000). The free energy barrier is much smaller than that found for glycerol permeation in Glpf using MSMD calculations (7.3 kcal/mol(Jensen et al., 2002)), as expected for the different nature of the permeating molecule.

3.3.5. Comparison with previous MD calculations on AQP-1. Although our computational setup differs from that of previous calculations, our results are similar to those already obtained by MD simulations on human AQP-1(de Groot and Grubmuller, 2001). Indeed, both systems feature an ordered single file of water molecules in the inner region of the pore, with a similar H-bond pattern. In particular, this pattern is interrupted in the narrowest part of the channel, where the key residues His182 and Arg197 face each other. In addition, in both simulations the water molecule dipoles reorient at the side chains of Asn78 and Asn194. Finally, our largest free energy barrier is located close to Asn78 and

Asn194 (Fig. 6), as that calculated from the residence frequency along the pore axis(de Groot and Grubmuller, 2001).

3.4. CONCLUSIONS

We have reported an MD study of the high resolution x-ray structure of bovine AQP-1(Sui et al., 2001). Our calculations establish that His182, a key residue lining the channel in its narrowest part, is protonated in the δ position of the imidazole ring. Water in the inner region of the channel is organized in a single file of H-bonded molecules which results highly ordered due to the electrostatic properties of the protein.

The energetics of water permeation has been calculated through the MSMD approach, adopting the procedure already succefully used by Schulten and co-workers for the permeation of glycerol in the glycerol channel(Jensen et al., 2002). To the best of our knowledge this is the first free energy calculation based on the new, high resolution x-ray structure of the channel. The translocation from the center of the pore, at the Asn78 and Asn194 side chains, to the adjacent hydrophobic regions, where a water molecule do not achieve full coordination, have been identified as the barriers to water permeation. The position of the free energy minima correlates fairly well with the location of the water molecules crystallographically observed. These minima correspond to the polar sites where H-bonds can form between water molecules and the protein. These findings are very similar to those recently reported based on MD calculations on human AQP-1(de Groot and Grubmuller, 2001) and reported by other MD studies of water diffusion through the *E. coli* glycerol facilitator(de Groot and Grubmuller, 2001;Tajkhorshid et al., 2002). Thus, MD simulations, in which different approximations for the model system, force fields

and membrane mimic were used, provide independently a fully consistent picture for water transport in these channels.

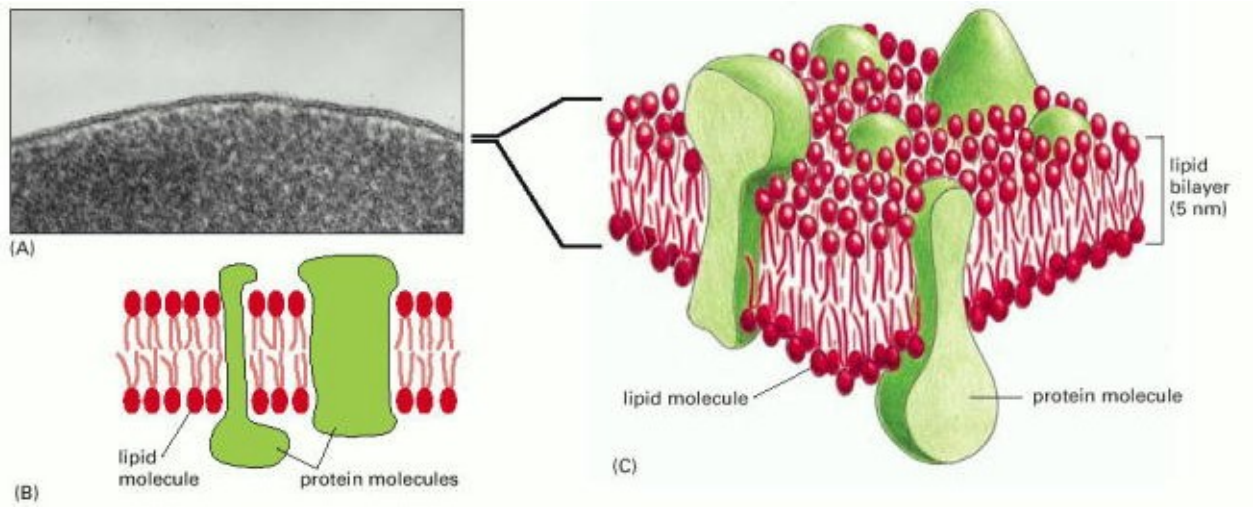


Figure 1. Views of a cell membrane. (A) An electron micrograph of a plasma membrane (of a human red blood cell) seen in cross section. (B and C) These drawings show two-dimensional and three-dimensional views of a cell membrane. © 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.

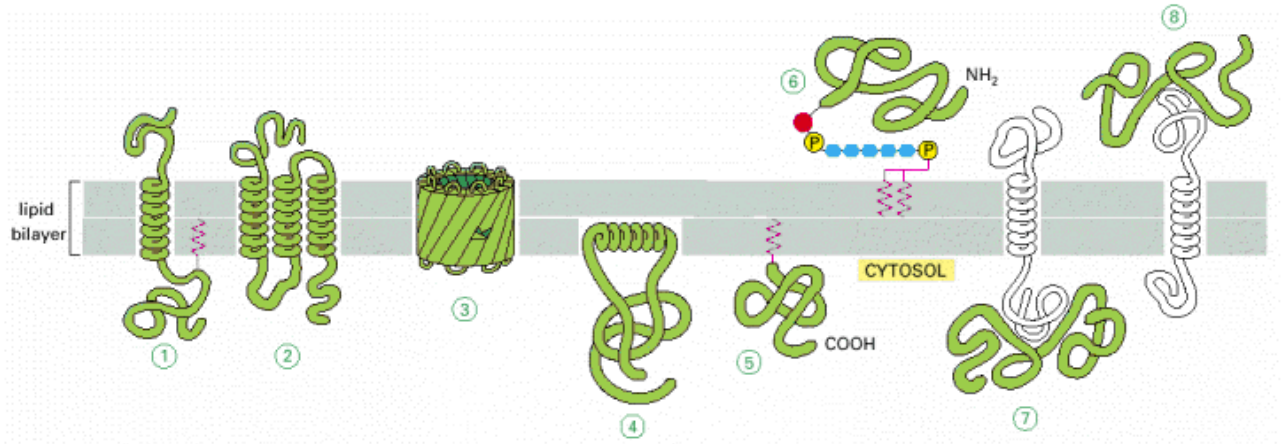


Figure 2. Ways in which membrane proteins associate with the lipid bilayer. Most trans-membrane proteins are thought to extend across the bilayer as (1) a single α helix, (2) as multiple α helices, or (3) as a rolled-up β sheet (a β barrel). Some of these "single-pass" and "multipass" proteins have a covalently attached fatty acid chain inserted in the cytosolic lipid monolayer (1). Other membrane proteins are exposed at only one side of the membrane. (4) Some of these are anchored to the cytosolic surface by an amphipathic α helix that partitions into the cytosolic monolayer of the lipid bilayer through the hydrophobic face of the helix. (5) Others are attached to the bilayer solely by a covalently attached lipid chain in the cytosolic monolayer or, (6) via an oligosaccharide linker, to phosphatidylinositol in the noncytosolic monolayer. (7, 8) Finally, many proteins are attached to the membrane only by noncovalent interactions with other membrane proteins.

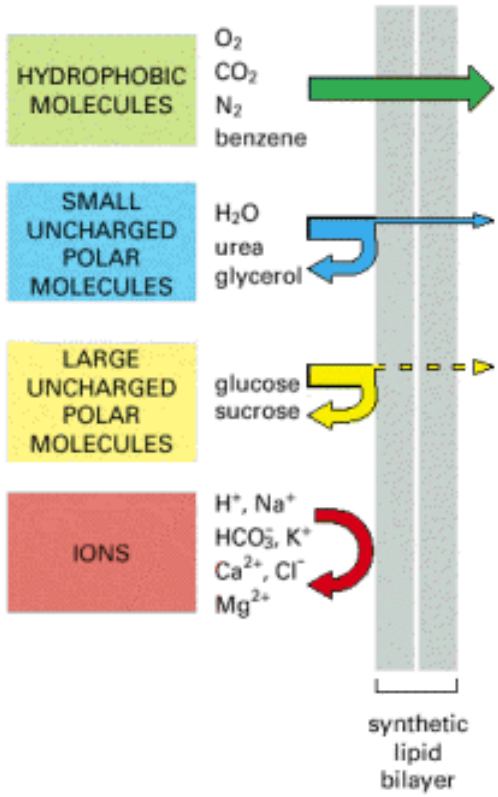


Figure 3. The relative permeability of a synthetic lipid bilayer to different classes of molecules. The smaller the molecule and, more importantly, the less strongly it associates with water, the more rapidly the molecule diffuses across the bilayer. © 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.

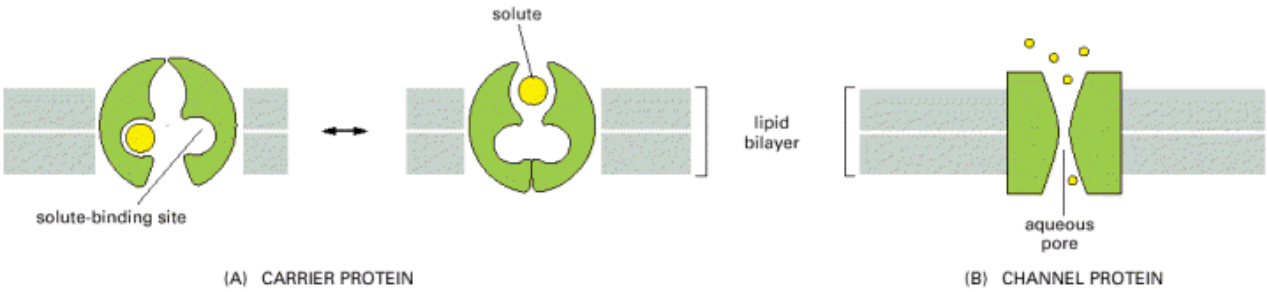


Figure 4. Carrier proteins and channel proteins. (A) A carrier protein alternates between two conformations, so that the solute-binding site is sequentially accessible on one side of the bilayer and then on the other. (B) In contrast, a channel protein forms a water-filled pore across the bilayer through which specific solutes can diffuse. © 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.

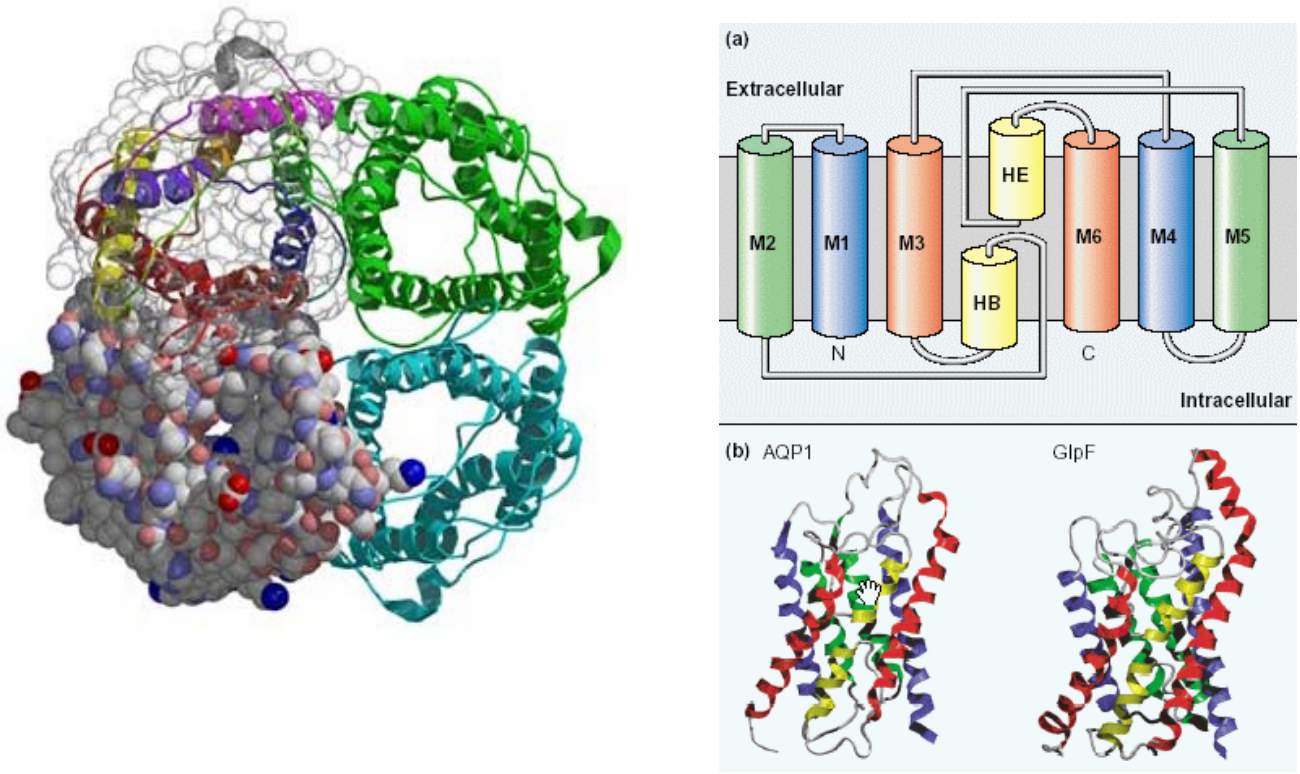


Figure 5. *Left*: top view of the AQP-1 tetramer (Sui et al., 2001). *Right*: a) AQP-1 topology, with cylinders representing α-helices. b) AQP-1 and Glpf monomers.

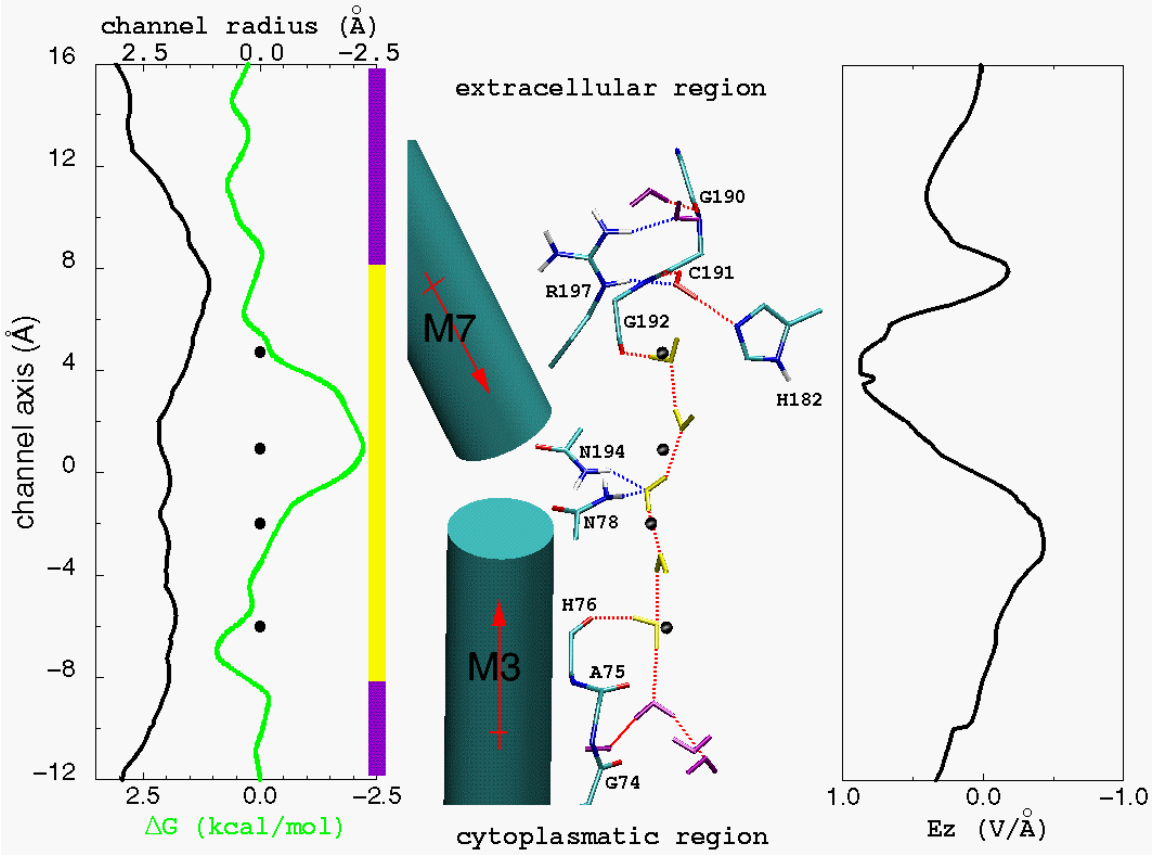


Figure 6. Water permeation in the channel. *Left:* free energy profile (green curve) and channel radius (black curve) along the channel axis. The zero of the axis coincides with the midpoint between the side chains of Asn78 and Asn194. The violet and yellow vertical bars indicate the vestibules and the inner region, respectively. The location of the water molecules, as detected in the X-ray structure (Sui et al., 2001), are marked as black dots. Error bars, which are ~ 0.2 kcal/mol, are not shown for the sake of clarity. *Center:* snapshot after ~ 2 ns MD reporting the ordered water single file structure in the channel. Water molecules of the inner region, in the constriction site (Arg197/His182) and in the vestibules are colored in yellow, red and violet, respectively. H-bonds are represented as dashed lines. Short helices M3 and M7 (Sui et al., 2001) are shown as cylinders. Their dipoles are represented as red arrows. The location of the water molecules, as detected in the X-ray experiment, are marked as black dots. *Right:* component of the electrostatic field along the channel axis (E_z), calculated from the coulomb expression of the AMBER force field (Cornell et al., 1995). Because of the very approximate form of the field used here, the calculated values are used here for a qualitative comparison. Structural and electrostatic properties have been calculated for the MD structure represented in the center panel.

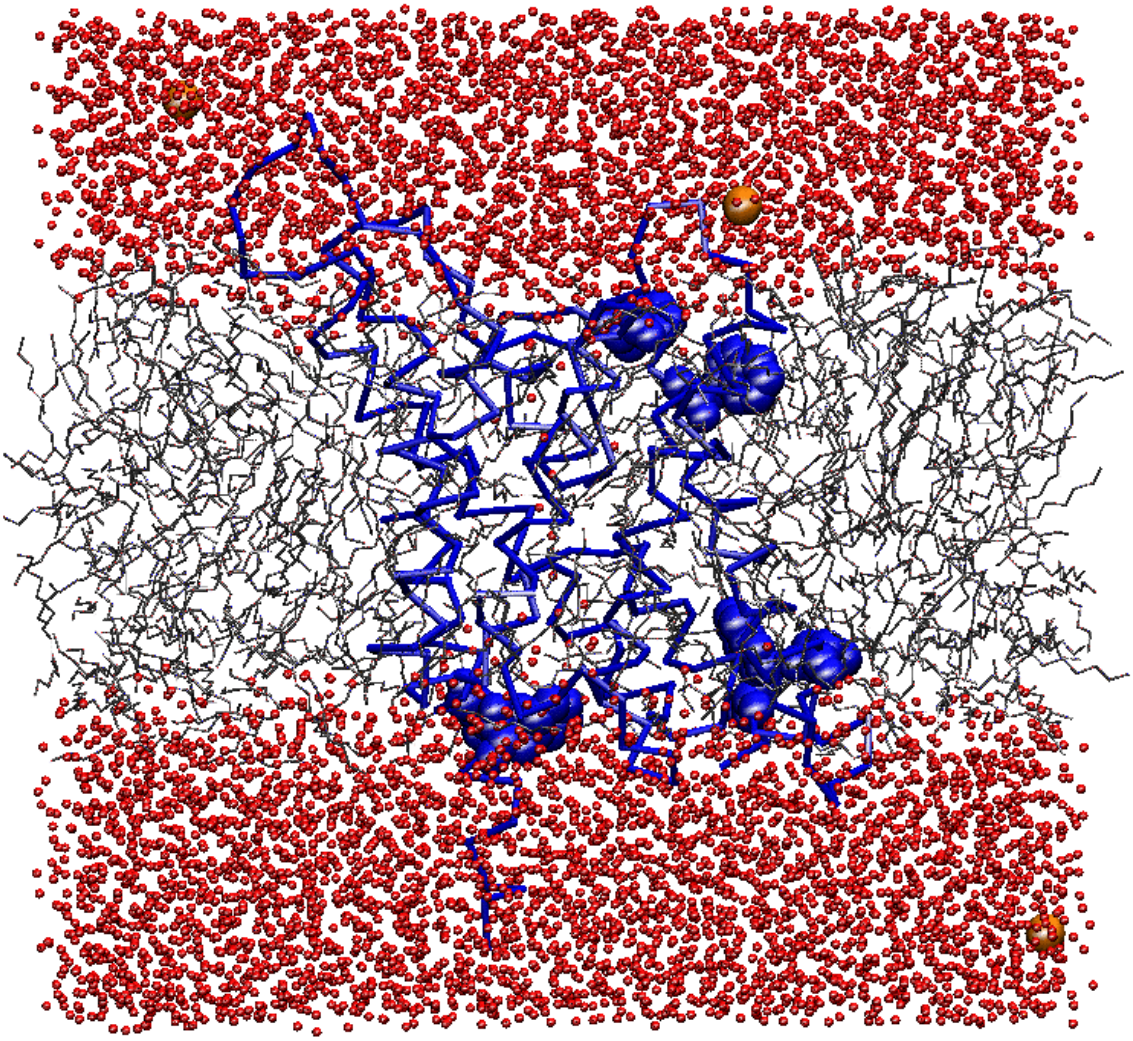


Figure 7. AQP-1 monomer structure. The protein is immersed in a water/membrane mimic water bilayer. The Cα trace is represented as a thick blue line. Water oxygen atoms are represented as red spheres, *n*-octane as thingray lines. The atoms forming the aromatic residues Trp11, Trp212, Trp247, Phe10, Phe208, Phe231, located at the water/*n*-octane interface, are represented as blue spheres. The Cl- counterions are represented as orange spheres. Hydrogen atoms are not displayed for the sake of clarity.

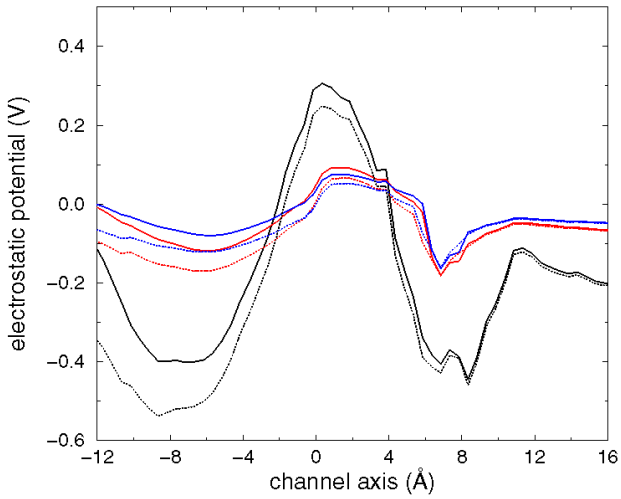
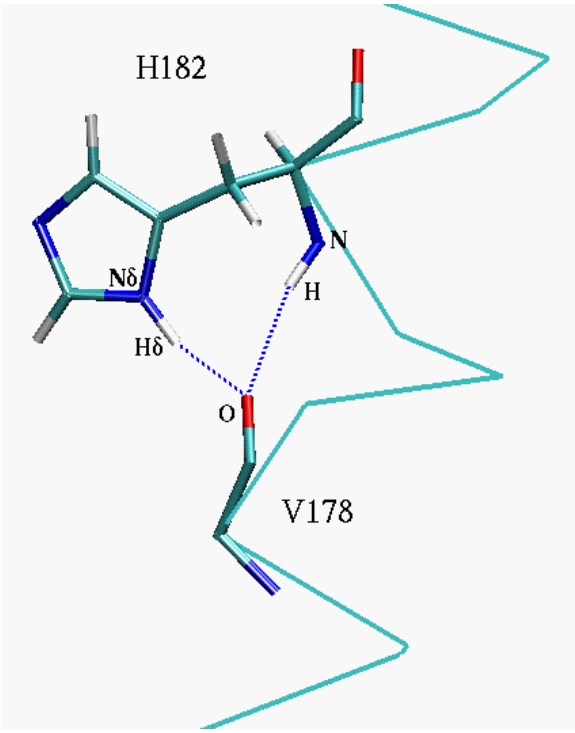


Figure 8. Electrostatic potential along the channel axis, calculated from the Poisson-Boltzmann equation on the x-ray structure (Sui et al., 2001) with (dotted line) or without (continuous line) residues 250-271, not detected in the x-ray experiment and added here via simulated annealing calculations (See Methods Section). Black, red and blue lines refer to dielectric constant ε set to 5, 50, 80 respectively. In this and the following figures the channel axis is that reported in Fig. 6.



Atoms	distance
V178_O - H182_N	3.2 ± 0.2 Å
V178_O - H182_Nδ	2.8 ± 0.1 Å
	angle
V178_O - H182_H - H182_N	151° ± 13°
V178_O - H182_Hδ - H182_Nδ	160° ± 10°

Figure 9. H-bond interactions between His182 and Val178 in model **A**. MD-averaged structural parameters are reported above.

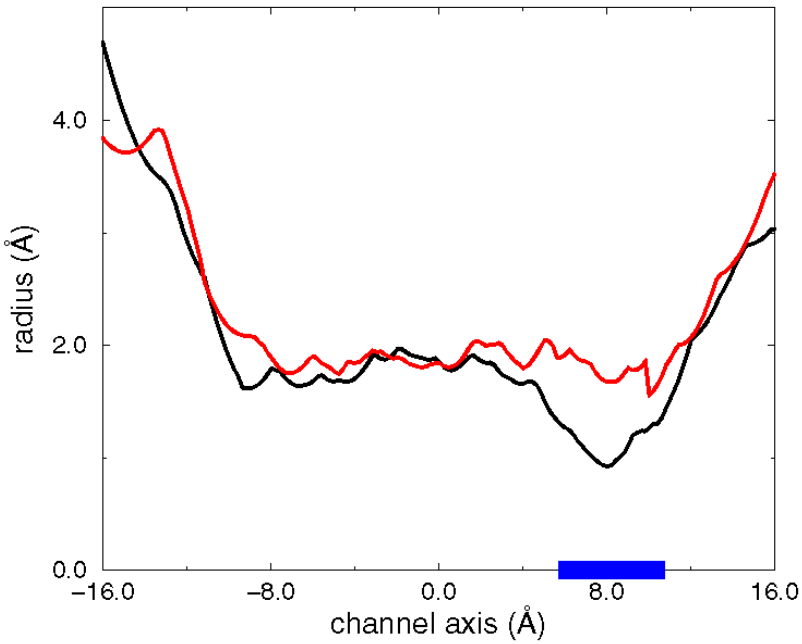


Figure 10. Model **C.** Channel radius as calculated by the program HOLE (Smart et al., 1993). The black line refers to the initial value (Sui et al., 2001). The red line refers to the structure after 1.4 ns MD. The blue bar along the horizontal axis denotes the location of residues His182 and Arg197.

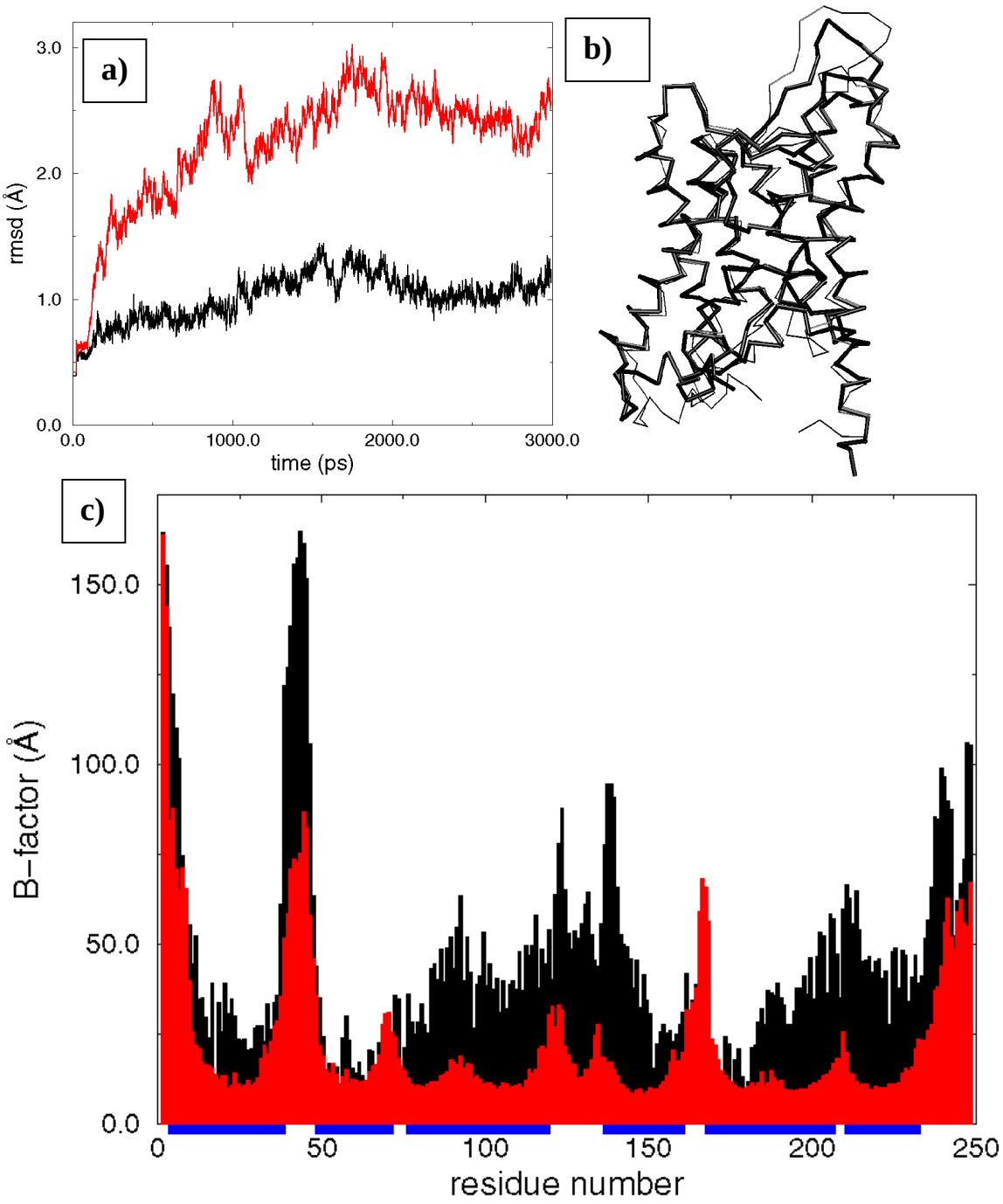
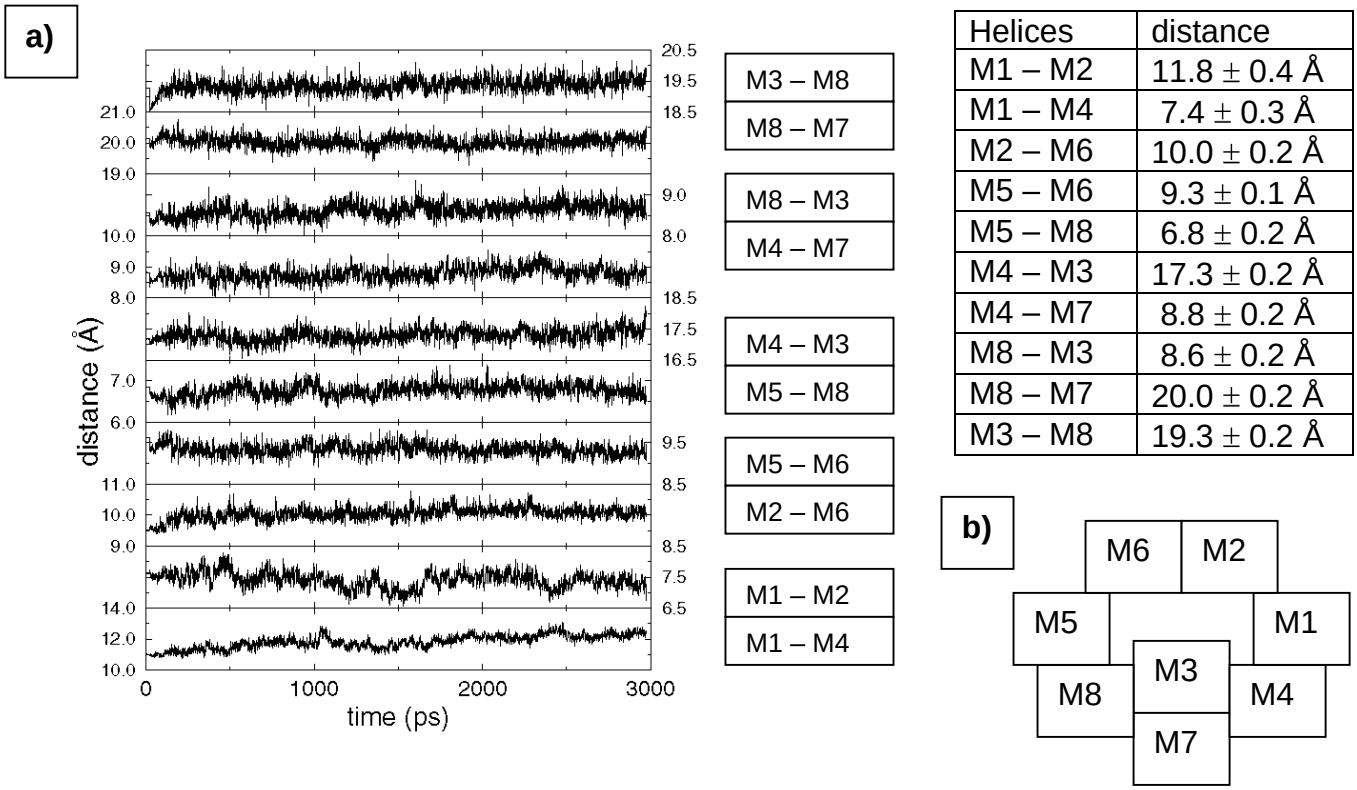
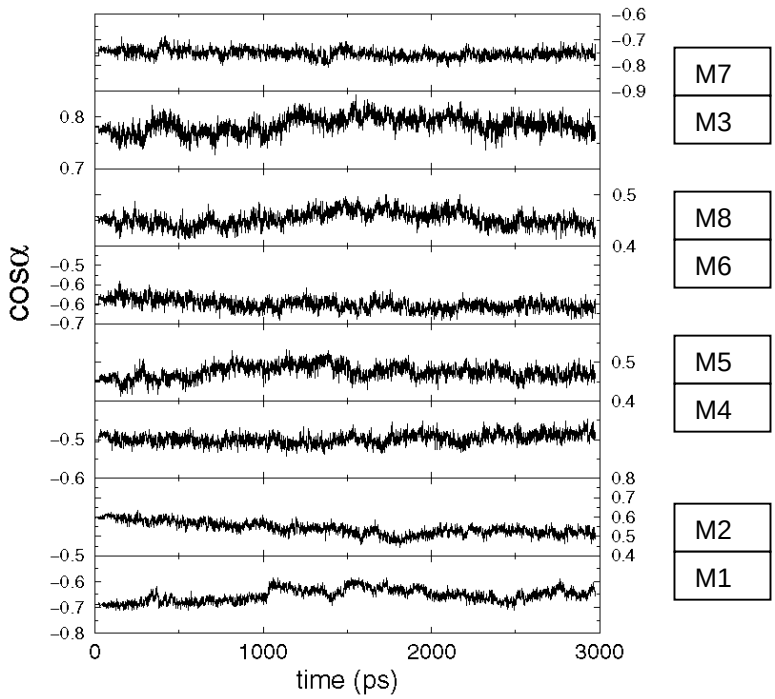


Figure 11. AQP-1 dynamics. **a**) rmsd of backbone atoms belonging to the entire protein (red line) and to the helix bundle (black curve) plotted as a function of time; **b**) superposition between the x-ray Cα trace (thin line) and that obtained after 3 ns of dynamics (thick line); **c**) experimental(Sui et al., 2001) (black) and calculated (red) Cα B-factors: horizontal blue bars indicate α-helices residues.



Helices	distance
M1 – M2	11.8 ± 0.4 Å
M1 – M4	7.4 ± 0.3 Å
M2 – M6	10.0 ± 0.2 Å
M5 – M6	9.3 ± 0.1 Å
M5 – M8	6.8 ± 0.2 Å
M4 – M3	17.3 ± 0.2 Å
M4 – M7	8.8 ± 0.2 Å
M8 – M3	8.6 ± 0.2 Å
M8 – M7	20.0 ± 0.2 Å
M3 – M8	19.3 ± 0.2 Å

Figure 12. Model **A. a)** Distances between centers of mass of adjacent helices plotted as a function of simulated time. The MD-averaged values are reported above. **b)** Topology of the AQP-1 subunit. Nomenclature of α-helices as in (Sui et al., 2001).



helix	<cos α>
M1	-0.65 ± 0.02
M2	0.54 ± 0.03
M3	0.73 ± 0.02
M4	-0.50 ± 0.01
M5	0.47 ± 0.02
M6	-0.65 ± 0.01
M7	-0.75 ± 0.02
M8	0.45 ± 0.01

Figure 13. Model **A.** Orientation of each α-helix relative to the channel axis (See Fig. 6). Nomenclature of α-helices as Fig. 12. The MD-averaged values are reported above.

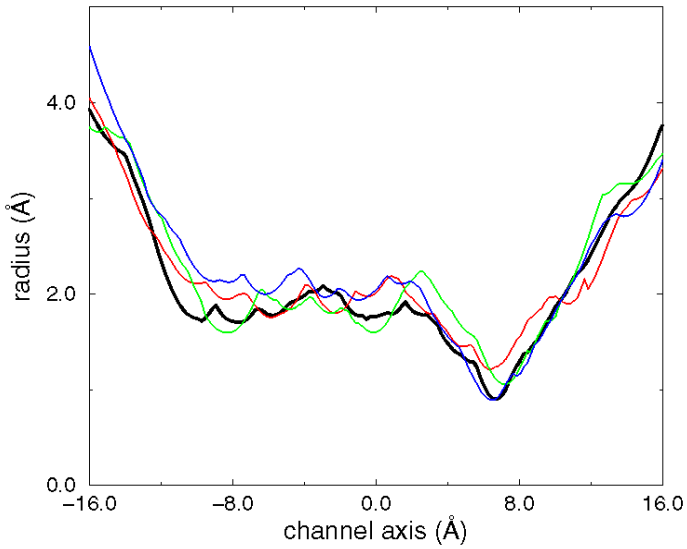


Figure 14. Model **A.** Channel radius calculated by the program HOLE (Smart et al., 1993) plotted as a function of the channel axis (see Fig. 6). The black, red, green, blue lines refer to the structure after 0, 1, 2, 3 ns MD, respectively.

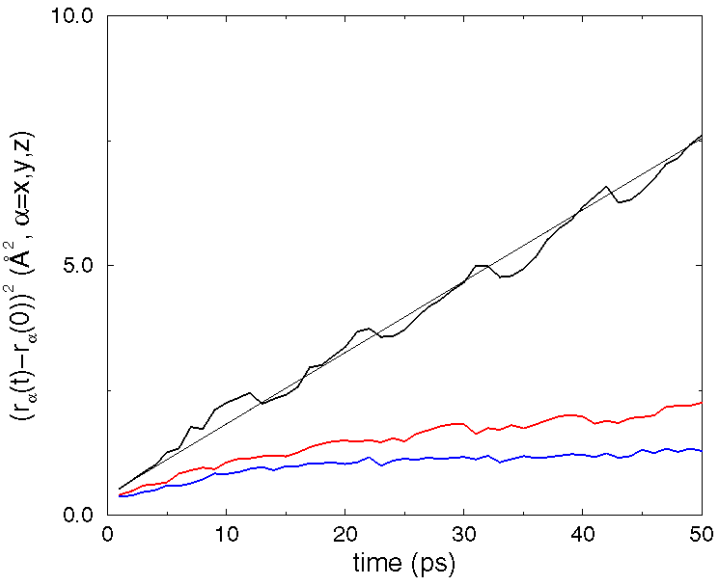


Figure 15. Model **A.** Average translational motion of water molecules in the channel along the x (red line), y (blue line) and z (black line) direction. The straight line is the linear least square fit to the diffusion along z.

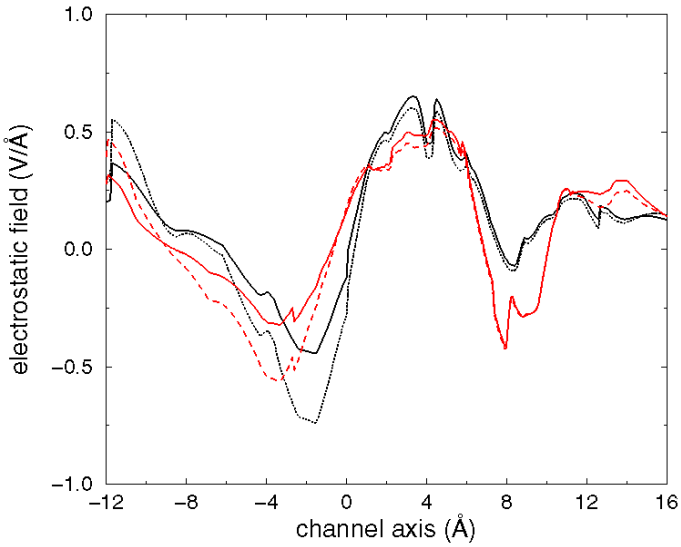


Figure 16. Model **A.** Component of the electrostatic field (E_z) along the channel axis (see Fig. 6) based on the AMBER force field (Cornell et al., 1995). Because of the very approximate form of the field, the values are here used only for comparison at the qualitative level. The black and red lines refer to the structure after 0 and 3 ns MD, respectively. The field of the entire protein is represented in solid line; that of helices M3 and M7 and of polar groups lining the pore (see Fig. 6) in dotted lines.

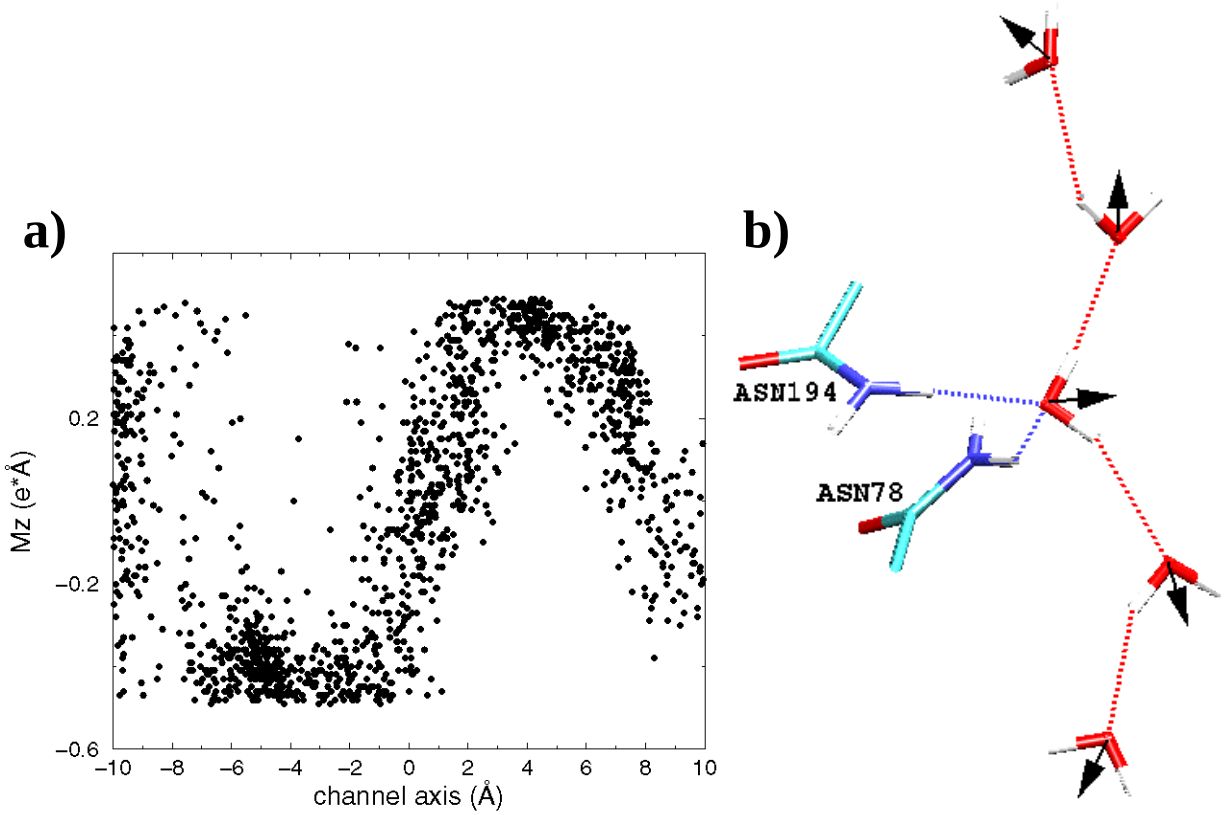


Figure 17. Model **A. a)** component of dipole along the channel axis of a water molecule permeating the pore; **b)** snapshot after ~ 2ns MD reporting the ordered water single file structure in the inner region. The water dipoles (black arrows) reorient at the Asn78 and Asn194 side chains. H-bonds are represented as dashed lines.

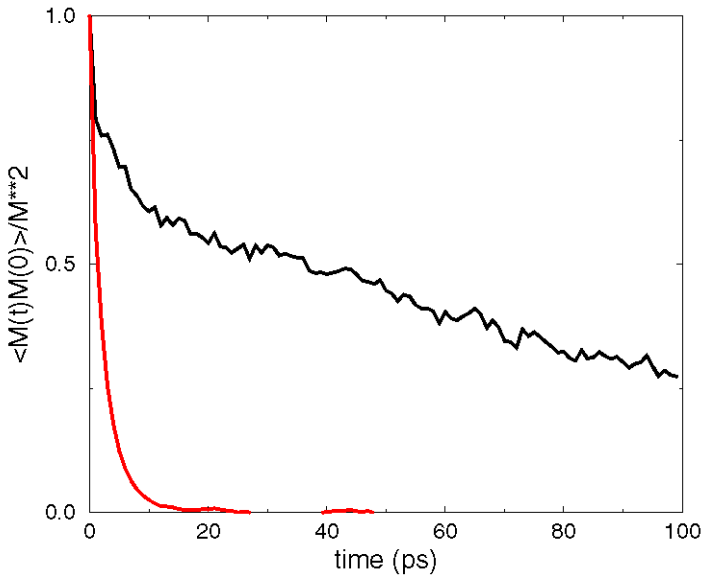


Figure 18. Autocorrelation function of the dipole moments (M) of water molecules in the pore (black line) and in the bulk phase (red line).

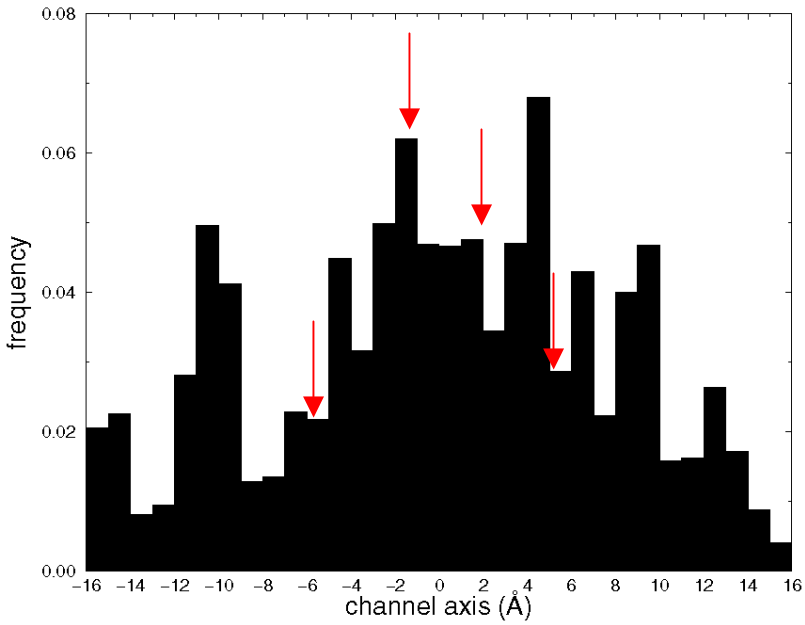


Figure 19. Water residence frequency along the channel axis (see Fig. 6), calculated during the MD simulation. The red arrows indicate the location of the crystallographic water oxygens (Sui et al., 2001).

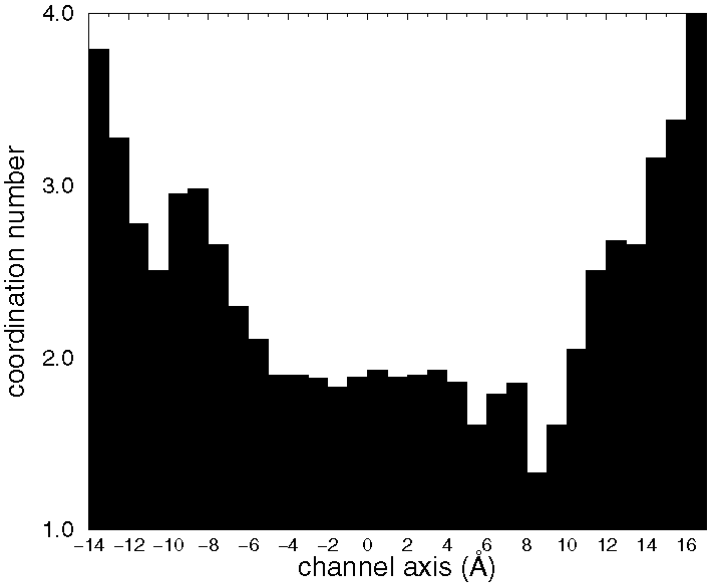


Figure 20. Water coordination number (calculated as an integral of the oxygen-oxygen radial distribution function up to 3.5 Å) along the channel axis.

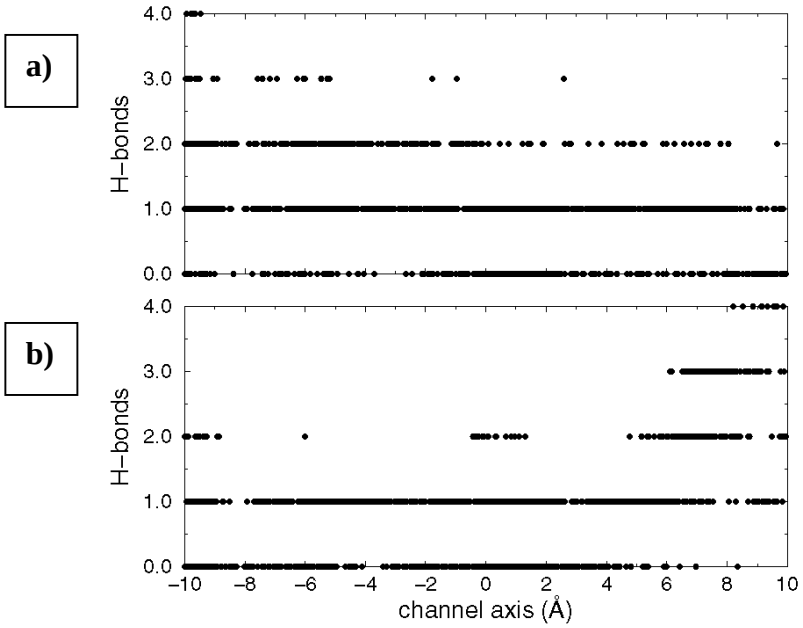


Figure 21. Number of H-bonds formed by a water molecule in the pore either with other water molecules (**a**) and or with the protein (**b**). H-bonding is here assumed if the distance between donor (X) and acceptor (Y) is lower than 3.0 Å and the angle X-H-Y is larger than 120º and lower than 180º.

3.5. REFERENCES

Alberts,B. et al. Molecular Biology of the Cell. 2002. Garland Science. London.

Agre,P. 2004. Aquaporin water channels (Nobel lecture). *Angewandte Chemie-International Edition* 43:4278-4290.

Borgnia,M., S.Nielsen, A.Engel, and P.Agre. 1999. Cellular and molecular biology of the aquaporin water channels. *Annu. Rev. Biochem.* 68:425-458.

Breed,J., R.Sankararamakrishnan, I.D.Kerr, and M.S.Sansom. 1996. Molecular dynamics simulations of water within models of ion channels. *Biophys. J.* 70:1643-1661.

Cornell,W.D., P.Cieplak, C.I.Bayly, I.R.Gould, K.M.Merz, Jr., D.M.Ferguson, D.C.Spellmeyer, T.Fox, J.W.Caldwell, and P.A.Kollman. 1995. A second-generation force field for the simulation of proteins and nucleic acids. *J.Am.Chem.Soc.* 117:5179-5197.

Costa,V. and P.Carloni. 2003. Calcium binding to the transmembrane domain of the sarcoplasmic reticulum Ca2+-ATPase: insights from molecular modeling. *Proteins* 50:104-113.

de Groot,B.L., A.Engel, and H.Grubmuller. 2003. The structure of the aquaporin-1 water channel: a comparison between cryo-electron microscopy and X-ray crystallography. *J. Mol. Biol.* 325:485-493.

de Groot,B.L. and H.Grubmuller. 2001. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* 294:2353-2357.

Fu,D., A.Libson, L.J.Miercke, C.Weitzman, P.Nollert, J.Krucinski, and R.M.Stroud. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290:481-486.

Fujiyoshi,Y., K.Mitsuoka, B.L.de Groot, A.Philippsen, H.Grubmuller, P.Agre, and A.Engel. 2002. Structure and function of water channels. *Curr. Opin. Struct. Biol.* 12:509-515.

Guidoni,L., V.Torre, and P.Carloni. 1999. Potassium and sodium binding to the outer mouth of the K+ channel. *Biochemistry* 38:8599-8604.

Heymann J.B. and Engel A. Aquaporins: Phylogeny, Structure, and Physiology of Water Channels. News Physiol Sci. 14, 187-193. 1999.

Heymann,J.B. and A.Engel. 2000. Structural clues in the sequences of the aquaporins. *J. Mol. Biol.* 295:1039-1053.

Jensen,M.O., S.Park, E.Tajkhorshid, and K.Schulten. 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A* 99:6731-6736.

kirkpatrick,S., C.D.Gelatt, and M.P.Vecchi. 1983. Optimization by Simulated Annealing. *Science* 220:671-680.

Laskowski R A, MacArthur M W Moss D S & Thornton J M. PROCHECK: a program to check the stereochemical quality of protein structures. *J.Appl.Cryst*. *26*, 283-291. 1993.

Leach,A.R. Molecular modelling. 2001. Pearson Education Limited. Harlow, Essex

Murata,K., K.Mitsuoka, T.Hirai, T.Walz, P.Agre, J.B.Heymann, A.Engel, and Y.Fujiyoshi. 2000. Structural determinants of water permeation through aquaporin-1. *Nature* 407:599-605.

Partenskii,M.B., V.Dorman, and P.C.Jordan. 1994. Influence of a channel-forming peptide on energy barriers to ion permeation, viewed from a continuum dielectric perspective. *Biophys. J.* 67:1429-1438.

Ren,G., V.S.Reddy, A.Cheng, P.Melnyk, and A.K.Mitra. 2001. Visualization of a water-selective pore by electron crystallography in vitreous ice. *Proc. Natl. Acad. Sci. U. S. A* 98:1398-1403.

Rocchia, W, Alexov, E, and Honig, B. Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J.Phys.Chem.B* 28, 6507-6514. 2001.

Saparov,S.M., D.Kozono, U.Rothe, P.Agre, and P.Pohl. 2001. Water and ion permeation of aquaporin-1 in planar lipid bilayers. Major differences in structural determinants and stoichiometry. *J. Biol. Chem* 276:31515-31520.

Settanni,G.C.A.a.P.C. 2003. Molecular Dynamics Simulations of the NGF-TrkA [AQ1] Domain5 Complex and Comparison with Biological Data. *Biophys. J*.

Shi,L.B., W.R.Skach, and A.S.Verkman. 1994. Functional independence of monomeric CHIP28 water channels revealed by expression of wild-type mutant heterodimers. *J. Biol. Chem* 269:10417-10422.

Smart,O.S., J.M.Goodfellow, and B.A.Wallace. 1993. The pore dimensions of gramicidin A. *Biophys. J.* 65:2455-2460.

Sui,H., B.G.Han, J.K.Lee, P.Walian, and B.K.Jap. 2001. Structural basis of water-specific transport through the AQP1 water channel. *Nature* 414:872-878.

Tajkhorshid,E., P.Nollert, M.O.Jensen, L.J.Miercke, J.O'Connell, R.M.Stroud, and K.Schulten. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 296:530.

Zeidel,M.L., S.V.Ambudkar, B.L.Smith, and P.Agre. 1992. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry* 31:7436-7440.

Zeidel,M.L., S.Nielsen, B.L.Smith, S.V.Ambudkar, A.B.Maunsbach, and P.Agre. 1994. Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin channel-forming integral protein in proteoliposomes. *Biochemistry* 33:1606-1615.

Zhong,Q., Q.Jiang, P.B.Moore, D.M.Newns, and M.L.Klein. 1998. Molecular dynamics simulation of a synthetic ion channel. *Biophys. J.* 74:3-10.

4. BINDING of PHOSPHINATES and PHOSPHONATES INHIBITORS to ASPARTIC PROTASES.

4.1. INTRODUCTION

4.1.1. The Human Immunodeficiency Virus and the HIV-1 Protease enzyme. In the 80's the Human Immunodeficiency Virus of type 1 (HIV-1) was identified as the causative agent of AIDS. HIV-1 is a retrovirus, i.e. the genetic information of the virus is stored on a double filament of RNA, in contrast to the other organisms which use DNA. The RNA filament is protected by a shell of proteins called capsid that is immersed in the cytoplasm(Pantaleo and Fauci, 1996) (Fig. 1). The viral genome codes for a total of 15 proteins of which only few are actually essential for the viral life cycle and can be exploited as targets for anti-AIDS therapy(Frankel and Young, 1998;Turner and Summers, 1999). One of these major targets is the HIV-1 protease enzyme (PR)(Turner and Summers, 1999;Seelmeier et al., 1988;Fitzgerald and Springer, 1991;Kohl et al., 1988).

HIV-1 PR catalyzes the reaction of hydrolysis of the peptide bond in the Gag-Pol polypeptide chain of aminoacids. This enzyme was subject of theoretical and experimental studies aimed at elucidating structural and functional aspects of its biochemistry (for a review see(Wlodawer and Gustchina, 2000;Wlodawer and Vondrasek, 1998). Here we summarize the main findings.

In the early stage of the viral life cycle HIV-1 PR is not active. It is capable to perform its activity only during the maturation phase. Premature activation leads to the production of uninfective immature virion particles(Krausslich, 1991). Thus, the regulation of the enzyme activity appears to be important for the correct development of new viral particles.

Due to its crucial importance as target for anti-AIDS therapies, the free enzyme and the complexes with inhibitors have been the subject of a large number of X-ray crystallography(Baldwin et al., 1995a;Miller et al., 1989b;Miller et al., 1989a;Silva et al., 1996;Rose et al., 1996;Navia et al., 1989) and nuclear magnetic resonance (NMR)(Wang et al., 1996a;Wang et al., 1996b;Freedberg et al., 1998;Yamazaki et al., 1994b;Yamazaki et al., 1994a) studies aimed at elucidating the 3D structure.

HIV-1 PR is a homodimeric aspartic protease(Wlodawer and Gustchina, 2000;Hong et al., 1998;Baca and Kent, 1993) composed of two identical subunits of 99 amino acids each. The sequence of residues is commonly numbered from 1 (N-terminus) to 99 (C-terminus) in subunit 1 and from 1' to 99' in subunit 2 (Fig. 2).

As in all the other aspartic proteases known so far, the active site is characterized by a pair of Aspartate(Asp)-Threoneine(Thr)-Glycine(Gly) sequences (Davies, 1990) (Fig. 2). The cleavage site is located at the interface between the two subunits, each subunit contributing with one triplet of residues. Only the two Asps (Asp dyad) are directly involved in the chemical catalysis. These groups are located very close to each other and almost coplanar. This peculiar conformation of the active site Asp dyad is a distinctive feature of aspartic proteases and is crucial for enzymatic activity and drug binding(Fitzgerald and Springer, 1991).

The highly conserved Thr26(26') and Gly27(27') residues(Davies, 1990) do not participate directly in the catalysis. The crystal structures of all aspartic proteases reveal that they are involved in a network of hydrogen bonds surrounding the active site (forming the so called *fireman's grip*, Fig. 3)(Navia et al., 1989) thus rendering this region rather rigid. This hydrogen bond network is characteristic of aspartic proteases and essential for enzymatic

activity(Loeb et al., 1989). In the free enzyme the active site is located at the bottom of a large water-filled cavity. When the enzyme is bound to a ligand (namely substrates or inhibitors molecules) the cavity is almost completely occupied by the ligand atoms. The Gibbs free energy of inhibitor binding is dominated by entropy changes(Velazquez-Campoy et al., 2001), a distinctive feature of a binding process driven by hydrophobic interactions. Indeed most of the residues surrounding the active site are hydrophobic.

On the opposite side of the Asp dyad, two large flaps delimit the active site. The flaps are flexible(Freedberg et al., 1998;Harte et al., 1990;York et al., 1993) and experience large rearrangements after inhibitor binding(Miller et al., 1989b;Wlodawer and Erickson, 1993); these rearrangements have large effects on the energetics and on the protease stability(Todd and Freire, 1999) as the flaps contribute to most of the binding energy for the inhibitors(Luque et al., 1998).

4.1.2. Drug design. Since the discovery that inhibition of the protease activity with pepstatin leads to the production of immature, non-infectious viral particles(Seelmeier et al., 1988), HIV-1 PR has become a fundamental target for the computer aided design of anti-AIDS drugs(Wlodawer and Erickson, 1993;Mehanna, 2003;Wlodawer and Vondrasek, 1998). As the enzyme is expected to strongly interact with transition-state mimics, most of the inhibitors are designed such as to resemble closely either the transition states or the reaction intermediate. Eight drugs are now routinely used in anti AIDS treatments.

Despite the relative successes obtained, the development of drugs against HIV-1 PR is continuously frustrated by the fast development of mutations that confer to the virus resistance against a particular pharmacological agent. Several of these mutations are known and have been studied(Baldwin et al., 1995b). Most of the characterized mutations

belongs to residues close to the active site and directly disrupt favorable enzyme-drug interactions. Nevertheless there are mutations that affect residues far from the active site. The mechanisms that lead to drug-resistance in these cases are often not known.

Moreover, it is now recognized that Protease inhibitors therapy is associated with serious side effects, likely due to inhibition of a number of unrelated molecules(Hruz et al., 2001;Carr, 2003).

Thus, a large effort is devoted to identify new inhibitors, which could possibly overcome these drawbacks.

The affinity of the FDA-approved drugs is caused mostly by the entropy associated to the transfer of partially ordered water in the active site of the enzyme to the bulk solvent(Velazquez-Campoy et al., 2001).

This is a result of the current paradigms in drug design, which usually generate highly hydrophobic and conformationally constrained ligands. Recently, it has been proposed a possible strategy to develop new inhibitors(Velazquez-Campoy et al., 2000). The main idea is to relax the conformational rigidity of the ligands and to compensate the loss of this favourable conformational entropy contribution through highly stabilizing interactions(Velazquez-Campoy A et al., 2001).

Among such kind of possible interactions are Low Barrier Hydrogen Bond (LBHB), as it has been suggested, based on ab initio quantum chemistry calculations, that these may provide stabilization energy as large as few tens of kcal/mol(Gerlt et al., 1997;Cleland and Kreevoy, 1994). Requirements for the formation of LBHB are i) short Donor-Acceptor

distances; ii) low polarity of the medium; iii) similar pKa between Donor and Acceptor groups.

A potent class of HIV-1 PR inhibitors is that of phosphinate and phosphonate derivatives (Fig. 4). In these compounds, the PO2 group is a noncleavable transition state analogue that binds at the AP active site(Wlodawer and Erickson, 1993;Bartlett et al., 1987;Bartlett and Giangiordano, 2003). These inhibitors form H-bonds in which the distance between H-bond donor and acceptor is as small as ~2.5 Å. The mode of binding of one of these inhibitors (**I1** in Fig. 4) has been provided by X-ray crystallography at atomic resolution(Abdel-Meguid et al., 1993). The phosphinic group binds to the outer aspartic dyad oxygens (d(O@Asp25-OP2)= 2.55 Å and d(O@Asp25'-OP1)= 2.84 Å, Fig. 4). Similar modes of binding are found in complexes of eukaryotic isoenzymes, in which d(O@Asp35-OP2)= ~2.4-2.5 Å in Fig. 4). However, in these complexes the phosphinic group binds to one outer and one inner oxygen of the Asp dyad (Fig. 4). The different H-bond network might be caused by the different geometry at the active site and/or the presence, only in eukaryotic AP's, of H-bonds with two side chain OH groups (Fig. 4).

1H-NMR is a powerful probe for detecting LBHB's in proteins, as the chemical shifts are characterized by highly downfield signals(Tobin et al., 1995;Frey et al., 1994). So far, experiments have been done for the I2/endothiapepsin complex(Coates et al., 2002). Two signals at 16.1 and 18.5 ppm in the 1H-NMR spectrum have been observed, suggesting that only two hydrogens are located at the ligand/Asp dyad interface.

4.1.3. Motivation and outline of present work. Due to its importance in modern pharmacology, HIV-1 PR is one of the most studied enzymes. As mentioned above, the demand to develop new inhibitors that are able to adapt to active site mutations is still

actual. A possible route is to identify strong protein ligand interactions. Here we use density functional theory (DFT) and QM/MM methods to investigate the interactions of the phosphinate and phosphonate transition state mimetic bound to the AP catalytic dyad. We have investigated eukaryotic enzymes together with HIV-1 PR. This was done in order to investigate how the polarity of the medium affects the nature of the H-bond, since in the eukaryotic enzymes a couple of conserved hydroxyl groups, not present in the viral enzyme, H-bond to the outer oxygens of the catalytic dyad. Moreover, a phosphinate and a phosphonate derivatives bound to eukayotic enzymes were investigated in order to asses how pKa differences might affect hydrogen location.

Car Parrinello ab initio molecular dynamics simulations are suitable to investigate the physical origin of the stability of the active site. In its implementation with BLYP exchange-correlation functional, DFT allows a reliable description of relatively large systems. Furthermore, temperature effects, of outmost importance for biological function, are fully included. Moreover, within the QM/MM approach, the entire protein may be included into the calculations.

4.2. MODEL SYSTEMS and METHODS

As for HIV-1 AP, we have used the only phosphinate complex for which experimental 3D structure is available(Abdel-Meguid et al., 1993). For the eukaryotic isoenzymes penicillopepsin and endothiapepsin, we have focused on the two structures solved at

highest resolution, namely that of I2 bound to endothiapepsin(Coates et al., 2002) (Fig. 4) and I3 bound to penicillopepsin(Khan et al., 1998) (Fig. 4).

As for the eukaryotic protein, we focus on the diprotonated state, which has been suggested by 1H-NMR experiments. In models of increasing complexity of the *indothiapepsin/I2 complex* (Scheme 1), the inhibitors were represented as dimethyl phosphinic acid, and the Asp dyad as a couple of acetic acids. The Asp-Thr-Gly residues, characteristic of all AP(Davies, 1990), and the Ser38/Thr222, conserved among the eukaryotic AP(Davies, 1990;Dunn et al., 2002), have been progressively included in the models (Scheme 1). The model of *penicillopepsin/I3 complex* contains the phosphonic moiety, the catalytic Asp side chains and the Thr-Gly-Ser/Thr backbone units and hydroxyl side chains (Fig. 5 and Scheme 1). In both complexes, Hydrogen atoms have been added using standard bond lengths and bond angles. For all possible protomers the position of the hydrogen atoms have been optimized at the DFT-BLYP level, while keeping heavy atoms at their crystallographic conformation. 1H-NMR shielding constants have been calculated for the endothiapepsin models. Notice that DFT-geometry optimizations along with QM/MM calculations[°] on the eukaryotic enzymes turned out not to be consistent with high-resolution structural data.

As no experimental information on the protonation state of HIV-1 AP is available, we consider two plausible protonation states for the viral protein (Fig. 5), the diprotonated, negatively charged (HIV(-) hereafter) and the neutral, triprotonated (HIV(0)) active site forms have been considered. For this system, QM/MM Car-Parrinello simulations were carried out. The QM region comprised the inhibitor phosphinic moiety and the catalytic Asp side chains. 2.8 and 2.2 ps of QM/MM MD simulations were performed for models HIV(-) and HIV(0), respectively, based on the X-ray conformation. Finally, 5.4 ps long QM/MM

[°] The QM region comprised the phosphinic moiety of the inhibitor, the Asp dyad side chains, the Thr/Gly peptide unit and the Ser38 and Thr222 side chains.

MD simulations for model HIV(-) have been performed based on the last snapshot of a 2 ns long classical MD simulation.

4.3. RESULTS

We first determine the location of protons in gas phase models of the endothiapepsin complex, for which NMR data are available(Coates et al., 2002). Then, the binding site protonation state of the other proteins is presented.

4.3.1. Endothiapepsin/I2 complex. In order to investigate the effect of the size of the model on the accuracy of the chemical shifts calculations, we optimized the position of hydrogen atoms in models of increasing complexity. Notice that fully geometry optimizations, along with QM/MM calculations turned out not to be consistent the with high-resolution structural data.

It has been long pointed out that the minimal model of the protein includes the Asp moiety and the Thr/Gly peptide groups, H-bonding to the inner dyad oxygens(Beveridge and Heywood, 1993). These groups, along the phosphinic moiety representing the ligand moiety binding to the cleavage site, are included in our smallest model, **I-A**. In the energy-minimized form, the two protons are located at OP2 of the phosphinic group and at O1 of ASP219. The calculated chemical shifts are slightly overestimated relative to experiment (Tab. 1) and they suggest that the most deshielded hydrogen is that located between the closest oxygen atoms, as has been suggested based on the known dependence of the proton shifts on the interoxygen distance(Coates et al., 2002).

I-B includes also the residues H-bonding to the Asp dyad, the side chains of Ser38 and Thr222 (Fig. 4 and Fig. 5). This protomer is very similar to **I-A**, with a slight increase (less

than 0.1 A, Tab. 1) of the O2@Asp35 – H2 distance. The calculated chemical shifts are better than those of **I-A,** presumably because of the decreased base strength of the carboxylic groups caused by the two interacting groups. Finally, **I-C**, which includes also the peptide units of Asp35 and Asp219, is very similar to **I-B** with a small difference in the calculated chemical shift by about 0.5 ppm, clearly caused by a shielding effect of the two added peptide units.

We conclude that (i) agreement is obtained between calculated and experimental structural and NMR data(Coates et al., 2002). (ii) The added groups on passing from **I-A** to **I-C** may contribute to the 1H shielding constants in two ways, with a direct effect on the magnetic properties of the system and with an indirect effect, as the calculated values depend on the location of the proton along the interoxygen axis.

4.3.2. Penicillopepsin/I3 complex. The binding mode of the phosphinic group to endothiapepsin is very similar to that of the phosphonic one to penicillopepsin: the rmsd of the common active site heavy atoms being 0.17 Å. However, the different acidity of the two inhibitors (pKa 3.0 and 1.5 for the phosphinate and phosphonate group, respectively) renders non-trivial the determination of the protonation state in this complex. We consider here a model equivalent in size to **I-B** (model **II-B**, Fig. 5 and scheme 1), as the latter appears already able to capture the structural and NMR properties of the binding mode of the endothiapepsin/I2 complex , along the models investigated.

The most stable protomer features the neutral phosphonic group hydrogen bonded to the negatively charged catalytic dyad (Table 1 and Figure 5). This is the same protonation pattern as in model **I-B.**

4.3.3. HIV-1 Protease/I1 complex. The binding mode is different than that of the eukaryotic isoenzymes (Fig. 4) and both the di- and triprotonated forms may be postulated.

In this case, hybrid QM/MM MD on both ionization states, (models HIV(-) and HIV(0), Fig. 5) have been performed. In this approach, the side chains of the catalytic dyad and the phosphinic moiety of the inhibitor have been described at the quantum level. These calculations revealed that the most stable ionization state is the diprotonated state in terms of structural parameters characterizing the binding mode (Tab. 2 and Fig. 6). The hydrogen bonding pattern is maintained during the dynamics. In particular, hydrogen atoms (H1 and H2 in Fig. 4) reside for all the dynamics on the outer aspartic oxygens. These two hydrogen bonds are equivalent on the time scale investigated, as can be inferred from the reported average H-bond parameters (Table 3). This result is consistent with the fact that both the ligand and the enzyme are C2 symmetric. In the X-ray structure two conformations of the inhibitor were detected, in which the distance between one Asp25 and the inhibitor's phosphinic group was either 2.55 Å or 2.67 Å (the distance to the other Asp was similar it the two conformations). Test QM/MM simulations were carried out for 2.1 ps in which the backbone units of Thr26/26'-Gly27/27'-Ala28/28' have been described at the quantum level (model HIV(-)B) because of their importance for the structural properties of the enzyme(Piana and Carloni, 2000). It turns out that the structural features are nearly identical to those of the smaller system (Tab. 3).

4.4. DISCUSSION

We have presented an investigation of the protonation state of eukaryotic endothiapepsin and penicillopepsin and HIV-1 Aspartic Proteases in complex with phosphinate and phosphonate inhibitors. Whereas QM/MM simulations could be performed on the viral

enzyme, DFT-geometry optimizations along with QM/MM calculations turned out not to be consistent with high-resolution structural data and for the eukarytic complexes only the optimization of the hydrogen position could be carried out. In the case of eukaryotic AP's, our molecular models include the entire Asp-Thr-Gly conserved sequence characteristic of aspartic proteases, which plays an important role for the structural properties(Beveridge and Heywood, 1993;Sirois et al., 2003). The most stable protomer features the neutral phosphinic group hydrogen bonded to the negatively charged catalytic dyad. The calculated values suggest that the downfield 1H chemical shift is characteristic of short hydrogen bonds. This hypothesis is consistent with that by Garcia-Viloca et al.(GarciaViloca et al., 1997), who, based on studies on gas phase calculations of dicarboxylic acids, postulated that protons involved in LBHB always display unusually downfield 1H-NMR chemical shifts, but the opposite is not necessarily true, thus suggesting that low-field proton signals in NMR spectra are not conclusive evidence for a LBHB.

In the case of HIV-1 PR, QM/MM calculations suggested that the protons are located at the outer oxygens of the Asp dyad, and that the two hydrogen bonds between the Asp dyad and the phosphinic group are equivalent on the time scale investigated (Table 3). This result is consistent with the fact that both the ligand and the enzyme are C2 symmetric. In the X-ray structure two conformations of the inhibitor were detected, in which the distance between one Asp25 and the inhibitor's phosphinic group was either 2.55 Å or 2.67 Å (the distance to the other Asp was similar it the two conformations). This discrepancy between theory and experiment could be caused by several factors: (i) the very short time scale investigated; (ii) uncertainties in the X-ray structure, solved at 2.3 A resolution. (iii) the BLYP approximation of the exchange-correlation functional.

In conclusions, our calculations provide information on the mode of binding of an important class of inhibitors, phosphinates and phosphonates, suggesting that the protons are

located at the catalytic Asp dyad in the viral complex, while in the eukaryotic complexes the proton shared by the closest oxygen atoms is located at the phosphinic/phosphonic group.

4.4.1. H-bond network in the catalytic mechanism. In this final section, we investigate whether in the enzymatic reaction catalyzed by AP's, the substrate can form H-bond structurally similar to those formed by the inhibitors investigated here. To this aim, we briefly summarize our recent work on the reaction mechanism of one viral and one eukaryotic AP's, HIV-1 Protease and human β-secretase(Piana et al., 2004;Cascella et al., 2005). These enzymes feature the two known folds of AP(Tyndall et al., 2005).

The overall reaction of HIV-1 PR(Piana et al., 2004) and the first chemical step of β-secretase(Cascella et al., 2005) were investigated using the same computational approach as that presented here, namely hybrid QM/MM Car-Parrinello MD simulations(Laio A. et al., 2002).

Our calculations, along with those of refs(van Gunsteren et al., 1998;Silva et al., 1996;Okimoto et al., 1999), suggest that the enzymes lower the free energy of the reaction by providing a proton donor and a proton acceptor group in the proper positions. Specifically, in the two chemical steps of the reaction of HIV-1 AP(Piana et al., 2004), a double proton transfer that involves the substrate and the two catalytic Asp is observed. The first chemical step of the reaction catalysed by β-secretase follows a similar mechanism(Cascella et al., 2005), although in this case the proton transfer is not concerted[+].

[+] The second step has not been investigated.

Tab. 4 reports the average H-bond distances within the active site upon the formation of the transition states (TS1 and TS2, Fig. 7) and the gem-diol intermediate (INT, Fig. 7). In both APs, short strong hydrogen bonds are formed during the reaction, with distances resembling those encountered in the inhibitor-enzyme complexes investigated here These are, for TS1 and TS2, between the substrate's carbonyl oxygen and one of the outer dyad oxygens (i). These might resemble the OP_2···H2···O2 interactions in the inhibitor/enzyme complexes (Fig. 4); for TS1 and INT, between the water oxygen and one outer dyad oxygen (ii). As a result of these short distances, a series of proton transfer events occur (for more details see refs (Piana et al., 2004;Cascella et al., 2005)). Thus, we conclude that structurally similar strong hydrogen bonds between one outer oxygen of the catalytic dyad and one oxygen ligand are formed in the transition states of the reaction and with the PO2 moiety of phosphinate and phosphonate inhibitors.

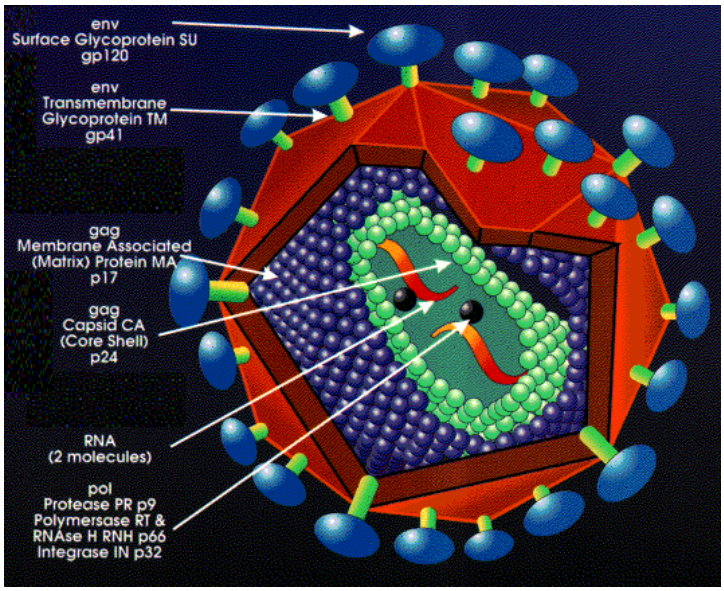


Figure 1. The Human Immunodeficiency Virus Type-1 (HIV-1). The HIV-1 protease is enclosed in a capside (green) along with RNA and the Reverse Trascriptase enzyme (RT).

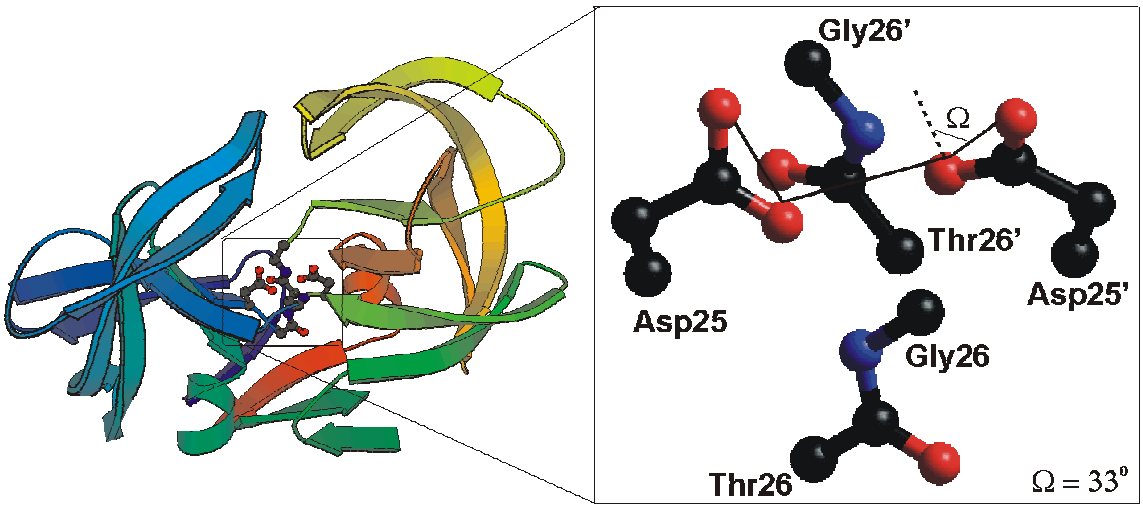


Figure 2. The active site of HIV-1 PR. The active site of HIV-1 PR is located at the interface between the subunits. According to the numbering in the text the cleavage site is composed by Asp25(25')-Thr26(26')-Gly27(27'). The two subunits are related by a C_2 symmetry axis.

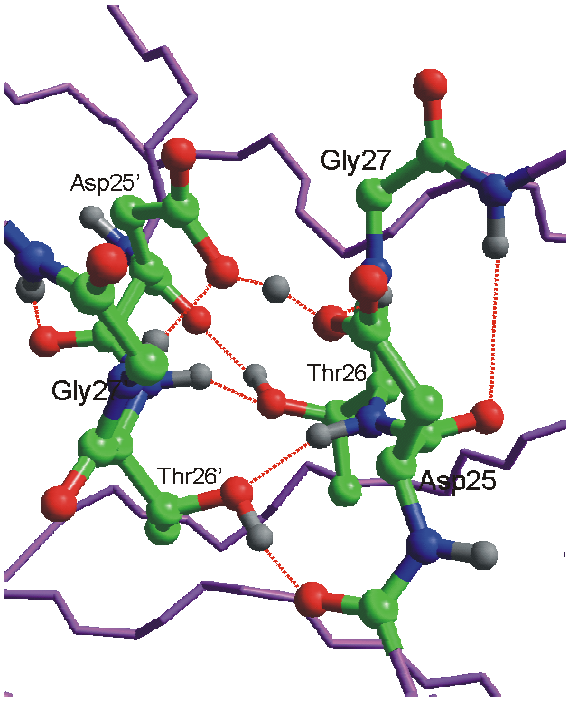


Figure 3. The "fireman's grip".

H-bonds are represented as dashed lines.

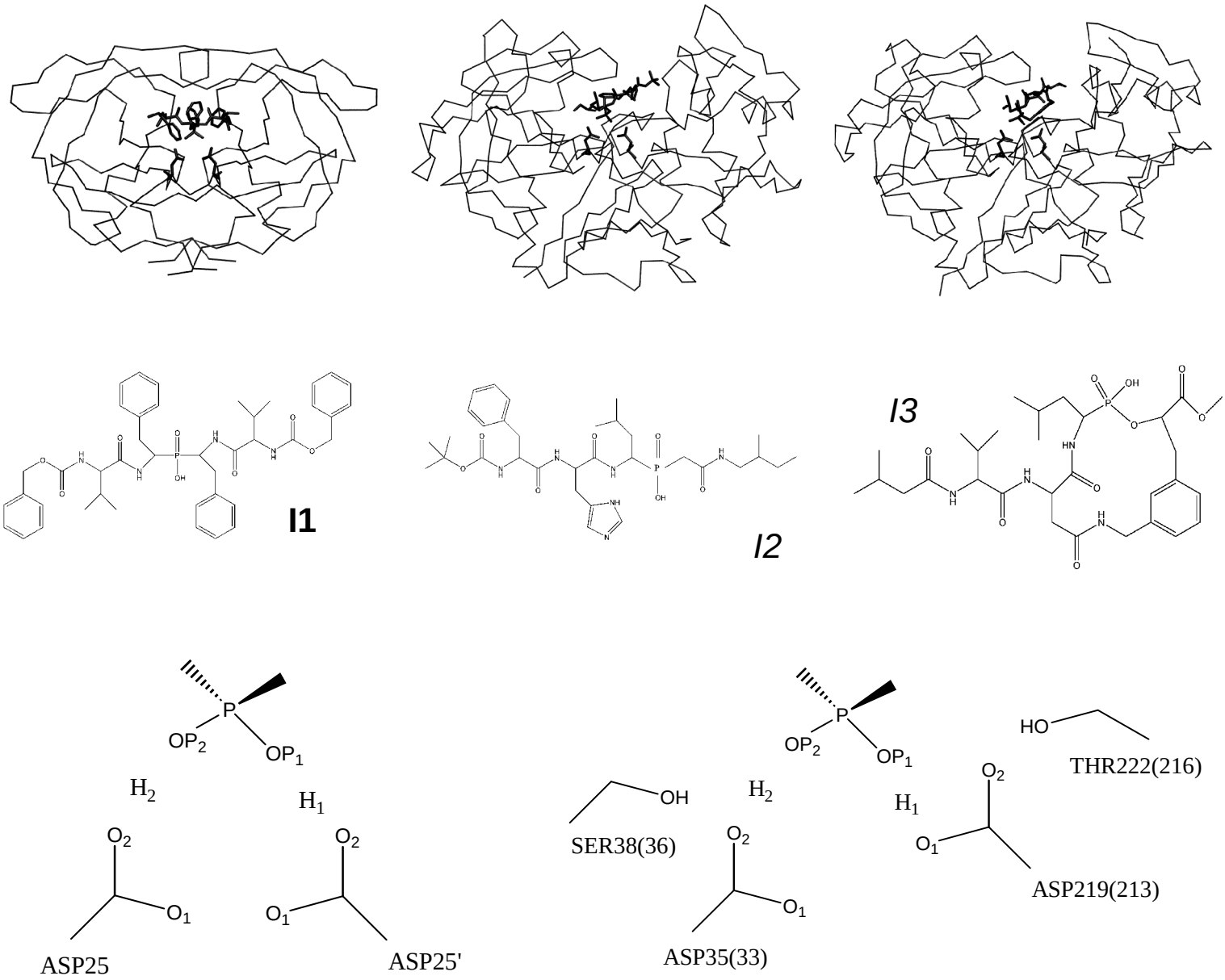


Figure 4. Structures of the aspartyl proteases considered in this work. *Top*: HIV-1 AP (left), endothiapepsin (middle) and penicillopepsin (right) X-ray structures. The viral and eukaryotic proteins feature a different fold: HIV-1 AP is a homodimer with the catalytic Asp dyad located on two loops at the monomer-monomer interface; endothiapepsin and penicillopepsin are α/β monomers composed of two asymmetric lobes, with the catalytic Asp dyad located at the lobe interface, the so called "pepsin-like" protease.Thin line: protein Cα trace; thick line: inhibitor and catalytic Asp dyad. *Middle:* HIV-1 AP (left), endothiapepsin (middle) and penicillopepsin (right) inhibitors. *Bottom*: binding site of HIV-1 AP/I1 (left) and endothiapepsin/I2 (right, penicillopepsin residues numbers in parentheses) complexes.

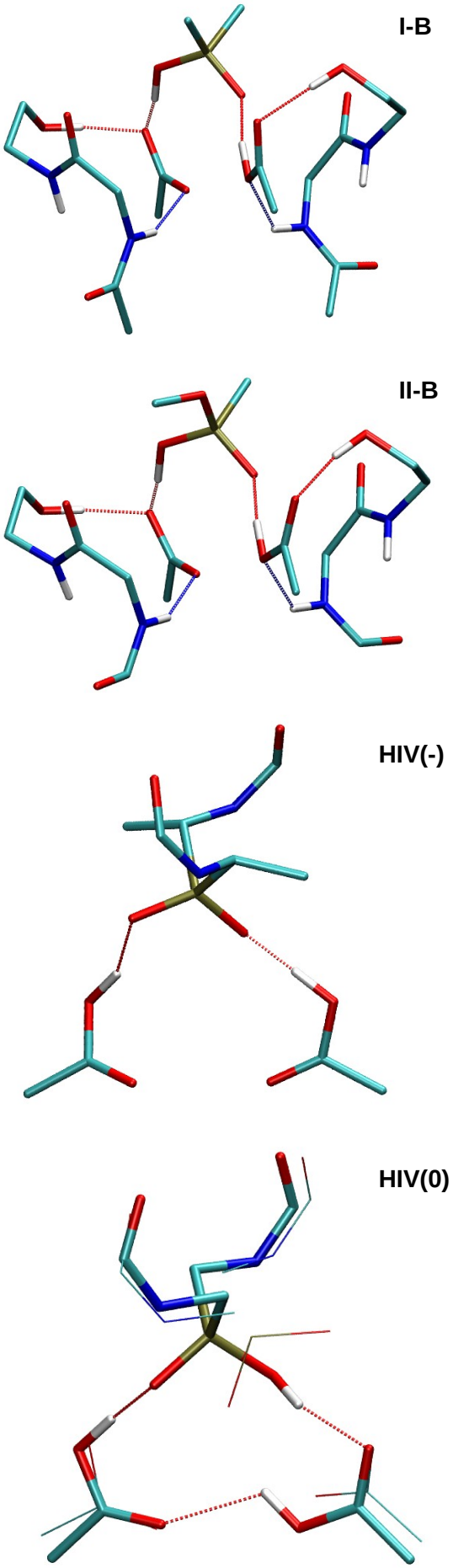


Figure 5. Representative QM models used in this work. *Top:* Endothiapepsin/I2 active site model **I-B**. *Middle:* Penicillopepsin/I3 active site: model **II-B**. *Bottom:* HIV-1 AP/I1 binding site models **HIV(-)** and **HIV(0)**, QM subsystems in the QM/MM calculations (See Text).

Thick line: crystallographic structure; *thin line*: QM/MM MD structure after ~2 ps. Only polar hydrogens have been represented. Dotted lines represent hydrogen bonds (donor-acceptor distance 3.2 Ang, and donor-hydrogen-acceptor angle 120-180 º).

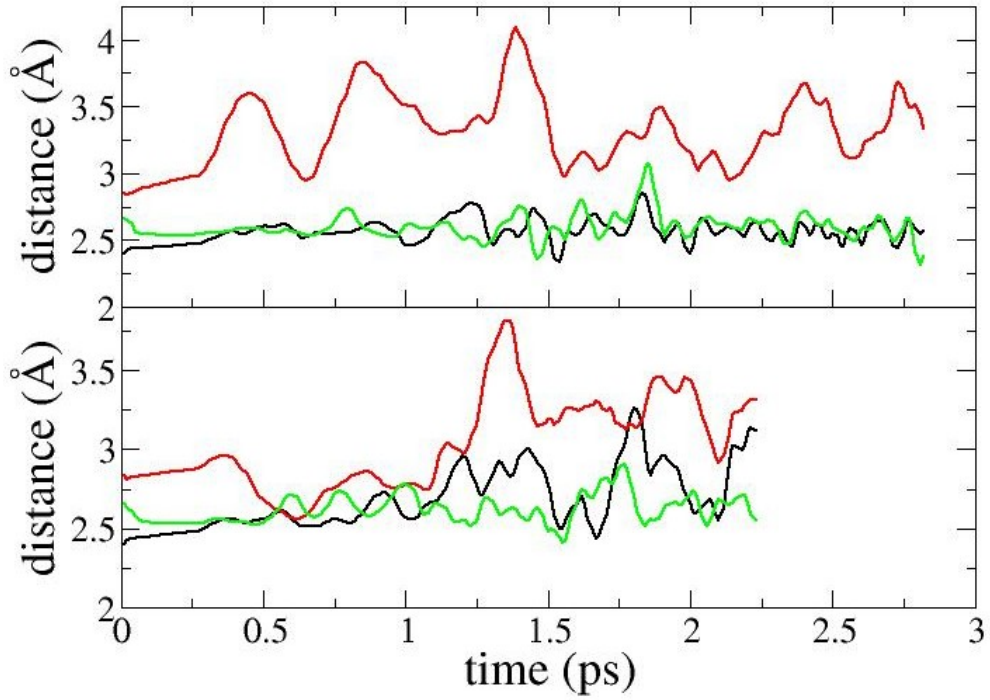


Figure 6. QM/MM MD: active site distances between oxygen atoms during the simulation of **HIV(-) (**top) and **HIV(0)** (bottom) active site forms. Red: O1@Asp25-O1@Asp25'; black: O2@Asp25'-OP1; green: O2@Asp25-OP2. Atom names as in Figure 4.

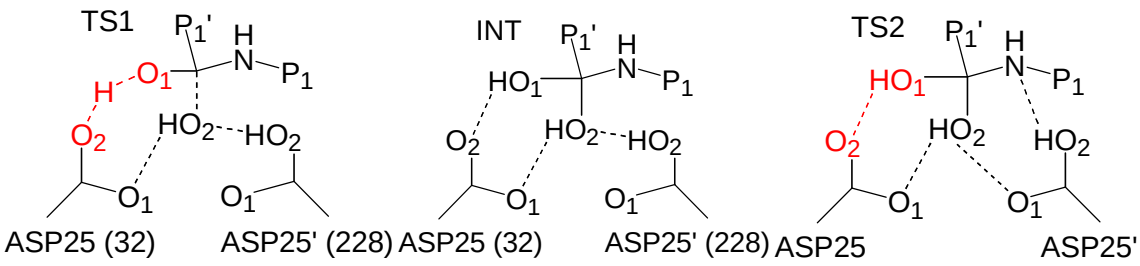


Figure 7. Scheme representing the adducts between AP's catalytic dyad and TS1, INT and TS2 species in the enzymatic reaction mechanism suggested by refs(Piana et al., 2004;Cascella et al., 2005). The substrates studied were: Thr-Ile-Met-Met-Gln-Arg(Piana et al., 2004) and Glu-Val-Asn-Leu-Ala-Ala-Glu-Phe(Cascella et al., 2005). The Met-Met and the Leu-Ala peptide bonds are cleaved during the reaction. O1@TS1/INT/TS2 is the substrate's carbonyl oxygen, while O2@TS1/INT/TS2 belongs to the catalytic water. The H-bonds featuring some similarities with those of the inhibitors investigated here (see text) are colored in red. The overall charge of the active site is the same as for the phosphinic/APs complexes investigated in this work. However, the number of hydrogen atoms is different and the interactions between the reactive species and the Asp dyad display some differences.

model	σH_1	dO_1O_1	dO_1@Asp219H_1	$dO_{1P}H_1$	σH_2	dO_2O_2	dO_2@Asp35H_2	$dO_{2P}H_2$
I-A	17.1	2.58	1.07	1.52	19.4	2.41	1.27	1.13
I-B	17.5	2.58	1.07	1.51	18.6	2.41	1.32	1.09
I-C	18.0	2.58	1.08	1.50	18.5	2.41	1.33	1.08
II-B		2.54	1.09	1.46		2.43	1.29	1.15

Table 1. **Structural and NMR data of eukaryotic aspartyl proteases complexes.** Calculated distances (Å) and calculated NMR chemical shifts (ppm). The H-bond angles are very similar in all structures (170º±5º) and are not reported here. Experimental chemical shifts(Coates et al., 2002): H1 16.1 ppm, H2 18.5 ppm. Atoms names as in Fig. 4.

Model	RMSD backbone (Å)	RMSD active site (Å)
HIV(-)	1.3	1.5
HIV(0)	1.2	2.0

Table 2. HIV-1 AP**:** RMSD of the final QM/MM MD conformation with respect to the X-ray structure after least square fit to the protein backbone.

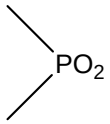
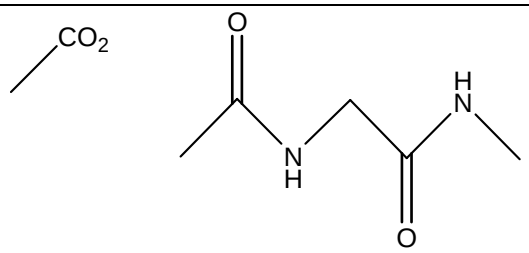
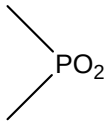
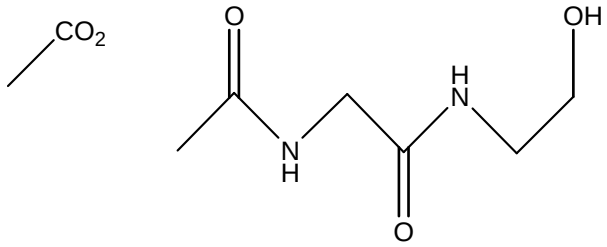
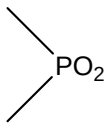
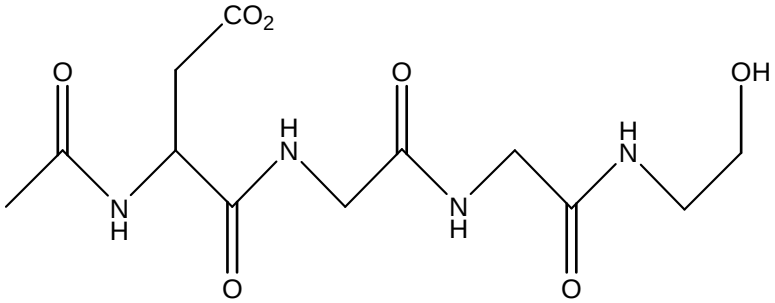
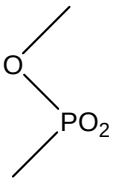
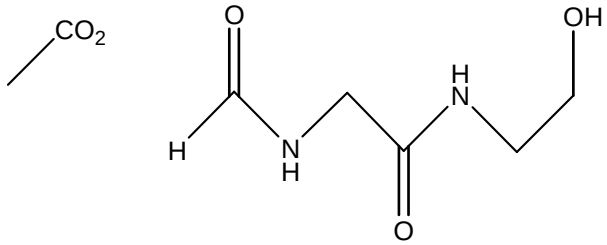
distances (Å)	model HIV(-)	model HIV(-)B
OP2-O2@Asp25	2.56 ± 0.09	2.60 ± 0.10
OP1-O2@Asp25’	2.55 ± 0.08	2.57 ± 0.10
O1@Asp25-O1@Asp25’	3.27 ± 0.19	3.05 ± 0.20
O2@Asp25-H2	1.05 ± 0.04	1.03 ± 0.04
OP2-H2	1.53 ± 0.11	1.60 ± 0.11
O2@Asp25’-H1	1.05 ± 0.03	1.04 ± 0.04
OP1-H1	1.51 ± 0.09	1.55 ± 0.12

Table 3. Average H-bond distances from QM/MM MD simulations of models **HIV(-)** and **HIV(-)B**. H-bond angles are very similar in both models (165º ± 10º) and are not reported. Atom names as in Figure 4.

	O1-O2@Asp25(32)	**O2-O1@Asp25(32)**	**O2-O2@Asp25’(228)**
β-secretase / TS	2.66 ± 0.17	2.59 ± 0.10	2.53 ± 0.08
β-secretase / INT	2.87 ± 0.17	2.71 ± 0.06	2.54 ± 0.07
HIV-1 / TS1	2.53 ± 0.11	2.87 ± 0.15	2.50 ± 0.08
HIV-1 / INT	2.68 ± 0.12	3.04 ± 0.45	2.61 ± 0.16
	O1-O2@Asp25	**O2-O1@Asp25**	**O2-O1@Asp25’**
HIV-1 / TS2	2.48 ±0.07	3.10 ± 0.27	2.70 ± 0.09

Table 4. QM/MM-averaged hydrogen bond distances observed in the reaction of HIV-1 PR and β-secretase with model substrates(Piana et al., 2004;Cascella et al., 2005).

Scheme 1. Chemical structures of the fragments in the models used in the DFT calculations of the euraryotic aspartyl proteases.

I.endothiapepsin complex	**inhibitor**	**protein**
A		 ASP35/219 THR36/220 GLY37/221 SER38/THR222
B		 ASP35/219 THR36/220 GLY37/221 SER38/THR222
C		 PHE34/ALA218 ASP35/219 THR36/220 GLY37/221 SER38/THR222
II.penicillopepsin complex		
B		 ASP33/213 THR34/214 GLY35/215 SER36/THR216

3.5 REFERENCES

Abdel-Meguid,S.S., B.Zhao, K.H.Murthy, E.Winborne, J.K.Choi, R.L.DesJarlais, M.D.Minnich, J.S.Culp, C.Debouck, T.A.Tomaszek, Jr. 1993. Inhibition of human immunodeficiency virus-1 protease by a C2-symmetric phosphinate. Synthesis and crystallographic analysis. *Biochemistry* 32:7972-7980.

Baca,M. and S.B.H.Kent. 1993. Catalytic Contribution of Flap-Substrate Hydrogen-Bonds in Hiv-1 Protease Explored by Chemical Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 90:11638-11642.

Baldwin,E.T., T.N.Bhat, S.Gulnik, B.S.Liu, I.A.Topol, Y.Kiso, T.Mimoto, H.Mitsuya, and J.W.Erickson. 1995a. Structure of Hiv-1 Protease with Kni-272, A Tight-Binding Transition-State Analog Containing Allophenylnorstatine. *Structure* 3:581-590.

Baldwin,E.T., T.N.Bhat, B.S.Liu, N.Pattabiraman, and J.W.Erickson. 1995b. Structural Basis of Drug-Resistance for the V82A Mutant of Hiv-1 Proteinase. *Nature Structural Biology* 2:244-249.

Bartlett,P.A., C.K.Marlowe, P.P.Giannousis, and J.E.Hanson. 1987. Phosphorus-containing peptide analogs as peptidase inhibitors. *Cold Spring Harb. Symp. Quant. Biol.* 52:83-90.

Bartlett,P. and M.Giangiordano. 2003. Transition state analogy of phosphonic acid peptide inhibitors of pepsin. *journal of organic chemistry* 61:3433-3438.

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

Beveridge,A.J. and G.C.Heywood. 1993. A quantum mechanical study of the active site of aspartic proteinases. *Biochemistry* 32:3325-3333.

Carr,A. 2003. Toxicity of antiretroviral therapy and implications for drug development. *Nat. Rev. Drug Discov.* 2:624-634.

Cascella,M., C.Micheletti, U.Rothlisberger, and P.Carloni. 2005. Evolutionarily conserved functional mechanics across pepsin-like and retroviral aspartic proteases. *Journal of the American Chemical Society* 127:3734-3742.

Cleland,W.W. and M.M.Kreevoy. 1994. Low-barrier hydrogen bonds and enzymic catalysis. *Science* 264:1887-1890.

Coates,L., P.T.Erskine, M.P.Crump, S.P.Wood, and J.B.Cooper. 2002. Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism. *J Mol Biol.* 318:1405-1415.

Davies,D.R. 1990. The structure and function of the aspartic proteinases. *Annu. Rev. Biophys Biophys Chem* 19:189-215.

Dunn,B.M. 2002. Retroviral proteases. *Chem. Rev. 102,* 4431-4458.

Fitzgerald,P.M.D. and J.P.Springer. 1991. Structure and Function of Retroviral Proteases. *Annual Review of Biophysics and Biophysical Chemistry* 20:299-320.

Frankel,A.D. and J.A.T.Young. 1998. HIV-1: Fifteen proteins and an RNA. *Annual Review of Biochemistry* 67:1-25.

Freedberg,D.I., Y.X.Wang, S.J.Stahl, J.D.Kaufman, P.T.Wingfield, Y.Kiso, and D.A.Torchia. 1998. Flexibility and function in HIV protease: Dynamics of the HIV-1

protease bound to the asymmetric inhibitor kynostatin 272 (KNI-272). *Journal of the American Chemical Society* 120:7916-7923.

Frey,P.A., S.A.Whitt, and J.B.Tobin. 1994. A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science* 264:1927-1930.

GarciaViloca,M., R.Gelabert, A.GonzalezLafont, M.Moreno, and J.M.Lluch. 1997. Is an extremely low-field proton signal in the NMR spectrum conclusive evidence for a low-barrier hydrogen bond? *Journal of Physical Chemistry A* 101:8727-8733.

Gerlt,J.A., M.M.Kreevoy, W.Cleland, and P.A.Frey. 1997. Understanding enzymic catalysis: the importance of short, strong hydrogen bonds. *Chem Biol.* 4:259-267.

Harte,W.E., S.Swaminathan, M.M.Mansuri, J.C.Martin, I.E.Rosenberg, and D.L.Beveridge. 1990. Domain Communication in the Dynamic Structure of Human Immunodeficiency Virus-1 Protease. *Proceedings of the National Academy of Sciences of the United States of America* 87:8864-8868.

Hong,L., J.A.Hartsuck, S.Foundling, J.Ermolieff, and J.Tang. 1998. Active-site mobility in human immunodeficiency virus, type 1, protease as demonstrated by crystal structure of A28S mutant. *Protein Science* 7:300-305.

Hruz,P.W., H.Murata, and M.Mueckler. 2001. Adverse metabolic consequences of HIV protease inhibitor therapy: the search for a central mechanism. *Am. J Physiol Endocrinol. Metab* 280:E549-E553.

Hyland,L.J., T.A.Tomaszek, and T.D.Meek. 1991a. Human Immunodeficiency Virus-1 Protease .2. Use of Ph Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* 30:8454-8463.

Hyland,L.J., T.A.Tomaszek, G.D.Roberts, S.A.Carr, V.W.Magaard, H.L.Bryan, S.A.Fakhoury, M.L.Moore, M.D.Minnich, J.S.Culp, R.L.DesJarlais, and T.D.Meek. 1991b. Human Immunodeficiency Virus-1 Protease .1. Initial Velocity Studies and Kinetic Characterization of Reaction Intermediates by O-18 Isotope Exchange. *Biochemistry* 30:8441-8453.

Khan,A.R., J.C.Parrish, M.E.Fraser, W.W.Smith, P.A.Bartlett, and M.N.G.James. 1998. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* 37:16839-16845.

Kohl,N.E., E.A.Emini, W.A.Schleif, L.J.Davis, J.C.Heimbach, R.A.F.Dixon, E.M.Scolnick, and I.S.Sigal. 1988. Active Human Immunodeficiency Virus Protease Is Required for Viral Infectivity. *Proceedings of the National Academy of Sciences of the United States of America* 85:4686-4690.

Krausslich,H.G. 1991. Human-Immunodeficiency-Virus Proteinase Dimer As Component of the Viral Polyprotein Prevents Particle Assembly and Viral Infectivity. *Proceedings of the National Academy of Sciences of the United States of America* 88:3213-3217.

Laio A., VandeVondele J., and Roethlisberger U. 2002. A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116:6941-6947.

Loeb,D.D., R.Swanstrom, L.Everitt, M.Manchester, S.E.Stamper, and C.A.Hutchison. 1989. Complete Mutagenesis of the Hiv-1 Protease. *Nature* 340:397-400.

Luque,I., M.J.Todd, J.Gomez, N.Semo, and E.Freire. 1998. Molecular basis of resistance to HIV-1 protease inhibition: A plausible hypothesis. *Biochemistry* 37:5791-5797.

Mehanna,A.S. 2003. Rationale Design of Anti-HIV Drugs. in *Burger's medicinal chemistry*

and drug discovery. Vol. 5: C*hemotherapeutic Agents.* Ed. Abraham, D.J. John Wiley &

Sons. Chapter 11, pp. 457-483.

Miller,M., M.Jaskolski, J.K.M.Rao, J.Leis, and A.Wlodawer. 1989a. Crystal-Structure of A Retroviral Protease Proves Relationship to Aspartic Protease Family. *Nature* 337:576-579.

Miller,M., J.Schneider, B.K.Sathyanarayana, M.V.Toth, G.R.Marshall, L.Clawson, L.Selk, S.B.H.Kent, and A.Wlodawer. 1989b. Structure of Complex of Synthetic Hiv-1 Protease with A Substrate-Based Inhibitor at 2.3-A Resolution. *Science* 246:1149-1152.

Navia,M.A., P.M.D.Fitzgerald, B.M.Mckeever, C.T.Leu, J.C.Heimbach, W.K.Herber, I.S.Sigal, P.L.Darke, and J.P.Springer. 1989. 3-Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus Hiv-1. *Nature* 337:615-620.

Okimoto,N., T.Tsukui, M.Hata, T.Hoshino, and M.Tsuda. 1999. Hydrolysis mechanism of the phenylalanine-proline peptide bond specific to HIV-1 protease: Investigation by the ab initio molecular orbital method. *Journal of the American Chemical Society* 121:7349-7354.

Pantaleo,G. and A.S.Fauci. 1996. Immunopathogenesis of HIV infection. *Annual Review of Microbiology* 50:825-854.

Piana,S., D.Bucher, P.Carloni, and U.Rothlisberger. 2004. Reaction mechanism of HIV-1 protease by hybrid carparrinello/classical MD simulations. *Journal of Physical Chemistry B* 108:11139-11149.

Piana,S. and P.Carloni. 2000. Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: An ab initio study on the free enzyme. *Proteins* 39:26-36.

Polgar,L., Z.Szeltner, and I.Boros. 1994. Substrate-Dependent Mechanisms in the Catalysis of Human-Immunodeficiency-Virus Protease. *Biochemistry* 33:9351-9357.

Poorman,R.A., A.G.Tomasselli, R.L.Heinrikson, and F.J.Kezdy. 1991. A Cumulative Specificity Model for Proteases from Human-Immunodeficiency-Virus Type-1 and Type-2, Inferred from Statistical-Analysis of An Extended Substrate Data-Base. *Journal of Biological Chemistry* 266:14554-14561.

Richards,A.D., L.H.Phylip, W.G.Farmerie, P.E.Scarborough, A.Alvarez, B.M.Dunn, P.H.Hirel, J.Konvalinka, P.Strop, L.Pavlickova, V.Kostka, and J.Kay. 1990. Sensitive, Soluble Chromogenic Substrates for Hiv-1 Proteinase. *Journal of Biological Chemistry* 265:7733-7736.

Rose,R.B., C.S.Craik, N.L.Douglas, and R.M.Stroud. 1996. Three-dimensional structures of HIV-1 and SIV protease product complexes. *Biochemistry* 35:12933-12944.

Seelmeier,S., H.Schmidt, V.Turk, and K.Vonderhelm. 1988. Human Immunodeficiency Virus Has An Aspartic-Type Protease That Can be Inhibited by Pepstatin-A. *Proceedings of the National Academy of Sciences of the United States of America* 85:6612-6616.

Silva,A.M., R.E.Cachau, H.L.Sham, and J.W.Erickson. 1996. Inhibition and catalytic mechanism of HIV-1 aspartic protease. *Journal of Molecular Biology* 255:321-340.

Sirois,S., E.I.Proynov, J.F.Truchon, C.M.Tsoukas, and D.R.Salahub. 2003. A density functional study of the hydrogen-bond network within the HIV-1 protease catalytic site cleft. *J Comput Chem* 24:1110-1119.

Tobin,J.B., S.A.Whitt, C.S.Cassidy, and P.A.Frey. 1995. Low-barrier hydrogen bonding in molecular complexes analogous to histidine and aspartate in the catalytic triad of serine proteases. *Biochemistry* 34:6919-6924.

Todd,M.J. and E.Freire. 1999. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins-Structure Function and Genetics* 36:147-156.

Turner,B.G. and M.F.Summers. 1999. Structural biology of HIV. *Journal of Molecular Biology* 285:1-32.

Tyndall,J.D.A., T.Nall, and D.P.Fairlie. 2005. Proteases universally recognize beta strands in their active sites. *Chemical Reviews* 105:973-999.

van Gunsteren,W.F., H.Liu, and F.Muller-Plathe. 1998. The elucidation of enzymatic reaction mechanisms by computer simulation: Human Immunodeficiency Virus protease catalysis. *Theochem-Journal of Molecular Structure* 432:9-14.

Velazquez-Campoy A, Luque I, and Freire E. 2001. The application of thermodynamic methods in drug design. *Thermochimica Acta* 280:217-227.

Velazquez-Campoy,A., Y.Kiso, and E.Freire. 2001. The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. *Arch. Biochem Biophys* 390:169-175.

Velazquez-Campoy,A., M.J.Todd, and E.Freire. 2000. HIV-1 protease inhibitors: Enthalpic versus entropic optimization of the binding affinity. *Biochemistry* 39:2201-2207.

Wang,Y.X., D.I.Freedberg, S.Grzesiek, D.A.Torchia, P.T.Wingfield, J.D.Kaufman, S.J.Stahl, C.H.Chang, and C.N.Hodge. 1996a. Mapping hydration water molecules in the HIV-1 Protease/DMP323 complex in solution by NMR spectroscopy. *Biochemistry* 35:12694-12704.

Wang,Y.X., D.I.Freedberg, T.Yamazaki, P.T.Wingfield, S.J.Stahl, J.D.Kaufman, Y.Kiso, and D.A.Torchia. 1996b. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* 35:9945-9950.

Wlodawer,A. and J.W.Erickson. 1993. Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem* 62:543-585.

Wlodawer,A. and A.Gustchina. 2000. Structural and biochemical studies of retroviral proteases. *Biochim. Biophys Acta* 1477:16-34.

Wlodawer,A. and J.Vondrasek. 1998. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure* 27:249-284.

Yamazaki,T., L.K.Nicholson, D.A.Torchia, S.J.Stahl, J.D.Kaufman, P.T.Wingfield, P.J.Domaille, and S.Campbellburk. 1994a. Secondary Structure and Signal Assignments of Human-Immunodeficiency-Virus-1 Protease Complexed to A Novel, Structure-Based Inhibitor. *European Journal of Biochemistry* 219:707-712.

Yamazaki,T., L.K.Nicholson, D.A.Torchia, P.Wingfield, S.J.Stahl, J.D.Kaufman, C.J.Eyermann, C.N.Hodge, P.Y.S.Lam, Y.Ru, P.K.Jadhav, C.H.Chang, and P.C.Weber. 1994b. Nmr and X-Ray Evidence That the Hiv Protease Catalytic Aspartyl Groups Are Protonated in the Complex Formed by the Protease and A Nonpeptide Cyclic Urea-Based Inhibitor. *Journal of the American Chemical Society* 116:10791-10792.

York,D.M., T.A.Darden, L.G.Pedersen, and M.W.Anderson. 1993. Molecular-Dynamics Simulation of Hiv-1 Protease in A Crystalline Environment and in Solution. *Biochemistry* 32:1443-1453.

5. DEUTERIUM ISOTOPE EFFECT in A:T and A:U BASE PAIRS

5.1. INTRODUCTION

Deuterium isotope effects (DIE's) on C-13 chemical shifts can provide very valuable information on the nature of H-bonding interactions(Reuben J., 1987;Smirnov et al., 1996;Perrin and Nielson, 1997;Abildgaard et al., 1998), being as large as 1 ppm if the H atom involved in the H-bond is the site of isotopic substitution. Recently, DIE measurements provided important information on base-pair H-bonding in nucleic acids(Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b).

First, it was shown that, in five DNA duplexes, A:T pairs DIE's are generally smaller than the corresponding A:U DIE's in the RNA homologues(Vakonakis and LiWang, 2004a). This could be taken as an indication of a larger strength of the A:U@RNA H-bonding relative A:T@DNA(Vakonakis and LiWang, 2004a). However, quantum chemical calculations at the DFT level pointed to an absence of correlation between H-bond strength and the chemical shifts(Swart et al., 2004), and, along with calculations at the ab initio correlated level, to a similarity in the energy of formation of A:T@DNA and A:U@RNA base pairs(Perez et al., 2005). Thus, factors other than H-bond strength must play a crucial role for the observed differences between DIE's of RNA and DNA.

Second, A:T@DNA DIE's turn out to vary up as much as to 13 ppb along the oligonucleotide(Vakonakis and LiWang, 2004b), indicating that the local conformation of the DNA duplex influences the transmission of the isotope effect across the hydrogen bonds. However, the key factors for this modulation are still to be understood.

Clearly, calculations of DIE's in nucleic acids, so far lacking, could help rationalize these data. Here, by perfoming DFT and QM/MM calculations, we show that the DIE's differences between A:T and A:U are caused by the difference in electronic structure of the two base pairs due to the replacement of a methyl group with a hydrogen atom. In addition, our calculations show that the base pair conformation affects the DIE, supporting previous hypothesis(Vakonakis and LiWang, 2004b). In particular, Shear, Stretch and Opening, the parameters defining the geometry of the H-bond, play a major role. Finally, our calculations provide evidence that electrostatic interactions between the base pair and the surrounding environment can have a large influence on the DIE.

5.2. MODEL SYSTEMS and METHODS

DFT calculations were performed on the A:T and A:U base pairs in vacuo. Calculations were also carried out for A:T embedded in DNA duplexes in aqueous solution, using a hybrid DFT/MM approach. The interface between the QM and MM subsystems was located at the C1'-N9@A and C1'-N1@T/U bonds. Dangling bonds were saturated by hydrogen atoms. Spurious electrostatic interactions between the capping hydrogens and close-by classical atoms were excluded from the QM/MM Hamiltonian according to ref. (Laio et al., 2004).

Classical Molecular Dynamics (MD) simulations were performed on the following oligonucleotides (Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b): d1: d(CGCGAATTCGCG)2; r1: r(CGCGAAUUCGCG)2; d2: d(CGTTTTAAAACG)2; r2: r(CGUUUUAAAACG)2; d3: d(CGAAAATTTTCG)2; r3: r(CGAAAAUUUUCG)2. d1 is an extensively studied dodecamer. The simulation of the others allows to investigate the conformational properties of several A:T(U) pairs in a single MD simulation. Notice that the sequence d4, investigated in the work of (Vakonakis and LiWang, 2004a;Vakonakis

and LiWang, 2004b), was not considered here because of the high number of ApT steps present, which have been reported to be poorly described by the current force fields(Sponer et al., 2001).

Initial conformations for the duplexes were constructed with the NUCGEN module of the AMBER suite of programs(Pearlman et al., 1995). DNA duplexes were modelled in the canonical B form, while the A form was used for RNA duplexes(Arnott et al., 1976). The duplexes were solvated with water molecules (4491 and 4990 for DNA and RNA, respectively) and 22 Na+ ions to neutralize the simulation cell.

Classical MD calculations were performed with the GROMACS program(Berendsen et al., 1995). The AMBER (version 99(Cornell et al., 1995;Cheatham et al., 1999)) and TIP3P(Jorgensen et al., 1983) force fields were used for the duplex and for water, respectively. The AMBER force field has recently proven to be successful in describing DNA duplexes structure(Dornberger et al., 2001), elastic properties(Lankas et al., 2000) and base pair deformability(Lankas et al., 2004;Lankas et al., 2003). Recently, the same simulation procedure was used to investigate the relative flexibility of duplexes d1 and r1, which turn out to be in line with a statistical analysis of high resolution X-ray structures(Perez et al., 2004;Noy et al., 2004).

The equilibration phase for each system was as follows: i) equilibration of the solvent by performing 1 ns of MD with the duplex fixed at the initial conformation; ii) energy minimization of the system with 250 steepest descent steps followed by 250 conjugate gradient steps, while the positions of the heavy atoms of the duplex were restrained; then 500 steepest descent steps followed by 500 conjugate gradient steps were performed with no restraints; iii) 30 ps of NPT MD simulation from 0 to 10 K on the entire system, 20 ps from 10 to 50 K, 20 ps for each 50 K increase of temperature from 50 K up to 298 K. ~10

ns of NPT MD simulation were carried out for each system and the last 7 ns were used for data analysis.

Helical parameters and base pair conformations along the classical MD trajectory were analyzed with the X3DNA program(Lu and Olson, 2003).

5.2.1. Calculation of NMR properties. Quantum chemical calculations at the Hartree-Fock level on small molecular systems have shown that DIE's are dominated by bond stretching(Jameson, 1996). Consistently, a normal mode analysis performed here on the Watson-Crick A:T pair[2] predicted that the mode involving the N3-H3@T bond is basically decoupled from the motion of the rest of the system. Indeed, this mode has a 98% projection on the N3-H3@T distance, and, in addition, the frequency of this mode changes by as much as 25% upon isotopic substitution.

Based on this facts, we made here a reasonable and simplifying assumption that the potential energy surface (PES) of the base pairs required for the calculation of DIE's can be reduced to the N3@T-H3@T and N3@U-H3@U stretching for A:T and A:U, respectively (see Scheme 1).

The PES can be calculated by fixing the N3-H3 distance (Scheme 1) at increasing values and by either relaxing the other degrees of freedom (relaxed scan) or keeping them fixed (rigid scan). Here we performed most of the DIE's calculations within the rigid scan approach, which allows to get a qualitative picture of the dependence of the DIE on base pair conformation, while reducing the cost of the calculation. However, when comparing DIE's between A:U and A:T, for which a semiquantitative estimate is deemed necessary, both approaches were used.

[2] the geometry of the pair is optimized at the BLYP level of theory.

Shielding constants are calculated through the variational approach of density functional perturbation developed by Sebastiani et al.(Sebastiani and Parrinello, 2001) for each point along the PES. The calculated energy and shielding values were fitted to a polynomial of 8^{th} degree and of 6^{th} degree, respectively, which provided a good fit. The 1D nuclear Shroedinger equation for the hydrogen motion was then solved variationally using the calculated PES. Harmonic oscillator basis set functions were used to describe the vibrational ground state wavefunction(Wilson et al., 1955). One hundred terms were enough to assure that the calculated DIE's were converged. The DIE's were calculated as the difference between the shielding expectation values over the proton and deuterium probability distributions(Jameson, 1996).

DIE's calculations of A:T and A:U in vacuo were carried out for the following conformations: (i) the optimized A:T and A:U structures. (ii) 13 different geometries of the A:T pair obtained by combining the optimized geometries of A and T, using the program X3DNA(Lu and Olson, 2003). In these conformations, only one base pair parameter at a time was varied; iii) 23 conformations of the A:T pair built with base pair parameters from MD simulations of base pair 5 of the d(CGAAAATTTTCG)2 oligonucleotide in aqueous solution.

QM/MM DIE calculations were performed for the following base pair embedded in the duplex environment: Bp5@d1, Bp4@d2, Bp8@d2, Bp6@d3, Bp6@d3. For each base pair, a snapshots was extracted from MD simulations. This underwent 500 QM/MM NVE-MD steps, followed by 1000 steps of QM/MM annealing (with scaling factor of 0.99). The final conformation was used for the calculation of the DIE. Comparison was made with calculations in vacuo of the DIE's of the base pairs only and of the base pair surrounded by its adjacent nucleotides treated at the DFT-BLYP level of theory.

5.3. RESULTS AND DISCUSSION

Here we use DFT and molecular dynamics simulations to investigate, at the qualitative level, i) why observed trans-hydrogen-bond DIE's on ^{13}C shieldings are (generally) larger in RNA than in DNA and ii) the dependence of the DIE's on the base pair conformation.

5.3.1. Differences between DIE's in DNA and RNA. The observed differences in DIE's may result from the different conformation of the DNA and RNA duplexes and/or to an intrinsic difference between the two base pairs. To assess the relative importance of these two contributions, we compare (i) the intrinsic differences in DIE's between the two pairs; (ii) the distributions of conformational parameters of the A:T and A:U pairs embedded in DNA and RNA duplexes.

The calculated DIE's in A:T and A:U pairs turn out to differ mostly for pyrimidine atoms, where larger differences are observed for atoms closer to the site of substitution (C5 in pyrimidines, Tab. 1). The difference between the calculated DIE at C2@A is ~1 ppb within the rigid PES scan, while it is 18 ppb within the relaxed scan. This can be compared with the experimental average difference of DIE at C2@A between DNA and RNA duplexes of about 5 ppb(Vakonakis and LiWang, 2004a).

The DIE's calculated in this work (i) reproduce the observed difference between A:T and A:U; (ii) the DIE on C2@A and that on C4@T are of opposite signs, as experimentally observed(Vakonakis and LiWang, 2004b). This is in line with reported calculated DIE's for *o*-hydroxy acyl aromatics, which generally correlate well with experimental DIE(Abildgaard et al., 1998), although a quantitative agreement is in general not achieved(Stare et al., 2004).

The calculated difference of the DIE between A:T and A:U may result either from a different PES or from different shielding surfaces. To assess the relative importance of these two contributions, we assume that the following approximation, based on a Taylor expansion(Abildgaard et al., 1998), holds:

$$DIE = (\partial\sigma(C)/\partial r)\ (\langle r\rangle_H - \langle r\rangle_D)$$

$\partial\sigma(C)/\partial r$ is the derivative of the shielding of atom C at the PES minimum, which clearly depends on the shielding surface. $\langle r\rangle_H$ and $\langle r\rangle_D$ are the average N3-H3@T/U distance over the proton and deuterium probability distributions, respectively. Their difference reflects differences in the PES. It turns out that ($\langle r\rangle_H$ - $\langle r\rangle_D$) is the same in A:U and A:T (Tab. 1), and indeed at the minimum the two PES are nearly identical both within the rigid or within the relaxed PES scans (Fig. 1). In contrast, $\partial\sigma(C)/\partial r$ differs in the two systems (Tab. 1), suggesting that the difference in DIE's between A:T and A:U mainly result from the magnetic properties of thymine relative to uracil.

Consistently with these findings, the calculated association energy of A:T is marginally smaller (by 0.05 kcal/mol) than A:U in their optimized conformation at the DFT-BLYP level. This result is also in line with previous calculations at the DFT and ab initio correlated level (Perez et al., 2005;Swart et al., 2004) of the base pairs in a variety of conformations. We conclude that both previous calculations and ours rule out a greater stability of A:U relative to A:T.

We now turn our attention to the dependence of DIE's on the conformational properties of the base pairs by performing classical MD simulations of several DNA and RNA duplexes.

We focus on those base pair parameters (Shear, Stretch, Stagger, Buckle, Propeller and Opening), which have been proposed to play a key role for the DIE's (Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b).

Although some of the calculated values differ for the two base pairs (Tab. 2), their distributions are completely or largely overlapping, for Shear/Opening and Stretch, Stagger, Buckle, Propeller, respectively (Fig. 2a). Stretch and Propeller are generally less negative for A:U pairs, while Stagger and Buckle are generally shifted towards negative values in A:U pairs with respect to A:T. Buckle is the most variable parameter. Although it is expected that the DIE's should depend on the base pair conformation, scatter plots of base pair parameters with respect to the experimental DIE do not reveal any appreciable degree of correlation (Fig. 2b), suggesting either that a combination of several parameters may contribute to the magnitude of the DIE, or that other effects are important in determining the observed DIE.

We conclude that the there is an intrinsic difference between DIE's in A:T and A:U base pairs that should be mainly attributed to the different magnetic properties of the T and U nucleobases.

5.3.2. DIE's in different A:T conformations. In order to quantify the dependence of the DIE's from the base pair conformation, the base pair parameters were varied from the ideal, planar Watson-Crick base pair conformation. The results of these calculations are presented in Figure 3 and suggest that: (i) the parameters that mostly influence the H-bond geometry (Shear, Stretch and Opening) are also the parameters that have the largest effect on the DIE's (Fig. 3a). Stagger has also a significant contribution, while Buckle and Propeller have a less pronounced effect (Fig. 3a). (ii) For Shear and Opening,

the trend depends on the atom where the DIE is calculated: higher Shear and Opening correspond to smaller DIE at C2@A, while the opposite trend is displayed by the DIE at C4@T (Fig. 3a). These opposing trends explain the weak correlation between DIE's at C4@T and C2@A observed in the NMR experiments(Vakonakis and LiWang, 2004a).

Consistently with previous conclusions(Swart et al., 2004), scatter plots of the DIE at C2@A and C4@T with respect to the base pair binding energy indicate that a large DIE does not necessarily correspond to a stronger interaction (Fig. 3b)[3]. Indicators of the strength of H-bonds such as shieldings of hydrogens involved in H-bonds(Cleland, 2000;Del Bene et al., 1999;Reuben J., 1987) (here H3@T) and distances between donors and acceptors atoms(Harris and Mildvan, 1999) (here d(N1@A-N3@T) correlate well with the DIE on C4@T (correlation coefficients 0.98 and 0.94, Fig. 3b), while the correlation is weaker with DIE on C2@A (correlation coefficients –0.80 and –0.71, respectively). As a given d(N1@A-N3@T) distance may result from different base pair parameters, the lower level of correlation between the DIE on C2@A and d(N1@A-N3@T) compared to the DIE on C4@T is the result of the opposing trends displayed by the DIE at these two atoms with respect to the Shear and Opening parameters.

So far, we have considered only isolated base pairs. To investigate the effects of the environment, the DIE's were compared with values calculated within a QM/MM approach where the A:T base pair is treated at the DFT level, while the rest is described with the AMBER force field(Cornell et al., 1995). The QM/MM DIE's differ from those calculated in

[3] However, no attempt is here made to partitioning the energy, and thus it is not possible to probe the dependence of the DIE on the strength of the N1@A…H3@T-N3@T H-bond only.

vacuo by –14 to +16 ppb (Tab. 3). These contributions, which clearly are related to the electrostatic field, are comparable to the effects of the conformation.

Finally, because our QM/MM calculations do not include shielding contributions arising from the stacked bases, we performed DFT calculations on the base pair of interest surrounded by its adjacent nucleotides (Tab. 3). The conformation of the surrounding base pairs were taken from the corresponding QM/MM calculation. The resulting DIE's increase by 10-30% relative to A:T (Tab. 3). This effect appears to be modulated by the overlap between contiguous base pairs: a smaller overlap corresponding to a smaller shift with respect to the isolated base pair (Fig. 4).

5.3.3. Analysis of classical MD simulations. All the duplex simulations are stable within the time scale investigated, as indicated by the radius of gyration and the RMSD with respect to the starting structure which oscillate around a constant during the dynamics (Tab. 4 and Fig. 5). The root mean square fluctuations (RMSF) with respect to the average structure is a measure of the local flexibility. RMSF were calculated for each base and are reported in Figure 6. Due to the symmetry of the duplexes investigated the values for the two strands of the same duplex should be identical. Indeed, the plots reported in Figure 6 are very similar, indicating that the sampling of the double helix conformation in MD simulations is converged. The terminal bases are generally the most flexible; amongst the internal sequences, the T rich regions are also remarkably flexible.

The average A:T base pair parameters calculated from the MD simulations were compared with those calculated from high resolution X-ray structures(Olson et al., 2001). Figure 7 shows that all average parameters are within 2 standard deviations from the average X-ray values, most being within 1 standard deviation. It is interesting to note that the A:U stretch value in the RNA duplexes is always larger than the average value observed in A-DNA structures(Olson et al., 2001). Base pair parameters calculated from

the MD simulation of duplex **d1** can be compared with parameters calculated from the high resolution X-ray and NMR structures of the same duplex(Drew et al., 1981;Shui et al., 1998b;Shui et al., 1998a;Holbrook et al., 1985;Westhof, 1987;Wu et al., 2003). Tab. 5 shows that the calculated values are not generally in agreement with the experimentally determined values, with only about 50% of MD average parameters being within 3 standard deviations of the X-ray average values.

5.4. CONCLUSIONS

The effect of electronic structure, base pair conformation and duplex environment on the DIE have been investigated by performing DFT calculations of DIE's in several A:T and A:U base pairs. Our results show that that a key contribution to the difference between DIE's in A:T and A:U pairs is the different intrinsic electronic properties of the base pairs.

DIE's are strongly influenced by the duplex conformation. Several factors are important in this respect, namely: i) the base pair conformation; ii) the electric field generated by the duplex and the counterions and, to a lesser extent, iii) the magnetic field generated by the adjacent base pairs. None of these three contributions is clearly dominant over the other two. These findings highlight the difficulties of directly relating calculated DIE's to a single structural parameter, like an H-bond distance.

Table 1. Calculated NMR properties of A:T and A:U in the ideal, planar Watson-Crick base pair conformation.

Base	**DIE's at A (ppb)**					**DIE's at U/T (ppb)**				**Δr^c (Å)**	**$\partial\sigma$(C2@A)/∂r (ppm/Å)**
Atom	**C2**	**C4**	**C5**	**C6**	**C8**	**C2**	**C4**	**C5**	**C6**		
A:T[a]	78.3	-15.6	-11.4	49.6	-42.3	-38.2	-73.6	8.7	46.4	0.021	4.8
A:U[a]	79.6	-16.1	-11.3	50.3	-42.9	-33.6	-83.5	3.3	48.7	0.021	15.3
A:T[b]	128.5	-56.0	-75.2	-28.5	-64.2	-196.4	-262.9	-27.5	24.1	0.021	4.9
A:U[b]	146.5	-56.5	-77.8	-23.8	-65.6	-195.8	-301.6	-42.9	17.2	0.021	17.1

[a] 1D *rigid* PES and shielding surface.

[b] 1D *relaxed* PES and shielding surface.

[c] $\Delta r = \langle r \rangle_H - \langle r \rangle_D$

Table 2. MD-averaged A:T and A:U base pair parameters. Standard error in parentheses.

Base pair parameter	**A:T**	**A:U**
Shear	0.11 (0.02)	0.07 (0.02)
Stretch	-0.086 (0.005)	-0.056 (0.011)
Stagger	0.003 (0.078)	-0.169 (0.106)
Buckle	6.1 (3.7)	-3.5 (3.7)
Propeller	-15.2 (3.3)	-14.1 (1.9)
Opening	0.25 (0.65)	0.67 (0.91)

Table 3. DIE (ppb) calculated for selected base pairs in different DNA duplexes for a conformation from MD simulation within different approaches.

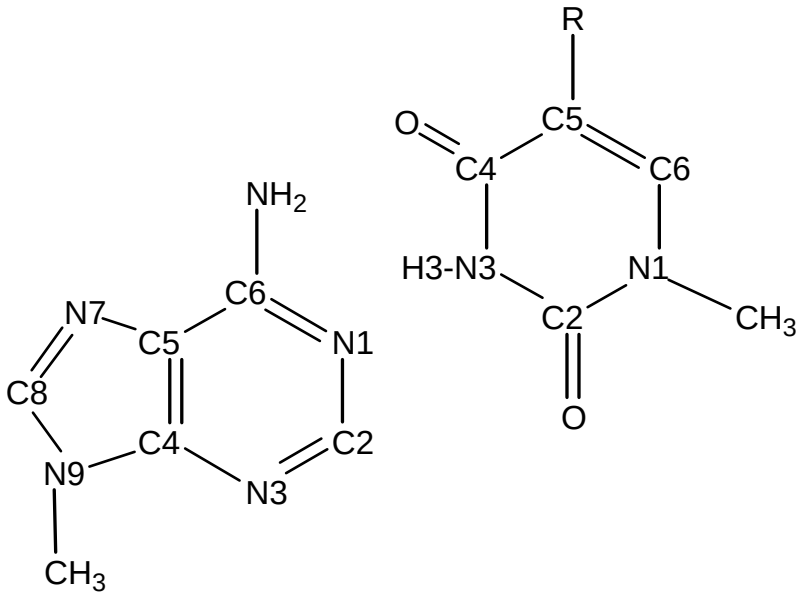
base	**A**					**T**			
atom	**C2**	**C4**	**C5**	**C6**	**C8**	**C2**	**C4**	**C5**	**C6**
Bp4@d2									
Single bp	57.6	-19.6	-12.6	57.5	-43.9	-39.4	-72.7	18.9	42.8
QM/MM	46.7	-19.6	-16.9	52.3	-44.1	-36.1	-73.7	19.7	41.1
triplet	74.3 (29%)	-26.3	-10.9	66.8	-48.8	-42.5	-85.1 (17%)	16.4	44.7
Bp6@d3									
Single bp	81.9	-20.1	-17.0	65.6	-48.4	-46.4	-93.6	22.0	50.3
QM/MM	68.0	-17.4	-21.2	61.7	-46.1	-40.1	-96.4	23.9	46.0
triplet	93.5 (14%)	-22.1	-12.7	64.5	-48.5	-45.7	-96.3 (3%)	18.1	48.4
Bp3@d3									
Single bp	52.4	-9.8	-16.4	31.1	-28.8	-21.4	-44.1	11.7	30.6
QM/MM	41.5	-10.4	-16.3	25.4	-26.3	-16.0	-41.7	13.6	28.4
triplet	57.4 (10%)	-11.2	-20.7	36.1	-29.9	-23.3	-44.3 (0%)	10.2	29.7
Bp5@d1									
Single bp	84.0	-19.7	-16.5	51.1	-42.9	-42.8	-74.6	14.0	47.2
QM/MM	90.8	-21.7	-21.7	53.7	-48.0	-46.4	-88.8	18.0	53.4
triplet	111.9 (33%)	-26.6	-18.3	70.6	-50.4	-49.5	-89.4 (20%)	18.2	55.5
Bp8@d2									
Single bp	97.9	-16.6	-18.2	50.7	-45.8	-46.0	-77.4	13.9	49.6
QM/MM	112.4	-21.5	-27.7	65.0	-57.7	-56.2	-108.6	22.7	62.7
triplet	120.1 (23%)	-23.0	-19.3	64.1	-54.4	-55.8	-91.1 (18%)	14.7	53.1

Table 4. Average radius of gyration of each duplex (standard deviations in parentheses).

duplex	**Radius of gyration (nm)**
d1	1.34 (0.02)
d2	1.35 (0.02)
d3	1.35 (0.02)
r1	1.31 (0.03)
r2	1.28 (0.03)
r3	1.32 (0.03)

Table 5. Average base pair parameters (standard deviations in parentheses) from x-ray (first line) and NMR (second line) structures(Drew et al., 1981;Shui et al., 1998b;Shui et al., 1998a;Holbrook et al., 1985;Westhof, 1987;Wu et al., 2003) and MD simulation (third line) for each base pair in duplex **d1**.

	Shear	**Stretch**	**Stagger**	**Buckle**	**Propeller**	**Opening**
Base pair 1	0.05(0.32)	-0.21(0.06)	0.10(0.04)	4.28(1.48)	-15.32(2.13)	-1.92(1.95)
	0.35(0.05)	-0.52(0.02)	-0.36(0.03)	-8.05(1.32)	-17.90(0.51)	1.39(1.04)
	0.19(0.30)	-0.15(0.11)	-0.02(0.47)	3.88(14.25)	-13.79(11.12)	-0.19(3.73)
Base pair 2	-0.16(0.12)	-0.29(0.07)	0.38(0.19)	2.65(6.55)	-11.82(1.85)	-3.45(1.22)
	-0.14(0.03)	-0.34(0.03)	0.35(0.03)	-2.04(1.07)	-14.55(1.14)	0.85(0.14)
	-0.13(0.26)	-0.16(0.10)	0.20(0.33)	-1.98(8.74)	-9.21(7.59)	-1.26(2.97)
Base pair 3	0.19(0.10)	-0.17(0.08)	0.23(0.07)	-5.76(1.47)	-5.01(1.41)	-0.30(1.34)
	0.41(0.08)	-0.47(0.04)	-0.05(0.11)	-3.69(0.43)	-14.18(1.53)	0.81(1.09)
	0.08(0.26)	-0.13(0.10)	0.15(0.32)	-3.72(8.46)	-1.92(7.68)	-0.86(2.88)
Base pair 4	-0.48(0.21)	-0.30(0.17)	-0.13(0.10)	8.65(1.53)	-8.99(1.14)	0.81(2.89)
	-0.19(0.08)	-0.39(0.02)	0.01(0.04)	0.32(0.35)	-13.36(1.10)	-1.26(0.98)
	-0.10(0.26)	-0.12(0.11)	0.04(0.36)	9.90(11.26)	-3.82(8.13)	-0.03(3.08)
Base pair 5	0.11(0.12)	-0.14(0.05)	0.03(0.08)	5.28(1.09)	-15.83(1.01)	1.48(1.64)
	-0.20(0.05)	-0.27(0.05)	-0.18(0.04)	-3.89(0.31)	-19.64(0.14)	-1.13(0.76)
	0.14(0.24)	-0.08(0.11)	-0.04(0.37)	8.49(9.36)	-16.16(6.93)	-0.27(4.89)
Base pair 6	0.06(0.11)	-0.04(0.06)	0.12(0.03)	2.07(1.12)	-16.83(1.54)	4.67(2.13)
	-0.06(0.01)	-0.23(0.06)	-0.24(0.05)	-1.81(0.83)	-22.94(0.17)	-2.76(0.55)
	0.11(0.24)	-0.09(0.11)	0.05(0.36)	2.74(8.45)	-19.40(6.67)	-0.06(4.88)
Base pair 7	0.13(0.15)	-0.08(0.12)	0.06(0.09)	0.24(1.00)	-17.06(0.46)	4.69(1.88)
	0.07(0.02)	-0.24(0.06)	-0.24(0.05)	1.76(0.77)	-22.96(0.10)	-2.68(0.64)
	-0.11(0.25)	-0.09(0.11)	0.07(0.36)	-3.10(8.38)	-19.42(6.59)	-0.34(4.94)
Base pair 8	0.06(0.17)	-0.20(0.11)	-0.03(0.07)	-0.97(0.83)	-16.86(1.11)	1.77(2.16)
	0.20(0.05)	-0.27(0.06)	-0.19(0.05)	3.97(0.47)	-19.47(0.35)	-1.00(0.99)
	-0.14(0.25)	-0.08(0.11)	-0.04(0.37)	-8.81(9.36)	-16.48(6.58)	-0.84(4.78)
Base pair 9	0.15(0.10)	-0.14(0.17)	-0.01(0.19)	-11.43(2.85)	-12.82(2.93)	-0.47(1.05)
	0.18(0.07)	-0.39(0.02)	0.03(0.03)	-0.34(0.19)	-13.31(1.03)	-1.33(0.92)
	0.10(0.26)	-0.12(0.10)	0.04(0.35)	-10.58(10.52)	-4.07(8.15)	-0.28(3.00)
Base pair 10	-0.05(0.07)	-0.20(0.15)	0.22(0.09)	3.88(1.46)	-5.15(2.72)	0.41(3.08)
	-0.40(0.07)	-0.46(0.03)	-0.05(0.08)	3.72(0.53)	-14.32(1.57)	0.77(0.88)
	-0.10(0.26)	-0.13(0.10)	0.17(0.32)	2.43(8.75)	-2.97(7.82)	-0.80(2.91)
Base pair 11	0.09(0.14)	-0.22(0.11)	0.44(0.13)	-3.07(1.87)	-16.78(1.03)	-4.06(1.49)
	0.15(0.01)	-0.34(0.02)	0.35(0.01)	2.12(0.62)	-14.59(0.85)	0.77(0.21)
	0.12(0.26)	-0.16(0.10)	0.19(0.34)	0.52(9.03)	-7.45(8.19)	-1.15(2.96)
Base pair 12	-0.23(0.18)	-0.16(0.08)	0.25(0.11)	5.68(1.45)	-4.19(5.16)	-3.25(1.32)
	-0.34(0.05)	-0.52(0.04)	-0.35(0.08)	8.06(0.92)	-17.86(0.56)	1.37(0.98)
	-0.18(0.30)	-0.14(0.11)	-0.01(0.47)	-1.49(15.16)	-11.83(11.68)	-0.31(3.66)



Scheme 1. A:T and A:U Watson-Crick base pairs. R=CH_3 in A:T, R=H in A:U.

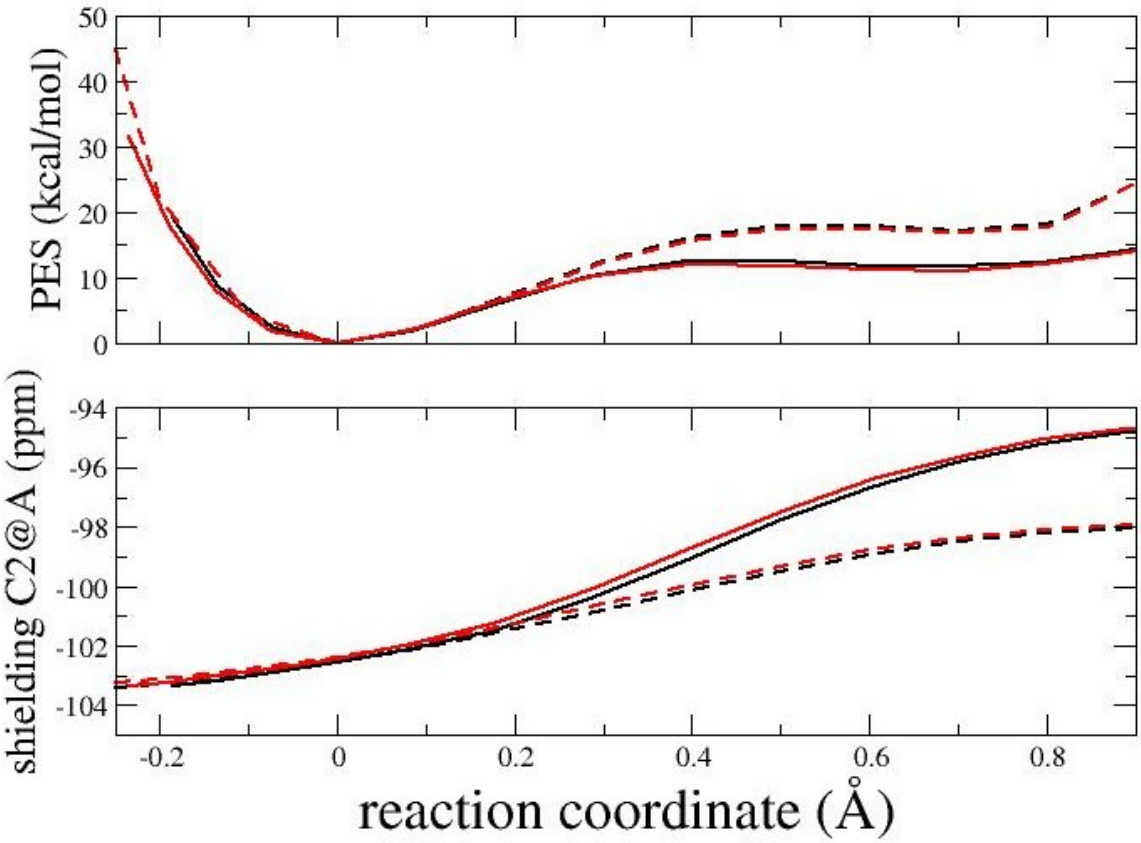


Figure 1. DFT calculations on A:T and A:U base pairs. 1D Potential Energy Surface (PES, top) and shielding surface for C2@A (bottom) for the A:T (black lines) and A:U (red lines) base pairs, within the *relaxed* (continuous lines) and the *rigid* (dotted lines) scan approaches (see text).

b)

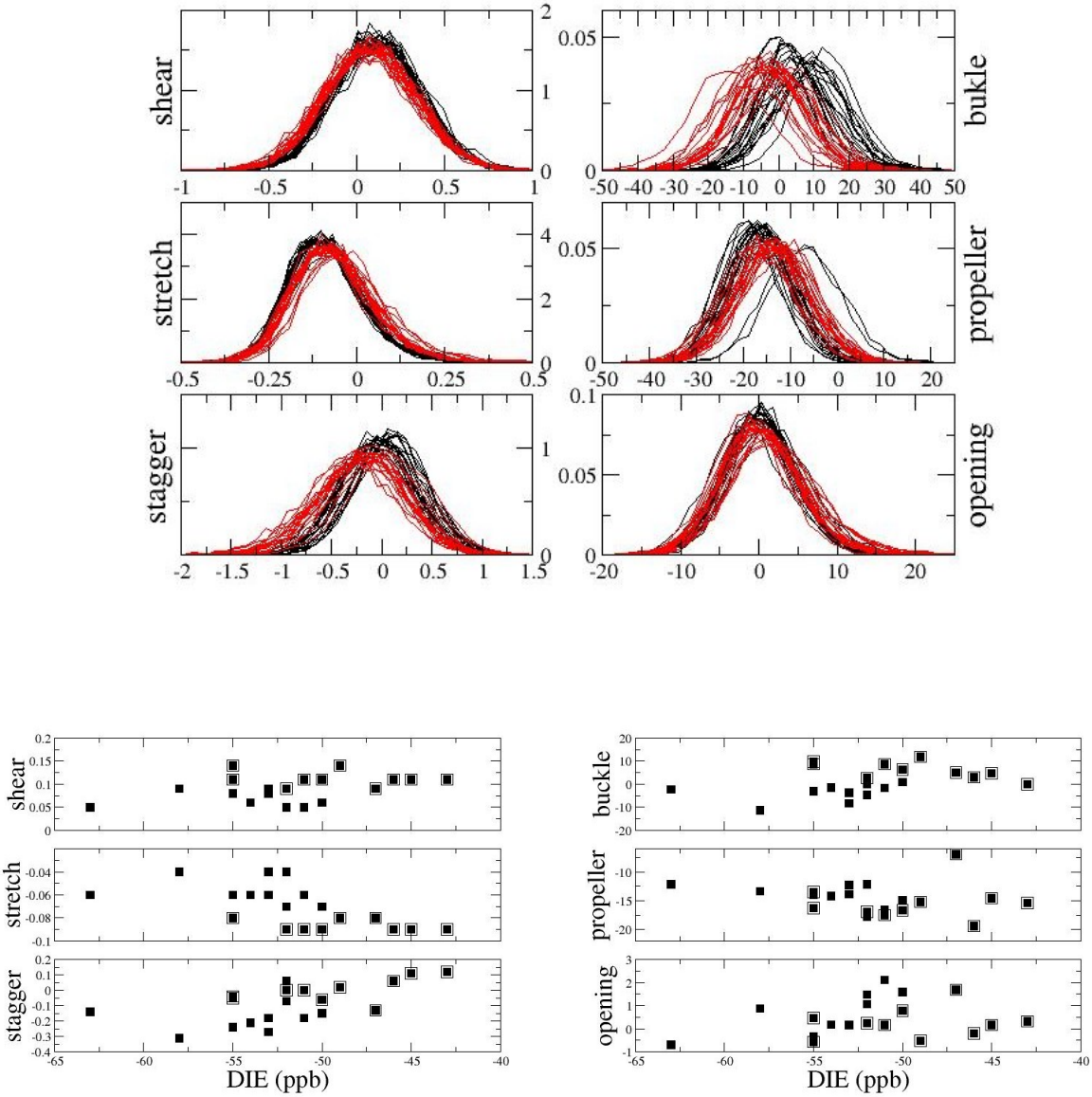


Figure 2. Classical MD simulations of the following oligonucleotides: d(CGCGAATTCGCG)2; r(CGCGAAUUCGCG)2; d(CGTTTTAAAACG)2; r(CGUUUUAAAACG)2; d(CGAAAATTTTCG)2; r(CGAAAAUUUUCG)2 (Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b). **a)** Distributions of base pair parameters for A:T (black lines) and A:U (red lines) pairs in the duplexes investigated. Notice that Buckle is the only parameter for which the distribution is not completely symmetric. **b)** Scatter plot of MD-averaged base pair parameters and experimental DIE's(Vakonakis and LiWang, 2004a;Vakonakis and LiWang, 2004b). Black frames highlight values for A:T pairs. Parameters for sequence related base pairs were averaged.

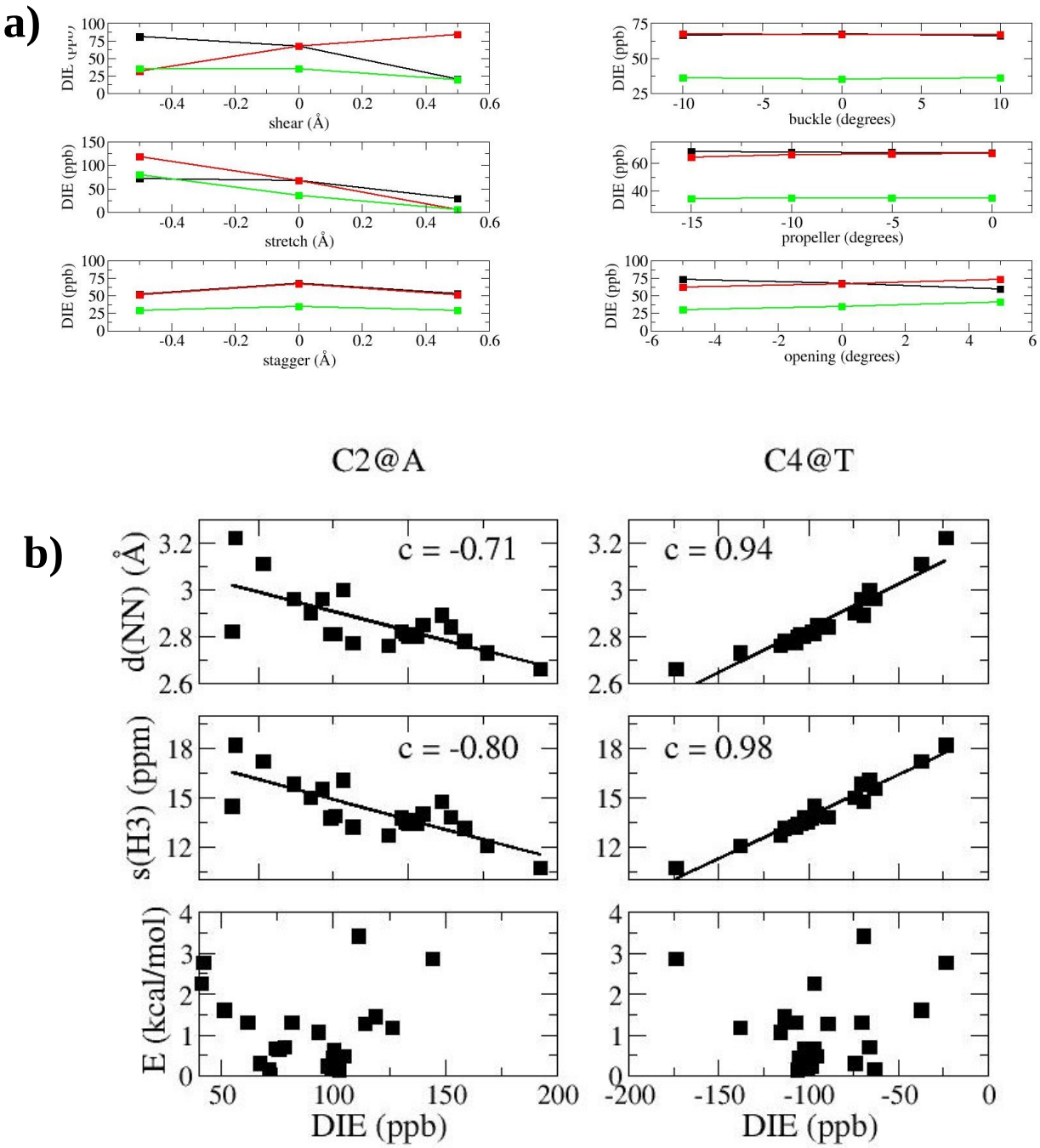


Figure 3. DFT calculations of A:T base pair. **a)** profile of C2@A (black line), C4@T (red line) and C2@T (green line) DIE's with variations of one base pair parameter. **b)** Scatter plots of the distance between N1@A and N3@T (d(NN), top), the shielding at H3 (s(H3), middle) and base pair energy (E, bottom) with respect to the DIE's at C2@A (left) and C4@T (right). Black lines are the least square fit to the data.

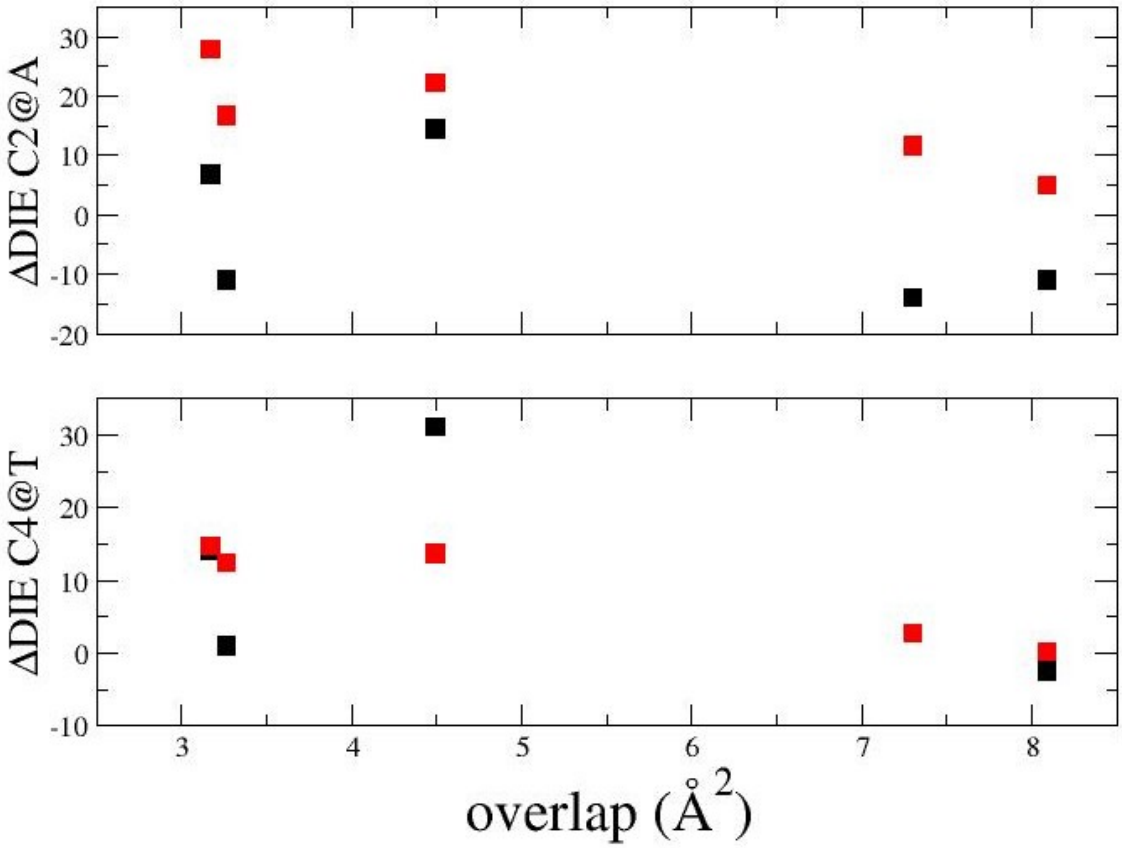


Figure 4. Difference (Δ) between the DIE's calculated within the QM/MM approach (black squares) or of DFT calculations of a triplet of base pairs (red squares) and the isolated base pair calculations. Δ are plotted as a function of the overlap between the central and the adjacent base pairs(Lu and Olson, 2003).

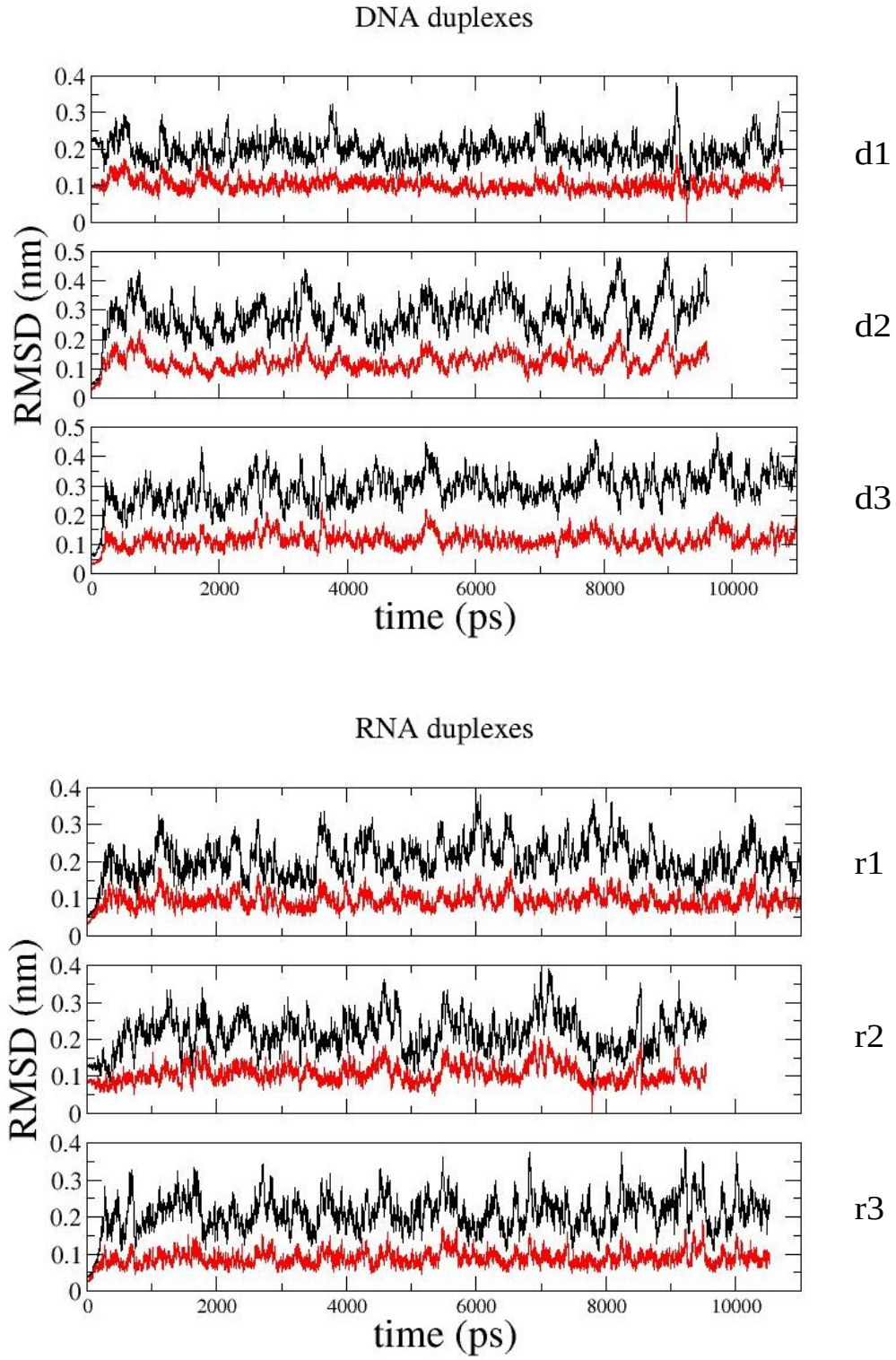


Figure 5. RMSD for backbone atoms (black line) and atoms in base pairs 3 to 10 (red line) for each duplex investigated.

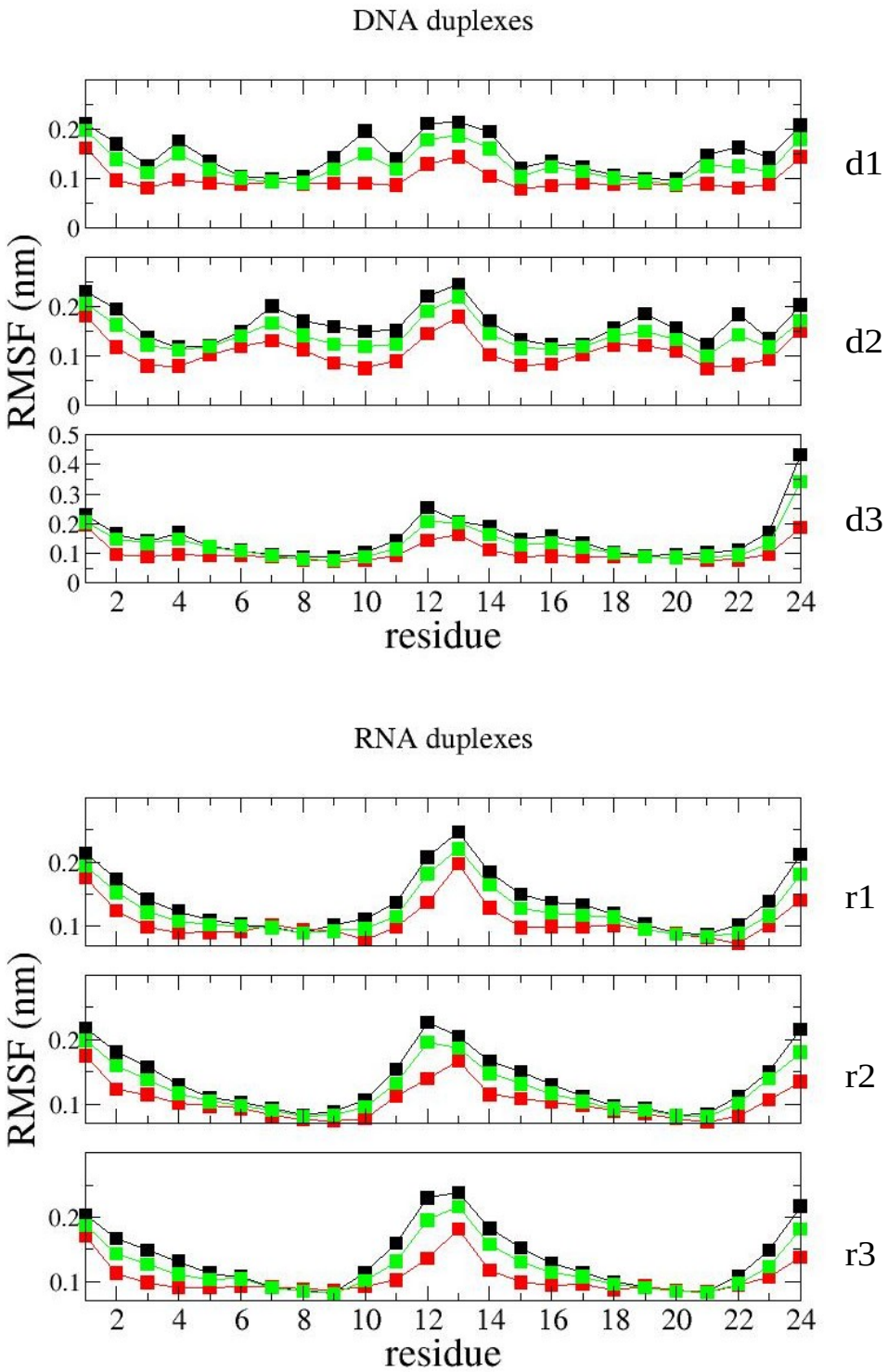


Figure 6. RMSF per residue for each duplex investigated. Black squares for backbone, red for base and green for sugar atoms.

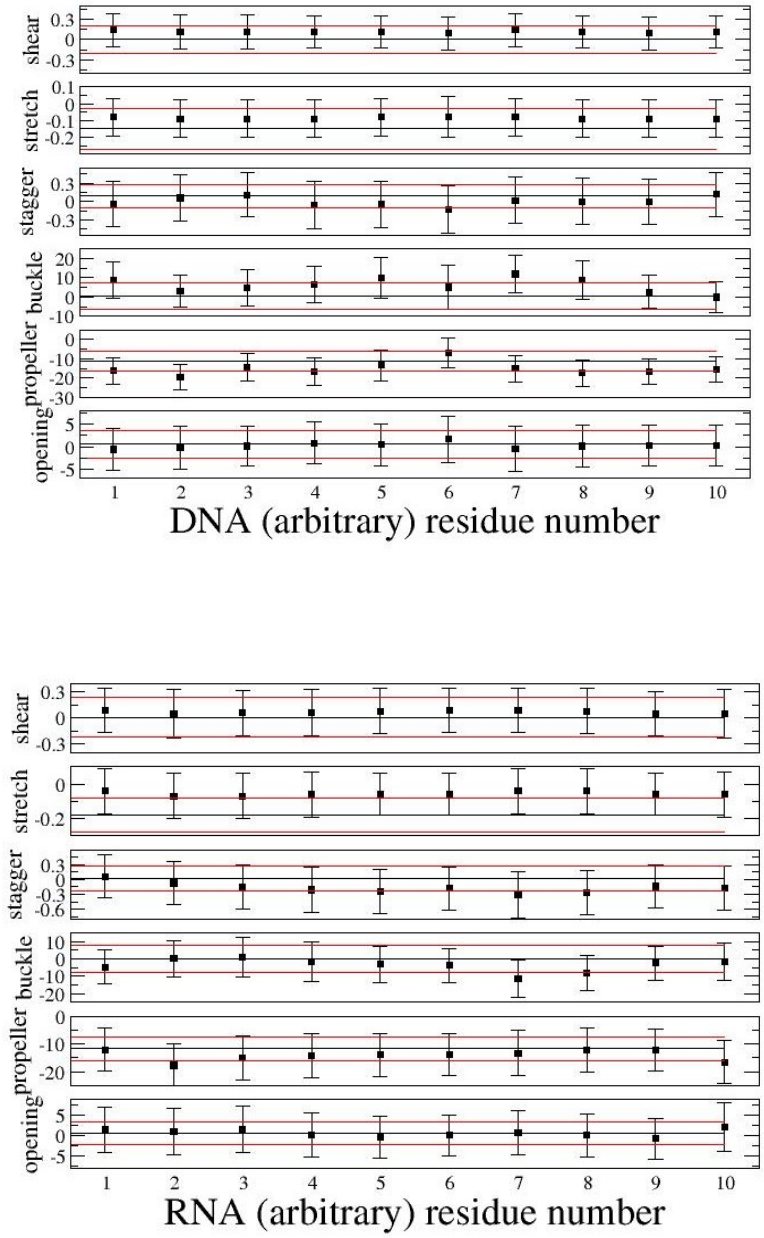


Figure 7. Comparison of the MD averaged base pair parameters (black squares, bars as standard deviation) and the averages from a statistical analysis of high resolution X-ray structures(Olson et al., 2001) (black line represents the average, red lines 1 standard deviation displacements from the average). Note that comparison for RNA is made with parameters from A-DNA structures(Olson et al., 2001).

5.5. REFERENCES

Abildgaard,J., S.Bolvig, and P.E.Hansen. 1998. Unraveling the electronic and vibrational contributions to deuterium isotope effects on C-13 chemical shifts using ab initio model calculations. Analysis of the observed isotope effects on sterically perturbed intramolecular hydrogen-bonded o-hydroxy acyl aromatics. *Journal of the American Chemical Society* 120:9063-9069.

Arnott,S., P.J.Bond, E.Selsing, and P.J.C.Smith. 1976. Models of Triple-Stranded Polynucleotides with Optimized Stereochemistry. *Nucleic Acids Research* 3:2459-2470.

Berendsen, HJC, Van der spoel, D, and Vandrunen, R. 1995. GROMACS - A message-passing parallel molecular-dynamics implementation. Comput.Phys.Commun. 91, 43-56.

Cheatham,T.E., P.Cieplak, and P.A.Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *Journal of Biomolecular Structure & Dynamics* 16:845-862.

Cleland,W.W. 2000. Low-barrier hydrogen bonds and enzymatic catalysis. *Archives of Biochemistry and Biophysics* 382:1-5.

Cornell,W.D., P.Cieplak, C.I.Bayly, I.R.Gould, K.M.Merz, Jr., D.M.Ferguson, D.C.Spellmeyer, T.Fox, J.W.Caldwell, and P.A.Kollman. 1995. A second-generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* 117:5179-5197.

Del Bene,J.E., S.A.Perera, and R.J.Bartlett. 1999. Hydrogen bond types, binding energies, and H-1 NMR chemical shifts. *Journal of Physical Chemistry A* 103:8121-8124.

Dornberger,U., N.Spackovj, A.Walter, F.A.Gollmick, J.Sponer, and H.Fritzsche. 2001. Solution structure of the dodecamer d-(CATGGGCC-CATG)2 is B-DNA. Experimental and molecular dynamics study. *J. Biomol. Struct. Dyn.* 19:159-174.

Drew,H.R., R.M.Wing, T.Takano, C.Broka, S.Tanaka, K.Itakura, and R.E.Dickerson. 1981. Structure of A B-Dna Dodecamer - Conformation and Dynamics .1. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 78:2179-2183.

Harris,T.K. and A.S.Mildvan. 1999. High-precision measurement of hydrogen bond lengths in proteins by nuclear magnetic resonance methods. *Proteins-Structure Function and Genetics* 35:275-282.

Holbrook,S.R., R.E.Dickerson, and S.H.Kim. 1985. Anisotropic Thermal-Parameter Refinement of the Dna Dodecamer Cgcgaattcgcg by the Segmented Rigid-Body Method. *Acta Crystallographica Section B-Structural Science* 41:255-262.

Jameson,C.J. 1996. Understanding NMR chemical shifts. *Annual Review of Physical Chemistry* 47:135-169.

Lankas,F., J.Sponer, P.Hobza, and J.Langowski. 2000. Sequence-dependent elastic properties of DNA. *J. Mol. Biol.* 299:695-709.

Lankas,F., J.Sponer, J.Langowski, and T.E.Cheatham, III. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85:2872-2883.

Lankas,F., J.Sponer, J.Langowski, and T.E.Cheatham, III. 2004. DNA deformability at the base pair level. *J. Am. Chem. Soc.* 126:4124-4125.

Lu,X.J. and W.K.Olson. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research* 31:5108-5121.

Noy,A., A.Perez, F.Lankas, F.J.Luque, and M.Orozco. 2004. Relative flexibility of DNA and RNA: a molecular dynamics study. *Journal of Molecular Biology* 343:627-638.

Olson,W.K., M.Bansal, S.K.Burley, R.E.Dickerson, M.Gerstein, S.C.Harvey, U.Heinemann, X.J.Lu, S.Neidle, Z.Shakked, H.Sklenar, M.Suzuki, C.S.Tung, E.Westhof, C.Wolberger, and H.M.Berman. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *Journal of Molecular Biology* 313:229-237.

Pearlman,D.A., D.A.Case, J.W.Caldwell, W.S.Ross, T.E.Cheatham, S.Debolt, D.Ferguson, G.Seibel, and P.Kollman. 1995. Amber, A Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* 91:1-41.

Perez,A., A.Noy, F.Lankas, F.J.Luque, and M.Orozco. 2004. The relative flexibility of B-DNA and A-RNA duplexes: database analysis. *Nucleic Acids Research* 32:6144-6151.

Perez,A., J.Sponer, P.Jurecka, P.Hobza, F.J.Luque, and M.Orozco. 2005. Are the hydrogen bonds of RNA (A:U) stronger than those of DNA (A:T)? A quantum mechanics study. *Chemistry-A European Journal* 11:5062-5066.

Perrin,C.L. and J.B.Nielson. 1997. Asymmetry of hydrogen bonds in solutions of monoanions of dicarboxylic acids. *Journal of the American Chemical Society* 119:12734-12741.

Reuben J. 1987. Isotopic multiplets in the carbon-13 NMR spectra of aniline derivatives and nucleosides with partially deuterated amino groups: effects of intra and intermolecular hydrogen bonding. *J. Am. Chem. Soc.* 109:316-321.

Sebastiani,D. and M.Parrinello. 2001. A new ab-initio approach for NMR chemical shifts in periodic systems. *J. Phys. Chem. A* 105:1951-1958.

Shui,X.Q., L.McFail-Isom, G.G.Hu, and L.D.Williams. 1998a. The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* 37:8341-8355.

Shui,X.Q., C.C.Sines, L.McFail-Isom, D.VanDerveer, and L.D.Williams. 1998b. Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* 37:16877-16887.

Smirnov,S.N., N.S.Golubev, G.S.Denisov, H.Benedict, P.SchahMohammedi, and H.H.Limbach. 1996. Hydrogen deuterium isotope effects on the NMR chemical shifts and geometries of intermolecular low-barrier hydrogen-bonded complexes. *Journal of the American Chemical Society* 118:4094-4101.

Sponer,J., J.Leszczynski, and P.Hobza. 2001. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* 61:3-31.

Stare,J., A.Jezierska, G.Ambrozic, I.J.Kosir, J.Kidric, A.Koll, J.Mavri, and D.Hadzi. 2004. Density functional calculation of the 2D potential surface and deuterium isotope effect on C-13 chemical shifts in picolinic acid N-oxide. Comparison with experiment. *Journal of the American Chemical Society* 126:4437-4443.

Swart,M., C.F.Guerra, and F.M.Bickelhaupt. 2004. Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. *Journal of the American Chemical Society* 126:16718-16719.

Vakonakis,I. and A.C.LiWang. 2004a. N1...N3 hydrogen bonds of A:U base pairs of RNA are stronger than those of A:T base pairs of DNA. *J. Am. Chem. Soc.* 126:5688-5689.

Vakonakis,I. and A.C.LiWang. 2004b. Trans-hydrogen bond deuterium isotope effects of A:T base pairs in DNA. *J. Biomol. NMR* 29:65-72.

Westhof,E. 1987. Re-Refinement of the B-Dodecamer D(Cgcgaattcgcg) with A Comparative-Analysis of the Solvent in It and in the Z-Hexamer D(5Brcg5Brcg5Brcg). *Journal of Biomolecular Structure & Dynamics* 5:581-600.

Wilson,E.B., J.C.Decius, and P.C.Cross. 1955. Molecular vibrations: the theory of infrared and Raman vibrational spectra. McGraw-Hill, New York.

Wu,Z.G., F.Delaglio, N.Tjandra, V.B.Zhurkin, and A.Bax. 2003. Overall structure and sugar dynamics of a DNA dodecamer from homoand heteronuclear dipolar couplings and P-31 chemical shift anisotropy. *Journal of Biomolecular Nmr* 26:297-315.

6. CONCLUSIONS

In this thesis I have presented molecular simulation applications to the calculation of free energy profiles and NMR properties. These quantities have been related to structural features of proteins and nucleic acids.

First, I investigated the permeation of water through the Aquaporin-1 channel. We have shown that water passes through the channel as a single file and must overcome a small energetic barrier due to the loss of H-bonding interactions. The water molecules are differently oriented in the two halves of the channel, due to the protein electrostatic field. The role of the two conserved Asn residues lining the channel was established as being the site where the water molecules reorient.

Next, we have focused on the calculations of NMR properties. We have characterized first the highly non-trivial chemical shifts in Aspartic Proteases/phosphorous containing inhibitors, providing with structural models of the H-bond pattern at the active site. Our calculations also confirmed that the chemical shift of a proton involved in a Short Strong Hydrogen Bond is highly downfield shifted with respect to normal hydrogen bonds. Then, our attention turned to nucleic acids. Hydrogen/Deuterium Isotope Effects (DIE's) on C-13 chemical shifts in A:T and A:U pairs in DNA and RNA oligomers were investigated to rationalize recent experimental data (Vakonakis and LiWang. 2004. *J. Biomol. NMR* 29:65. *J. Am. Chem. Soc.* 126:5688). Our calculations showed that the observed differences between the DIE's in A:U and A:T are due to the different magnetic properties of the base pairs, and that the range of the observed DIE's is modulated by the base pair conformation and by the electric and magnetic fields generated by the duplex and counterions.

In conclusion, we have shown that our atomistic simulations help to relate structural features to the function of biomolecules and to assign and interpret non-trivial experimental biophysical data.