



# ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

## Theoretical investigations on ligand- and RNA-protein complexes of pharmaceutical relevance

Thesis submitted for the Degree of *Doctor Philosophiae*

Candidate:

Sergio Fabian Pantano Gutierrez

Supervisor:

Prof. Paolo Carloni

Cosupervisor:

Prof. Mauro Giacca

October 2001

**SISSA - SCUOLA  
INTERNAZIONALE  
SUPERIORE  
DI STUDI AVANZATI**

TRIESTE  
Via Beirut 2-4

**TRIESTE**



SISSA



ISAS

SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

# **Theoretical investigations on ligand- and RNA-protein complexes of pharmaceutical relevance**

Thesis submitted for the Degree of *Doctor Philosophiae*

Candidate:

Sergio Fabian Pantano Gutierrez

Supervisor:

Prof. Paolo Carloni

Cosupervisor:

Prof. Mauro Giacca

October 2001





## Table of contents

<b>FOREWORD</b>	<b>5</b>
<b>CHAPTER I</b>	
<b>SUMMARY</b>	<b>7</b>
<b>1.1 NADH Interactions with wt- and S94A-Acyl Carrier Protein Reductase from Mycobacterium tuberculosis:</b>	<b>8</b>
<b>1.2 Characterization of HIV-1 Tat protein from molecular dynamics simulations.</b>	<b>9</b>
<b>1.3 Amino acid modification in the HIV-1 Tat basic domain: insights from molecular dynamics and in vivo functional studies</b>	<b>11</b>
<b>1.4 HIV-1 Tat-TAR complex: Insights from MD studies.</b>	<b>13</b>
<b>CHAPTER II</b>	
<b>METHODS</b>	<b>15</b>
<b>2.1 Classical (Force Field based) molecular dynamics simulations.</b>	<b>15</b>
2.1.1 Empiric force fields	15
2.1.2 The integration algorithms	17
2.1.3 Periodic Boundary Conditions	19
2.1.4 Long Range Electrostatics evaluation	19
2.1.5 The SHAKE method	22
2.1.6 Temperature and pressure control	23
2.1.7 Calculated properties	25
<b>2.2 Electronic structure calculations</b>	<b>27</b>
2.2.1 Density Functional Theory and Kohn-Sham formulation	28
2.2.2 Local Density and Gradient Corrected Approximations	30

2.2.3	Plane waves	31
2.2.4	Pseudopotentials	32
2.2.5	Environment effects.	33
2.2.6	Maximally localized Wannier Functions	34
2.2.7	Car-Parrinello Molecular Dynamics	35
2.2.8	Calculated properties	36
<b>2.2</b>	<b>Experimental Methods:</b>	<b>37</b>
2.2.1	Site-directed mutagenesis experiment.	37
 <b>CHAPTER III</b>		
<b>ENZYME LIGAND INTERACTIONS IN DRUG RESISTANT INHA FROM</b>		
<b><i>MYCOBACTERIUM TUBERCULOSIS</i></b>		<b>39</b>
<b>3.1</b>	<b>Introduction: Mycobacterium tuberculosis and Isoniazid target</b>	<b>39</b>
<b>3.2</b>	<b>Structural Models</b>	<b>42</b>
3.2.1	Quantum Chemistry Calculations	42
3.2.1	Classical MD calculations	43
<b>3.3</b>	<b>Computational Details</b>	<b>44</b>
3.3.1	Quantum Chemistry Calculations	44
3.2.1	Force field based Calculations	45
<b>3.3</b>	<b>Results</b>	<b>46</b>
3.3.1	Ab initio MD calculations on wt.	46
3.3.2	WAT1 Polarization.	48
3.3.3	Comparison with S94A.	50
3.3.4	Classical MD calculations on wt.	51

<b>3.4</b>	<b>Discussion</b>	<b>54</b>
------------	-------------------	-----------

## **CHAPTER IV**

	<b>STRUCTURAL AND DYNAMICAL CHARACTERIZATION OF TAT PROTEIN.</b>	<b>57</b>
4.1	Introduction: Human Immunodeficiency Virus Type 1(HIV-1)	57
4.2	The Tat protein in the HIV-1 life cycle.	58
4.3	Structural model	61
4.4	Computational Details	61
4.5	Results	62
4.5.1	Electrostatic analysis	63
4.5.2	Motional analysis:	64
4.5.3	Site directed mutagenesis experiments	69
4.6	Discussion	70

## **CHAPTER V**

	<b>AMINO ACID MODIFICATION IN THE HIV-1 TAT BASIC DOMAIN</b>	<b>73</b>
5.1	<b>Introduction: The Arginine Rich Domain</b>	<b>73</b>
5.2	<b>Structural models</b>	<b>74</b>
5.2.1	K50R, K50A, AcK50, K50A and K51R Tat.	74
5.3	<b>Computational Details</b>	<b>75</b>
5.4	<b>Results</b>	<b>76</b>
5.4.1	MD simulations.	76
5.5	<b>Biochemical characterization of the Tat mutants.</b>	<b>83</b>

<b>5.6 Discussion</b>	<b>84</b>
<b>CONCLUSIONS</b>	<b>87</b>
<b>BIBLIOGRAPHY</b>	<b>89</b>

## Foreword

In the last twenty years, force-field based molecular dynamics simulations (MD) have emerged as very powerful tool to investigate structural, energetic and dynamical aspects of biomolecules and their interactions with ligands (Wang et al. 2001). The technique have more recently been extended the domain its applications to nucleic acid- and protein- protein complexes and to time-scales which approach those relevant for several biophysical processes, so that direct contact with experimental measurements can be made.

In this work, MD techniques, accompanied by *ab initio* calculations, have been used to explore structure/function relationships on fundamental targets for pharmaceutical relevance.

The first study focuses on the interactions between the anti- tubercular drug Isoniazid and its target, the acyl carrier protein reductase. Isoniazid inhibits the production of long chain fatty acids used by the bacteria to construct the cell membrane, determining a progressive weakening of the cell wall that eventually leads to bacterial death (Bradford and Daley 1998). Our MD studies and *ab initio* calculations on the protein complexed with NADH and drug helps to explain the small changes in free energy acyl carrier protein reductase on passing from the wild-type enzyme to a mutant causing drug-resistance, S94A (Dessen et al. 1995). Furthermore, they call for a review of the refinement procedure of the x-ray determined crystallographic structure.

The second study is devoted to a structural and dynamical characterization of the transactivation of transcription (Tat) protein of the human immunodeficiency virus type 1 (HIV-1) and the adduct with its viral RNA partner, a hairpin like segment of the viral long terminal repeat, the Trans Activation Response region (TAR). Tat has been recognized since its discovery, about fifteen years ago (Sodroski et al. 1985), as a very promising target for therapeutic intervention against AIDS. Unfortunately, due to its large flexibility and aggregation properties very little structural information is available.

MD simulations based on the NMR structure (Bayer et al. 1995) of Tat have been carried out to investigate structure/function relationships of wild type Tat and several point mutants have been investigated. The *in silico* findings have been confronted with *in vivo* experiments simultaneously carried out in the laboratory of Molecular Medicine directed by Prof. Mauro Giacca at the International Centre for Genetic an Engineering and Biotechnology (ICGEB).

Our results provide a rationale for the reduced biological activity of some mutants and they have allowed to engineer new site-directed mutagenesis experiments.

As final step of our investigation we have dealt with the binding of Tat with its mRNA target TAR, which plays a crucial role for viral transcription (Karn 1999). Using MD simulations, we have constructed a structural model of Tat/TAR based on the isolated components, whose NMR structures are known (Bayer et al. 1995; Aboul-ela, Karn, and Varani 1995; Aboul-ela, Karn, and Varani 1996). This structural model constitute an important step for rational drug design leading to complexes capable of inhibit the normal function of Tat protein and hence viral replication.

# Chapter I

## Summary

The following chapter provides a brief overview of the main findings of my thesis work.

The first paragraph is devoted to the description of a study on the enzyme-substrate-drug interaction in InhA of *Mycobacterium tuberculosis*. Theoretical results of both, *ab initio* and classical molecular dynamics are confronted against experimental data on dissociation constants and crystallographic data from the literature. The second paragraph is referred to the structural characterization through classical molecular dynamics of Tat, a viral protein of HIV-1, which is crucial for viral replication, and constitute a very attractive possibility for anti AIDS chemotherapeutic intervention. The third paragraph concerns the study of several mutants of the Tat protein in residues that are very recently been reported as target of post-translational modification. Finally, and within the same framework, we have also explored the interaction of Tat protein with its RNA target. A very simple methodology using vacuum MD allowing to partially overcome the problem of limited time sampling is presented. The theoretical model obtained matches most of the experimental data available in literature (Calnan et al. 1991a; Calnan et al. 1991b; Puglisi et al. 1992; Sumner-Smith et al. 1991; Wang and Rana 1998; Naryshkin et al. 1997; Puglisi et al. 1992; Sumner-Smith et al. 1991; Wang and Rana 1998). In the later two cases theoretical studies have been conducted simultaneously with experiments, working in collaboration with Prof. Mauro Giacca's Laboratory of Molecular Medicine at the International Centre for Genetic Engineering and Biotechnology (ICGEB).

## 1.1 *NADH Interactions with wt- and S94A-Acyl Carrier Protein Reductase from *Mycobacterium tuberculosis*:*

Isonicotinic acid hydrazide (isoniazid, Inh) has been a first-line chemotherapeutic agent in treating *Mycobacterium tuberculosis* aggression (Bloom and Murray C.J.L. 1992) from the early fifty's. It targets the InhA bacterial enzyme (Bradford and Daley 1998) that catalyzes the NADH dependent reduction of unsaturated, long chain,  $\beta$ -branched fatty acids (mycolic acids), which are vital for bacterial cell wall synthesis. Inhibition of the enzyme may lead to an increased bacterial vulnerability towards external oxidative attacks and eventually to bacterial death.

Drug benefit is severely limited by the emergence of resistance, that is the ability of specific mutants developed by the bacterium to overcome drug's action while retaining InhA enzymatic activity. Clarifying the effect of InhA point mutations on drug action at the molecular level has shown to be difficult.

An example of this difficulty is given by the S94A mutant, which causes resistance towards the widely-used drug isoniazid (isonicotinic acid-NADH, Inh•NADH). This mutation causes the loss of two water-mediated H-bonds with the drug, as shown by the crystal structures of drug/substrate wt (Rozwarski et al. 1998) and substrate S94A (Dessen et al. 1995) complexes (Fig. 3.1). However, the structural changes are accompanied by a decrease of affinity to NADH bind of only  $\sim 0.7$  kcal/mol (Dessen et al. 1995; Parikh, Xiao, and Tonge 2000), that is much less than what expected from the changes in the structural features.

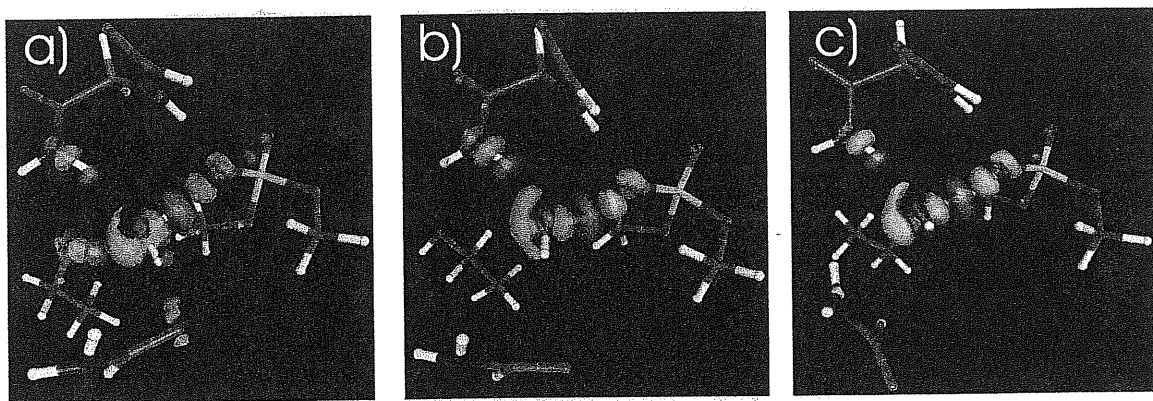
We address this by carrying out classical molecular dynamics simulation on the entire protein based on the GROMOS96 force field (Scott and van Gunsteren 1995) in presence of explicit solvent. Comparison is made with ab initio MD methods on relatively large models representing wt- and S94A-Inh•NADH interactions at the 94 site.

Force field based calculations suggest that the water-mediated H-bond interactions between Ser94 side chain and NADH, present in wt InhA X-ray structure, can be lost during the dynamics.

The quantum-chemical calculations provided further support for the fluctuational behavior of Ser94.



This conformational change in Ser94 is accompanied by a structural rearrangement of Gly14. The calculated structure of wt is remarkably similar to the X-ray structure of the S94A mutant in terms of geometrical parameters and chemical bonding. Electronic structure results of both, the protein components and the water confined in the enzymatic cavity highlight the importance of quantum calculations since force field based approach (lacking of polarization effects) are not capable to properly describe the physico-chemical interactions within the enzyme pocket.



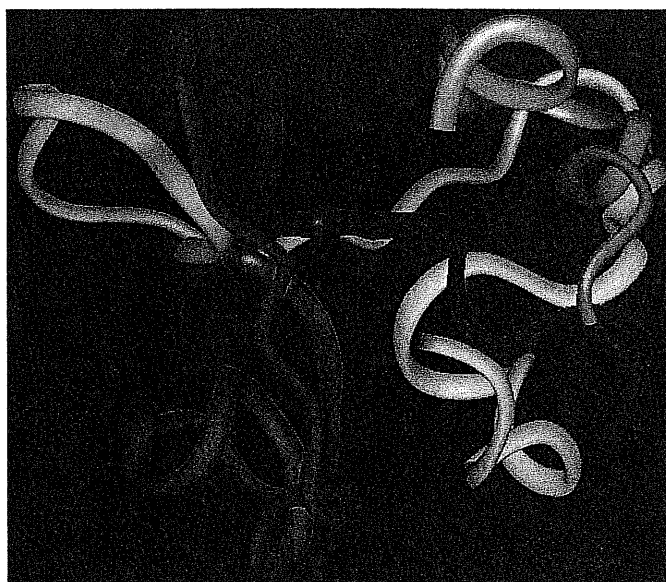
**Figure 1.1:** Comparison between quantum models of wt and S94A InhA.  $\Delta\rho$  indicating polarization effects on water molecule inside the enzymatic pocket of InhA. Initial structure (a). Conformer with partially disrupted water mediated interaction between protein and phosphate moiety (b). S94A InhA enzyme (c). Isosurfaces contour at 0.0035 a.u.<sup>3</sup> Yellow (Pink) corresponds to negative (positive) values.

The previously unrecognized high mobility of Ser94 provides a rational of the small change in free binding energies on passing from wt to S94A InhA.

## 1.2 Characterization of HIV-1 Tat protein from molecular dynamics simulations.

The gene expression of Human Immunodeficiency Virus (HIV) is regulated by the small nuclear protein Tat (transactivator protein). Therefore, Tat has been identified as a promising target for anti HIV therapeutic intervention since inhibition of Tat-TAR binding would eventually lead to the block of viral replication. Based in amino acid relative abundance in the

sequence, and of their conservation with homologous Tat proteins from other lentiviruses (Derse et al. 1991; Carroll, Martarano, and Derse 1991), it is possible to identify six regions, expressed by two exons in the viral genome.

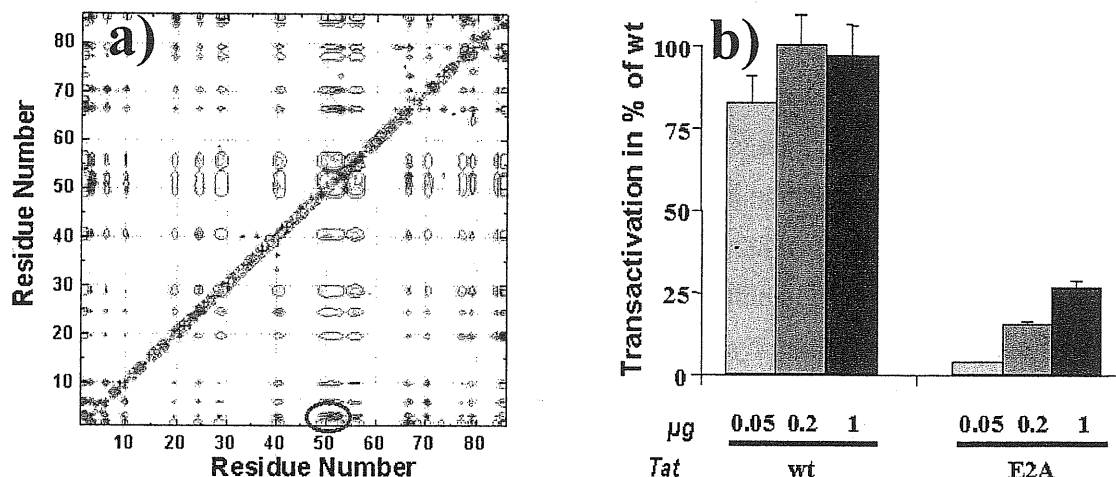


**Figure 1.2.** Ribbon representation of HIV-1 Tat backbone (Zaire 2 isolate) (Bayer et al. 1995). The N-terminus (Residues 1-20, gray color) domain. (ii) Cysteine rich domain (Residues 20-34, yellow). (iii) The Core region (Residues 35-48, green). (iv) The basic domain (Residues 49-57, red). (v) The glutamine rich domain (Residues 58-72, blue). (vi) The second exon expressed C terminal (Residues 73-86, pink).

This classification, however, has been object of some criticism, as the polypeptide segments do not constitute independent motional domains or clearly defined functional units. Based on the NMR structure (Bayer et al. 1995) we have analyzed dynamical aspects of Tat protein. By carrying out 3ns of molecular dynamics simulation of wt Tat in presence of explicit solvent in the npt ensemble with particle mesh Ewald treatment for long range electrostatics.

Electrostatic analysis together with an essential dynamics and correlation studies on the MD trajectory allowed us to identify crucial interactions that provide structural stability to the protein.

Correlation matrix and essential dynamics and analysis shows that there is some small degree of correspondence between biochemically defined domains and independent motional regions in the protein



**Figure 1.3:** Electrostatic energy map calculated per residue (a). Color code varies according to the spectrum continuously from blue (negative or stabilizing energy) to red (positive or destabilizing energy). Red ovals highlights the deep minimum corresponding to the Asp2-basic domain stabilizing interaction.

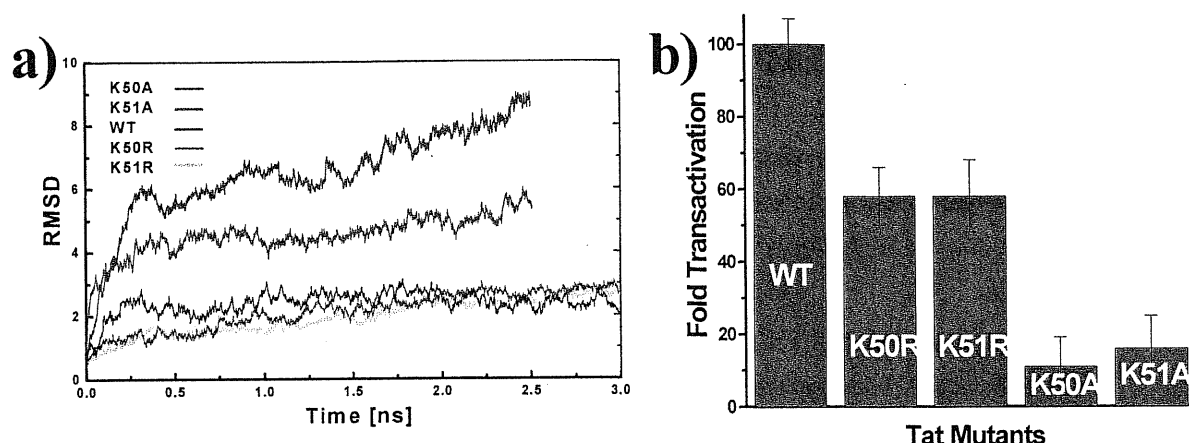
In vivo transactivation of Tat mutant at position 2 (b). Measurement assays were performed 48 hours after transfection. Shown are means and standard deviations of three independent transfections.

Comparison of theoretical prediction is made by measuring the in vivo biological activity of mutants at position 2 either with glutamine or with alanine introduced by recombinant PCR techniques.

### 1.3 *Amino acid modification in the HIV-1 Tat basic domain: insights from molecular dynamics and in vivo functional studies*

Tat is an unstructured protein. It contains a very basic region ( $R^{49}KKRRXRRR^{57}$ ) which is crucial for a number of key functions. In particular, this domain is needed for the interaction of Tat with other cellular macromolecules and viral RNA. Recent experimental evidence indicates that Lys50 and 51 of the basic domain are also targets for post-translational modification through acetylation (Kiernan et al. 1999; Col et al. 2001; Ott et al. 1999; Deng et al. 2000), suggesting the relevance of these amino acids for protein function. Mutation studies of these

residues have indicated that substitution of Lys with Ala strongly affects transactivation while a conservative substitution with Arg has a more modest effect (Kiernan et al. 1999). Starting from the NMR structure of Bayer et al. (Bayer et al. 1995), we have here undertaken a molecular dynamics (MD) study of Ala and Arg mutants at positions 50 and 51 to understand the effects of these modifications at the structural level. Similarly, we have used a MD simulation of acetylated Lys 50 to discriminate the consequences of charge neutralization from



those caused by amino acid length.

**Figure 1.4:** Root mean square deviation [Å] for Tat wt and mutants (a) during dynamics. In vivo transactivation of Tat wt and mutant at positions 50 and 51. Measurement assays were performed 48 hours after transfection. Shown are means and standard deviations of three independent transfections. Notice the correlation between the structure conservation (panel a) and the corresponding biological activity (panel b)

We find that replacement with Ala causes a profound structural rearrangement of the protein, which may eventually, result in unfolding. Furthermore, we observe that the structural determinants of the acetyl derivative of Lys 50 are remarkably similar to those of both wild-type and Lys 50 Arg mutant, suggesting that the change of chemical properties on passing from Lys to Ala plays a key role for unfolding.

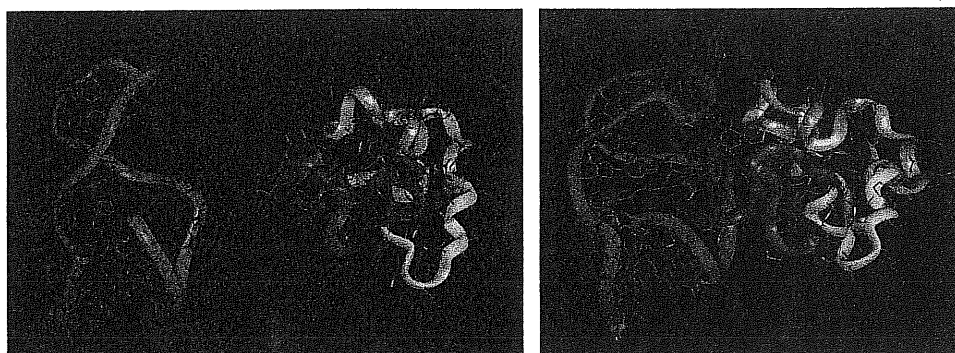
Finally, to experimentally substantiate these theoretical results, for all the mutants we provide a homogeneous set of biological data (including transactivation properties and cellular localization), which turn out to be consistent with the theoretical predictions.

## 1.4 HIV-1 Tat-TAR complex: Insights from MD studies.

HIV1-Tat promotes and enhances viral gene transcription by specifically binding, through its basic domain, a hairpin mRNA element (trans-activation-responsive region; TAR) present at the 5' end of viral transcripts (nucleotide position +1 to +59). Inhibition of Tat-TAR binding blocks the viral replication cycle. Therefore, a detailed description of the contact points in the protein-RNA interface is crucial for the development of pharmacological agents targeted to inhibit Tat-TAR binding.

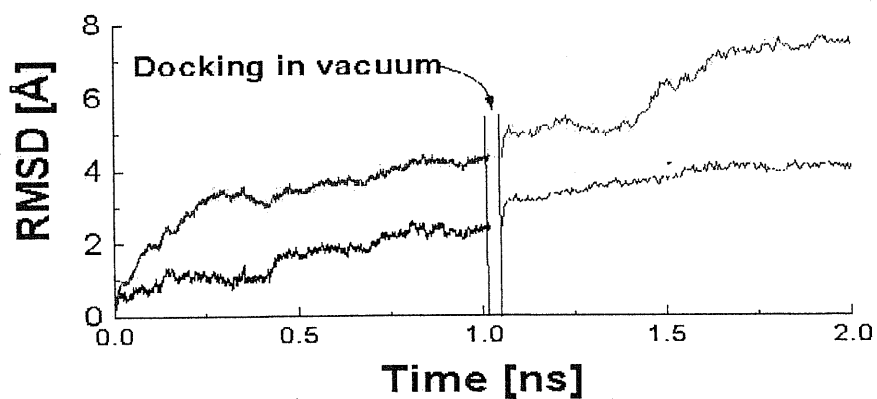
Using as initial point the NMR derived structures of unbound Tat (Bayer et al. 1995) and TAR (Aboul-ela, Karn, and Varani 1996), a 3-step methodology in which Molecular dynamics is used as a fully flexible tool for protein-RNA docking is presented.

In a **first step**, both components were simulated separately by 1 ns in npt ensemble until they reach dynamical stabilization. In a **second step**, both molecules are appropriately placed one in front of the other and vacuum the short MD simulation is performed on the dual system. Since both objects are highly charged with opposite sign (-27 for RNA and +11 for the protein) they attract one to the other. So, no intermolecular constraints are needed to bring them together. Absence of solvent viscosity allows for a much faster mutual recognition and induced fit. After 50 ps vacuum simulation is interrupted. In a **third step**, the adduct is immersed in a water simulation box with explicit solvent, neutralizing counterions and PME for long range electrostatics. Under these conditions, Molecular dynamics is performed on the unconstrained complex until dynamical stabilization is reached (Fig. 1.5).



**Figure 1.5.** Initial (left) and final (right) configurations of the TAR-Tat complex

The final structure of the adduct shows large adaptive rearrangement as can be seen in Fig.1.6.



**Figure 1.6** .RMSD behavior of the complex components during the three steps of the docking procedure.

Black: RNA TAR, red Tat protein.

Experimental confront with previously published results (Aboul-ela, Karn, and Varani 1995; Wang and Rana 1998; Puglisi et al. 1992; Sumner-Smith et al. 1991; Calnan et al. 1991a; Calnan et al. 1991b) convalidate the resulting structure. Furthermore, it provide hints to rationalize the reduced biological activity found in Lys→Arg mutants within the basic domain of Tat.

## Chapter II

### Methods

This chapter presents the theoretical background, methods and computational tools used in the calculations. For general references see (Allen and Tildesley 1987; Leach 1996; Parr and Yang 1989). Computational details such as specific parameters and building of initial configuration of molecular models are given in the respective chapters.

#### ***2.1 Classical (Force Field based) molecular dynamics simulations.***

In classical molecular dynamics framework, atoms are treated within the Born-Hoppenheimer approximation and the electronic degrees of freedom are averaged out under this assumption. For a system of  $N$  particles of masses  $\{m_i\}$  and positions  $\{\mathbf{r}_i\}$ , we may write the Newton's equations of motion as:

$$\mathbf{m}_i \ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i} E \quad i = 1, \dots, N \quad (2.1)$$

The potential energy  $E$  acting on the atomic nuclei may be described by simple functions of the ionic positions themselves  $E = E(\{\mathbf{r}_i\})$ .

The force-field expression is obtained from analytical functions derived of experimental data fitting. In this section the force-field parameterization is introduced as well as computational tools used in our simulations. Details about the structural models and the specific parameters used in the calculations are reported in the next chapter.

##### **2.1.1 Empiric force fields**

From the previous section, it is clear that at the heart of any molecular dynamics simulation lies the functional form and parameterization potential or force field.

A characteristic feature of force fields for biomolecular simulations is the use of a two-body additive force field. Such a model attempts to include implicitly many body effects in the parameterization. A general force field hamiltonian has the following form:

$$\begin{aligned}
 E &= E_{\text{bond}} + E_{\text{nonbond}} \\
 E &= E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} + E_{\text{van der Waals}} + E_{\text{electrostatic}} = \\
 &= \sum_{\text{bonds}} K_r (\mathbf{r} - \mathbf{r}_{\text{eq}})^2 + \sum_{\text{angles}} K_{\theta} (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}, n} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \\
 &\quad + \sum_{i < j} \left[ \frac{A_{ij}}{(\mathbf{r}_i - \mathbf{r}_j)^{12}} - \frac{B_{ij}}{(\mathbf{r}_i - \mathbf{r}_j)^6} + \frac{\mathbf{q}_i \mathbf{q}_j}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \right]
 \end{aligned} \tag{2.2}$$

### 2.1.1.1 Bonded Interactions:

$E_{\text{bonds}} + E_{\text{angles}}$ : bond stretching and angle bending interactions are taken into account by using harmonic energy terms involving the covalently bound atoms. The equilibrium bond and angle values  $r_{\text{eq}}$ ,  $\theta_{\text{eq}}$ , as well as the spring strengths  $K_r$  and  $K_{\theta}$  are adjusted to reproduce experimentally determined normal mode frequencies.

$E_{\text{dihedrals}}$ : dihedral parameters are calibrated on small model compounds, comparing the energies with those obtained by *ab initio* quantum chemical calculations. So called “Improper dihedral” angles are inserted between non linked bonds to preserve planarity in aromatic rings, restrict tetrahedral conformations, etc. Introduction of this restrictions is also useful since it may help the system to avoid sampling of non physical conformations.

### 2.1.1.2 Non Bonded Interactions:

$E_{\text{van der Waals}}$ : Lennard-Jones potential is used to mimic the van der Waals interactions. Parameters are assigned so as to reproduce chemical-physical properties in simple organic liquids (densities, enthalpies of vaporization, free energies of solvation, etc).



$E_{electrostatic}$ : The electrostatic energy is obtained from the evaluation of the interaction between atomic point charges. All force fields attempt to make the electrostatics in such a way that interactions are balanced both with the other terms in eq. 2.2 and with simple water models that reproduce macroscopic properties of bulk water. This is normally accomplished by using rather large partial charges.

All but AMBER force field use an empiric procedure for partial charge determination. AMBER charges are derived from Hartree-Fock calculations with a 6-31G\* basis set. The electrostatic potential  $\{V_j\}$  is calculated on  $M$  grid points outside the molecule. The point charges  $\{q_i\}$  are fitted minimizing  $\chi^2_{resp}$ :

$$\chi^2_{resp} = \sum_{j=1}^M (V_j - \bar{V}_j)^2 + \sum_{i=1}^N [(q_i^2 + a^2)^{1/2} - b] \quad (2.3)$$

where  $\{\bar{V}_j\}$  is the electrostatic potential generated by the point charges,  $a$  and  $b$  are appropriate parameters (Singh and Kollman 1986).

The above quoted treatment of non bonded interactions works very well to reproduce either intermolecular or non close atoms within the same molecule, however, it fails in reproducing covalently or 1-3 bound interactions since it give place to huge repulsion values of the energies because it does not take into account quantum interactions. This problem is overcome calculating non bonding terms only between atoms in different molecules or for atoms in the same molecule separated by at least three bonds.

### 2.1.2 The integration algorithms

For any given energy function of the atomic and molecular interactions, the trajectory of the system is determined by its initial conditions (e.g. atomic positions and velocities). Since the Newton's equations of motions cannot be solved analytically the integration of the equations of motion is achieved using finite difference methods. The time variable is discretized by assuming that no variations in the force acting on the atoms occurs within the relatively small period of time step  $\delta t$ . Several algorithms have been established for the numerical integration, differing for accuracy and computational speed (Leach 1996; Allen and Tildesley 1987). The next two paragraphs shortly describe the two algorithms used in this work.

### 2.1.2.1 The Verlet algorithm:

The velocity Verlet algorithm propagates the positions  $\mathbf{r}$  and velocities  $\mathbf{v}$  from time  $t$  to time  $t+\delta t$  using the following set of equations(Swope et al. 1982).

$$\begin{cases} \mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) \\ \mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{2} \delta t [\mathbf{a}(t) + \mathbf{a}(t+\delta t)] \end{cases} \quad (2.4)$$

where  $\mathbf{a}$  is the acceleration.

### 2.1.2.2 The Leap-Frog algorithm:

This algorithm, calculates the velocities  $\mathbf{v}$ , at time  $t+1/2\delta t$ , which are then used to calculate the positions,  $\mathbf{r}$ , at time  $t+dt$ .

$$\begin{cases} \mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \frac{\delta t}{2}) \\ \mathbf{v}(t + \frac{\delta t}{2}) = \mathbf{v}(t - \frac{\delta t}{2}) + \mathbf{a}(t) \delta \end{cases} \quad (2.5)$$

It is like velocities *leap* over the positions, and then the positions *leap* over the velocities.

Since velocities are not calculated at the same instant  $t$  than positions, they can be approximated as:

$$\mathbf{v}(t) = \frac{1}{2} \left[ \mathbf{v}(t - \frac{\delta t}{2}) + \mathbf{v}(t + \frac{\delta t}{2}) \right] \quad (2.6)$$

### 2.1.3 Periodic Boundary Conditions

In molecular dynamics, periodic boundary conditions are often used to minimize boundary effects. In this scheme, a particle interacts with all other  $N-1$  particles into the simulation box which is surrounded by identical translated images of it self in an infinite array of periodic cells (Fig 2.1) (Allen and Tildesley 1987). Clearly this imposes a periodicity that can be overcome by chosen a large enough system.

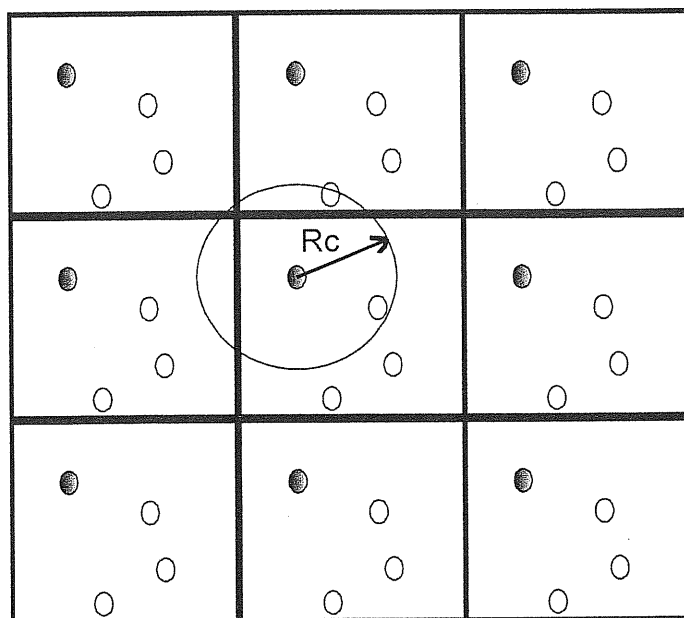


Fig 2.1 Periodic Boundary Conditions

In this scheme, a generic atom interacts with atoms in the box or its images in the surrounding replicas that lie within a cut off radius  $R_c$  that fulfills the condition  $R_{\text{box}} > 2 R_c$ , ensuring that any atom can not simultaneously interact with other atom and its image.

Non Bonding (e.g. Electrostatic and van der Waals )interactions are commonly evaluated using this method with a typical cut off radius ranging from  $\sim 9$  to  $15 \text{ \AA}$  with an obvious increase of computational effort for higher values.

### 2.1.4 Long Range Electrostatics evaluation

In the next two paragraphs we briefly describe two of the most commonly implemented methods to take into account long range electrostatic effects in molecular dynamic simulations. Since electrostatics are very slowly decaying, special treatment is needed to accurately reproduce the interactions governing biological systems.

#### 2.1.4.1 Poisson-Boltzmann reaction field correction

The reaction field correction consists basically in a modification to the interatomic force due to electrostatic interaction, so that:

$$|\mathbf{F}_{i,j}| = q_i q_j \left( \frac{1}{r_{i,j}^3} - \frac{C_{rf}}{R_{rf}^3} \right) \quad (2.7)$$

where the correction term represents the force on the  $i^{\text{th}}$  atom due to the reaction field induced by the  $j^{\text{th}}$  atom considering them embedded in a electrostatic continuum media of dielectric permittivity  $\epsilon_2$  and a inverse Debye length  $\kappa$  outside the reaction field cutoff  $R_{rf}$ .  $C_{rf}$  reads:

$$C_{rf} = \frac{(2 - 2\epsilon_2)(1 + \kappa R_{rf}) - \epsilon_2 (\kappa R_{rf})^2}{(1 + 2\epsilon_2)(1 + \kappa R_{rf}) + \epsilon_2 (\kappa R_{rf})^2} \quad (2.8)$$

Reaction field methods with the use of a long enough cut off radius offer a computationally cheap description of non bonded electrostatics suitable for most protein systems. However, highly charged proteins and nucleic acids may require a more convenient treatment as the Ewald summation method.

#### 2.1.4.2 The Ewald summation method.

The Ewald summation method makes uses of the periodic boundary conditions considering the system as an infinite crystal.

The electrostatic potential energy  $V$  in an infinite lattice of charges  $N q_i$  can be written as:

$$V = \frac{1}{8\pi\epsilon_0} \sum_{|n|=0}^{\infty} \sum_{i=1}^N \sum_{\substack{j=1 \\ (i \neq j \text{ if } n=0)}}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (2.9)$$

where  $\mathbf{r}_{ij}$  and  $\mathbf{n}$  are the position and direct lattice vectors respectively .

This is and infinite series is conditionally convergent and thus is computationally impracticable. A solution to this problem is divide in two parts. For each charge  $q_i$  in the

system at position  $r_i$ , a gaussian charge distribution of opposite total charge  $(-q_i)$ , is added ( $\rho_i^s$ ) and subtracted ( $\rho_i^G$ ), so that the total charge distribution of the system  $\rho(r)$  reads:

$$\begin{aligned}\rho(\mathbf{r}) &= \rho(\mathbf{r}) + \rho^s(\mathbf{r}) - \rho^G(\mathbf{r}) = \\ &= \sum_{i=1}^N q_i \delta(\mathbf{r} - \mathbf{r}_i) + \sum_{i=1}^N q_i (\alpha / \sqrt{\pi})^3 e^{-\alpha^2(\mathbf{r} - \mathbf{r}_i)^2} - \sum_{i=1}^N q_i (\alpha / \sqrt{\pi})^3 e^{-\alpha^2(\mathbf{r} - \mathbf{r}_i)^2}\end{aligned}\quad (2.10)$$

Let now consider  $V^s$ , which depends on  $\rho(\mathbf{r}) + \rho^s(\mathbf{r})$  with the electrostatic potential, and  $-\rho^G(\mathbf{r})$  with the potential  $V^G$ .

$$V^s = \frac{1}{8\pi\epsilon_0} \sum_{|\mathbf{n}|=0}^{\infty} \sum_{i=1}^N \sum_{\substack{j=1 \\ (i \neq j \text{ if } \mathbf{n}=0)}}^N \frac{q_i q_j \text{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (2.11)$$

where erfc is the complementary error function:  $\text{erfc}(\mathbf{x}) = \frac{2}{\pi} \int_{\mathbf{x}}^{\infty} e^{-t^2} d\mathbf{t}$ .

$V^s$  converges rapidly because the opposite charged  $\rho^s(\mathbf{r})$  distribution shields the point charges, allowing to neglect the electrostatic contribution for large values of  $|\mathbf{r}_{ij} + \mathbf{n}|$ . The rate of convergence increases with the width of the gaussian, e.g. with the  $\alpha$  parameter, that should be chosen so that only interactions within the cut off radius  $R_c$  are included (i.e., those for which  $|\mathbf{n}|=0$ ).

The term  $V^G$ , which depends on  $\rho^G(\mathbf{r})$ , can be fastly evaluated in the reciprocal (Fourier) space:

$$V^G = \frac{1}{4\pi\epsilon_0} \frac{1}{2\pi L^3} \sum_{\mathbf{k} \neq 0} \sum_{i=1}^N \sum_{j=1}^N q_i q_j \frac{4\pi^2}{k^2} e^{-k^2/4\alpha^2} \cos(\mathbf{k} \cdot \mathbf{r}_{ij}) \quad (2.12)$$

The convergence is much more rapid than the original point charge sum. As can be seen from eqs. 2.11 and 2.12 a compromise must be reached to adjust the  $\alpha$  parameter that modulates the convergence of both  $V^s$  and  $V^G$ .

An efficient Particle mesh Ewald implementation consists in interpolating the reciprocal space potential on a grid using smooth function (Darden and York 1993; Essman et al. 1995). Therefore, the sum can be evaluated by using order  $N \ln(N)$  fast Fourier transform algorithms.

### 2.1.5 The SHAKE method

In MD simulations the highest characteristic frequency motions are the bond vibrations. In particular light atoms, like hydrogen, are involved in the fastest bond stretching vibrations with characteristic frequencies of  $\sim 10^{14}$ - $10^{15}$  Hz. These frequencies are often of little or none interest in classical simulations of biological systems. The use of specific constraints allows to keep these bond lengths fixed during the simulation. The time step can be, therefore, increased without decreasing the accuracy of the simulation.

The set of  $N_c$  molecular constraints  $\sigma_k$ , one for each constrained bond, has the following form:

$$\sigma_k(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathbf{r}_{k_1 k_2}^2 - \mathbf{d}_{k_1 k_2}^2 = 0; \quad k = 1, \dots, N_c \quad (2.13)$$

where  $\mathbf{d}_{k_1 k_2}$  are the constraint and  $\mathbf{r}_{k_1 k_2}$  are the actual distances between the atoms  $k_1$  and  $k_2$ . An extra term is added to the potential energy function, and the equation of motions become:

$$\mathbf{m}_i \ddot{\mathbf{r}}_i = -\nabla_{\mathbf{r}_i} \mathbf{E} - \nabla_{\mathbf{r}_i} \sum_{k=1}^{N_c} \lambda_k(t) \sigma_k(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad i = 1, \dots, N \quad (2.14)$$

where  $\{\lambda_k\}$  are time dependent Lagrange's multipliers that have to be determined to simultaneously satisfy eq. 2.14.

After the time step  $\delta t$ , the system coordinate set  $\mathbf{r}'_{k_1}(t + \delta t)$  will differ by a quantity  $\delta \mathbf{r}_i$  with respect to that of the unconstrained system  $\mathbf{r}_{k_1}(t + \delta t)$ :  $\mathbf{r}'_{k_1}(t + \delta t) = \mathbf{r}_{k_1}(t + \delta t) + \delta \mathbf{r}_{k_1}$ . Therefore, corrections due to constraints  $\mathbf{d}_{k_1 k_2}$  are applied on atoms  $k_1$  and  $k_2$  in the direction of the  $\mathbf{r}_{k_1 k_2}$  but in opposite sense as seen in Fig 2.2 (Ryckaert, Ciccotti, and Berendsen 1977).

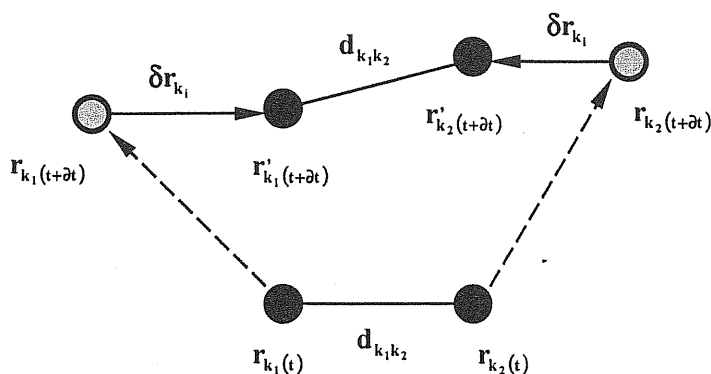


Fig 2.2. The SHAKE method

The SHAKE algorithm used in the simulations allows satisfying the set of  $k$  holonomic constraints by using an iterative procedure (Ryckaert, Ciccotti, and Berendsen 1977). For each atom involved in the constraints, an additional positional shift  $\delta r$  in the bond direction is added into the Verlet algorithms in order to subsequently satisfy all the  $k$  constraints. The fulfillment of the  $k$ th constraint partially destroys the previous ones. The procedure is therefore iterated up to the desired convergence.

### 2.1.6 Temperature and pressure control

Temperature can be kept close to the desired target value  $T_0$  using the *weak coupling method* originally proposed by Berendsen *et al.* (Berendsen et al. 1984).

This method virtually couple the system to a thermal bath at a target temperature  $T_0$ . The atomic equations of motions are modified step by step so that the system is driven towards a reference temperature.

$$\frac{dT_{(t)}}{dt} = \tau_T^{-1} [T_0 - T_{(t)}] \quad (2.15)$$

The temperature  $T_{(t)}$  of a set of  $N_{df}$  degrees of freedom is defined in terms of the kinetic energy  $E_{kin}(t)$  of those degrees of freedom by the relation:

$$E_{kin}(t) = \sum_{i=1}^N \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N_{df} k_b T(t) \quad (2.16)$$

where the summation over the atoms is restricted to the set of  $N_{df}$  degrees of freedom. A change in temperature  $\Delta T$  will relate with a change in kinetic energy  $\Delta E_{kin}$  through the heat capacity per degree of freedom  $c_v^{df}$  as:

$$\Delta E_{kin}(t) = N_{df} c_v^{df} \Delta T(t) \quad (2.17)$$

and a change in kinetic energy can be obtained by scaling the atomic velocities with a factor  $\lambda$

$$\Delta E_{kin}(t) = [(\lambda(t))^2 - 1] \frac{1}{2} N_{df} k_b T(t) \quad (2.18)$$

If now we discretize  $dt$  in eq. (2.15) using the MD time step  $\Delta t$  and solving the *temperature scaling factor*  $\lambda(t)$  it yields:

$$\lambda(t) = \left[ 1 + \frac{2c_v^{df}}{k_b} \frac{\Delta t}{\tau_T} \left[ \frac{T_0}{T(t)} - 1 \right] \right]^{\frac{1}{2}} \quad (2.19)$$

The  $\lambda(t)$  factor is used to scale the velocities  $v_i(t)$  at each MD step relaxing the temperature  $T(t)$  towards the target temperature  $T_0$ . The relaxation rate is controlled by the ratio of the heat capacity per degree of freedom  $c_v^{df}$  and the chosen *temperature relaxation time*  $\tau_T$ .

Analogously to the temperature control, it is possible to couple the system to a hydrostatic bath. A change of volume  $v$  per time unit is given by:

$$\frac{1}{v} \frac{dv}{dt} = \beta \tau_P^{-1} [P_0 - P(t)] \quad (2.20)$$

Where  $\beta$  is the isothermic compressibility and  $\tau_P$  the coupling constant. To achieve a realistic variation of the volume of the simulation box and reproduce its effects on the molecules comprising the system, a uniform rescaling of the atomic position is needed.

$$\Delta \mathbf{v} = \mathbf{v}(\mu^3 - 1) \quad (2.21)$$

Introducing (2.19) into the discrete expression of eq. (2.20) we get:

$$\mu = \sqrt[3]{1 + \beta \frac{\Delta t}{\tau_P} (P_0 - P)} \quad (2.22)$$

The values of  $c_v^{df}$  and  $\beta$  may not be accurately known, but this is no problem, since  $\tau_T$  and  $\tau_P$  are adjustable parameters.

The values of  $\tau_T$  and  $\tau_P$  should be chosen sufficiently small (strong coupling) to achieve the required average temperature and pressure, but, on the other hand sufficiently large (weak



coupling) to avoid disturbance of the physical properties of the system by the coupling to the baths.

## 2.1.7 Calculated properties

### 2.1.7.1 Root mean square displacement.

The root mean square displacements (rmsd's) of a set of snapshots  $\{r_i\}$  with respect a referring structure  $r^*$  is calculated using:

$$\text{rmsd} = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_{\text{ref}} - r_i)^2} \quad (2.23)$$

where  $N$  is the number of configurations and  $r_{\text{ref}}$  is the initial (minimized) reference configuration.

#### 2.1.1.1 Energy maps

The inter residue electrostatic (non bonded + 1-4) interaction energy matrix was calculated as

$$E_{i-j} = \text{NBE}_{\text{Ri-Rj}} + \text{BE}_{\text{Ri-Rj}} \quad (2.24)$$

Where NBE and NB correspond to non bonded interactions and bonded (1-4) electrostatic interactions respectively for all the residues ( $R_{i,j}$ ) in the protein. The maps are obviously symmetric. Diagonal matrix elements provide the electrostatic energy of the individual residues. Off-diagonal elements provide the long range electrostatic interactions between charged residues.

Energy maps were calculated on corresponding minimized MD snapshots in presence of the explicit solvent and using a cutoff radius off 50 Å, since it includes the whole protein. Test calculations with higher cut off radius does not show differences in the calculated values.

#### 2.1.1.2 Essential Dynamics

Large-scale (or essential) displacements (Amadei, Linssen, and Berendsen 1993; Garcia 1992) were calculated based on the covariant matrix  $C$  of the  $C\alpha$  atoms along the dynamics. A matrix element of  $C$  reads:

$$C = \left\langle (\chi_i - \langle \chi_i \rangle)(\chi_j - \langle \chi_j \rangle)^T \right\rangle \quad (2.25)$$

where  $\langle \rangle$  brackets indicate temporal average and  $\chi$  indicates the instantaneous position of the  $C\alpha$  atoms. The symmetric matrix  $C$  can always be diagonalized by an orthogonal coordinate transformation  $T$ :

$$x - \langle x \rangle = Tq \text{ or } q = T^T(x - \langle x \rangle) \quad (2.26)$$

which transforms  $C$  into a diagonal matrix  $\Lambda = \langle qq^T \rangle$  of eigenvalues  $\lambda_i$ :

$$C = T\Lambda T^T \text{ or } \Lambda = T^T C T \quad (2.27)$$

The  $i$ th column of  $T$  is the eigenvector belonging to  $\lambda_i$ . When a sufficient number of independent configurations (at least  $3N+1$ ) are available to evaluate  $C$  there will be  $3N$  eigenvalues, of which at least 6 representing overall translation and rotation are nearly zero. When a number of configurations,  $S$ , less than  $3N+1$ , is analyzed the total number of nonzero eigenvalues is at most covariance matrix will have not full rank.

The matrix has  $C$  the property of being always connected to the system constraints. A subspace which is forbidden for the motion is always fully defined by a subset of eigenvectors of the matrix  $C$  with zero or nearly zero eigenvalues. It is also important to note that the probability distribution of the displacements along the eigenvectors, although linearly uncorrelated, is not necessarily statistically independent. On the other hand, if a linear orthogonal transformation defines a subset of statistically independent generalized coordinates, then the unit vectors corresponding to this subset will always be eigenvectors of covariance matrix  $C$ . The total positional fluctuation  $\sum_i \langle (\mathbf{x}_i - \langle \mathbf{x}_i \rangle)^2 \rangle$  can be thought to be built up from the contributions of the eigenvectors:

$$\sum_i \langle (\mathbf{x}_i - \langle \mathbf{x}_i \rangle)^2 \rangle = \langle (\mathbf{x} - \langle \mathbf{x} \rangle)^T (\mathbf{x} - \langle \mathbf{x} \rangle) \rangle = \langle \mathbf{q}^T \mathbf{T}^T \mathbf{T} \mathbf{q} \rangle = \langle \mathbf{q}^T \mathbf{q} \rangle = \sum_i \langle \mathbf{q}_i^2 \rangle = \sum_i \lambda_i. \quad (2.28)$$

Translational and rotational motions were eliminated by performing a rmsd fit between the MD snapshots and the first conformer (Amadei, Linssen, and Berendsen 1993).

The amount of motion covered by a subset of the first  $n$  (essential) eigenvectors can be calculated by a summation over the first  $n$  (essential) eigenvalues ( $\lambda_i$ ) sorted in decreasing order as:

$$\sum_i^n \left\langle (x_i - \langle x_i \rangle)^2 \right\rangle = \sum_i^n \lambda_i \quad (2.29)$$

This sorting of the eigenvectors allows to immediately recognize the most important motional contributions since it turns out that most of the total motion is spanned by the first two eigenvectors.

To single out the contribution of different protein domains (CD) to an essential eigenvector, we average the scalar product between the eigenvectors  $\delta$  and the displacement vectors  $v$  of the  $N_{AD}$  (number of C $\alpha$  atoms per domain) C $\alpha$  atoms belonging to each domain:

$$CD = \frac{1}{N_{AD}} \sum_i^{N_{AD}} v_i \cdot \delta \quad (2.30)$$

where  $v_i$  is the displacement vector ( $v_i = x_{\text{final}} - x_{\text{initial}}$ ) of the C $\alpha$  atoms belonging to a given domain.

## 2.2 Electronic structure calculations<sup>1</sup>

The properties of the time independent ground state  $|\Psi\rangle$  of a system of interacting ions and electrons for any atomic system is provided by the time independent Schrödinger equation:  $H|\Psi\rangle = E|\Psi\rangle$

given a  $N$  electrons system, its Hamiltonian  $H$  can be written as:

$$H = T + V_{ee} + V_{ext} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \frac{1}{2} \sum_{i \neq j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^N V_{ext}(\mathbf{r}_i) \quad (2.31)$$

where  $T$  is the kinetic energy,  $V_{ee}$  the electron electron interaction potential and  $V_{ext}$  is the external potential applied to electrons, including the ion-electron electrostatic terms.

---

<sup>1</sup> For a sake of simplicity, atomic units are used in this section.  $R_{\text{ion}}$  denotes nuclei position while  $r_i$  denotes electronic positions.

In the next sections we review the basic concepts of the methods and the approximations and practical implementation used. Because in the present work we are interested only close-shell systems we limit the treatment to this special case.

### 2.2.1 Density Functional Theory and Kohn-Sham formulation

Density Functional Theory (DFT) has been shown to be a very useful framework to solve the quantum problem. Within this formalism, the Schrödinger equation for the  $N$ -electron wave function is replaced with a simpler and more intuitive scheme in which the electron density  $\rho(\mathbf{r})$  plays the fundamental role. DFT is based in two fundamental theorems by Hohenberg and Kohn (Hohenberg and Kohn 1964) that stay:

- 1- The external potential  $V_{ext}$  is determined, within a trivial additive constant, by the electron density  $\rho(\mathbf{r})$ . Therefore, all the properties of the system can be derived by the knowledge of the density  $\rho(\mathbf{r})$  alone. In particular, the total energy  $E = \langle \psi | H | \psi \rangle$  may be written as a functional of  $\rho(\mathbf{r})$ :

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{ext}[\rho] = T[\rho] + V_{ee}[\rho] + \int v_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (2.32)$$

The first two terms are called Hohenberg-Kohn (HK) functional ( $E_{HK}[\rho] = T[\rho] + V_{ee}[\rho]$ ) and are completely independent on the external potential.

- 2- For a given density  $\rho'$ , such that  $\rho'(\mathbf{r}) \geq 0$  and  $\int d\mathbf{r}\rho'(\mathbf{r}) = N$ , the ground state energy  $E_0$  of the system is the minimum of the total energy functional  $E[\rho']$  with respect to variations of  $\rho$ . This theorem is analogous to the variational principle for wave functions.

This variational principle requires that the ground state density satisfy the following stationary principle

$$\frac{\delta}{\delta \mathbf{n}} \left[ E[\rho] - \mu \int d\mathbf{r}\rho(\mathbf{r}) \right] = 0, \quad (2.33)$$

where  $\mu$  is a Lagrange multiplier introduced to impose the normalization condition.

To solve eq. (2.33) Kohn and Sham (KS) introduced (Kohn and Sham 1965) a solution scheme based on an orthonormal set of  $N$  independent single particle wave functions, the KS orbitals  $\psi_i$ . The basic idea is to choose a basis where the kinetic term of the HK functional may be

written in a simpler way in terms of these orbitals, and handling separated the residual corrections. They associated to (2.31) a corresponding non-interacting reference system with the following hamiltonian(Kohn and Sham 1965):

$$\hat{H}_s = \sum_{i=1}^N \left( -\frac{1}{2} \nabla_i^2 \right) + \sum_{i=1}^N v_{\text{eff}}(\mathbf{r}_i) \quad (2.34)$$

where the  $N$  occupied orbitals are the lowest eigenstates of the one-electron hamiltonian:

$$\left[ -\frac{1}{2} \nabla_i^2 + v_{\text{eff}}(\mathbf{r}) \right] \psi_i = \epsilon_i \psi_i, \quad (2.35)$$

and  $v_{\text{eff}}$  is a an oportune one-body effective potential. Kohn and Sham provided a procedure for the calculation of  $v_{\text{eff}}$ . In this system simple formulas for  $\rho(r)$  and for  $T[\rho]$  may be written:

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2 \quad (2.36)$$

$$T_s[\rho] = \sum_{i=1}^N \langle \psi_i | -\frac{1}{2} \nabla^2 | \psi_i \rangle \quad (2.37)$$

The HK functional may be rewritten as a function of  $T_s$ :

$$E_{\text{HK}}[\rho] = T_s[\rho] + \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{\text{xc}}[\rho] \quad (2.38)$$

Note that the first term  $T_s$  (the kinetic energy of a *non-interacting* system, i.e. without  $V_{ee}$ ) differs with respect  $T$ . The second term is the “classical” part of the electron interaction energy  $V_{ee}$ . The third term  $E_{\text{xc}}$  is the exchange correlation energy. It includes the difference between  $T$  and  $T_s$ , and the non-classical part of  $V_{ee}$ .

The Euler-Lagrange eq. (2.33) becomes:

$$\begin{cases} \mu = \frac{\delta E_{\text{HK}}[\rho]}{\delta \rho(\mathbf{r})} = v_{\text{eff}}(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho(\mathbf{r})} \\ v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{\text{xc}}(\mathbf{r}) \end{cases} \quad (2.39)$$

A density satisfying the above equation may be obtained by solving the set of  $N$  one-electron eq. (2.35).

Equations (2.35) and (2.39) may be solved in a self-consistent way: for a given set of starting orbitals  $\psi_i$ , we evaluate  $v_{eff}$  using eq. (2.39), and solve eq. (2.35) to determine a new set of orbitals. The cycle is repeated unless convergence in the KS orbitals (and therefore in density) is achieved.

However no exact expression is known for the exchange and correlation terms in the above equations and much effort has been devoted to find an accurate approximation for them.

## 2.2.2 Local Density and Gradient Corrected Approximations

A very popular approximation for the exchange and correlation energy  $E_{xc} = \int d\mathbf{r} v_{xc}(\mathbf{r})$  was originally proposed by Kohn and Sham (Kohn and Sham 1965). It is the so called Local Density Approximation (LDA). This functional is calculated based in the exchange-correlation energy of a homogeneous electron gas. For such a system the exchange energy component of  $\epsilon_x^{hom}[\rho]$  can be exactly calculated for an homogeneous electron gas (Parr and Yang 1989) as:

$$\epsilon_x^{hom}[\rho] = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3} \rho^{1/3} \quad (2.40)$$

and, the exchange-correlation energies per particle of a uniform interacting electron gas at different densities  $\epsilon_c^{hom}[\rho]$  define completely the LDA functional as:

$$E_{xc}^{LDA}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \epsilon_{xc}^{hom}[\rho(\mathbf{r})] \quad (2.41)$$

Of course, this approximation holds exactly only in the limit of uniform densities, however, it has been revealed to be a good starting point for the determination of many ground state properties of atoms and molecules, despite these systems are much distant to be homogeneous. Unfortunately LDA is not very much suitable for the description of biological systems since it encounters problems in the description of H-bonding.

A further improvement to  $E_{xc}$  include corrections involving the gradient of the density that have been shown to properly describe h-bonded systems (Carlioni and Roethlisberger 2000; Sulpizi and Carlioni 2000; Piana and Carlioni 2000). Becke (Becke 1988) has proposed a functional of the form:

$$E_x^B = E_x^{LDA} - \beta \int \rho^{4/3} \frac{\mathbf{x}^2}{(1 + 6\beta \sinh^{-1} \mathbf{x})} d\mathbf{r}, \quad \text{with : } \mathbf{x} = \frac{|\nabla \rho|}{\rho^{4/3}} \quad (2.42)$$

where the parameter  $\beta$  was obtained by fitting exchange energies values obtained from atomic Hartree-Fock calculations on noble gas atoms (Becke 1988).

While for the correlation energy we have used the parameterization introduced by Lee, Yang and Parr (LYP)(Lee, Yang, and Parr 1988).

$$E_c = -a \int \frac{1}{1+d\rho^{-1/3}} \left\{ \rho + b\rho^{-2/3} \left[ C_F \rho^{5/3} - 2t_w + \left(\frac{1}{9}t_w + \frac{1}{18}\nabla^2\rho\right) \right] e^{-c\rho^{-1/3}} \right\} d\mathbf{r} \quad (2.43)$$

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

### 2.2.3 Plane waves

The solution of the Kohn-Sham equations, requires an expansion of the wave function in terms of some finite basis set. A suitable formulation, for periodic systems is the use of plane wave basis set. In a crystalline array, for each Bloch function  $\psi_i^k$  corresponding to a band  $i$  and momentum  $\mathbf{k}$ , we can write:

$$\psi_i^k(\mathbf{r}) = \sum_{\mathbf{G}} c_i^{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (2.44)$$

where  $\mathbf{k}$  belongs to the first Brillouin zone of the crystal, and  $\mathbf{G}$  is a reciprocal lattice vector.

Because the non-local nature of the plane waves, usually a large amount of basis functions is needed. The use of plane waves has the advantage that the basis set does not depend on the atom geometry in the cell. Additionally, the use of fast Fourier transform algorithms made this choice computationally efficient.

Periodic Boundary Conditions, however, are usually not adequate to simulate biological complexes since they impose an artificial crystalinity on the system. In our calculations the procedure proposed by Barnett and Landmann (Barnett and Landman 1993) was followed. This method enforce non-periodic boundary conditions by expanding the one electron wavefunctions in plane waves within the cell with the constraint that  $\psi_i^k(\mathbf{r}) = 0$  for  $\mathbf{r}$  outside the cell and the boundary condition that both  $\psi_i^k(\mathbf{r})$  and  $\nabla\psi_i^k(\mathbf{r})$  are negligibly small at the cell borders.

### 2.2.4 Pseudopotentials

The dimension (and accuracy) of the basis set is determined by the choice of a maximum cutoff  $E_{cut}$  for the kinetic energy  $T^{PW}$  of the plane waves by the condition:

$$T^{PW} = \frac{1}{2} |k + G|^2 \leq E_{cut} \quad (2.45)$$

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

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

We used the ab initio pseudopotentials of Martins-Troulliers (MT)(Troullier and Martins 1991) type. This particular class fulfills the following requisites:

- i) Smoothness, e.g., they contain no nodes.
- ii) *Norm-conserving*, e.g., 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  $R_l^{AE}$

$$R_l^{PP}(r) = R_l^{AE}(r) \text{ for } r > r_{cl} \quad (2.46)$$

- iii) the valence pseudopotential eigenvalues are equal to the valence all electron eigenvalues;  $\epsilon_l^{PP} = \epsilon_l^{AE}$ .

- iv) *Charge-conserving*, e.g., the total charge enclosed within  $r_{cl}$  is equal to the all electron charge;

$$\int_0^{r_{cl}} |R_l^{PP}(r)|^2 r^2 dr = \int_0^{r_{cl}} |R_l^{AE}(r)|^2 r^2 dr \quad (2.47)$$

- v) at  $r = r_{cl}$  the pseudo wavefunction and its first four derivatives are continuous;
- vi) the pseudo potential should have zero curvature at the origin. The radial part follows the general form

$$R_l^{PP}(r) = \begin{cases} R_l^{AE}(r); & r > r_{cl} \\ r^l e^{p(r)}; & r \leq r_{cl} \end{cases} \quad (2.48)$$



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

where the seven coefficient of the polynomial  $p(r)$  are obtained by imposing iii), v) and vi). The local and non local parts of the pseudopotential can be fully separated using the Kleinman-Bylander approach.

### 2.2.5 Environment effects.

State of the art ab initio (DFT) calculation are computationally expensive and are limited to a relatively small number of atoms, typically in the order of 100. In biological systems, however, far charged moieties or solvent have important effects on the specific part one is interested in. The electrostatic potential  $\Phi$  generated by the protein/substrate/solvent on the ab initio complexes, which was calculated assuming a dielectric constant  $\epsilon_0 = 1$  as:

$$\Phi_{(r)} = \sum_i q_i \left( \frac{|r_c^n - r_i^n|}{|r_c^{n+1} - r_i^{n+1}|} \right) \quad (2.49)$$

where  $q_i$  are the atomic point charges from some force field parameterization, at the atomic positions  $r_i$ . The cut radius ( $r_c$ ) is adopted so as  $r_c > r_i - r_j$  for any atoms  $i, j$  of the system.  $n$  was set to 4. This form of the potential helps prevent the problem of 'spill out' of the electrons (Laio, Vande Vondele, and Roethlisberger 2001).

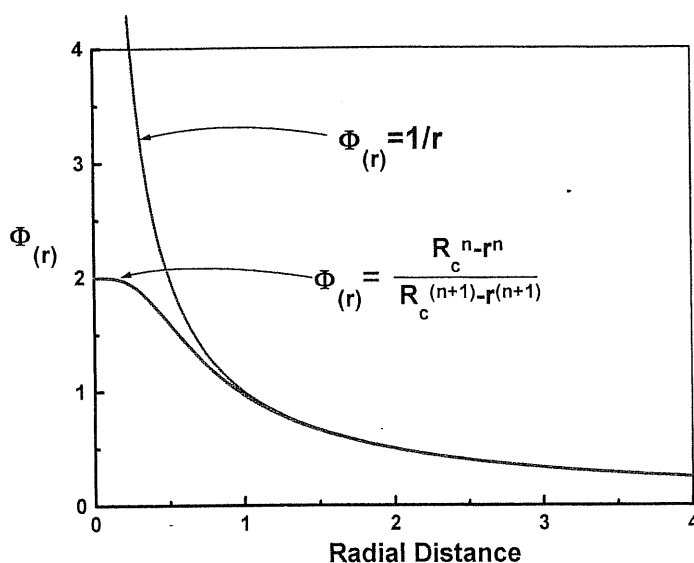


Fig 2.3: Radial dependence for modified Coulomb potential

## 2.2.6 Maximally localized Wannier Functions

The Wannier functions  $w_n(r - R)$  constitute an alternative representation to Bloch orbitals of the electronic states in a solid and are defined as a unitary transformation of the Bloch function  $\psi_n^k(r)$ . This representation has the advantage that recovers the very intuitive ideas of organic chemistry such as covalent bonds, bond order and electron lone pairs

For solid, the Wannier function associated with band  $n$  are written as:

$$w_n(r - R) = \frac{V}{(2\pi)^3} \int e^{-ik \cdot R} \psi_n^k(r) dk \quad (2.50)$$

Where there is an arbitrariness in the phase factor of each Bloch orbital is reflected in the non-uniqueness of the Wannier orbital definition.

Among the infinite possible sets of localized orbitals Marzari and Vanderbilt (Marzari and Vanderbilt 1997) showed that the centers  $\langle r_n \rangle$  of the maximally localized Wannier orbitals (WO) obtained by minimization of the functional:

$$\Omega = \sum_n \left( \langle r^2 \rangle_n - \langle r_n \rangle^2 \right) \quad (2.51)$$

provide an immediate and strikingly clear picture of electron localization. They furnish also both quantitative (Pauling 1988) as well as a pictorial description of chemical bonding.

## 2.2.7 Car-Parrinello Molecular Dynamics

The Car-Parrinello approach treat the electronic degrees of freedom as classical particles of mass  $\mu$ . The system composed by both the nuclei and electrons is described by a Lagrangian of the general form:

$$L = K_N + K_e + E_{\text{pot}} \quad (2.52)$$

where  $K$  are the kinetic energies of nuclei and electrons respectively, while  $E_{\text{pot}}$  correspond to the potential energy coming from both the ionic ( $\mathbf{R}_{\text{ion}}$ ) and electronic  $\Psi_i$  configuration.

The Lagrangian proposed by Car and Parrinello can be written in function of the one particle Kohn-Sham orbitals as:

$$L = \sum_i \frac{1}{2} \mu \int |\dot{\psi}_i(\mathbf{r})|^2 d\mathbf{r} + \sum_{\text{ion}} \frac{1}{2} M \dot{\mathbf{R}}_{\text{ion}}^2 - E[\{\psi_i\}, \{\mathbf{R}_{\text{ion}}\}] + \sum_{ij} \lambda_{ij} (\langle \psi_i(\mathbf{r}) | \psi_j(\mathbf{r}) \rangle - \delta_{ij}) \quad (2.53)$$

With  $\lambda_{ij}$  are the Lagrange multipliers associated with the constraint necessary to ensure orthonormality of the one electron wavefunctions during the dynamics.

$$\int \psi_i(\mathbf{r}) \psi_j^*(\mathbf{r}) d\mathbf{r} = \delta_{ij} \quad (2.54)$$

The corresponding equations of motion for both the electronic and the ionic degrees of freedom are obtained from the Euler-Lagrange equations

$$\begin{aligned} \frac{\delta}{\delta t} \frac{\delta L}{\delta \dot{\mathbf{R}}_{\text{ion}}} &= \frac{\delta L}{\delta \mathbf{R}_{\text{ion}}} \\ \frac{\delta}{\delta t} \frac{\delta L}{\delta \dot{\psi}_i^*} &= \frac{\delta L}{\delta \psi_i^*} \end{aligned} \quad (2.55)$$

$$\begin{cases} \mu \ddot{\psi}_i(\mathbf{r}) = -\mathbf{H}_{\text{KS}} \psi_i(\mathbf{r}) + \sum_j \lambda_{ij} \psi_j(\mathbf{r}) \\ \mathbf{M}_{\text{ion}} \ddot{\mathbf{R}}_{\text{ion}} = -\frac{dE[\{\psi_i\}, \{\mathbf{R}_{\text{ion}}\}]}{d\mathbf{R}_{\text{ion}}} \end{cases} \quad (2.56)$$

The  $\mu$  parameter is a fictitious variable that describes how electronic degrees of freedom evolve in time, the quotient  $M_{ion}/m$  characterizes how fast electronic variables propagate respect to nuclear variables.

The value of  $\mu$  is chosen such that  $\mu \ll M_{ion}$ . In this case the frequency spectra of ions  $\omega_{ion}$  and electrons  $\omega_i$  are separated.

$$\omega_i \propto \sqrt{\frac{1}{\mu}} \text{ and } \omega_{ion} \propto \sqrt{\frac{K_{ion}}{M_{ion}}} \quad (2.57)$$

where  $K_{ion}$  is the kinetic energy of the nuclei.

Under these conditions there is no energy transfer between the electronic and the ionic degrees of freedom and the resulting dynamic is adiabatic.

## 2.2.8 Calculated properties

### 2.2.8.1 Bond polarity indices

The centers of maximally localized Wannier orbitals (Silvestrelli et al. 1998), provide the so-called Bond ionicity indices (BI), which relate the difference in Pauling electronegativity between two atoms A and B forming a chemical bond (Alber, Folkers, and Carloni 1999) are defined as:

$$BI_{AB}(A) = \left( \frac{d_A}{d_{AB}} + \frac{r_A}{r_A + r_B} \right). \quad (2.58)$$

Where  $d_A$  ( $d_B$ ) are the distances between the atom A (B) and its WO center along the AB bond,  $d_{AB}$  is the bond length and  $r_{A(B)}$  the covalent radius of the corresponding atoms.

#### 2.1.1.1 $\Delta\rho$

The difference in electron density upon inclusion of a molecule in a complex, were calculated as (Gu, Kar T, and Scheiner 1999):

$$\Delta\rho = \rho_{whole} - \rho_p - \rho_{mol} \quad (2.59)$$

where  $\rho_{whole}$ ,  $\rho_p$  and  $\rho_{mol}$  denote the electron density of the whole complex, of the complex without the molecule mol respectively. The volume integral  $\Delta q = \int |\Delta\rho| d\tau$  provided a

quantitative estimation of the polarization effect due to the presence (absence) of the molecule within the complex.

## **2.2 *Experimental Methods:***

Here we briefly summarize the experimental methods used for the construction of wt and Tat mutants and the measuring of their corresponding biological activity. Experiments have been conducted by Dr. Mudit Tyagi at the Prof. Giacca's Laboratory of Molecular Medicine in the International Centre for Genetic Engineering and Biotechnology (ICGEB).

### **2.2.1 Site-directed mutagenesis experiment.**

#### **2.2.1.1 Plasmids.**

Starting from plasmid pcDNA3-Tat86, expressing the wt Tat protein of HIV-1 strain HXB2 (86 amino acids) from the CMV promoter in the pcDNA3 (Invitrogen, Carlsbad, CA) vector backbone, two mutated derivatives were obtained, bearing substitutions on glutamic acid at position 2 either with Gln (E2Q) or with Ala (E2A). All mutations were introduced by recombinant PCR techniques using appropriate primers pairs; the nucleotide sequence of the final constructs was verified by direct sequencing. Notice that a conservative mutation (Glu2→Asp) occurs on passing from HXB2 to HIV1Z2.

#### **2.1.1.1 Cell transfections and CAT assay.**

HIV-1 LTR transactivation assays were performed by transfection of HL3T1 cells, kindly donated by B. Felber (Felber and Pavlakis 1988), a HeLa derivative cell line containing an integrated LTR-CAT (chloramphenicol acetyl transferase) construct. HL3T1 cells were grown in DMEM and HAM'S F10 medium, respectively, supplemented with 10% fetal calf serum, 2 mM glutamine and 50 µg/ml gentamicin. Cells were transfected by the standard calcium-

phosphate procedure. All transfections were adjusted to the same content of transfected plasmid (and of CMV promoter sequences) by addition of the appropriate amounts of pcDNA3. CAT assays were performed 48 hours after transfections. All transactivation results represent the average values obtained in at least three independent transfections.

Biological activity of Tat and its mutant derivatives is evaluated by measured the enzymatic activity of CAT since the amount of mRNA coding CAT produced by the cell directly depends on the transactivation capacity of the Tat protein or its mutants.

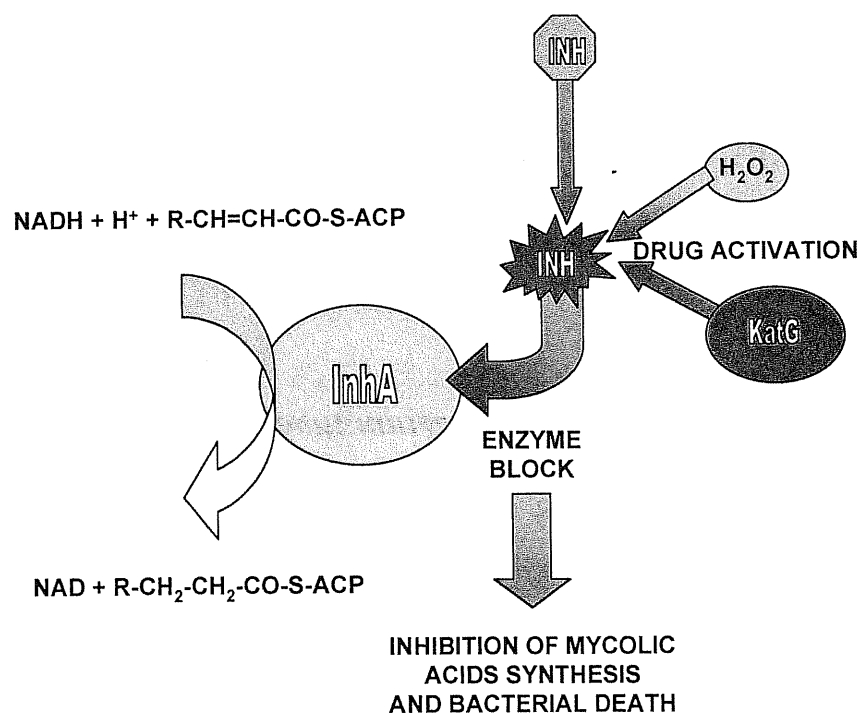
## Chapter III

### Enzyme ligand Interactions in drug resistant InhA from *Mycobacterium tuberculosis*

#### 3.1 Introduction: *Mycobacterium tuberculosis* and Isoniazid target

Among bacteria, the genus *Mycobacterium* is distinguishable by its complex layered structure composed by high molecular weight mycolic acids and waxes. They constitute a very resistant cell wall that protects the mycobacteria from acids, oxidative attacks and macrophage.

Isonicotinic acid hydrazide (isoniazid, Inh) has been a first-line chemotherapeutic agent in treating *Mycobacterium tuberculosis* aggression (Bloom and Murray C.J.L. 1992; Bradford and Daley 1998) from the early fifty's. Inh action mechanisms as well as processes conferring the bacteria Inh resistance are rather complex and not yet completely understood. Once isoniazid enters into the cell, it is first activated by the KatG protein, a catalase-peroxidase dual function bacterial enzyme (Reddy et al. 1992; Quemard et al. 1996; Rozwarski et al. 1998) and an electron sink (hydrogen peroxide, H<sub>2</sub>O<sub>2</sub>) (Shoeb et al. 1985; Quemard et al. 1996; Rozwarski et al. 1998). Subsequently, in its activated form (either an anion or a radical) Inh targets an enoyl-acyl carrier protein reductase (ACPR), the InhA enzyme (Quemard et al. 1995).



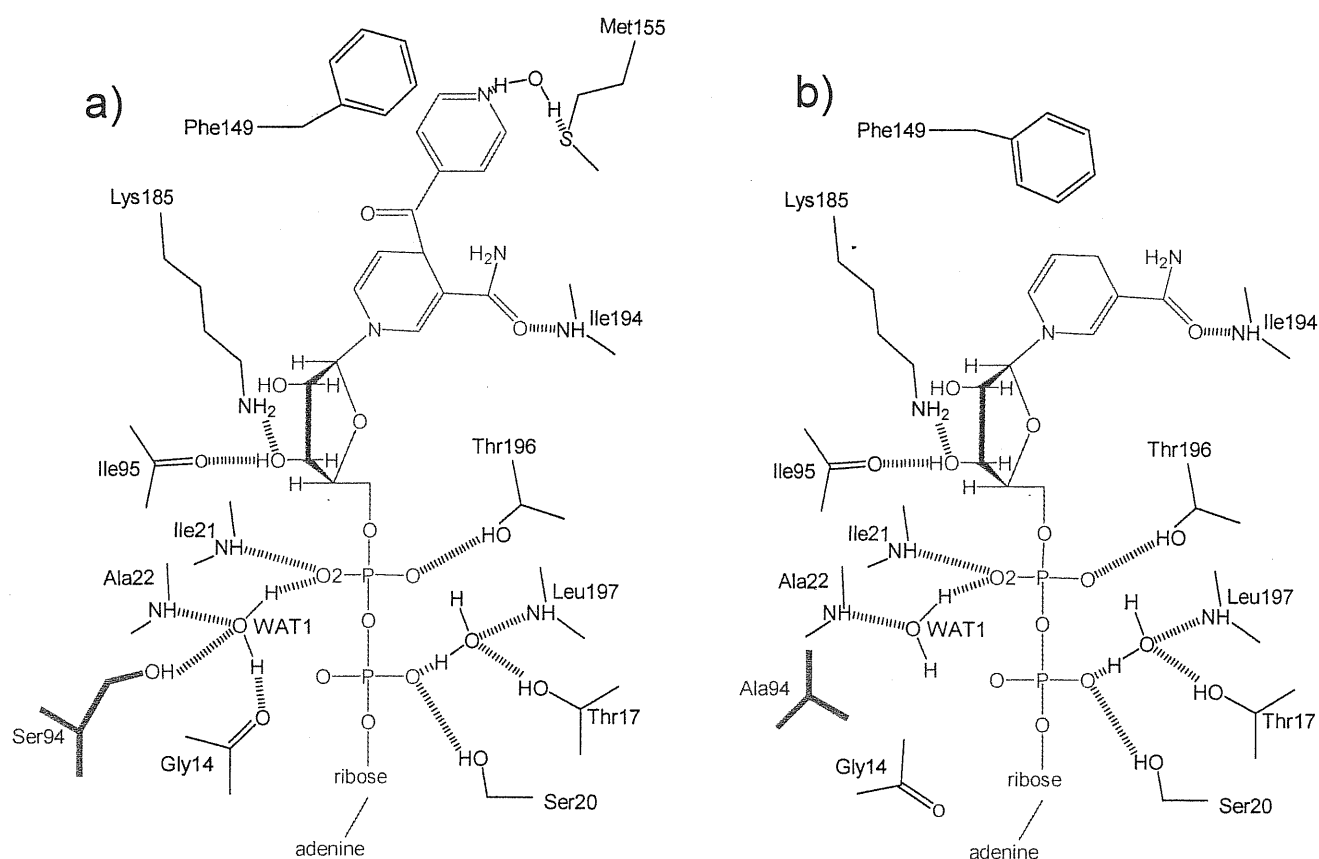
**Chart 3.1** Simplified cycle of action of isoniazid in *M. tuberculosis*

InhA catalyzes the reduction of elongating fatty acids chains linked to the carrier protein. The reaction products are then converted to very long chain  $\beta$ -branched fatty acids (mycolic acids) (Cole S.T. 1994), which are vital for cell wall building. Therefore, InhA inhibition lead to an increased vulnerability to external oxidative attacks and eventually to bacterial death (Chart 3.1).

Drug benefit, however, is severely limited by the emergence of resistance, that is the ability of specific mutants developed by the bacterium to overcome drug's action while retaining InhA enzymatic activity. The effect of InhA point mutations on drug action at the molecular level is difficult to clarify. Inh is covalently bound to the NADH cofactor within the InhA active site pocket as showed in Fig. 3.1

S94A mutant causes resistance towards the widely used drug isoniazid (isonicotinic acid-NAD, Inh•NAD). This mutation provokes the loss of two water-mediated H-bonds with the drug, as shown by the crystal structures of drug/substrate wt (Rozwarski et al. 1998) and substrate S94A (Dessen et al. 1995) complexes (Fig. 3.1).





**Figure 3.1.** Chemical structures of wt/Inh•NAD and S94A/NADH InhA (black) complexes (Adapted from (Rozwarski et al. 1998; Dessen et al. 1995)). The cofactor and the residue in 94 position are marked in green and magenta respectively. H-bonds are indicated in blue dashed lines. In wt/Inh•NAD (a),  $\beta$ -phosphate belonging to the substrate interacts via water molecule WAT1 (red) with Ser94, Gly14 and Ala22. In S94A (b), Ser94 and Gly14 H-bonds are lost and Gly14 backbone rearranges significantly.

However, the structural changes are accompanied by a decrease of affinity to NADH bind of only  $\sim 0.7$  kcal/mol (Tab. 3.I) (Dessen et al. 1995; Parikh, Xiao, and Tonge 2000), that is much less than what expected from the changes in the structural features.

Clearly, a theoretical description of the drug-target interactions would be of great value in explaining these apparently conflicting results.

**Table 3.I.** wt and S94 Kinetic parameters from(Dessen et al. 1995).

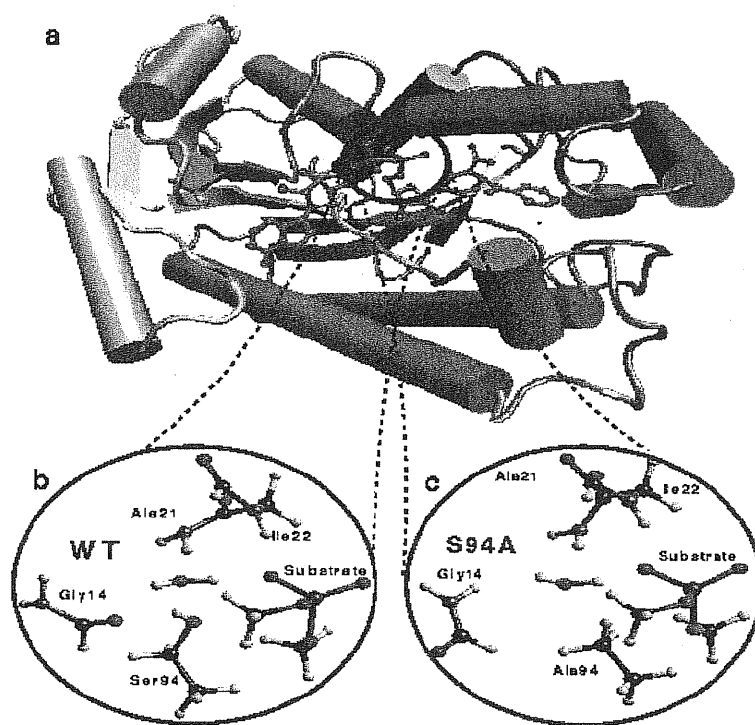
Kinetic parameters for wild-type and S94A mutant InhA measured by following NADH oxidation at 340 nm with a Uvikon 93310 spectrophotometer (Kontron Instruments). All reactions were done in 30 mM Pipes buffer (pH 6.8) at 25°C. Standard reactions for the determination of the  $K_m$  of NADH contained 10  $\mu$ M 2-trans-octenoyl-ACP, variable NADH (10 to 40  $\mu$ M), and 7.5 nM InhA. Standard reactions for the determination of the maximum velocity and  $K_m$  for 2-trans-octenoyl-ACP contained variable 2-trans-octenoyl-ACP (4 to 25  $\mu$ M), 100  $\mu$ M NADH, and 7.5 nM InhA.

Enzyme	$K_m(\mu\text{M})$		Vmax ( $\mu\text{mol min}^{-1} \text{mg}^{-1}$ )
	2-trans-octenoyl-ACP	NADH	
Wild-Type	2 $\pm$ 1	8 $\pm$ 0	2.2 $\pm$ 0.4
S94A	3 $\pm$ 1	38 $\pm$ 2	3.1 $\pm$ 0.6

## 3.2 Structural Models

### 3.2.1 Quantum Chemistry Calculations

Initial ab initio three dimensional models, complexes **I** and **II** (Fig. 3.2), were based on the X-ray structures of wt (Rozwarski et al. 1998) - and S94A (Dessen et al. 1995)- Inh•NAD complexes (PDB database entry codes: 1ZID and 1ENZ, respectively). The models focused on the region around 94 position; they included : i) Inh•NAD  $\beta$ -phosphate, which was modeled as dimethylphosphate.  $\beta$ -phosphate was assumed to be anionic as the NADH- $\beta$ -phosphate moiety is expected to be deprotonated (Saenger 1983) at physiological pH. (ii) Ile21, Ala22 and Gly14 backbone units. (iii) Ser94 (complex **I**, Fig. 3.3a) or Ala94 (complex **II**, Fig. 3.3c) for wt and S94A, respectively. (iv) Water molecule WAT1 (Fig. 3.1). Complexes **I** and **II** included the following hydrogen bond interactions: O2( $\beta$ -phosphate)–H3(Ile21), O2( $\beta$ -phosphate)–H2(WAT1), Ow(WAT1)–H4(Ala22), O(Gly14)–H1(WAT1) and Ow(WAT1)–H5(Ser94) (only in wt).



**Figure 3.2:** a) Cartoon representation of InhA enzyme (1ZID). Close up on the region of interest. Quantum models. b) wt (1ZID) and c) S94A (1ENZ)

Position constraints (De Santis and Carloni 1999; Piana and Carloni 2000) were imposed on DMP methyl carbons, on all carbon atoms in Ala22 and Ile21 fragments, on the backbone nitrogen in Gly14 and in the hydrogen atoms replacing carbons in Gly14 and Ser94 fragments. This was meant to mimic the presence of the protein and substrates frames.

### 3.2.1 Classical MD calculations

The initial structural model was the X-ray structure of the complex between wt InhA enzyme and Inh•NAD (Rozwarski et al. 1998). The water molecules present in the crystal structure were included. The complex was immersed in a 74.8 Å length truncated octahedron box containing 5812 and four Na<sup>+</sup> counterions, which neutralize the negative charge of the protein.

The Gromos force field (Scott and van Gunsteren 1995) and the SPC/E model (Berendsen, Grigera, and Straatsma 1987) were used for the protein/substrate and water, respectively.

### 3.3 *Computational Details*

#### 3.3.1 Quantum Chemistry Calculations

The quantum problem was solved within first principles density functional theory (DFT). Gradient-corrected exchange and correlation functionals of the Becke (Becke 1988) and Lee-Yang-Parr (Lee, Yang, and Parr 1988) (BLYP) type were adopted. The BLYP parameterization has shown to reliably describe structural and dynamical properties of biochemical systems (Carloni and Roethlisberger 2000). We used a plane-wave basis sets up to an energy cutoff of 70 Ry, the interactions between ionic cores and valence electrons being described by Troullier-Martins (Troullier and Martins 1991) pseudopotentials. The complexes were treated as isolated systems as in ref. (Barnett and Landman 1993). This computational setup has already been proven to describe hydrogen bonded systems (Silvestrelli and Parrinello 1999; Silvestrelli, Bernasconi, and Parrinello 1997) with high accuracy and it has also been extensively tested in the description of biological systems (Roethlisberger and Carloni 1999; Piana and Carloni 2000; De Santis and Carloni 1999; Roethlisberger et al. 2000; Roethlisberger et al. 2000; Piana and Carloni 2000; De Santis and Carloni 1999), including also phosphate binding proteins (Alber and Carloni 2000).

Geometry optimizations of hydrogen atoms and crystallographic water oxygen (WAT1) in complex **II** were performed using the method of direct inversion of the iterative sub-space up to an energy gradient of  $10^{-4}$  and  $10^{-6}$  for the ions and electrons respectively. The calculations provided the same H-bond pattern as that postulated by the crystallographers (Rozwarski et al. 1998; Dessen et al. 1995).

Density functional theory-based molecular dynamics (MD) simulations were carried out for complex **I** using the Car-Parrinello approach. A timestep of 0.121 fs and a fictitious electron mass of 500 a.u. were used. Since the initial configuration (**Ia** hereafter) evolves to complex **Ib** (see Figs. 3.3a and 3.3b and Results section), structural properties were calculated separately

for the two complexes by averaging the MD trajectory for the first 1.25 ps (**Ia**) and for the last 1.25 ps (**Ib**). Constant temperature simulations were achieved by coupling the system to a Nose (Nose 1984)-Hoover (Hoover 1985) thermostat at a frequency of  $500\text{ cm}^{-1}$ . 3 ps of MD were collected (Piana and Carloni 2000).

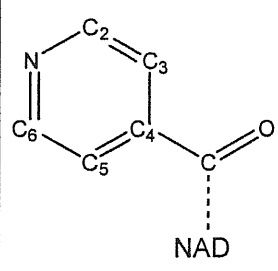
The calculations were carried out with the CPMD V3.0h program (Hutter et al. 1996).

The electrostatic potential  $\Phi$  generated by the protein/substrate/solvent on the ab initio complexes, which was calculated with the GROMOS96 atomic point charges (van Gunsteren et al. 1996).

### 3.2.1 Force field based Calculations

Classical MD simulations were carried out with GROMOS96 force field and GROMOS96(van Gunsteren et al. 1996) suite of programs. Periodic boundary conditions were applied. A cutoff of  $15\text{ \AA}$  was used for the non-bonded interactions. Reaction field corrections for the long-range electrostatic interactions (Tironi, Sperb, and van Gunsteren 1995) were used. All bonds were constrained using the SHAKE (Ryckaert, Ciccotti, and Berendsen 1977) algorithm. A time step of 2 fs was used. Constant temperature and pressure conditions were achieved through a Berendsen bath (Berendsen et al. 1984) with 0.2 ps and 1 ps relaxation times for temperature and pressure respectively.

**Table 3.II.** Partial Charge distribution for the INH segment used for the classical MD calculation.\* see Gromos96 manual (van Gunsteren et al. 1996), IAC = Integer Atom Code

	Atom Name	IAC*	Group Charge*	Partial Charge
	C	11	0	0.38
	O	1	1	-0.38
	C2	16	0	0.1
	C3	16	0	0.0
	C4	14	0	0.0
	C5	16	0	0.0
	C6	16	0	0.1
	N	8	1	-0.2

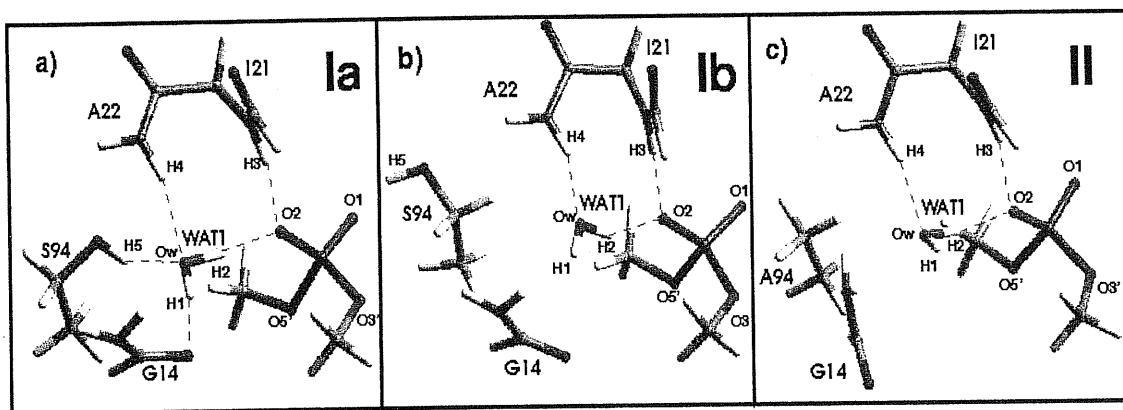
The system underwent energy minimization and stabilization by coupling the system to a 1 atm pressure bath and to a 150 K thermal bath by 0.5 ns and to a 300K bath by 0.1 ns. Finally, 1 ns MD simulation at 300 K and 1atm was performed.

Inh parameterization (see Table 3.II) was based on the existent force field of NADH available in the Gromos96 database(van Gunsteren et al. 1996).

### 3.3 Results

In this section, we describe the results from our ab initio MD calculations on the complex representing Inh•NAD/InhA interactions around the 94 position (Fig. 3.5).

The complex includes all the molecular interactions involving WAT1 Ser94. Subsequently, we compare our results with the structural properties of the S94A mutant, for which the X-ray structure is available(Dessen et al. 1995) (Fig. 3.5c). Finally, comparison is also made with classical molecular dynamics calculations on the entire protein bound to Inh•NAD (Fig. 3.4 ).



**Figure 3.3.** Quantum chemical models. Initial structure (a) and structure after 1.4 ps of ab initio MD (b) of wt/Inh•NAD. (c) S94A/NADH X-ray structure.

#### 3.3.1 Ab initio MD calculations on wt.

During the dynamics, the *trans-gauche* conformation of  $\beta$ -phosphate from the X-ray structure is well maintained (Table 3.III).

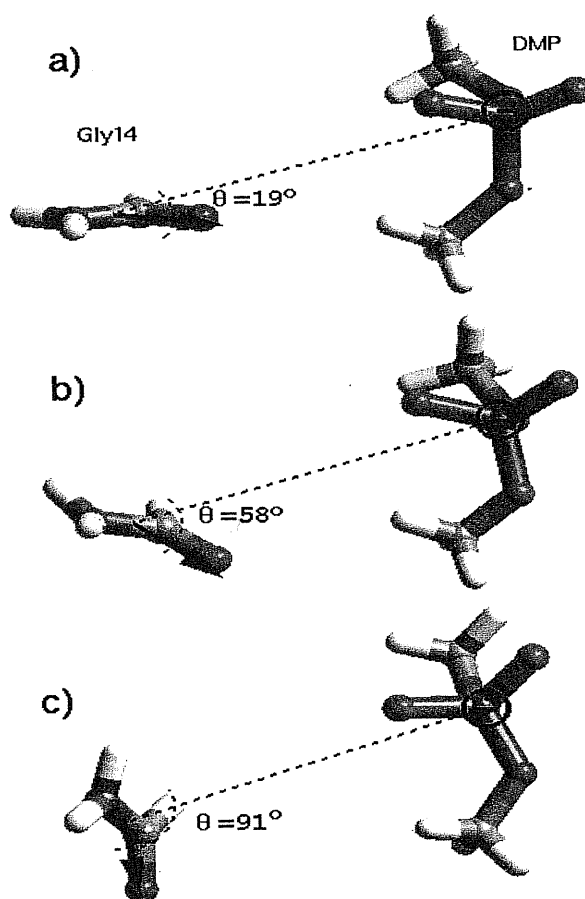
Differences between MD and X-ray structures are instead found for Ser94 side chain. During the first 1.4 ps (Fig. 3.5a), Ser94 hydroxyl points towards the  $\beta$ -phosphate oxygen O2 (Ser94 C $\alpha$ -C $\beta$  torsional angle  $\chi_1 = 171^\circ$ , complex **Ia**) and it H-bonds to water molecule WAT1, as being found in the crystallographic structure (Fig. 3.5a).

**Table 3.III.** Selected structural parameters of  $\beta$ -phosphate in complex **I** MD-averaged structure, and in wt(Rozwarski et al. 1998) and S94A (Dessen et al. 1995) InhA/Inh $\bullet$ NAD X-ray structures. Standard deviations are given in parentheses.

Bond Lengths [Å]			
	wt MD	X-ray wt	X-ray S94A
$d(\text{P-O3}')$	1.65(3)	1.59	1.58
$d(\text{P-O5}')$	1.66(3)	1.61	1.68
$d(\text{P-O1})$	1.49(3)	1.51	1.56
$d(\text{P-O2})$	1.53(3)	1.47	1.52
Angles [°]			
$\angle(\text{O3}'\text{-P-O5}')$	99(4)	101	102
$\angle(\text{O3}'\text{-P-O1})$	109(4)	112	100
$\angle(\text{O3}'\text{-P-O2})$	108(4)	101	112
$\angle(\text{O5}'\text{-P-O1})$	112(4)	109	107
$\angle(\text{O5}'\text{-P-O2})$	106(4)	108	107
$\angle(\text{O1-P-O2})$	120(4)	122	127
Torsional Angles [°]			
$\angle(\text{C-O3}'\text{-P-O5}')$	148(5)	119	156
$\angle(\text{C-O5}'\text{-P-O3}')$	-67(5)	-82	-76

The H-bond pattern is the same as that postulated on the basis of the X-ray structure of wt (Rozwarski et al. 1998).

Subsequently, the hydroxyl group rotates around the C $\alpha$ -C $\beta$  bond ( $\chi_1 = -71^\circ$ , complex **Ib**)(Fig. 3.5b). As a result, it neither points towards  $\beta$ -phosphate nor interacts with WAT1. This conformation is stable during the rest of the dynamics, suggesting that the barrier upon rotation of Ser94 hydroxy group is in the order of  $k_B T$  and therefore that this flipped conformation is energetically accessible at room conditions.



**Figure 3.4.** Gly14 dipole/phosphate charge interactions in **Ia**, **Ib**, and **II**. Although a quantitative estimate of this interaction is difficult (Leach 1996), it is clear that dipole/charge interactions are more favorable on passing from a) to c).

Loss of Ser94-WAT1 interaction leads to a rearrangement of the Gly14 backbone dipole so as to minimize its unfavorable dipolar interaction with the phosphate negative charge (Fig. 3.4). This dipole reorientation breaks a further hydrogen bond between Gly14 and WAT1 (Table 3.IV). Finally, WAT1-Ala22 H-bond is fully maintained during the dynamics (Fig 3.3 a-b) (Table 3.IV).

### 3.3.2 WAT1 Polarization.

WAT1 is highly polarized by the protein frame. In **Ia**, the electron density shifts from hydrogen to oxygen upon complex formation (Fig 3.5). Consistently, the difference in Pauling

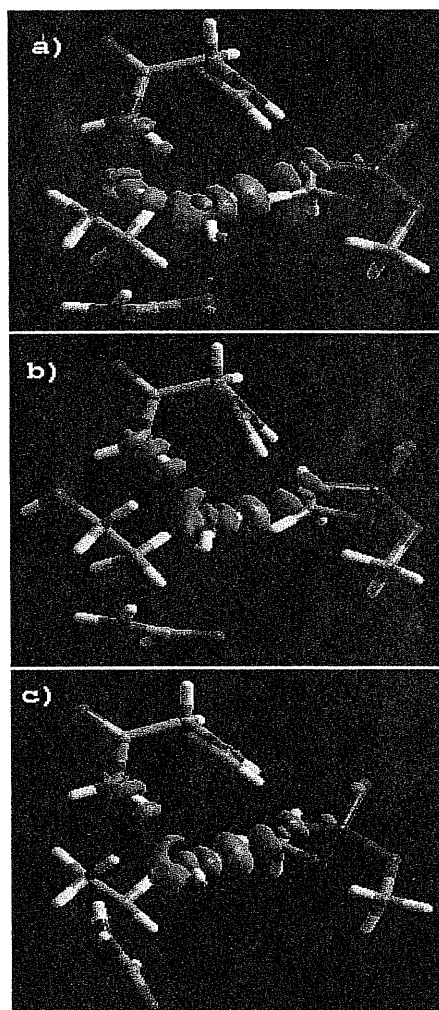


electronegativity ( $\Delta\chi$ ) between Ow and H2 atoms belonging to WAT1 is larger than that of water in the bulk and, more, in the gas phase (Table 3.IV).

**Table 3.IV.** Selected properties of the quantum-chemical complexes. Electronic properties are calculated from representative MD snapshots; structural data are averaged during the dynamics (Standard deviations are given in parenthesis).  $\Delta\chi(\text{O-H})$  is the difference in Pauling Electronegativity between O and H atoms in water in different chemical environments (Values in the presence of the protein field in waved brackets).  $\Delta\chi(\text{O-H})$  values in the gas and bulk phase are 1.2 and 1.5, respectively.  $\Delta q$  is the polarization upon formation of the complex in electrons.

	Ia	Ib	II
<b><i>POLARITY OF WAT1 OH BONDS</i></b>			
$\Delta\chi(\text{O-H1})^*$	1.4 {1.5}	1.4{1.5}	1.4{1.5}
$\Delta\chi(\text{O-H2})$	1.8 {1.7}	1.8 {1.7}	1.8{1.7}
<b><i>CHARGE POLARIZATION</i></b>			
$\Delta q$	0.43	0.34	0.36
<b><i>H-BOND CONTACTS</i></b>			
$d(\text{H1(WAT1)}-\text{O(Gly14)})$	2.5(4)	3.1 (5)	3.9
$d(\text{H2(WAT1)}-\text{O2(DMP)})$	1.8(2)	1.8(2)	1.62
$d(\text{H3(Ile21)}-\text{O2(DMP)})$	1.7(1)	1.8(5)	1.88
$d(\text{H4(Ala22)}-\text{Ow(WAT1)})$	2.3(3)	2.3(3)	1.99
$d(\text{H5(Ser94)}-\text{Ow(WAT1)})$	2.4(4)	---	---

Thus, WAT1 OwH2 bond is highly polarized towards the oxygen atom because of its interaction with the negatively charged phosphate group. Inclusion of the protein electric potential reduces the electric field by  $\approx 10\%$  on the region occupied by WAT1. As a result, the polarity of the bond is also reduced (Table 3.IV). However, the polarization of WAT1 is still larger than that of water in the liquid phase (Table 3.IV).



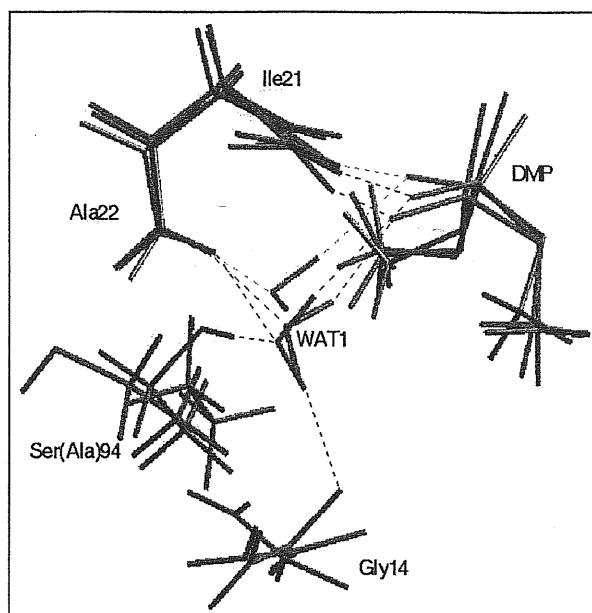
**Figure 3.5.** Isodensity contours of the electron density differences  $\Delta\rho$  in **Ia** (a), **Ib** (b) and **II** (c). Orange and light blue contour values correspond to  $3.5 \cdot 10^{-3} \text{ e/a.u.}$  and  $-3.5 \cdot 10^{-3} \text{ e/a.u.}$ , respectively.

WAT1 OH1 bond, which interacts with Gly14, is less polarized:  $\Delta\chi$ , in the presence of the protein field, is as polar as that of bulk water (Table 3.IV). Interestingly, the same features are found in the conformation **Ib** for WAT1, indicating that the dominant interaction is due to the negatively charged phosphate moiety.

### 3.3.3 Comparison with S94A.

The H-bond network of WAT1 and the orientation of the Gly14 backbone of complex **Ia** are different from those in the mutant (rmsd between backbone atoms = 0.5 Å) (Fig. 3.6). The structural discrepancies are associated to differences in chemical bonding: the polarization

effects on WAT1 and the rearrangement of the charge density upon formation of the complex are significantly larger in **Ia** than in **II** (Fig. 3.5 and Table 3.IV).



**Figure 3.6.** Comparison between the ab initio MD conformers of wt/ Inh•NAD (**Ia**, blue and **Ib**, red) with S94A/NADH X-ray structure (**II**, green).

**Ib** is instead structurally similar to **II** (Fig. 3.6), the rmsd between the backbone atoms being as small as 0.3 Å. In the two complexes, the H-bond pattern of WAT1 is the same and Gly14 conformational properties are similar: the loss Gly14-WAT1 hydrogen bond (Fig. 3.6) is counterbalanced by a reduction of repulsive Gly14 backbone dipole/ NADH-  $\beta$ -phosphate relative to **Ia**. Consistently, polarization effects and WAT1 electronic structure are similar (Figs. 3.6b, c and Table 3.IV).

### 3.3.4 Classical MD calculations on wt.

Our ab initio calculations suggest that Ser94 side chain can freely rotate during the dynamics. To investigate the barrier to the rotation in the presence of the protein frame, we have carried out a *classical* MD simulation based on the GROMOS96 force field (van Gunsteren et al. 1996) of the entire protein. The structure equilibrates well already within the first 0.2 ns, as shown by a plot of the rmsd as a function of simulated time; the overall folding is well-maintained during

the time-scale explored (1 ns), as it can be seen by a superposition of initial and final MD structures (Fig 3.7).

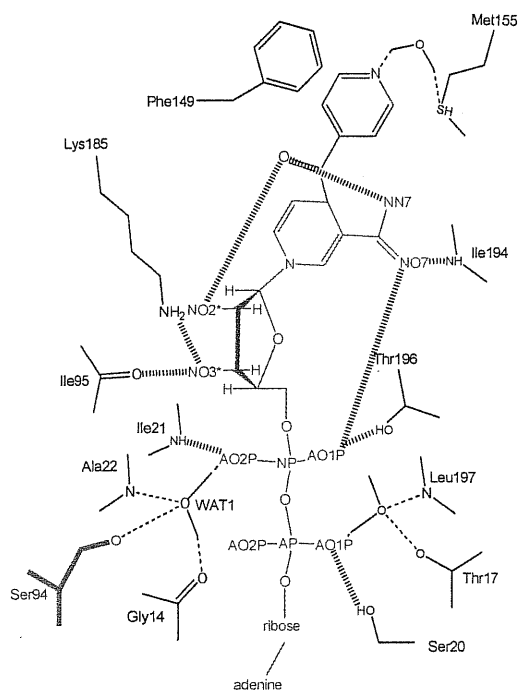


**Figure 3.7.** Classical MD on wt/Inh•NAD. Comparison between the initial, crystal structure and the last snapshot of the MD simulation (blue). The X-ray structure is showed with semitransparent secondary structure elements (helixes in red, sheets in blue, loops in green and substrate in gray).

During the dynamics, drug/substrate complex maintains most of its interactions with the protein (Table 3.V).

The most interesting feature is that during the dynamics Ser94 side chain is very flexible. In fact, It rotates its hydroxyl group around the C $\alpha$ -C $\beta$  bond few times in the timescale explored (Fig 3.8). As a result it the coordination of the ordered water in the enzymatic cavity that binds the NADH-  $\beta$ -phosphate is disrupted, breaking. By rotating, Ser94 OH group forms and breaks a weak H-bond with Ser13 carbonyl ( $d(\text{H-O}) = 2.2(0.5) \text{ \AA}$ ).

**Table 3.V:** Selected distances of Inh/NAD – InhA (Direct Inh-NAD / Protein Contacts as reported in fig. 4 ref (Rozwarski et al. 1998)). Inh is red, NAD is green, Protein and water molecules are black. Reported distances appear in blue dash line. Distances in Å, standard deviations are given in parenthesis. Atom names correspond to the figure.

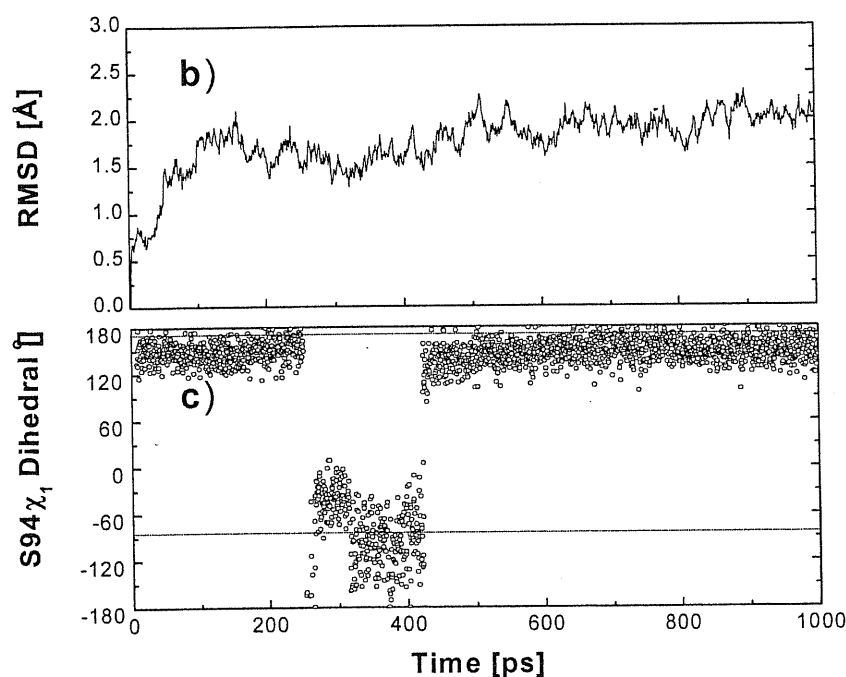


	MD SIMULATION	ROZWARSKI ET AL. 1998
AO1P - O $\gamma$ (Ser20)	2.6 (0.3)	2.8
NO1P - O $\gamma$ (Thr196)	3.4 (0.3)	3.5
NO2P - N(Ile21)	3 (0.2)	2.9
NO3* - O(Ile95)	3.5 (0.3)	3.2
NO3* - NZ(Lys185)	3.6 (0.5)	2.9
NO7 - N(Ile194)	3 (0.3)	2.9
NO7 - NO1P	4.5 (0.4)	3.7
O - NO2*	4 (0.3)	3.8
O - NN7	3.1 (0.3)	2.8

Values averaged over the last 0.8 ns of simulation

Gly14 is also highly flexible, its Ramachandran angles fluctuate around their average values ( $\Phi=90^\circ$ ,  $\Psi=4^\circ$ ) by more than  $20^\circ$ , however, there is no clear tendency pointing to a structural rearrangement of the Gly14 backbone while Ser94 is rotated.

This result further suggest that the barrier upon rotation of Ser94 hydroxy group is in the order of  $k_B T$  and therefore that this flipped conformation is energetically accessible at room conditions.



**Figure 3.8** RMSD Inha/Inh•NADH as function of time. c) Ser94  $\chi_1$  torsional angle plotted. Dashed lines correspond to the X-ray determined  $\chi_1$  torsional dihedral angles for 1ZID (upper,  $\chi_1 = 171^\circ$ ) and 1BVR, chain E (lower,  $\chi_1 = -84^\circ$ ).

### 3.4 Discussion

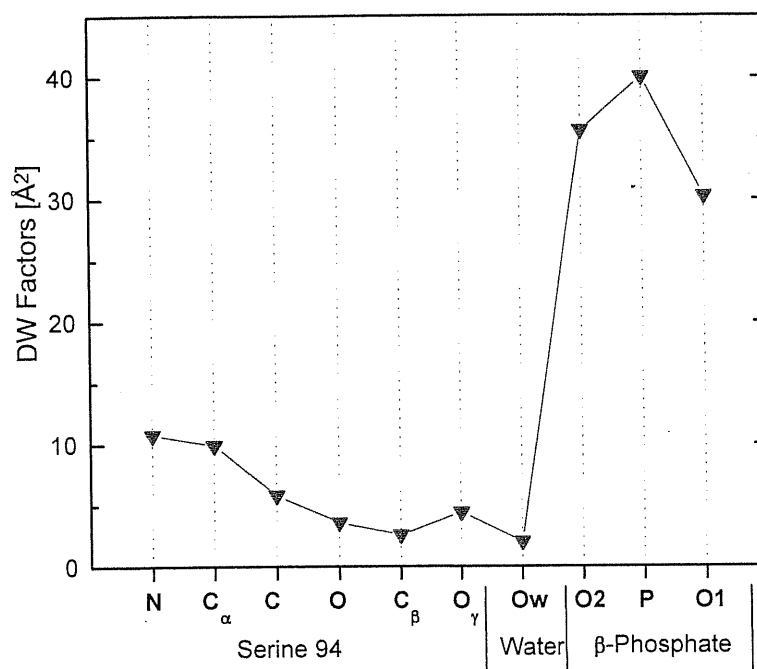
We have presented a theoretical study on the interactions between the inhibitor Inh•NAD and its target InhA at the 94 site (Fig. 3.1). Our ab initio MD calculations suggest that WAT1 water molecule, which mediates the interaction between Ser94 and the substrate, is highly polarized by the protein frame, its electronic properties being clearly different from those of bulk water (Table 3.IV). Classical MD calculations, which usually reproduce the properties of bulk water, may encounter difficulties to accurately describe the structural and dynamical properties of this water molecule and, more in general, waters confined in protein cavities (Alber and Carloni 2000; Cavalli and Carloni 2001).

Our *ab initio* calculations suggest that the Ser94 side chain is highly mobile: during the dynamics, the initial conformation **Ia** evolves spontaneously to **Ib**, where the water mediated Ser94/ Inh•NAD interaction is lost. In contrast to the X-ray structure (conformer **Ia**, Fig. 3.2a), the final MD structure (conformer **Ib**) turns out to exhibit similar structural and electronic properties of S94A mutant, which causes resistance to isoniazid (**II** in Fig. 3.2c). In particular, both complexes exhibit only one water-mediated H-bond between NADH and the protein frame.

It must be pointed out that, due to the very short time-scale investigated, we could not observe an equilibrium between the two conformers. Furthermore, the (relatively) small gas phase model may not account for all the interactions at the active site. Thus, the *ab initio* calculations alone might not be able to establish the high mobility of the serine residue. However, inclusion of the protein field does not appear to affect dramatically the properties of the complex. Furthermore, inspection of the X-ray crystal structure (Rozwarski et al. 1998) indicates that this rotation is sterically possible also in the presence of the protein environment; indeed this rotated conformation is experimentally observed in the InhA complexed with NADH (that is, the cofactor *not* bound to Inh, ref. (Rozwarski et al. 1999), PDB entry code 1BVR): in one of the six macromolecular complexes present in the asymmetric unit (chain E),  $\chi_1(\text{Ser94}) = -84^\circ$  (Fig 3.8). Finally, and most importantly, classical molecular dynamics simulation on the full protein provide an independent evidence for the high mobility of Ser94 side chain. Indeed, this residue is found to rotate few times in the time scale investigated (1 ns), forming and breaking a weak H-bond with Ser13 backbone. We notice that the frequency of the rotation events is much shorter in the *ab initio* calculations. This may result from the size of the model used in the quantum chemical calculations.

The mobility of Ser94 side-chain and of water molecule WAT1, which emerge from our calculations, is apparently not consistent with the extremely small Debye-Waller thermal factors (DW) reported for these two groups (entry 1ZID of the PDB (Rozwarski et al. 1998)). Here we argue that the reason of the discrepancy may be due to the crystallographic refinement procedure. This is illustrated by the following facts: (i) the value of the DW's reported for WAT1 ( $2.0 \text{ \AA}^2$ , Fig. 3.9) is the minimum default value of the program (XPLOR (Brunger 1992)) used during the refinement protocol (Rozwarski et al. 1998). (ii) The low values of Ser94 side chain are significantly larger than those of the side chain. (iii) The DW's along the hydrogen bonding network O $\gamma$  (Ser94) O $\omega$  (WAT1)-O2 ( $\beta$ -phosphate) exhibit no continuity as

it would be expected for a stable hydrogen bond. (iv) Ow's DW is smaller than those of O $\gamma$  and O2 (Fig. 3.9).



**Figure 3.9.** Crystallographic Debye-Waller thermal factors (DW) of selected atoms of wt X-ray structure (Rozwarski et al. 1998). WAT1 corresponds to HOH 404 in the PDB file.

In summary, our calculations suggest that two conformations of wt may exist at room temperature. One is the X-ray structure. The other, characterized by a rotation of Ser94 side chain and, probably, a small conformational change of Gly14, is structurally and electronically similar to S94A.

These findings may help to explain the very small difference in free binding energies of NADH binding to wt and to S94A. Indeed, our calculations suggest that Ser94- and Gly14 - WAT1 interactions are weaker than what one can expect from the X-ray structure, as they can form and break during the dynamics. Furthermore, the loss of the H-bonds may be partly compensated by a decrease of Gly14 dipole/phosphate charge repulsive interactions. Hence, the difference in energetics of binding between wt and S94A are expected to be rather small.

It would be highly interesting to test our proposal by expressing the S94T mutant and measuring its affinity for the substrate. As threonine is bulkier (and hence less mobile) than serine, we expect this mutant to have more affinity for NADH than wt.



## Chapter IV

### Structural and Dynamical Characterization of Tat protein.

#### 4.1 Introduction: *Human Immunodeficiency Virus Type 1 (HIV-1)*

The HIV retrovirus (Fig 4.1) was identified in the early '80 as the cause of the Acquired Immuno Deficiency Syndrome (AIDS).

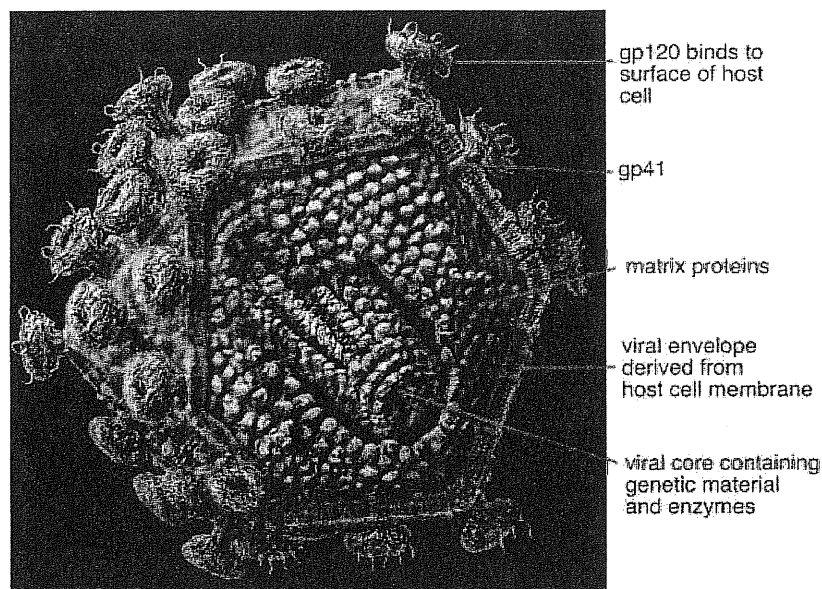


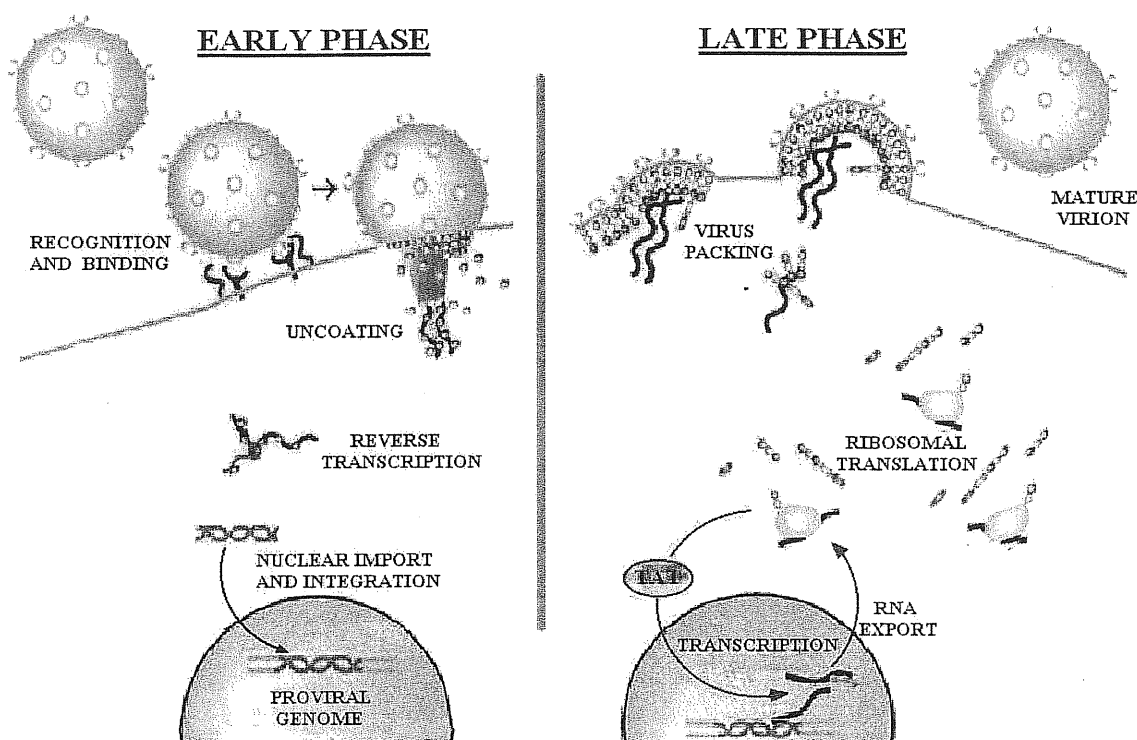
Fig 4.1 Structure of the HIV-1 virus

It is an encapsulated virus, that contains two identical copies of RNA inside the capsid that code for fifteen different proteins (Fig.4.2); among them only a few are essential for virus life



Early in the study of HIV, it was observed that a protein was required for activate viral transcription, and hence for efficient production of viral proteins (Sodroski et al. 1985). Differently to other transcription factors, Tat (TransActivator of Transcription) protein of the HIV-1 acts directly on RNA in stead of on the DNA promoter. In absence of Tat the product of RNA Polymerase II tend to be short transcripts and in any case they are produced in very reduced amounts (see Fig 4.3). This small amount of mRNA is exported to the cytoplasm where it is translated<sup>2</sup> into viral proteins, and in particular for Tat. Once generated, Tat is imported into the cell nucleus where it recruits a **Positive Transcription Elongation Factor** complex (pTEF-b). Once the complex Tat-pTEF-b is formed, Tat binds a nascent viral mRNA hairpin called **TransActivation Response Region (TAR)**. The formation of this complex positions pTEF-b near the **Carboxy Terminal Domain (CTD)** of the RNA Polymerase II which is phosphorylated, drastically increasing the production rate of mRNA establishing a efficient synthesis level. The resulting mRNA is again exported to the cytoplasm and more Tat is generated, imported into the nucleus and used to further amplify the mRNA production rate giving place to a sort of self-catalytic process. Production of Tat is the main factor separating the early from the late phase in viral infection (Fig 4.3).

58



**Figure 4.3.** Tat protein in the replication cycle of the HIV-1. Left: Early stage in infection leads to leading to virus entry and integration of proviral genome into the host genome. Right: Efficient viral transcription requires Tat, which is synthesized at very low levels during the early phase. Once imported to the nucleus, it generate a sort of positive feedback process that increasingly stimulates further rounds of transcription.

Tat is a relatively small protein composed of 86 or 101 amino acids (aa) and is expressed by two exons in the HIV-1 genome (Fig 4.2). No crystal structure of Tat has been obtained up to date, structural prediction and data from NMR spectroscopy (Bayer et al. 1995) indicate that the protein has a highly flexible structure and does not exhibit clear-cut secondary structure elements.

On the basis of amino acid distribution in the sequence, and of their conservation with homologous proteins from other viruses (Carroll, Martarano, and Derse 1991; Derse et al. 1991), it is possible to identify six regions (Figure 4.4), although this regions are frequently referred as domains this term is not rigorously correct since it has not been proved that they constitute independent domains of motion.



**Fig 4.4.** Primary sequence and ribbon representation of the Tat protein backbone. Coloring code as indicated in the text.

The first exon (aa 1-72) contains an **N terminal** (NT, Residues 1-21, grey color) domain, rich in Asp groups, of which only Asp 2 is structurally important (Reddy et al. 1992) as it binds to positive residues of the basic domain. The domain is rich also in prolines. (ii) The solvent exposed **Cysteine Rich Domain** (CRD, Residues 22-37, yellow) is thought to be involved in Zn binding and interaction with the CyclinT1 protein (Frankel, Brecht, and Pabo 1988). (iii) the highly conserved **Core region** (Residues 38-48, green) is rich in hydrophobic residues crucial for Tat transactivation (Kuppuswamy et al. 1989). (iv) The highly conserved **Arginine Rich Domain** (ARD, Residues 49-57, red) is rich in arginines and lysines. It is present in a rather extended conformation (Gregoire and Loret 1996). (v) The **Glutamine Rich Domain** (QRD, Residues 58-72, blue), located in between ARD, Core and CT, is supposed to provide structural stabilization to the protein. (vi) The **C terminal** (CT, Residues 73-86, pink) is expressed by the second exon, which is totally solvent exposed, appears to not be important for transactivation.

Because of the flexibility of the protein, it may be possible that the *dynamical* properties (as opposed to the *structural* ones) play a role for the intra- or interdomain interaction that could result in a coordinated motional fashion in some regions of the molecule.

Given the importance of Tat and in particular of its basic domain, it provides great scope for therapeutic intervention in HIV-1 infection. Despite the pharmacological potentiality of this

protein, very few structural data are available in literature. In this chapter, we use molecular dynamics calculations with analysis of essential modes and intramolecular electrostatic energy to address the fundamental issue whether these domains can be characterized on the basis of their dynamical behavior. Our simulation establish a simple correlation between some of the above described regions and motional domains. Furthermore, we identify crucial residues essential to maintain protein structure as position 2. In particular, we confirm this prediction by experimentally testing the effects of amino acid substitution at this position. Experimental procedures and methods are detailed in Apendix I

### 4.3 *Structural model*

We used as initial structure the NMR spectroscopy determined structure of HV1Z2 Tat (human immunodeficiency virus type 1, Zaire isolate 2, aa 1-86,PDB entry 1TIV (Bayer et al. 1995). The model used in these calculations correspond to the first out of ten structures deposited in the PDB data bank<sup>3</sup> which better reproduces the NOE constraints. All protonation states are considered as found in the NMR structure(Bayer et al. 1995). No disulphide bonds have been assumed since experiments where performed in reducing conditions so as to prevent dimerization. Cl<sup>-</sup> counterions were added to neutralize the protein positive charge (+11). The TIP3P (Jorgensen et al. 1986) water model was used as explicit solvent. The model was immersed in a water box of ~ 70 x 55 x 45 Å edges. The final system comprised ~ 17,000 atoms.

### 4.4 *Computational Details*

MD calculations were performed with the SANDER module of AMBER 5.0 (Case et al. 1997) suite of programs. Standard AMBER'94 (Cornell, Cieplak, and Bayly 1995) force field parameters was used for Cl<sup>-</sup> and all the protein. The cutoff for van der Waals parameters was set to 10 Å. The long-range electrostatic interactions were calculated using the Ewald particle mesh summation (Darden and York 1993). A dielectric constant equal to 1 was adopted.

---

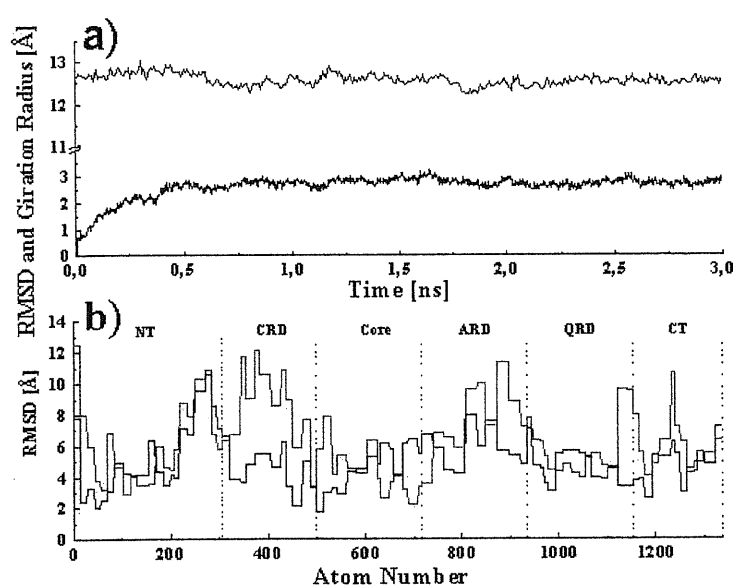
<sup>3</sup> No average structures has been deposited.

During dynamics all bonds were constrained by using the SHAKE (Ryckaert, Ciccotti, and Berendsen 1977) algorithm. The integration time step was set to 1 fs. Simulations at room temperature and pressure were achieved by coupling the systems to a thermal and a pressure Berendsen's baths (Berendsen et al. 1984).

The systems underwent first, energy minimization; then, solvent and counter ions were allowed to move for 0.02 ns. Subsequently, the systems were gently heated up from 0 K up to 300 K and achieved pressure stabilization during 0.12 ns. Finally, 3 ns of NPT MD simulation at 300 K and 1 atm pressure were performed on WT and 5 above quoted mutant proteins.

## 4.5 Results

The rmsd in function of time stabilizes after about 1.5 ns of simulation and reaches a maximum value of about 2.9 Å (Fig. 4.5). Thus, the structure shows, as expected, a high degree of flexibility because of the absence of secondary structure elements. During the stabilization time the MD structure shows a slight tendency to adopt a more compact conformation than the NMR structure, as reflected by the small decrease in the gyration radius of about 0.4 Å.



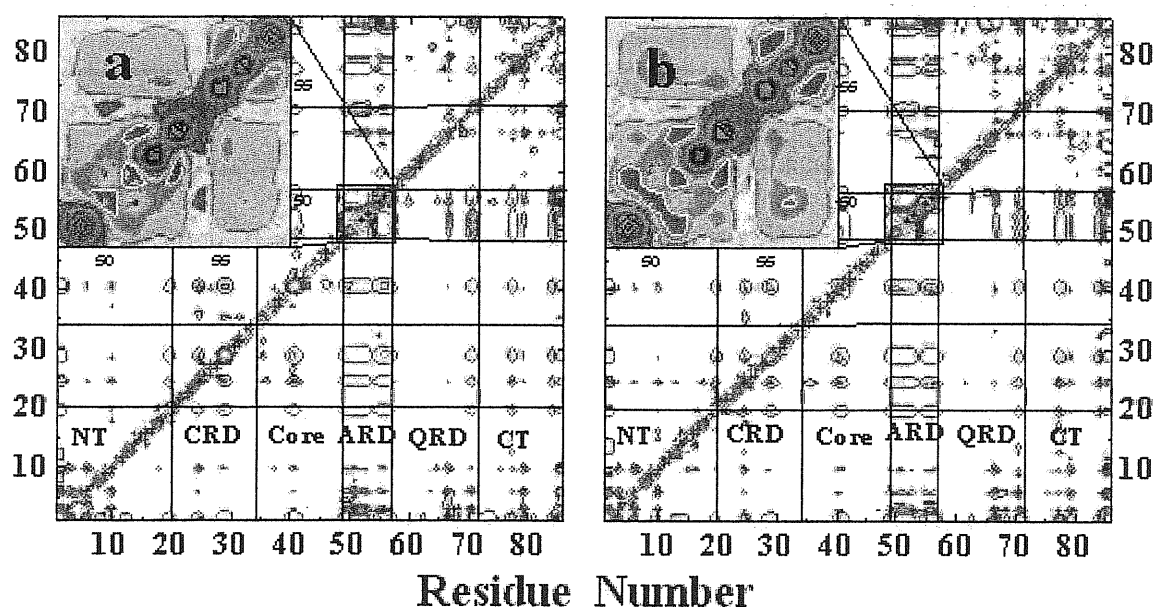
**Figure 4.5.** (a) Gyration radius (up) and root mean square deviation (down) of Tat protein as function of time, (b) rmsd calculated: per residue along 3 ns dynamics (black) and calculated among the 10 structures of the NMR family (red).

Comparison of the rmsd calculated per residue both, during dynamics and among the NMR family of structures (Fig. 4.5b), indicate that the mobility displayed the protein trajectory is compatible with the experimental determination. Specially if we consider that the pair wise calculated rmsd on the backbone atoms of the NMR family is 6.6 Å.

#### 4.5.1 Electrostatic analysis

In this paragraph we analyze the intraresidue electrostatic interactions, and in particular of the ARD, in presence of the explicit solvent.

Due to the elevated charge of the molecule (+11 e<sup>-</sup>) and to the very high density of charged residues (in average, 1 out of 4 residues is charged) electrostatics is by far the dominant interaction. Drawing of the electrostatic profile in a three dimensional map allow for a quick identification of the hottest spots of structural stability. Fig. 4.6 depicts the electrostatic inter residue interaction energy map. While diagonal elements provide the internal electrostatic energy of the individual residues, off-diagonal elements, instead, provide electrostatic interactions between charged residues.



**Figure 4.6.** Electrostatic interaction energy maps per residue at time = 0 (a) and time=3 ns (b). Color ranges continuously according to the spectrum from blue (negative) to red (positive). The scale has been normalized to the maximum and minimum peaks. The fill contour on the corners are close ups of the basic domain highlighting very strong intradomain interactions.

Comparison between initial and final configurations ( $t = 0$  and 3ns minimized, Fig. 4.6 a-b respectively) indicates that, despite the high flexibility displayed by the protein during the dynamics, the overall energetic profile remains almost unaltered. Differences consists mainly in a shallowing of the positive (repulsive) peaks due to a rearrangement in the disposition of the lateral chains and an improving in the solvation of charged residues upon dynamics as can be seen, for instance, in the interaction between Lys71 and Lys51 and 53. Very strong stabilizing interactions are verified between: (i) the first acidic residues Asp2, Asp5 (in the NT) and Lys51, Arg53 (ARD), (ii) Asp67 (QRD) and Arg55, 57 (ARD), (iii) Glu86 and Arg55, 57, (iv) Lys71 and Asp5, (v) Met1 and Asp67. Furthermore, and surprisingly, the most depleted areas in the map correspond to interactions with both CRD and Core region due to its low charge content. As counterpart, ARD intra e interdomains are the most pronounced. The close ups on the ARD evidence the high repulsion between basic residues is reduced by disposing the side chains in an alternated fashion, perpendicular to the backbone and highly extended, so that hydration is maximized.

**Table 4.I.** Hydration of basic residues within the ARD.

Residue	Arg49	Lys50	Lys51	Arg52	Arg53	Arg55	Arg56	Arg57
Hydration	5.59	5.54	4.14	6.12	4.64	4.26	4.72	5.59

Table 4.I shows the hydration number around the residues of the basic domains averaged over the last half of the simulation time. Lys51, Arg53 and Arg55 are less hydrated since they form strong hydrogen bond interaction with acidic residues (Table 4.II). The most hydrated residue in this domain is Arg52, which is found always solvent exposed.

#### 4.5.2 Motional analysis:

We turn now our attention to the study of the motion features exhibited by Tat during the simulated time.

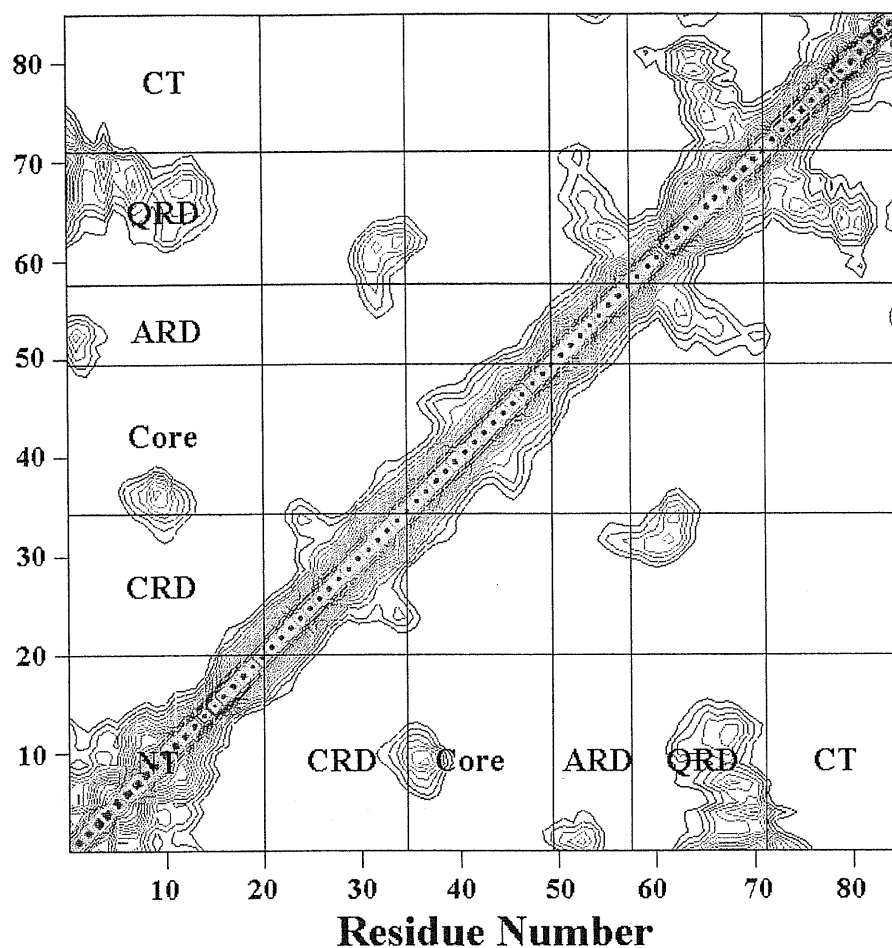
In spite of the lack of secondary structure elements, it is impossible to establish a priori from the structure whether the motion of a given residue in the chain can exhibit a (anti)correlated motion with non neighboring residues or not. To this propose, it results necessary the knowledge of the dynamic evolution of the system.



**Table 4.II.** Hbond distances for Met1 and ARD residues. (Occupancies signaled with 100\* indicate that two Hbonds are formed). Hbonding is calculated for time > 1.5 ns. Only occupancies > 20 % are reported.

H Donor - H Acceptor	Av. dist. [Å].	% occupancy
Met1 (N)-Asp67 (O)	2.9	73
Met1 (N)-Ile69 (O)	2.8	95
Met1 (N)-Lys71 (O)	2.8	93
Arg49 (NH1)-Tyr47 (O)	3	34
Lys51(NZ)-Asp2 (O)	3.1	33
Lys51 (NZ)-Asp2 (Oδ1-2)	3	98
Arg53 (NH1-2)-Asp2 (Oδ1-2)	3.2	48
Arg53 (NH1-2)-Asp67 (Oδ1-2)	3.3	67
Arg53 (NH1)-Gln72 (O)	3.1	42
Gln54 (Nδ)-Ile45 (O)	3.9	82
Arg55 (N)-Arg53 (O)	3.1	44
Arg55 (Nε)-Gln60 (Oε1)	3.1	21
Arg55 (NH1-2)-Gln60 (Oε1)	3.1	41
Arg56 (N)-Gln64 (O)	3.2	21
Arg56 (NH1-2)-His33 (Nε2)	3.1	42
Arg56 (NH1-2)-Gln54 (Oε)	2.9	20
Arg57 (NH1-2)-Gln60 (Oε)	3.1	40

In addition, Calculation of the covariance matrix along dynamic simulation of proteins have been shown to be a very clarifying and powerful tool to rationalize their collective motion.

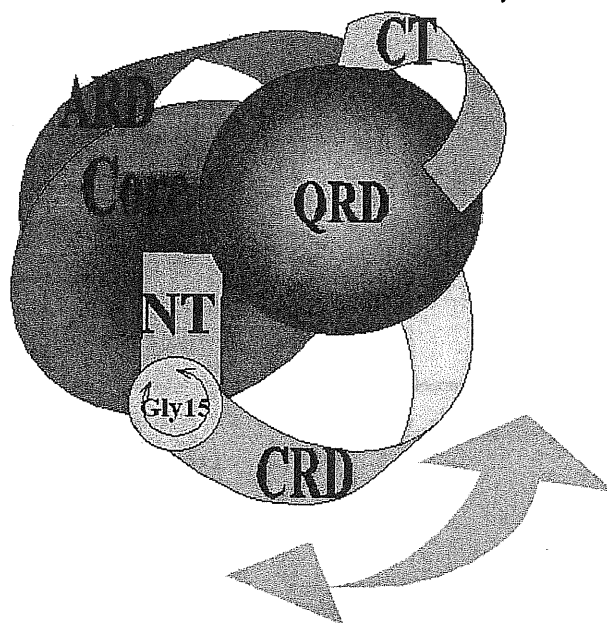


**Figure 4.7.** Covariance matrix map of  $C_{\alpha}$  atoms over 3ns of simulation time. Color code varies according to the spectrum continuously from blue (correlation= 0.25) to red (correlation=1)

The covariance matrix of the  $C_{\alpha}$  carbons calculated over the 3 ns of simulation is shown in Fig. 4.7 The map shows that the correlations among the domains is not pronounced: most off-diagonal peaks are lower than 0.6 with a complete absence of significative anticorrelated displacements, probably due to the absence of secondary structure. Interestingly, it is possible to recognize several regions of correlation map that approximately correspond with the main peaks of the electrostatic interaction as are the intersection area of NT - ARD, NT - QRD, ARD - QRD and ADR - CT.

NT is the only domain exhibiting intradomain correlated motion. Indeed, NT exhibits two clearly distinct regions. The motion of the first 14 aa are fairly correlated whereas the remaining 7 are not correlated. The rational for this behavior is simply that the first part of NT lies buried between the Core and the QRD and is, in addition rich in prolines that limit the

backbone degree of freedom. In addition, Gly15 acts as a pivot for the rest of the region (Fig 4.8), that is solvent exposed and highly flexible as can be seen from Fig. 1b. Other important interaction leading to correlated motion are the hydrophobic interaction between Trp11 (located near the geometrical center of the protein) with Val36, Phe38 and Pro68 (Avg. C $\alpha$ -C $\alpha$  distances 4.7, 6.9 and 6.1 Å respectively) and Ile8 with Pro68 (4.8 Å) and Pro6 with Pro70 (6.4 Å).



**Figure 4.8.** Cartoon representation of Tat indicating CRD mobility.

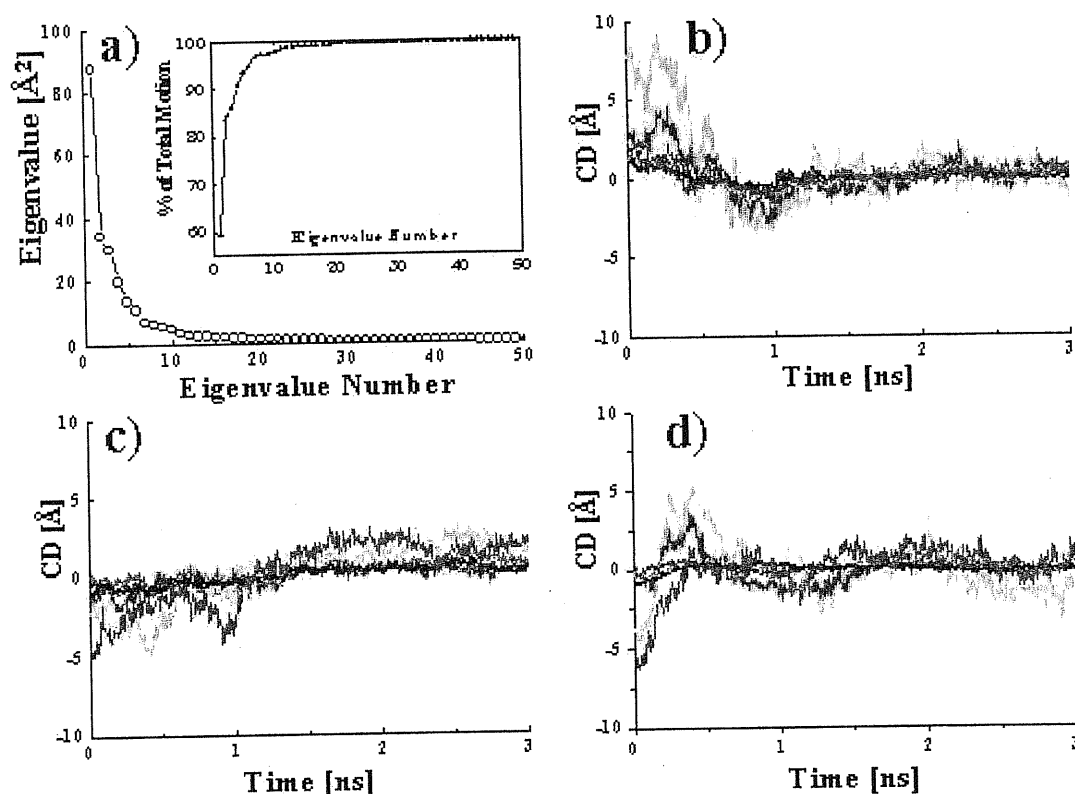
The most correlated region appears to be QRD which exhibits correlation maximums corresponding to regions spanning all the regions in the protein. CT tail exhibits correlated motion only with its neighboring region QRD and a shallow maximum corresponding to the strong interaction of Glu86 with Arg55 and Arg57 within the ARD (Table II).

The large-scale motion of the protein can be followed by an analysis of the essential vectors (Amadei, Linssen, and Berendsen 1993). It offers a very powerful tool to discriminate between effective displacements during dynamics and unimportant fluctuations and/or oscillations around equilibrium positions.

Fig. 4.9a shows a good convergence of the eigenvalues indicating that just the firsts eigenvectors are enough to describe the main motional features. It turns out that the first three

eigenvectors span more than 85% to the total motion (Fig. 4.9a inner region), thus we confine our to these three modes.

Fig. 4.9b-d plots the contribution of the different domains to the corresponding three first eigenvectors.



**Figure 4.9.** Essential dynamics analysis. Eigenvalues convergence (a) and % of total motion spanned by the corresponding eigenvectors (inner frame). (b) and (c) Projected contributions per domains for first and second essential eigenvectors respectively. Colors according to Fig. 4.4.

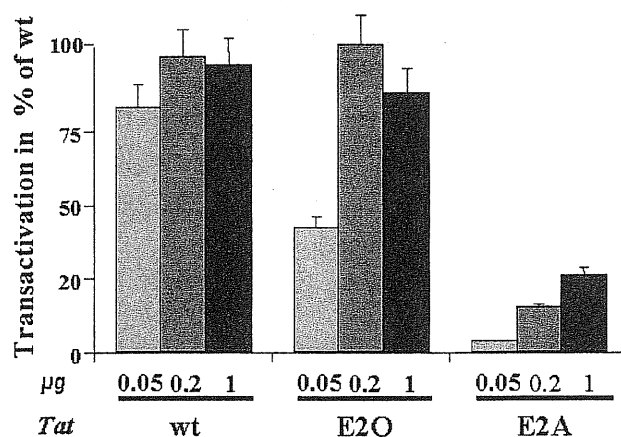
Under the light of the results of the covariance matrix analysis, we shifted here the NT-CRD border region from position 21-22 to 14-15. This domain redefinition shows a clear cut difference between the contributions of the NT tail and CRD to the corresponding eigenvectors. The largest motional contribution corresponds to the redefined CRD. In second place, we find an important contribution of the Core region to the total motion mainly attributed (Bayer et al. 1995) to the presence of three Gly residues (42, 44 and 48) in this region. In third place we find the CT. Finally, the relatively most rigid domains are the

shortened NT tail (see discussion above), ARD possibly because of its high charge density, and QRD.

Projection of the second and third essential vector on the real displacement vector (Fig. 4c and d) show a similar fashion to the first one, but with a much more important mobility of the CT domain.

### 4.5.3 Site directed mutagenesis experiments

Under the light of the above presented theoretical results we have been able to identify critical points for structural stability that permitted us to anticipate the impaired biological activity of a point mutant. Asp2 seem to be the only negative residue within the NT region make a salt bridge with both Lys51 and Arg53 in the ARD for long periods during the time scale investigated. Electrostatic interaction and correlation maps indicate that a negative residue at position 2 (unlike Asp5 and Glu9) is crucial for keeping the NT domain bound to the ARD and the rest of the protein. In fact, this residue is frequently mutated to Glu among several isolates, speculatively to facilitate this interaction, since glutamic acid has a larger side chain. To test the predictive power theoretical methods, we expressed and measured the biological activity of HXB2 Tat mutants at position 2, Glu→Ala and Glu→Gln.

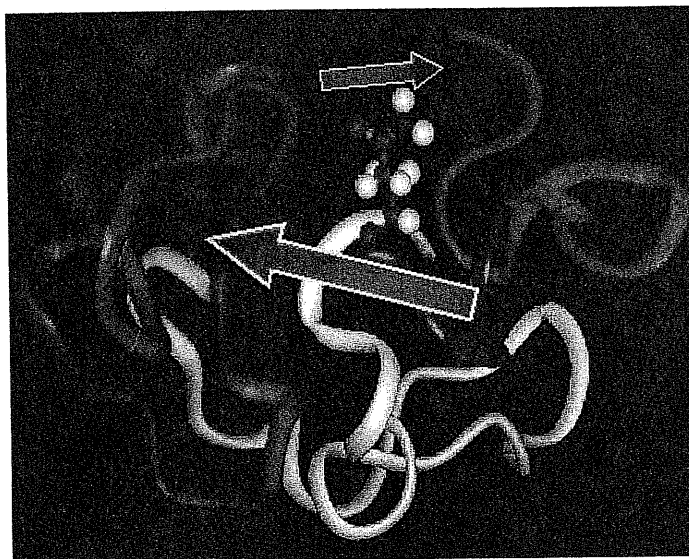


**Figure 4.10.** Transactivation of Tat mutants at position 2. HL3T1 cells (containing an integrated LTR-CAT construct) were transfected with the indicated amounts of plasmid constructs expressing wt and Tat mutants at position 2. CAT assays were performed 48 hours after transfection. Shown are means and standard deviations of three independent transfections.

Notice that a conservative mutation (Glu2→Asp) occurs on passing from HXB2 to HIV1Z2.

As predicted, the E2A mutant shows a strikingly reduced activity, while the transactivation levels supported by E2Q are not completely abolished (Fig. 4.10).

We hypothesize that, since Glutamine has a CO moiety, it must still hbond Lys50 and/or Arg53. Replacement of an amine group instead of O<sub>δ</sub> carbonyl oxygens in the aspartic acid gives a much more favorable charge-dipole and dipole-dipole interaction for a Gln oriented with the O<sub>δ</sub> towards the basic domain.



**Figure 4.11.** Most stable configuration for the E2Q mutant. The relative position of the Gln2 side chain. CO moiety is oriented towards the very positively charged ARD (red). Arrows indicate the favorable dipole-dipole interaction between the polar group of Gln2 and the total dipolar moment of the protein.

The absence of the charge, however, should render the hbond interaction weaker than in WT allowing for a higher flexibility of the NT that could eventually lead to a locally distorted molecular conformation responsible for the observed phenotype.

## 4.6 Discussion

We have performed a 3ns long molecular dynamics simulation of full length WT Tat protein. Analysis of the MD trajectories and electrostatic helped to establish fundamental interactions that give conformational stability of the protein which are not obvious from the NMR data.

During the dynamics the protein exhibited a high overall flexibility, due the absence of a well defined secondary structure. However comparison of global (Fig. 4.5a ) or per residue (Fig. 4.5b) rmsd behavior indicate that our results are well compatible with experimental determination.

Comparison of electrostatic interaction and correlation map indicates as could be expected that the molecular conformation of Tat is mainly dictated by its unusually high charge.

Electrostatic profile (Fig. 4.6) and hydration (Table 4.I) analysis within the basic domain provide further support for the hypothesis that Arg52 is the residue involved in the specific binding interaction with G26 in RNA TAR hairpin(Aboul-ela, Karn, and Varani 1995) since it is positioned as the more exposed and solvent accessible arginine within the basic domain.

Correlation matrix and essential dynamics analysis, indicate that there is a little correspondence between of the biochemically defined domains and their mobility(Fig. 4.7). Redefinition of the NT-CRD functional domain gives a very good separation of motional behavior of both domains (Fig. 4.8). The longer CRD becomes largely the most flexible domain in the protein consistently with the proposal that a high flexibility is required to this arm to bind Zn ions (Frankel, Bredt, and Pabo 1988;Frankel, Biancalana, and Hudson 1989;Quemard et al. 1996). without significant structural changes upon divalent metal ions binding or dimerization.

Core domain shows a some what independent motion just with small correlation with CRD which is the most flexible region of the protein. The shortened NT domain exhibit high correlation with QRD domain (Fig. 4.7), they are the more stiff region in the protein and they could be seen as one single, relatively rigid, structural body. QRD displays correlated motion with all the biochemical domains in agreement with the proposal that function of QRD is to provide structural stability to the protein. The second exon expressed CT domain shows slightly correlated motion only with it neighboring domain QRD (Fig 4.7). However, it has an important contributions to the essential movement. This result, added to the above mentioned NT-QRD strong interaction suggest that the first exon expressed structure of Tat (aminoacids1-72) should not be modified by the presence of the second exon expressed part. In fact, strong interaction such as:Arg53-Gln72, Met1 -Asp67, -Ile69 and -Lys71 hbonds are established during simulation(Table 4.II), that may stabilize the 72 amino acids structure. It presents interesting biological implications since it has been proven that only the first exon expressed

part of Tat is important for transactivation (Demarchi, Gutierrez, and Giacca 1999;Derse et al. 1990).

On the other hand, essential dynamics study shows a rapid convergence of the eigenvalues (Fig 4.8a), suggesting that, despite the high flexibility of the protein, within the simulation time, one is able to individuate a few essential components of the motion while all the others can be ascribed to positional fluctuations.

Asp2 (or Glu), which make a salt bridge with Lys51 and Arg53 is crucial for keeping the NT domain bound to the ARD and the rest of the protein. The calculations predict that mutation to any other non-negative residue hampers biological activity (Fig. 4.10). In this respect, it is worth observing that a number of experimental evidences obtained in the last 15 years by truncation or point mutations of the first 20 amino acids of Tat has led to the conclusion that this domain is absolutely essential for protein function (Reddy et al. 1992;Kuppuswamy et al. 1989). Since, up to our knowledge, no relevant Tat partner has been obtained to date interacting with this domain, it might be assumed that the essential function of the N-terminal portion of Tat is to maintain proper conformation of the whole protein by interacting with the basic domain.



## Chapter V

### Amino acid modification in the HIV-1 Tat basic domain

#### 5.1 *Introduction: The Arginine Rich Domain*

As already mentioned in the previous chapter, despite the fact that Tat does not present secondary structure elements, it is possible to identify six functional regions in the protein (Fig. 4.4). Among these domains, the basic region (amino acids 49-57) is crucial for a number of key functions. First, this domain is required for TAR activation by interacting with a bulge sequence in the TAR RNA (Berkhout, Silverman, and Jeang 1989; Berkhout and Jeang 1989). Second, it contains the nuclear localization signal of the protein (Schwarze et al. 1999; Hauber, Malim, and Cullen 1989). Additionally, this domain is needed for the interaction of Tat with other cellular macromolecules, leading to cellular internalization of the protein when present in the extracellular milieu (Garber et al. 1998; Frankel, Bredt, and Pabo 1988; Mann and Frankel 1991), to the activation of NF- $\kappa$ B (Deng et al. 2000), and to the recruitment of p300/CBP to the long terminal repeat (LTR) (Jeyapaul, Reddy, and Khan 1990).

Recent experimental evidence indicates that Lys50 and 51 of the basic domain are also targets for post-translational modification through acetylation (Ott et al. 1999; Kiernan et al. 1999; Col et al. 2001), suggesting the relevance of these amino acids for protein function. Site-directed mutagenesis of these residues have indicated that substitution of Lys with Ala, which removes one charge per residue, strongly affects transactivation, while a conservative substitution with Arg has a more modest effect (Kiernan et al. 1999). However, at what extent the structural rearrangements introduced in the protein by these modifications might account for the observed phenotypes has not been investigated so far. Furthermore, on the basis of these data, it is difficult to estimate whether the dramatic effect on function of the Ala mutants is caused by charge neutralization, change of chemical properties, or both.

Starting from the NMR structure of Bayer et al.(Bayer et al. 1995), we have here undertaken a molecular dynamics study of Ala and Arg mutants at positions 50 and 51 to understand the effects of these modifications at the structural level. Similarly, we have used a MD simulation of acetylated Lys 50 to discriminate the consequences of charge neutralization from those caused by amino acid length. We find that replacement with Ala causes a profound structural rearrangement of the protein, which may eventually result in unfolding. Furthermore, we observe that the structural determinants of the acetyl derivative of Lys 50 are remarkably similar to those of both wild-type and Lys 50 Arg mutant, suggesting that the change of chemical properties on passing from Lys to Ala plays a key role for unfolding.

Finally, to experimentally substantiate these theoretical results, for all the mutants we provide a homogeneous set of biological data (including transactivation properties and cellular localization), which turn out to be absolutely consistent with the MD predictions.

Experimental procedures and methods are detailed in Appendix I

## **5.2 Structural models**

### **5.2.1 K50R, K50A, AcK50, K50A and K51R Tat.**

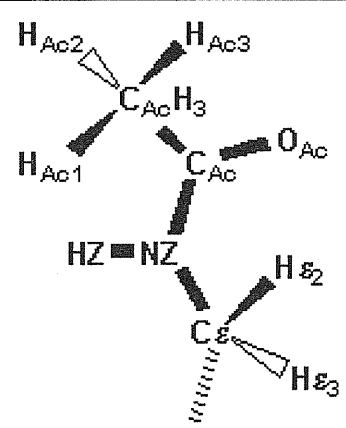
Starting from the NMR derived structure of the 86 amino acids of wt Tat from HV1Z2 (human immunodeficiency virus type 1, Zaire isolate 2)(Bayer et al. 2000) K50A, K50R, AcK 50, K51A and K51R Tat structures were constructed by replacing Lys 50 or 51 with the corresponding groups using the molecular graphics program Cerius2 of MSI. The mutating residues were placed in a similar orientation to that found in the wt structure without exhibiting bad contacts with the adjacent residues. Cl<sup>-</sup> counterions were added to neutralize the protein positive charge (+11 for K50R/K51R Tat, +10 for K50A/K51A/AcK 50 Tat).

### 5.3 Computational Details

For the protein mutants and Cl counterions AMBER(Cornell, Cieplak, and Bayly 1995) force field parameters were used. The parameterization of the non-standard residue AcK 50. was constructed as follows: (i) the van der Waals parameters and the partial atomic charges were taken as those of the ACE (Acetyl) and LYN (Neutral Lysine) standard groups present in the AMBER force field, except for NZ and HZ charges, which were slightly adapted from those of the peptide NH group so as to insure the electroneutrality of the residue (Table 5.I); (ii) bond stretching, angle bending and torsion parameters were those of a peptide NH group.

The MD calculations were performed with the SANDER module of AMBER 5.0(Case et al. 1997) suite of programs. Three ns of molecular dynamics in the npt ensemble and PME for long range electrostatics have been performed for Lys 50,51→Arg 50,51 mutants and AcK 50, while since structure is not conserved (see Results section), Lys 50,51→Ala 50,51 simulation where interrupted at 2,5 ns. MD protocol was identical to that reported for wt Tat in the previous chapter (section 4.4)

**Table 5.I:** Partial charges of AcK50 terminal atoms. The charges of NZ and HZ, which are different from the corresponding ones in ACE and LYN(Case et al. 1997), are reported in bold face.

	Atom Name	Charge [e <sup>-</sup> ]
	H <sub>Ac1-3</sub>	0.01
	C <sub>Ac</sub> H <sub>3</sub>	0.142
	C <sub>Ac</sub>	0.616
	O <sub>Ac</sub>	-0.504
	<b>NZ</b>	<b>-0.53563</b>
	<b>HZ</b>	<b>0.2719</b>
	H <sub>ε2</sub> , H <sub>ε3</sub>	-0.03358
	C <sub>ε</sub>	0.32604

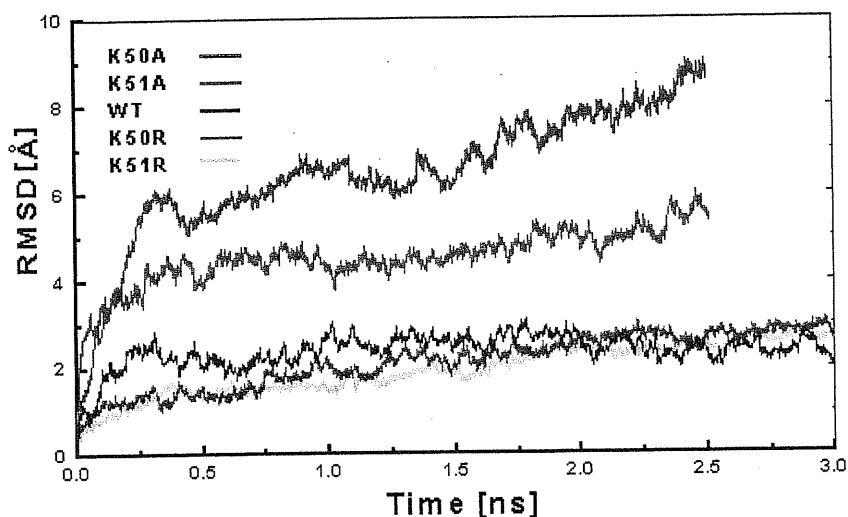
## 5.4 Results

### 5.4.1 MD simulations.

We present here the results obtained from our MD simulation on Ala and Arg Tat mutants at positions 50 and 51 and on the acetylated Lys 50 (AcK 50) derivative. Comparison is made with findings from our previous MD study on the wt protein (see Chapter 4), which are briefly summarized.

#### 5.4.1.1 Wt Tat.

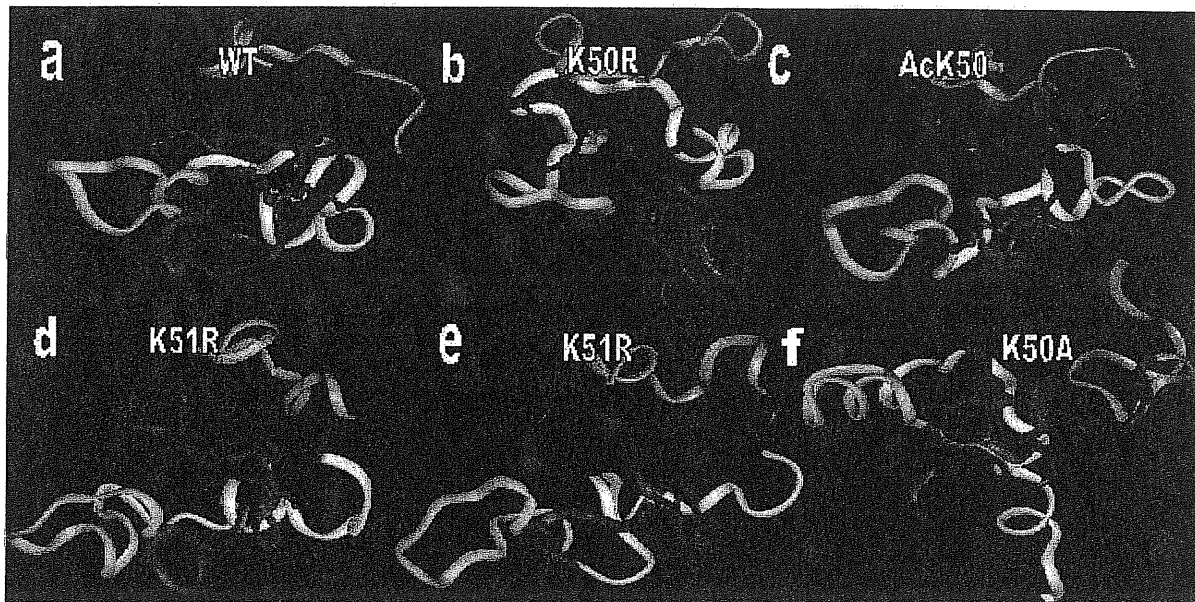
The wt protein turned out to exhibit a very high degree of flexibility, as evinced by a plot of the rmsd as a function of time (Fig. 5.1).



**Figure 5.1.** Rmsd [Å] of backbone atoms of wt Tat and its derivatives investigated here plotted as function of simulation time. K50A and K51A simulations have been interrupted after 2.5 ns because of a partial unfolding of the structure).

However, the overall fold of the NMR structure turned out to be well maintained. In particular, the basic domain stays in extended configuration and is widely exposed to the solvent as in the NMR structure. This is probably a consequence of the high electrostatic repulsion between its

charged residues. Lys 50 and Gln 54 side chains are oriented so as to avoid solvent/hydrophobic core interactions. Lys 51 forms hydrogen bonds with Asp 2 (belonging to the N-term domain); Arg 53, located in the middle of the basic domain, strongly interacts with the charged moieties of Asp 2, Asp 67 and the carbonyl groups of Gln 72, the latter two residues belong to the Gln-rich domain (Table 5.II).



**Figure 5.2.** wt (a), K50R (b), AcK 50 (c), K51R (d), K51A (e) and K50 (f) final MD structures. In wt, K50 and K51 are represented as ball-and-sticks. In all the other structures, the residue either mutated or acetylated in ball-and-sticks.

#### 5.4.1.2 K50R

Replacement of Lys 50 with Arg does not affect the overall folding of the protein, as evidenced by the rmsd of the backbone (Fig. 5.1) and comparison of the final molecular configuration (Fig. 5.2 b). Arg50 side chain is bent on itself, forming an very stiff H-bond network with Lys51, Gln54 (from ARD) and His33 (from CRD) (Fig 5.3a and Table 5.II). In addition, its hydrophobic portion (namely CG and CD) interact tightly to Ile45 and in a lesser extent to Pro3 (Table 5.III). This disposition of Arg50 lateral chain ultimately results in a compact conformation of the Core and slightly more exposed of QRD and CT domain.

**Table 5.II.** Hbond distances for Met1 and ARD residues for time > 1.5 ns. (Occupancies marked with 100\* indicate that double Hbond is established).

H Donor - H Acceptor	Av. dist. [Å]	% occupancy	H Donor - H Acceptor	Av. dist. [Å]	% occupancy
<b>wt Tat</b>			<b>K51R Tat mutant</b>		
Arg49 (NH1)-Tyr47 (O)	3	34	Arg55 (NH1) – Glu86 (O)	2.8	99
Lys51(NZ)-Asp2 (O)	3.1	33	Arg55 (NH2) – Ser59 (Oγ)	2.9	76
Lys51 (NZ)-Asp2 (Oδ1-2)	3	98	Arg55 (NH1) – Glu86 (O)	2.8	94
Arg53 (NH1-2)-Asp2 (Oδ1-2)	3.2	48	Arg56 (N) – Gln54 (O)	3.4	21
Arg53 (NH1-2)-Asp67 (Oδ1-2)	3.3	67	Ser59 (Oγ) - Arg57 (O)	3.0	90
Arg53 (NH1)-Gln72 (O)	3.1	42	<b>K50R Tat mutant</b>		
Gln54 (Ne)-Ile45 (O)	3.9	82	Arg49 (NH1)-Tyr47 (O)	3	20
Arg55 (N)-Arg53 (O)	3.1	44	Arg50 (NH1-2)-His33 (Ne2)	3.1	93
Arg55 (Ne)-Gln60 (Oε1)	3.1	21	Arg50 (NH2)-Lys51 (O)	2.8	100
Arg55 (NH1-2)-Gln60 (Oε1)	3.1	41	Arg50 (NH2)-Arg53 (O)	3.7	37
Arg56 (N)-Gln64 (O)	3.2	21	Arg50 (NH1-2)-Gln54 (Oε1)	3.1	21
Arg56 (NH1-2)-His33 (Ne2)	3.1	42	Lys51 (NZ)-Asp2 (O)	3.2	22
Arg56 (NH1-2)-Gln54 (Oε)	2.9	19	Lys51 (NZ)-Asp2 (Oδ1-2)	3.1	100 *
Arg57 (NH1-2)-Gln60 (Oε)	3.1	40	Arg55 (NH1-2)-Glu86 (Oε1-2)	3.1	100 *
<b>K51R Tat mutant</b>			Arg56 (N)-Gln61 (O)	3.3	30
Ile45 (N) – Arg49 (O)	3.0	98	Arg56 (NH1-2)-Gln63 (Oε1)	2.9	100
Arg49 (N) – Tyr47 (O)	3.5	90	Arg56 (NH1-2)-His33 (Ne2)	3.3	80
Arg51 (N) - Arg49 (O)	3.5	12	Arg56 (NH1-2)-Gln54 (Oε)	3.2	52
Arg51 (N) – Gln54 (Oε1)	3.0	78	Arg57 (Ne)-Glu86 (O-Oxt)	3	100 *
Arg51 (NH1-2) - Asp2 (Oδ1)	2.9	15	Arg57 (NH2)-Lys85 (O)	3.4	51
Arg51 (NH1) – Tyr47 (OH)	2.9	89	Arg57 (NH2)-Glu86 (O-Oxt)	3.1	100*
Arg53 (Ne) – Asp67 (Oδ2)	2.8	36	<b>AcK 50 Tat</b>		
Arg53 (NE) – Gln72 (O)	3.1	28	Gly44 (N)-AcK 50 (AcO)	3.1	69
Arg53 (NH1) – Glu86 (Oε1-2)	2.7	100*	AcK 50 (H)-Gly48 (O)	3.1	54
Arg53 (NH1) – Asp67 (Oδ2)	3.1	62	Arg53 (NH1)-Tyr47 (OH)	3.1	71
Arg53 (NH2) – Asp67 (Oδ2)	3.7	28	Arg53 (NH1-1)-Met1 (O)	3.1	70
Arg53 (NH2) – Glu86 (Oε1-2)	2.9	100*	Arg53 (NH1-2)-Asp2 (Oδ1-2)	3	100*
Gln54 (N) - Arg51 (O)	2.8	100	Gln54 (Ne)-AcK 50 (AcO)	3.1	20
Gln54 (N) - Arg52 (O)	3.0	37	Arg55 (N)-Gln54 (Oε1)	3	80
Gln54 (Ne1-2) – Gly42 (O)	3.5	41	Arg55 (NH1-2)-Glu86 (Oε1-2)	3.3	25
Gln54 (Ne22) – Gly44 (O)	3.7	19	Arg55 (NH1-2)-Glu86 (O-Oxt)	3.1	100*
Arg55 (Ne) – Arg57 (O)	3.6	16	Arg57 (Ne)-Ser59 (Oγ)	3.5	30
Arg55 (Ne) – Ser59 (Oγ)	3.1	83	Arg57 (Ne)-Glu86 (O-Oxt)	3.6	40
Arg55 (NH1) – Glu86 (Oε1)	3.1	35	Arg57 (NH1-2) Glu86 (O-Oxt)	3.1	100*

#### 5.4.1.3 AcK50.

In the acetylated species, the folding (Fig. 5.2c) and rmsd (Fig. 5.1) are well maintained with respect to WT. AcK50 side chain lies in an extended conformation between Core and ARD backbones (Fig. 5.2b). However the bonding interactions of AcK50 are weaker and shorter in time than those of K50R case (Table 5.II), probably because of the absence the charge.

**Table 5.III:** Hydrophobic interaction for residues at position 50. Occupancies have been calculated for time >1.5ns. Only occupancies higher than 20% are reported

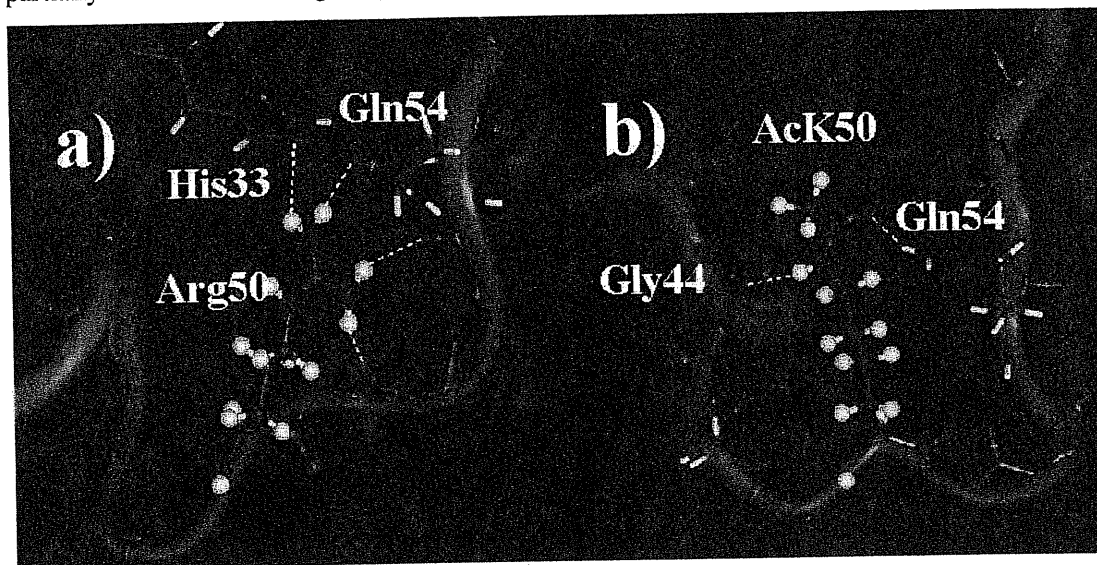
<b>Lys50 (wt)</b>	<b>% occupancy</b>	<b>AcK 50</b>	<b>% occupancy</b>
Lys50 (C $\beta$ ) Ile45 (C $\delta$ 1)	21	AcK 50 (C $\gamma$ ) Leu43 (C $\delta$ 1)	37
Lys50 (C $\beta$ ) Lys51(C $\alpha$ )	81	AcK 50 (C $\gamma$ ) Ile45 (C $\gamma$ 2)	61
Lys50 (C $\gamma$ ) Lys51 (C $\alpha$ )	51	AcK 50 (C $\gamma$ ) Gly44 (C $\alpha$ )	77
Lys50 (C $\gamma$ ) Arg52 (C $\alpha$ )	20	AcK 50 (C $\delta$ ) Ile45 (C $\gamma$ 1)	77
Lys50 (C $\gamma$ ) Gln54 (C $\beta$ )	29	AcK 50 (C $\delta$ ) Ile45 (C $\gamma$ 2)	77
Lys50 (C $\gamma$ ) Gln54 (C $\gamma$ )	22	AcK 50 (C $\delta$ ) Gly44 (C $\alpha$ )	77
<b>Arg50 (K50R)</b>		AcK 50 (C $\epsilon$ ) Leu43(C $\beta$ )	38
Arg50 (C $\beta$ ) Lys51 (C $\alpha$ )	23	AcK 50 (C $\epsilon$ ) Leu43(C $\gamma$ )	45
Arg50 (C $\gamma$ ) Ile45 (C $\gamma$ 1)	81	AcK 50 (C $\epsilon$ ) Leu43(C $\delta$ 1)	47
Arg50 (C $\gamma$ ) Ile45 (C $\delta$ 1)	83	AcK 50 (C $\epsilon$ ) Leu43(C $\delta$ 2)	52
Arg50 (C $\delta$ ) Ile45 (C $\gamma$ 1)	54	AcK 50 (C $\epsilon$ ) Leu43(C $\alpha$ )	58
Arg50 (C $\delta$ ) Pro3 (C $\beta$ )	32	AcK 50 (C $\epsilon$ ) Ile45(C $\gamma$ 2)	64
<b>AcK 50</b>		AcK 50 (C $\epsilon$ ) Gly44 (C $\alpha$ )	77
AcK 50 (C $\beta$ ) Lys51 (C $\alpha$ )	37	AcK 50 (C $_{AcH_3}$ ) Leu43(C $\delta$ 1)	47
AcK 50 (C $\beta$ ) Ile45 (C $\delta$ 1)	51	AcK 50 (C $_{AcH_3}$ ) Leu43(C $\beta$ )	55
AcK 50 (C $\beta$ ) Ile45 (C $\gamma$ 1)	63	AcK 50 (C $_{AcH_3}$ ) Leu43(C $\gamma$ )	68
AcK 50 (C $\beta$ ) Ile45 (C $\delta$ 2)	67	AcK 50 (C $_{AcH_3}$ ) Gly44(C $\alpha$ )	72
AcK 50 (C $\gamma$ ) Leu43 (C $\delta$ 2)	23	AcK 50 (C $_{AcH_3}$ ) Leu43(C $\alpha$ )	72

During the dynamics, AcK acetyl group H-bonds to Gly44, Gly48, Gln54 or Arg56; the CH carbons of AcK50 side chain form van der Waals contact mainly with Leu43, Gly45 and Ile45, that further stabilize the interaction of the residue with the protein scaffold. Table 5.III show that AcK50 side-chain exhibit optimal interactions Core and ARD, possibly because the long side-chain may optimally interact with both the Core hydrophobic region and the very basic ARD.

#### 5.4.1.4 K50A

The absence of the long Lys 50 side chain generates a hydrophobic cavity in the core-basic neighboring region, which is accessible to the solvent (Fig. 5.4a). The presence of water molecules in this cavity, which interact with hydrophobic residues at the surface, locally destabilizes the protein. During the dynamics, the hydrophobic repulsion is reduced by a conformational change in the core that approaches to the basic domain, which after about 1 ns

partially covers the core region (Fig 5.4b). Total energy interaction of the core domain with the

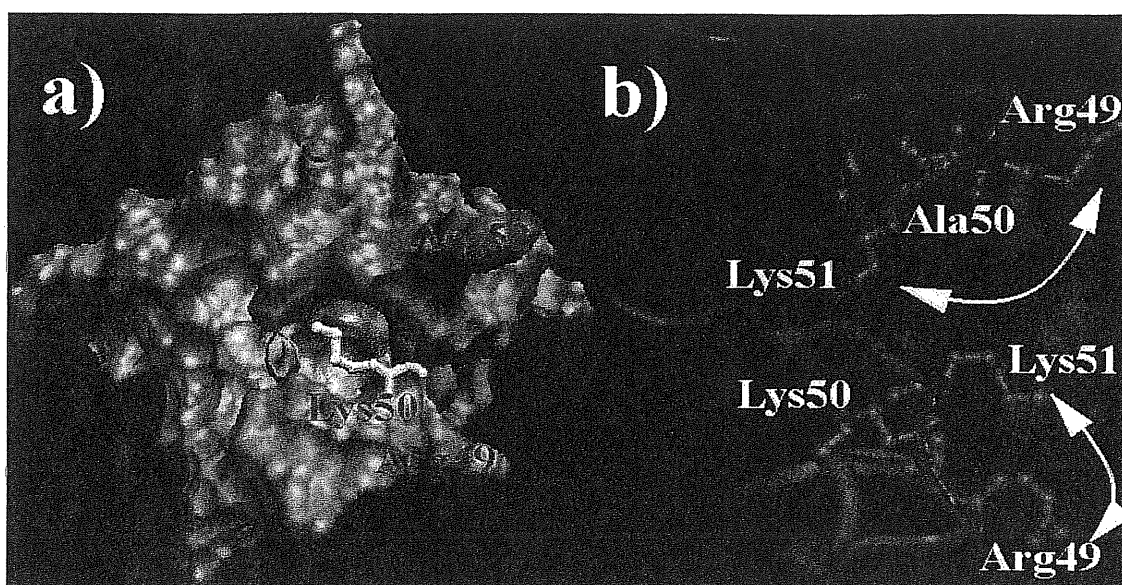


**Figure 5.3.** Intramolecular interactions of Arg50 in K50R (a) and AcK 50 (b) from representative snapshots.

rest of the system continuously decreases by about 40% its initial value during simulation time. This causes: (i) A displacement of Gln and C-term domains, thus disrupting Arg 55- and Arg 57- Glu 86 interactions (present in the wt protein and in the other mutants investigated here) and ultimately the separation of the C-term arm from the protein frame. (ii) A rearrangement of the N-term tail, thus disrupting Lys 51- and Arg 53-Asp 2 interactions with separation of the N-term arm from the rest of the protein. The absence of the Lys side chain also induces a rearrangement of the basic domain provoking an increase of the Arg 49-Lys 51 distance and a consequent dramatic decrease of electrostatic repulsion between these residues (see fig 5c)<sup>4</sup>. The overall result of the process is a partial defolding of the structure (Figs. 5.1 and 5.2e). Already after 1 ns the molecule is distorted to a great extent. The simulation of this mutant was interrupted at 2.5 ns as the description of the final folded state is clearly beyond the time-scale investigated here.

<sup>4</sup> Electrostatic energy in Arg 49-Lys 51 interaction calculated at time = 1 ns decreases to 75% of the initial conformation.

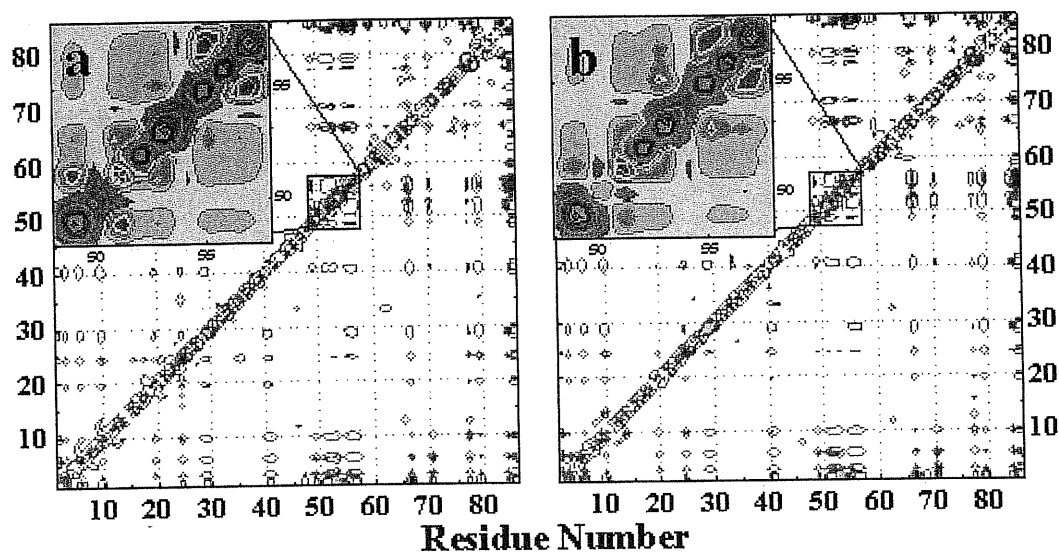




**Figure 5.4.** K50A mutant. (a) Molecular surface of the region around Ala50 at time=0 ns. The cavity formed by the absence of the long Lys sidechain is colored in red. (b) Comparison between K50A (top) and wt (bottom) at time = 1ns. Ribbon color as Fig. 1. Arrows indicate the mutual conformation of Arg 49 and Lys 51.

The calculated electrostatic interaction energies of residues at position 50 in the AcK 50 and K50A mutants (at initial time) with the rest of the protein differ by less than 10%. In fact as can be seen from Fig. 5.5 the electrostatic profile presented by the two molecules is practically the same. Thus, although electrostatic interactions are clearly dominant in Tat, the defolding process is mainly driven by a rearrangement of the hydrophobic core region that minimizes its destabilizing interaction with the solvent.

In conclusion, it is the *absence* of a long side chain, and not the *suppression* of the charge (as in AcK 50), which leads to the defolding of the protein in the K50A mutation.



**Figure 5.5.** Electrostatic interaction energy maps per residue at time = 0 for Ack50 (a) and K50A (b). Close ups on the basic domain highlight the similarity in the electrostatic profile of both molecules.

#### 5.4.1.5 K51A

Lys 51 strongly interacts with Asp 2 in the wt protein and in the other mutants studied. Its replacement with Ala removes this stabilizing interaction which in turn causes a much higher mobility in the N terminal tail relatively to wt (Fig. 5.1). The final structure (Fig. 5.2e) shows a small rearrangement of the core and that the acidic N-term domain is separated from the protein body and is exposed to solvent. Although in the explored time-scale this rearrangement is not propagated to the entire protein, the rather large rmsd (larger than all the systems investigated here except K50A) suggests that the structure might unfold on a longer time-scale.

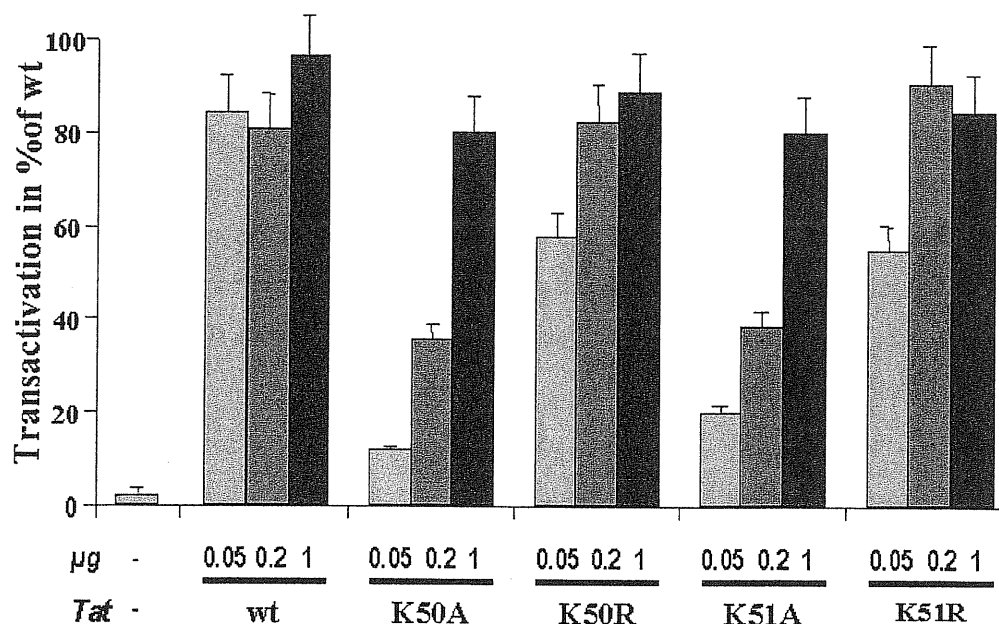
#### 5.4.1.6 K51R

The rmsd is the smallest of all the mutants investigated here (Fig. 5.1), indicating that replacement of Lys 51 with Arg does not cause dramatic changes in the protein structure. Arg 51 conformation is similar to that of Lys 51 in the wt protein. Because of the more delocalized charge relative to Lys, Arg can be better solvated and, consequently, Arg 51 has a more

flexible behavior than Lys 51 and its direct H-bond interaction with Asp 2 is more fluctuational than in both wt, AcK 50 and K50R mutants (table 1).

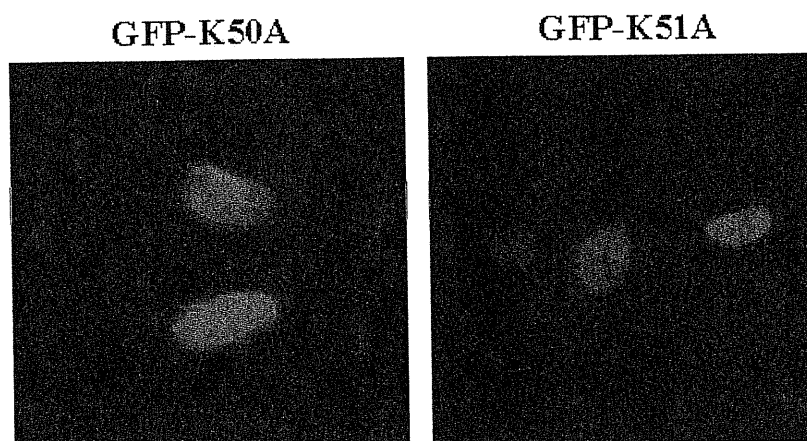
### 5.5 Biochemical characterization of the Tat mutants.

To correlate the observed MD features of the investigated Tat mutants with their transactivation properties, we obtained expression vectors for Tat K50A, K50R, K51A, and K51R and transfected HL3T1 cells, which bear an integrated LTR-CAT (chloramphenicol acetyl transferase) construct. The transactivation capacity of the Tat derivatives is therefore obtained by measuring the catalytic activity of CAT enzyme under the same conditions for all the mutants.



**Figure 6.** Biological characterization of Tat mutants. (a) Transactivation assay. HL3T1 cells (containing an integrated LTR-CAT construct) were transfected with the indicated amounts of plasmid constructs expressing Tat and Tat mutants. CAT assays were performed 48 hours after transfection. Shown are means and standard deviations of three independent transfections.

As compared to wt Tat, both, the K50A and K51A variants are found to be remarkably impaired in directing reporter gene expression, especially when tested at low concentration. In contrast, transactivation of the respective Arg mutants is only slightly decreased (Fig. 5.6).



**Figure 5.7.** Sucellular localization. Plasmid vectors expressing Tat mutants K50A and K51A as C-terminal fusions to the enhanced green fluorescent protein (GFP) were transfected in human U2OS cells. Fluorescence was visualized and recorded 48 hrs after transfection. Both mutants clearly show a prominent nuclear localization.

To ascertain that the observed diminished efficiency of the Ala mutants was not related to altered nuclear transport (ref), we also obtained plasmids expressing the K50A and K51A proteins tagged with the green fluorescent protein (GFP). Both these mutants are found to localize inside the nucleus(Fig. 5.7).

## 5.6 Discussion

This chapter presents a combined theoretical and biochemical study on mutants of Tat at positions 50 and 51, two lysine residues which lie in a domain which is critical for Tat function. It offer a molecular interpretation based on conformation stability to explain the functional behavior of these mutants. In particular, substitution of Lys 50 and 51 with Ala leads to a remarkable degree of defolding of the native structure. For the K51A mutant, defolding is most probably a consequence of the loss of a stabilizing electrostatic interaction. In the case of K50A, it is the absence of the Lys long side chain (Fig.5.4a) that determines local destabilization of protein structure that eventually leads to the defolding process (Fig. 5.1), resulting from repulsion of the charged residues at position 49 and 51 and closing of the core region to the basic domain backbone (Fig. 5.4b). This conclusion is strongly supported by

the MD simulation on variant AcK50. In this case, acetylation leads to loss of charge but maintenance of a long amino acid side chain. As a consequence, structure of this mutant is not significantly different from that of the wt protein. Thus, it is amino acid side *chain length* and not *charge* at this position that is essential for proper protein folding.

Substitution of Lys at positions 50 and 51 with Arg do not cause significant rearrangement of the backbone structure (Figs. 5.1, 5.2). Local rearrangement in the side chain disposition is expected to generate changes in the external electrostatic profile, thus possibly modifying surface recognition of other proteins interacting with Tat through this region or formation of the Tat-TAR complex.

The theoretical findings are compared with a homogeneous set of two distinct types of biochemical measurements: Transactivation properties and cellular localization. Our experimental data are fully consistent with the theoretical results. Arg mutants at position 50 and 51 show a modestly reduced biological activity compared with wt; while the K50A and K51A mutants are remarkably impaired in their capacity of transactivation of the HIV LTR. In spite of the important structural rearrangements obtained from the simulations, we asserted that the observed phenotypes are not due to defective nuclear import properties of the mutants by performing confocal microscopy of cell transfected with GFP tagged proteins. Both Ala mutants were found to markedly localize in the nucleus (Fig. 5.7). Indicating that Ala mutations affect the interaction of Tat with any of its RNA or protein complex partners inside the nucleus.

This observations have the important consequence of calling for a review of the “Ala scanning” experiments performed on Tat and its peptides, since this procedure can comport severe structural modifications in the protein.

Structural prediction on the AcK 50 variant deserves more comment. As noted above, the overall structure of this protein is not remarkably different from that of the wt, thus ruling out a major effect of acetylation at this residue in determining significant structural changes. In this respect, it is worth mentioning that the role and timing of Tat acetylation inside the cells are still largely unclear (Kiernan et al. 1999; Ott et al. 1999; Col et al. 2001; Deng et al. 2000; Kiernan et al. 1999; Col et al. 2001; Ott et al. 1999). It might be assumed that acetylation might be a transient modification that is essential to modulate a specific molecular event during transactivation. Since it is not expected that this results from a change in protein conformation, this modification most likely alters recognition of the Tat basic domain by TAR or other cellular proteins. Acetylation might be involved in the regulation of binding of Tat to TAR, or

formation of a stable complex among Tat, TAR and Cyclin T1. Alternatively, it may modulate the interaction between Tat and other factors such as TBP or RNA polymerase II (Ott et al. 1999; Kiernan et al. 1999; Carloni and Roethlisberger 2000). Further biochemical investigations are clearly required to clarify this issue.

## Conclusions

We have presented a molecular dynamics study, accompanied by ab initio calculations on two very important targets for pharmaceutical intervention.

The first has focused on complex between acyl carrier protein reductase InhA from *M. tuberculosis* and the drug isonicotinic acid hydrazide-NADH. We focused on wild-type (wt) InhA and a mutant causing drug-resistance, S94A, for which structural information is available (Rozwarski et al. 1998). Using force-field based classical molecular dynamics we have investigated the dynamics of the entire protein on the ns time scale at room conditions. An ab initio molecular dynamics on a model complex of the active site in the ps time-scale has provided complementary information on the chemical bonding. The two approaches provide a consistent picture of the target/drug interactions. The water-mediated H-bond interactions between Ser94 side chain and NADH, present in WT InhA X-ray structure, can be lost during the dynamics. This conformational change is accompanied by a structural rearrangement of Gly14. Both classical- and ab initio MD structures of the active site are remarkably similar to the X-ray structure of the S94A. The high mobility of Ser94 can provide a rational of the small change in free binding energies on passing from WT to S94A InhA.

The second part of the thesis has dealt with the HIV-1 Tat protein, a promising target for anti-AIDS therapy. Based on NMR structural information (Bayer et al. 1995), classical MD calculations have first provided a characterization of the large-scale motion and of the energetics of the protein. Electrostatics provides by far the major contribution to the energetics (the overall charge of the protein is + 11), therefore a rigorous treatment for long-range electrostatics, with the Ewald summation method, has been required. Our calculations show that the regions undergoing correlated motion may be associated to specific biological functions (such as the basic domain and the cysteine-rich domain). In particular, the C-terminal region (residues 73-86) turns out to be rather uncorrelated to the rest of the protein, suggesting that region of Tat expressed by the first exon (residues 1-72) is structurally independent from that expressed by the second exon. This may be consistent with the fact that the C-terminal of the protein is not required for transactivation (Tan and Frankel 1992; Truant and Cullen 1999). The calculations further suggest that the structure may partially unfold upon

mutation at position 2, which is fully consistent with *in vivo* experiments performed in infected human cells by Prof. M. Giacca in ICGEB (Trieste).

As a second step of our investigation of Tat, we have focused on modification and mutations on the basic region. This motif has been recently reported a target of post transcriptional modifications whose biological significance is currently been matter of debate (Ott et al. 1999;Kiernan et al. 1999). Unfortunately, however, no structural information is available. Our MD calculations have pointed to a striking correlation between structural rearrangements of mutants of Tat basic domain and biological activity. The calculations also show that Lys50 in the basic domain plays a crucial role for the stability of the protein, as it separates the hydrophobic core to the highly hydrophilic basic domains.

In the final step of our study, we have used MD simulation to construct a structural model of the Tat-TAR adduct, for which no structural information is available. Our model shows is consistent with cross-link and chemical modification assays reported in literature (Sumner-Smith et al. 1991;Puglisi et al. 1992;Wang and Rana 1998;Naryshkin et al. 1997). Furthermore, the construct offers a rational for the reduced transactivation capabilities encountered for some basic domain Tat mutants. Finally, it provides a template for the rational design based on drug-screening and docking procedures (Bohm 1992;Katchalski-Katzir et al. 1992;Eisen et al. 2001;Hilbert, Bohm, and Jaenicke 1993;Bohm 1992) of small ligands to the tat/TAR complex. These ligands in turn could tested in the laboratory to assess whether they are capable to inhibit Tat functionality and thus HIV replication.

In conclusion, our work has applied state-of-the-art computational chemistry techniques to pharmacologically relevant problems, using *in vivo* and *in vitro* data that have been generated by the Laboratory of Prof. Giacca as well as from the recent literature. We believe that this combination of highly interdisciplinary areas represents an important challenge for pharmacological research and for structural biology.



## Bibliography

- Aboul-ela,F., J.Karn, and G.Varani "The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein." *J. Mol. Biol.* 253(October 1995): 313-32.
- Aboul-ela,F., J.Karn, and G.Varani "Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge." *Nucleic Acids Res.* 24(October 1996): 3974-81.
- Alber,F. and P.Carloni "Ab initio Molecular Dynamics Studies on HIV-1 Reverse Transcriptase Binding Site: Implications for Nucleoside Analog Drug Resistance." *Protein Sci.*(2000).
- Alber,F., G.Folkers, and P.Carloni "Dimethyl Phosphate: Stereoelectronic versus Enviroments Effects." *J. Phys. Chem. B* 103(1999): 6121-6.
- Allen,M.P. and D.J.Tildesley *Computer Simulation of Liquids*. New York: Oxford University Press, 1987.
- Amadei,A., A.B.Linssen, and H.J.Berendsen "Essential dynamics of proteins." *Proteins* 17(December 1993b): 412-25.
- Barnett,R.N. and U.Landman "Born-Oppeneimer molecular dynamics simulations of finte systems: Structure and dynamics of (H<sub>2</sub>O)<sub>2</sub>." *Phys. Rev. B* 48(1993): 2081-97.
- Bayer,P. et al. "Structural studies of HIV-1 Tat protein." *J. Mol. Biol.* 247(April 1995): 529-35.
- Becke,A.D. "Density-functional exchange-energy approximation with corrcet asymptotic behavior." *Phys. Rev. A* 38(1988): 3098-100.
- Berendsen,H.J.C., J.R.Grigera, and T.P.Straatsma "The Missing Term in Effective Pair Potentials." *J. Phys. Chem.* 91(1987): 6269-71.

- Berendsen,H.J.C. et al. "Molecular Dynamics with Coupling to an External Bath." *J. Chem. Phys.* 81(1984b): 3684.
- Berkhout,B. and K.T.Jeang "Trans activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the trans- acting-responsive hairpin: a quantitative analysis." *J. Virol.* 63(December 1989): 5501-4.
- Berkhout,B., R.H.Silverman, and K.T.Jeang "Tat trans-activates the human immunodeficiency virus through a nascent RNA target." *Cell* 59(October 1989): 273-82.
- Bloom,B.R. and Murray C.J.L. "Tuberculosis: Commentary on a reemergent killer." *Science* 257(1992): 1055-64.
- Bohm,H.J. "The computer program LUDI: a new method for the de novo design of enzyme inhibitors." *J Comput Aided Mol Des* 6(February 1992): 61-78.
- Bradford,W.Z. and C.L.Daley "Multiple drug-resistant tuberculosis." *Infect. Dis. Clin. North Am.* 12(March 1998): 157-72.
- Brunger,A.T. X-PLOR, A System for X-Ray Crystallography and NMR. [3.1]. 1992. New Haven, CT, Yale Univ. Press. Computer Program
- Calnan,B.J. et al. "Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition." *Genes Dev.* 5(February 1991b): 201-10.
- Calnan,B.J. et al. "Arginine-mediated RNA recognition: the arginine fork." *Science* 252(May 1991a): 1167-71.
- Carlioni,P. and U.Roethlisberger "Simulations of enzymatic systems: perspectives from Car-Parrinello molecular dynamics simulations." *Theoretical Biochemistry - processes and properties of biological systems.* Eriksson L., ed.Elsevier Science, 2000.
- Carroll,R., L.Martarano, and D.Derse "Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras." *J. Virol.* 65(July 1991): 3460-7.

- Case,D.A. et al. AMBER 5, University of California, San Francisco. 1997.  
Ref Type: Generic
- Cavalli,A. and P.Carloni "Enzymatic GTP Hydrolysis:Insights from an ab initio Molecular Dynamics Study." *submitted*(2001).
- Col,E. et al. "The histone acetyltransferase, hGCN5, interacts with and acetylates the HIV transactivator, Tat." *J Biol Chem* 276(July 2001): 28179-84.
- Cole S.T. "Mycobacterium Tuberculosis: drug-resistance mechanisms." *Trends. Microbiol.* 2(1994): 411-5.
- Cornell,W.D., P.Cieplak, and C.I.Bayly "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." *J. Am. Chem. Soc.* 117(1995): 5179-97.
- Darden,T.A. and D.York "Particle Mesh Ewald: an N log(N) method for Ewald sums in large systems." *J Chem Phys* 98(1993): 10089-94.
- De Santis,L. and P.Carloni "Serine proteases: an ab initio molecular dynamics study." *Proteins* 37(December 1999): 611-8.
- Demarchi,F., M.I.Gutierrez, and M.Giacca "Human immunodeficiency virus type 1 tat protein activates transcription factor NF-kappaB through the cellular interferon-inducible, double- stranded RNA-dependent protein kinase, PKR ." *J. Virol.* 73(August 1999): 7080-6.
- Deng,L. et al. "Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones." *Virology* 277(November 2000): 278-95.
- Derse,D. et al. "A minimal lentivirus Tat." *J. Virol.* 65(December 1991): 7012-5.
- Derse,D. et al. "Structure and expression of the equine infectious anemia virus transcriptional trans-activator (tat)." *Dev. Biol. Stand.* 72(1990): 39-48.
- Dessen,A. et al. "Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis." *Science* 267(March 1995): 1638-41.

- Eisen,M.B. et al. "HOOK - A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding-site." *Proteins:Structure Function And Genetics* 19(2001): 199-221.
- Essman,U. et al. "A smooth particle mesh Ewald methods." *J. Chem. Phys. B.* 103(1995): 8577-93.
- Felber,B.K. and G.N.Pavlakakis "A quantitative bioassay for HIV-1 based on trans-activation." *Science* 239(January 1988): 184-7.
- Frankel,A.D., S.Biancalana, and D.Hudson "Activity of synthetic peptides from the Tat protein of human immunodeficiency virus type 1." *Proc. Natl. Acad. Sci. U. S. A* 86(October 1989): 7397-401.
- Frankel,A.D., D.S.Bredt, and C.O.Pabo "Tat protein from human immunodeficiency virus forms a metal-linked dimer." *Science* 240(April 1988): 70-3.
- Garber,M.E. et al. "The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein." *Genes Dev.* 12(November 1998): 3512-27.
- Garcia,A.E. "Large-amplitude nonlinear motions in proteins." *Phys. Rev. Lett.* 68(April 1992): 2696-9.
- Gregoire,C.J. and E.P.Loret "Conformational heterogeneity in two regions of TAT results in structural variations of this protein as a function of HIV-1 isolates." *J. Biol. Chem.* 271(September 1996): 22641-6.
- Gu,Y., Kar T, and S.Scheiner "Fundamental properties of the CH center dot center dot center O interaction: Is it a true hydrogen bond?" *J. Am. Chem. Soc.* 121(1999): 9411-22.
- Hauber,J., M.H.Malim, and B.R.Cullen "Mutational analysis of the conserved basic domain of human immunodeficiency virus tat protein." *J. Virol.* 63(March 1989): 1181-7.
- Hilbert,M., G.Bohm, and R.Jaenicke "Structural relationships of homologous proteins as a fundamental principle in homology modeling." *Proteins:Structure Function And Genetics* 17(October 1993): 138-51.

- Hohenberg,P. and W.Kohn "Inhomogeneous electron gas." *Phys. Rev. B* 136(1964): 864-71.
- Hoover,W.G. "Canonical dynamics - equilibrium phase - subspace distributions." *Phys. Rev. A* 31(1985): 1695-7.
- Hutter,J. et al. CPMD. MPI fur Festkorperforschung and IBM Zurich Research Laboratory[3.0h]. 1996. Computer Program
- Jeyapaul,J., M.R.Reddy, and S.A.Khan "Activity of synthetic tat peptides in human immunodeficiency virus type 1 long terminal repeat-promoted transcription in a cell-free system." *Proc. Natl. Acad. Sci. U. S. A* 87(September 1990): 7030-4.
- Jorgensen,W.L. et al. "Computer simulations of organic reactions in solution." *Ann. N. Y. Acad. Sci.* 482(1986): 198-209.
- Karn,J. "Tackling Tat." *J. Mol. Biol.* 293(1999): 235-54.
- Katchalski-Katzir,E. et al. "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques." *Proc Natl Acad Sci U S A* 89(March 1992): 2195-9.
- Kiernan,R.E. et al. "HIV-1 tat transcriptional activity is regulated by acetylation." *EMBO J.* 18(November 1999): 6106-18.
- Kohn,W. and L.J.Sham "Self-consistent equation including exchange and correlation effects." *Phys. Rev. A* 140(1965): 1133-8.
- Kuppuswamy,M. et al. "Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis." *Nucleic Acids Res.* 17(May 1989): 3551-61.
- Laio,A., J.Vande Vondele, and U.Roethlisberger "A Hamiltonian Electrostatic Coupled Scheme for Hybrid Car - Parrinello Molecular Dynamics Simulations." *submitted*(2001).
- Leach,A. *Molecular Modeling, Principles and Applications*. Singapore: Longman, 1996.
- Lee,C., W.Yang, and R.G.Parr "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density." *Phys. Rev. B* 37(1988): 785-9.

- Mann,D.A. and A.D.Frankel "Endocytosis and targeting of exogenous HIV-1 Tat protein." *EMBO J* 10(July 1991): 1733-9.
- Marzari,N. and D.Vanderbilt "Maximally-localized Wannier functions for composite energy bands." *Phys. Rev. B* 56(1997): 12847-65.
- Naryshkin,N.A. et al. "Chemical cross-linking of the human immunodeficiency virus type 1 Tat protein to synthetic models of the RNA recognition sequence TAR containing site-specific trisubstituted pyrophosphate analogues." *Biochemistry* 36(March 1997): 3496-505.
- Nose,S. "A unified formulation of the constant temperature molecular dynamics methods." *J. Chem. Phys.* 81(1984): 511-9.
- Ott,M. et al. "Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity." *Curr. Biol.* 9(December 1999): 1489-92.
- Parikh,S., G.Xiao, and P.J.Tonge. Inhibition of InhA, the Enoyl Reductase from *Mycobacterium tuberculosis*, by Triclosan and Isoniazid. *Biochemistry* 39, 7645-50. 2000.
- Parr,R.G. and W.Yang *Density-Functional theory of atoms and molecules*. New York: Oxford University Press, 1989.
- Pauling,L. *General Chemistry*. New York: Dover Publications Inc., 1988.
- Piana,S. and P.Carloni "Conformational flexibility of the catalytic asp dyad in HIV-1 protease: An ab initio study on the free enzyme [In Process Citation]." *Proteins* 39(April 2000): 26-36.
- Puglisi,J.D. et al. "Conformation of the TAR RNA-arginine complex by NMR spectroscopy." *Science* 257(July 1992): 76-80.
- Quemard,A. et al. "Binding of Catalase-Peroxidase-Activated Isoniazid to Wild-Type and Mutant *Mycobacterium tuberculosis* Enoyl-ACP Reductases." *J. Am. Chem. Soc.* 118(1996): 1561-2.

- Quemard,A. et al. "Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*." *Biochemistry* 34(July 1995): 8235-41.
- Reddy,M.V. et al. "Functional analysis of the N-terminal domain of Tat protein of the human immunodeficiency virus type 1." *Oncogene* 7(September 1992): 1743-8.
- Roethlisberger,U. and P.Carloni "Ab initio molecular dynamics studies of a synthetic biomimetic model of galactose oxidase." *Int. J. Quantum Chem.* 73(1999): 209-18.
- Roethlisberger,U. et al. "A comparative study of galactose oxidase and active site analogs based on QM/MM Car-Parrinello simulations." *J. Biol. Inorg. Chem.* 5(2000): 236-50.
- Rozwarski,D.A. et al. "Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*." *Science* 279(January 1998): 98-102.
- Rozwarski,D.A. et al. "Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD<sup>+</sup> and a C16 fatty acyl substrate." *J. Biol. Chem.* 274(May 1999): 15582-9.
- Ryckaert,J.-P., G.Ciccotti, and H.J.C.Berendsen "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints:Molecular Dynamics of n-Alkanes." *J. Comput. Phys.* 23(1977): 327.
- Saenger,W. *Principles of Nucleic Acid Structure*. New York: Springer-Verlag, 1983.
- Schwarze,S.R. et al. "In vivo protein transduction: delivery of a biologically active protein into the mouse." *Science* 285(September 1999): 1569-72.
- Scott,W. and W.F.van Gunsteren "Methods and techniques in computational Chemistry." *METECC-95*. E.Clementi and G.Corongiu, eds., 397-434. Cagliari: 1995.
- Shoeb,H.A. et al. "Peroxidase-mediated oxidation of isoniazid." *Antimicrob. Agents Chemother.* 27(March 1985): 399-403.
- Silvestrelli,P.L. et al. "Maximally-localized Wannier functions for disordered systems: application to amorphous silicon." *Solid State Commun* 107(1998): 7.

- Silvestrelli, P.L., M. Bernasconi, and M. Parrinello "Ab initio infrared spectrum of liquid water." *Chem. Phys. Lett.* 277(1997): 478-82.
- Silvestrelli, P.L. and M. Parrinello "Water molecule dipole in gas and in the liquid phase." *Phys. Rev. Lett.* 82(1999): 3308-11.
- Singh, U.C. and P.A. Kollman *J. Comp. Chem.* 7(1986): 718.
- Sodroski, J. et al. "Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat." *Science* 227(January 1985): 171-3.
- Sulpizi, M. and P. Carloni "Cation - pi versus OH interactions in proteins: a density functional theory study." *J. Phys. Chem. B* 104(2000): 10087-91.
- Sumner-Smith, M. et al. "Critical chemical features in trans-acting-responsive RNA are required for interaction with human immunodeficiency virus type 1 Tat protein." *J Virol* 65(October 1991): 5196-202.
- Swope, W.C. et al. *J Chem Phys* 76(1982): 637.
- Tan, R. and A.D. Frankel "Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine." *Biochemistry* 31(October 1992): 10288-94.
- Tironi, I.G., R. Sperb, and W.F. van Gunsteren "A Generalized Reaction Field Method for Molecular Dynamics Simulations." *J. Chem. Phys.* 102(1995): 5451-9.
- Troullier, N. and J.L. Martins "Efficient pseudopotentials for plane-wave calculation." *Phys. Rev. B* 43(1991): 1943-2006.
- Truant, R. and B.R. Cullen "The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals." *Mol. Cell Biol.* 19(February 1999): 1210-7.
- van Gunsteren, W.F. et al. *Biomolecular Simulation: the Gromos 96 manual and user guide.* Zurich: Vdf Hochschulverlag AG and der ETH Zurich, 1996.



Wang, W. et al. "Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions." *Annu Rev Biophys. Biomol. Struct* 30(2001): 211-43.

Wang, Z. and T.M. Rana "RNA-protein interactions in the Tat-trans-activation response element complex determined by site-specific photo-cross-linking." *Biochemistry* 37(March 1998): 4235-43.

