

UNIVERSITÀ DEGLI STUDI DI NAPOLI “FEDERICO II”

DOTTORATO DI RICERCA IN SCIENZE CHIMICHE – XXII CICLO

Physico-chemical characterization of the interaction
between thyroid transcription factor 1 (TTF-1) and DNA

Dottoranda

Paola Carullo

Tutore

Prof. Pompea Del Vecchio

Relatore

Prof. Piero Pucci

Coordinatore

Prof. Lucio Previtiera

CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Forces involved in protein and nucleic acid interaction.....	4
1.2 DNA binding motif.....	7
1.3 TTF-1 and Pax-8 transcription factors.....	11
1.4 References.....	14

CHAPTER 2: METHODS

2.1 Circular Dichroism.....	16
2.2 Fluorescence.....	18
2.3 Differential scanning calorimetry.....	20
2.4 Isothermal Titration Calorimetry.....	23
2.5 Molecular Dynamics methods.....	26
2.6 References.....	29

CHAPTER 3: CONFORMATIONAL STABILITY AND DNA BINDING ENERGETICS OF THE THYROID TRANSCRIPTION FACTOR 1 HOMEODOMAIN (TTF-1HD)

3.1 Abstract.....	30
3.2. Introduction.....	30
3.3. Experimental.....	32
3.3.1. Materials.....	34
3.3.2. Differential Scanning Calorimetry.....	35
3.3.3. Circular Dichroism.....	36
3.3.4. Isothermal Titration Calorimetry.....	36
3.3.5 Fluorescence.....	36
3.3.6. Construction of Complexes and Molecular Dynamics Simulations.....	
3.4. Results.....	37
3.4.1. Conformational Stability.....	37
3.4.2 Fluorescence measurement.....	42
3.4.3. Binding Energetics.....	42
3.4.4. Molecular Dynamics Simulations.....	47

3.5. Discussion.....	50
3.6. References.....	53
3.7 Supplementary Material - Figure of Chapter 3.....	57

CHAPTER 4: CHARACTERIZATION OF TTF-1 MUTANT FORMS

4.1 Introduction.....	62
4.2 Experimental.....	63
4.2.1 Materials.....	63
4.2.2 Circular Dichroism.....	65
4.2.3 Isothermal Titration Calorimetry.....	65
4.3 Results.....	66
4.3.1 Single mutants: Q50A, Y54A and W48A.....	66
4.3.2 Deleted mutant: desTTF-1HD.....	72
4.3.3 Recombinant fusion protein : TTF-1NH ₂ HD.....	73
4.4. References.....	76

CHAPTER 5: THERMAL STABILITY OF PAX-8 PRD AND DNA BINDING ENERGETICS

5.1 Introduction.....	78
5.2 Experimental.....	79
5.2.1 Materials.....	79
5.2.2 Circular Dichroism.....	80
5.2.3 Isothermal Titration Calorimetry.....	80
5.3 Results.....	81
5.3.1 Thermal stability of the protein, the DNA and the complex in different buffer conditions..	81
5.3.2 Protein structural changes upon DNA interaction.....	85
5.3.3 Energetics of the protein/DNA interaction.....	85
5.4 Discussion.....	87
5.4.1 Analysis of the enthalpic and the entropic contributions to the binding.....	88
5.5 References.....	90

CONCLUSIONS AND PERSPECTIVES.....	92
-----------------------------------	----

PUBLICATIONS.....	95
-------------------	----

CHAPTER 1

Introduction

1.1 Forces involved in protein and nucleic acid interaction

Specific binding is fundamental for the molecular organization of living matter. The molecular interactions determine and regulate the cellular machinery. Virtually, all biological phenomena depend on molecular recognition, which either is intermolecular as in ligand binding to a macromolecule and in the formation of macromolecular complexes, or intramolecular as in protein folding. It is clear that association of several independent molecular units into a single complex is driven by the same secondary forces responsible of protein folding, van der Waals contacts, oriented hydrogen bonds, electrostatic and hydrophobic interactions.

Characterization of these interaction necessarily involves investigation of the inter-relationship between function, structure (including dynamics), kinetics, and energetics (i.e. thermodynamics) of a system under defined physico-chemical conditions. Since all reversible biomolecular interactions are based on a redistribution of non-covalent bonds, the measurement of thermodynamic parameters is important. The thermodynamic quantity occurring when a protein goes from the free (unbound) to the bound state is the heat (enthalpy) uptake or release.

The structure and function of site-specific DNA binding proteins are matters of considerable interest because of the crucial roles of these proteins in the expression and regulation of genetic information. From a purely structural viewpoint, it is now evident that nature has devised a wide range of strategies to build proteins that recognize a specific DNA base sequence. Protein recognition of a DNA sequence may involve nonpolar contacts and direct or indirect (i.e., water-mediated) hydrogen bonds with the DNA bases as well as direct and water-mediated contacts with the sugars and phosphates of the DNA backbone. These recognition contacts may take place in the DNA major groove, in the minor groove, or in both grooves simultaneously and may be made by α helices, β sheets, or loops [1].

The quantitative description of forces that govern the formation of biomolecular complexes is of primary importance to understand what drive the molecules to interact with each other. Understanding the forces driving the formation of protein/DNA complexes requires measurement of the Gibbs energy of binding, $\Delta_b G^\circ$, and its enthalpic, $\Delta_b H^\circ$, and entropic, $\Delta_b S^\circ$, contributions. Binding constants, K_b (i.e. $\Delta_b G^\circ$), are best obtained by optical methods such as fluorescence while isothermal titration calorimetry (ITC) is the only technique currently available that can be used to direct measure the interaction enthalpy. To perform a site-specific biological function, a protein

must bind significantly more tightly to its DNA recognition site than to competing non-specific DNA. Equilibrium binding constants (intrinsic K_b) for nonspecific DNA are generally in the range of 10^3 – 10^6 M^{-1} [2], while for specific DNA are about 1000-fold higher. At the opposite end of the scale, binding affinity must not be excessive in order that binding may be reversible.

The formation of the recognition contacts, that are, hydrogen-bonds, ion-pairs, and nonpolar contacts between protein and DNA, is expected to be enthalpically favorable for all systems, although the magnitude of this favorable $\Delta_b H^\circ$ certainly varies somewhat for structurally different interfaces. Electrostatic forces are interactions between groups of opposite charge and they are called usually “salt bridges”. They are formed between the ionized phosphates of nucleic acids and either the ϵ -ammonium group of lysine, the guanidinium group of arginine, or the protonated group of imidazole of histidine in the protein. Compared to other forces between proteins and nucleic acids, salt bridges are relatively long range and rather insensitive to the relative orientation of the charges. They therefore are considered to contribute primarily to non specific protein-DNA binding [3]. The force of dipole-dipole interaction is the physical basis of hydrogen bonds that are formed between a partially negatively charged oxygen or nitrogen (bond acceptor) and a partially positively charged hydrogen atom (bond donor). Nucleic acids present numerous functional groups that can serve as hydrogen bond donors or acceptors. These include oxygen atoms and amide groups of the bases as well as the phosphodiester oxygen atoms of the DNA backbone. Regarding the protein, the appropriate groups for hydrogen bonding can be provided by amides and carbonyls of the peptide backbone as well as by most amino acid side chains. Hydrogen-bonding interactions can also be mediated by water molecules whose position and orientation are fixed by simultaneous hydrogen-bonding to both the protein and the DNA. The strength of the hydrogen bond declines with the inverse third power of the donor-acceptor distance and also decreases greatly if the bond is bent, if there is an angle between the dipoles. These features account for the sensitivity of the hydrogen bonding interactions to the conformation and flexibility of the DNA and, furthermore, for its contribution to sequence-specific DNA recognition. Hydrogen bonding is one of the most abundant and important interaction in protein-DNA complexes, providing both sequence specific and non-sequence-specific contacts between protein and DNA.

Thermodynamics strategies for site-specific DNA-binding proteins are very well described in literature [2]. Interactions between protein and DNA involve large changes in the organization of water at the surfaces of these macromolecules. These changes are caused by water-macromolecule and water-water interactions. Water associated with polar surfaces is hydrogen-bonded to donor or acceptor groups on the surface. Removal of such water requires the breaking of these hydrogen bonds and their replacement with hydrogen bonds to other water molecules. Polar surfaces apposed

at protein-DNA interfaces may entrap water molecules or may be relatively “dry” [4] (i.e., desolvated), this desolvation of polar groups would make a significant unfavorable contribution to $\Delta_b H^\circ$. The desolvation of nonpolar surfaces makes only a small contribution to $\Delta_b H^\circ$ at 298 K [5], although it is a dominant source of favorable $\Delta_b S^\circ$. The entropy changes caused by the displacement of bound water constitute an even larger contribution to the thermodynamics of many protein-DNA complexes. Water molecules near nonpolar surfaces are more constrained in motion and orientation than water molecules in bulk solution. Thus, removal of nonpolar surfaces from solution through complex formation releases water molecules into bulk solution, resulting in higher entropy. This effect promotes the association of hydrophobic groups and is therefore called the hydrophobic effect [6]. The hydrophobic effect plays a role in protein folding and stability as well as in molecular recognition. The strength of hydrophobic interactions does not arise from direct forces involving non-polar molecules but rather from the thermodynamics involved in the introduction of a non polar molecule in a polar solvent. Favorable $\Delta_b S^\circ$ in protein-DNA binding is primarily due to the release of water from nonpolar surfaces [7] and, in a smaller proportion, due to water release from polar surfaces and the redistribution of salt ions that were associated with the free DNA and protein [8]. The water-release contribution thus differs among systems because of variation in the amounts of buried nonpolar and polar surfaces and/or the residual retention of solvent at polar regions between protein and DNA. In opposition to these favorable $\Delta_b S^\circ$ terms, there is an unfavorable $\Delta_b S$ from the loss of translational and rotational freedoms of protein and DNA.

Many DNA-binding domains are either very flexible, or partly unstructured or even fully unfolded in the free state and become folded only when bound to the specific DNA target site [9]. Since DNA-binding domains are often partially unfolded at physiological temperatures, the ITC-observed enthalpy of binding may need to be corrected for the negative contribution from protein refolding. This correction is obtained by differential scanning calorimetric melting of the free DNA-binding domain. Induced folding of the protein upon binding, resulting in a favorable enthalpic binding contribution and in unfavorable reduction in conformational entropy. The DNA itself sometimes undergoes large structural deformation upon complex formation, e.g. bending [10]. In energetic terms, the view that binding specificity is simply the accumulation of specific favorable interactions between rigid binding partners is frequently not the case and conformational changes in the two components, as well as solvent exclusion and rearrangement in the binding site play an important role. There is thus a complicated energetic profile involving changes in $\Delta_b G^\circ$, $\Delta_b H^\circ$, and $\Delta_b S^\circ$ on going from the free components to the final complex. Systems that strongly distort the DNA nevertheless have net unfavorable $\Delta_b H^\circ$ as the result of molecular strain, primarily associated with the base pair destacking. In these systems coupled protein folding is less common and the strained

interface suffers less immobilization, so $\Delta_b S^\circ$ is net favorable. By contrast, systems with little DNA distortion have net favorable $\Delta_b H^\circ$, which must be counterbalanced by net unfavorable $\Delta_b S^\circ$, derived from loss of vibrational entropy and from coupling between DNA binding and protein folding. At one extreme, the low-DNA-distortion systems have very favorable $\Delta_b H^\circ$ and unfavorable $\Delta_b S^\circ$, whereas at the other extreme, the high-DNA distortion systems use favorable $\Delta_b S^\circ$ to pay for the enthalpic cost (unfavorable $\Delta_b H^\circ$) of DNA strain [10].

To characterize the thermodynamics of a binding reaction, it is necessary to determine the association Gibbs free energy, $\Delta_b G^\circ$, and its enthalpic and entropic components, $\Delta_b H^\circ$ and $\Delta_b S^\circ$, at a given reference temperature. An additional thermodynamic parameter that can readily be gain from ITC experiments, is the heat capacity change of the binding, $\Delta_b C_p$. This parameter is determined from the temperature dependence of the $\Delta_b H^\circ$. The heat capacity increment on binding, $\Delta_b C_p$, is a fundamental thermodynamic quantity which reflects the nature of forces involved in the association process (polar, non-polar) and contains information about the conformational changes in the free components on association [11]. Furthermore, $\Delta_b C_p$ is required to predict the change of these three quantities with temperature, according to the general thermodynamic relationships, thereby allowing values to be obtained at experimentally inaccessible temperatures. Experimental studies of the formation of specific DNA/protein complexes have shown that the enthalpy of this process depends on temperature, decreasing as temperature increases, i.e. the heat capacity change of association, $\Delta_b C_p$, is negative. This has been interpreted as evidence that complex formation is associated with the formation of contacts between the non-polar groups of protein and DNA, since it is known that the dehydration of these groups results in a significant heat capacity decrease. Consequently, it is widely assumed that hydrophobic forces play a dominant role in the stabilization of protein-DNA complexes.

A molecular approach to describe the biological macromolecular interaction needs to be faced taking into account both the structural description of the interaction and the energetic aspects. Combining together the two class of information is possible to understand how the system works in the expression and regulation of genetic information and what is the best strategy to efficiently interrelate with it. It is of great practical significance to modulate their activity and reverse the effect caused by mutations that are involved in human diseases.

1.2 DNA binding motif

Transcription factors (TFs) are the major regulators controlling critical cellular processes and responses to environmental conditions. TFs regulate the expression of their target genes by binding to regulatory elements in a sequence-specific manner. Thus, TFs and their DNA-binding sites are of

central importance for gene regulation, and intensive efforts have been invested in identifying TFs binding sites. TFs contain at least two parts: one directs DNA binding and the other, called the activating region, that presumably interacts with some component of the transcriptional machinery. In some cases these domains can be interchanged between proteins. This modularity allows the study of that domain independently [12].

Although it is not possible to cover the whole spectrum of DNA-binding proteins in this brief introduction, considering DNA-binding motifs as families simplifies the understanding of the structural features involved in protein-DNA binding and thus helps the classification of various types of protein-DNA recognition. On the basis of structural studies and comparisons of primary sequences, several families of DNA-binding proteins have been identified; whose members share the same DNA-binding motifs. Large, well-established families include helix-turn-helix (HTH) proteins, the homeodomains (HDs), zinc finger proteins, the steroid receptors, leucine zipper proteins and the helix-loop-helix proteins (Fig 1.1).

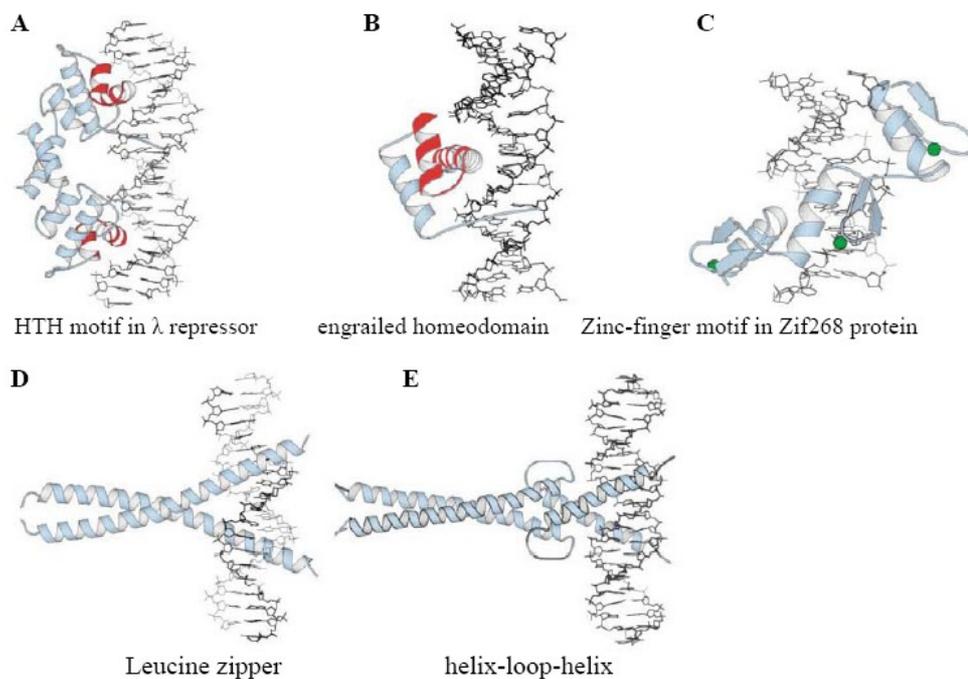


Figure 1.1 DNA-binding motifs. HTH motifs are highlighted. Small dark spheres in the Zinc-finger motif are Zn atoms

The helix-turn-helix motif is the most common and studied motif found in DNA-binding proteins in both prokaryotes and eukaryotes. This motif contains about 20 amino acids forming two α -helices interspaced by a short four-residue turn, which keeps the helices at a relatively fixed angle. The

second of the two α -helices, referred to as the recognition helix, inserts into the major groove and forms both base and sugar-phosphate backbone contacts (Fig. 1.2).

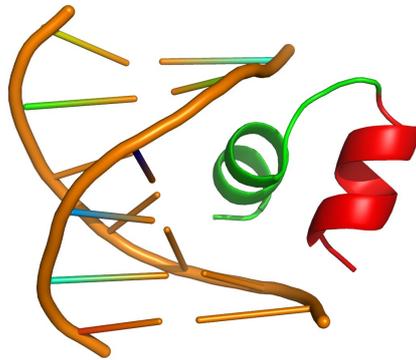


Figure 1.2 Example of a DNA-binding helix-turn-helix motif as found in the lambda repressor (PDB entry 1LMB): the recognition helix is green.

The first helix, while not embedded in the major groove, in some cases makes additional DNA contacts. It is worth noting that the HTH motif, unlike many other motifs, is not a separate stable unit. The HTH motif cannot fold or function by itself but always occurs as part of a larger DNA-binding domain. A good example of such a domain is the homeodomain, which is sometimes considered a separate family of DNA-binding motifs. The homeodomain typically has 60 amino acids and folds into three α -helices, of which the second and third helices resemble the HTH motif (Fig.1.3).

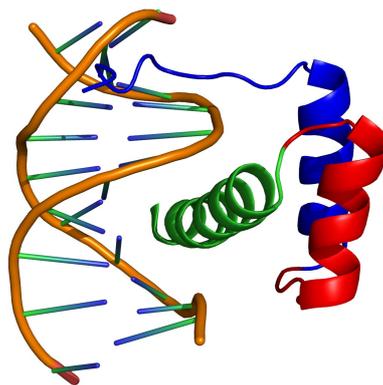


Figure 1.3 Example of a DNA-binding homeodomain motif as found in the vnd/NK-2 homeodomain-DNA complex (PDB entry 1NK3): the recognition helix is green.

Unlike the isolated HTH unit, the homeodomain forms a stable folded structure and can bind DNA by itself [13]. The third helix, which is more extended, is positioned into the major groove of DNA, making most of the specific contacts. The rest of the critical contacts are achieved by the extended

N-terminal arm, which embraces the DNA and fits into the minor groove. Another example of such a domain is the more elaborate paired domain: its contains two structurally independent, globular sub-domains that fold in two independent HTH motif. (Fig 1.4)

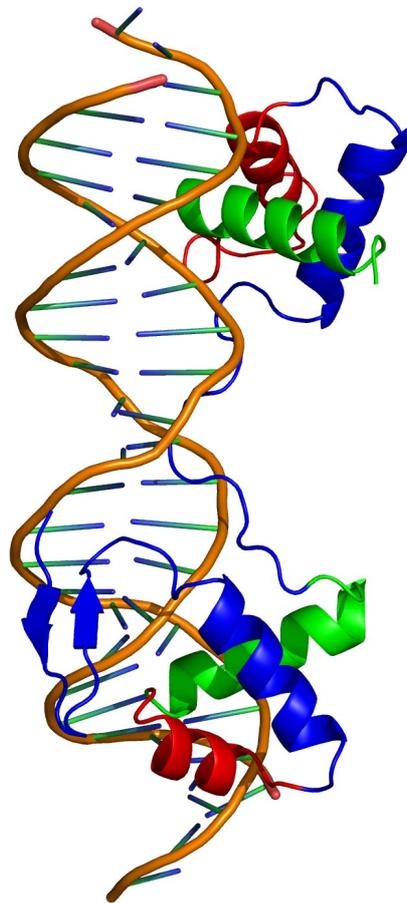


Figure 1.4 Example of a paired domain-DNA complex as found in the human Pax6 (PDB entry 6PAX).

Diversity of known DNA-binding motifs and contacts suggests that there are no simple rules or patterns for describing site-specific recognition. Nevertheless, it is possible to make some generalizations comparing known structures of protein-DNA complexes. Specific recognition always involves a set of contacts with the bases and with the DNA backbone [14]. Most of critical contacts are achieved by the protein side chains via hydrogen bonds, which are very important for site-specific recognition although other types of interaction occur. Contacts with DNA backbone also involve salt bridges with the phosphodiester oxygens. These contacts may serve as “reference marks” that stabilize the protein against the bases and thereby enhance the specificity of the side chain-base interactions. There is no simple recognition “code” consisting of a one-to-one correspondence between protein residues and the bases of DNA they contact. Many amino acid

side-chains can interact with more than one type of base and any given type of base can be contacted by different side-chains. Often, more than one side-chain contacts a given base, and in other instances a single side-chain may contact more than one base pair simultaneously. The DNA-binding motifs not only have a conserved folding, but in many cases also incorporate a conserved docking mechanism and conserved set of contacts [15], it is thought that folding and docking of entire protein can help to control the purpose that any particular side chain may have for specific DNA recognition [16]. This may explain why most of DNA-binding proteins dock into the major groove of DNA and although a number of proteins exist that specifically contact the DNA in the minor groove only.

1.3 TTF-1 and Pax-8 transcription factors

The thyroid transcription factor 1 (TTF-1) is a tissue-specific transcription factor involved in the development of thyroid and lung. TTF-1 activates the transcription of the thyroglobulin (Tg) and thyroperoxidase (TPO) genes in follicular thyroid cells but, very little is known about this mechanism. TTF-1 is a 371 amino acids protein that exhibits two independent transcriptional activation domains, located at the N- (N domain of 159 amino acids) and C-terminal (C domain of 145 amino acids) regions with respect to the DNA-contacting homeodomain (HD) that binds specifically the DNA sequence (Fig 1.5) [17].

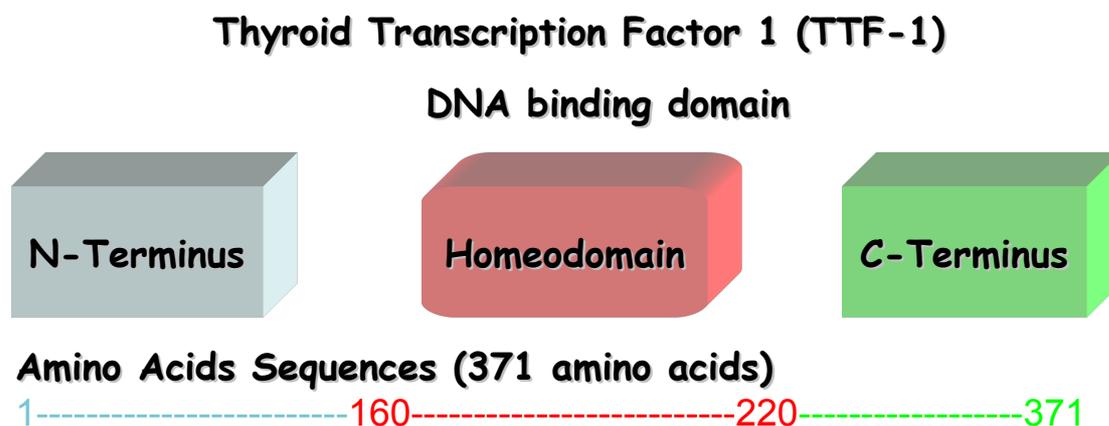


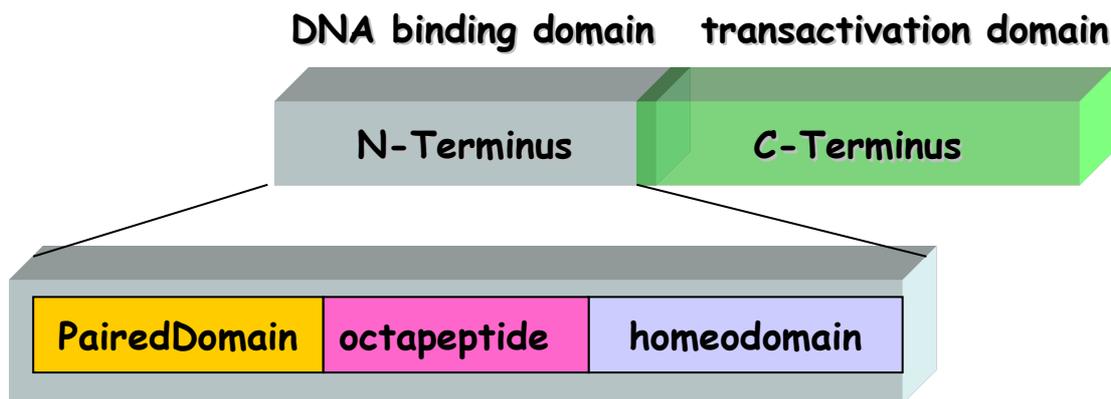
Figure 1.5 Schematic representation of TTF-1 protein organization. This protein consist of a two transcriptional activation domain, (N-terminal and C-terminal) with respect to homeodomain.

Although the N- and C- domains appear to be partially redundant, a functional hierarchy occurs between them. In fact, competition experiments revealed that the N domain, but not the C- domain, when present in a form unable to bind the target promoter, is able to repress both its own

transcriptional activity and that of the C-domain. These data would indicate that the N-domain interacts with factor(s) essential for the activation of the basal transcriptional machinery [18].

In thyroid follicular cells, three transactivator have been described: TTF-1, TTF-2 and Pax-8 a paired-domain transcription factor. Pax-8 is a transcription factor belonging to the PAX genes superfamily and its crucial role has been proven both in embryo and in the adult organism. The Pax family shares a bipartite functionality consisting of a N-terminal binding region and a C-terminal transactivation region. The N-terminal region is usually comprised of three domains, namely a Paired Box (Prd) domain, a conserved octapeptide, and a further non-functional homeodomain. (Fig 1.6) The Prd domain consists of a well conserved 128-residue-long region, formed by two distinct subdomains known as PAI (N-terminal) and RED (C-terminal). DNA binding studies have demonstrated that the two subdomains of the Pax-8 Prd domain bind DNA independently [19].

Paired-domain Transcription Factor (Pax-8)



Amino Acids Sequences (450 amino acids)

9-----135 Paired Domain

Figure 1.6 Schematic representation of Pax-8 protein organization. This protein has a dualistic organization, with a N-terminal DNA binding region and a C-terminal transactivation region. In the second cartoon is shown a magnification of the N-terminal region, showing the paired box domain, the evolutionary conserved octapeptide, and the non-functional homeodomain.

Despite the presence of Pax-8 and TTF-1 also in a few other tissues, the simultaneous expression of the two transcription factors occurs only in the thyroid, supporting the idea that Pax-8 and TTF-1 might cooperate to influence thyroid-specific gene expression. Pax8 and TTF-1 when co-expressed in HeLa cells synergistically activate Tg and TPO genes transcription. By mutational analysis was

determined the regions of Pax-8 and TTF-1 proteins that are involved in the functional cooperation. The synergism requires the N-terminal activation domain of TTF-1, and study on deleted forms of Pax-8 indicate that the C-terminal domain of the protein is involved [20]. Furthermore, it was recently claimed that Pax-8 and TTF-1 might cooperate in the regulation of thyroid-specific gene expression. Interestingly, the binding site of Pax-8 on the Tg promoter overlaps with the one of the TTF-1-binding sites and that, at least in vitro, the two factors cannot bind together to the same DNA region. The results provide the evidence of a direct interaction between Pax-8 and TTF-1, which occurs via protein-protein interaction in the absence of DNA. The molecular mechanisms leading to the synergistic activity of Pax-8 and TTF-1 implicate a region of the Tg promoter, containing 5'-CAAG-3' sequence, which is where the binding sites of the two factors overlap. It is, however, still unclear if the two proteins are both binding to the DNA or if only one of them is binding and thus recruits the other one to the promoter (Fig. 1.7) [21].

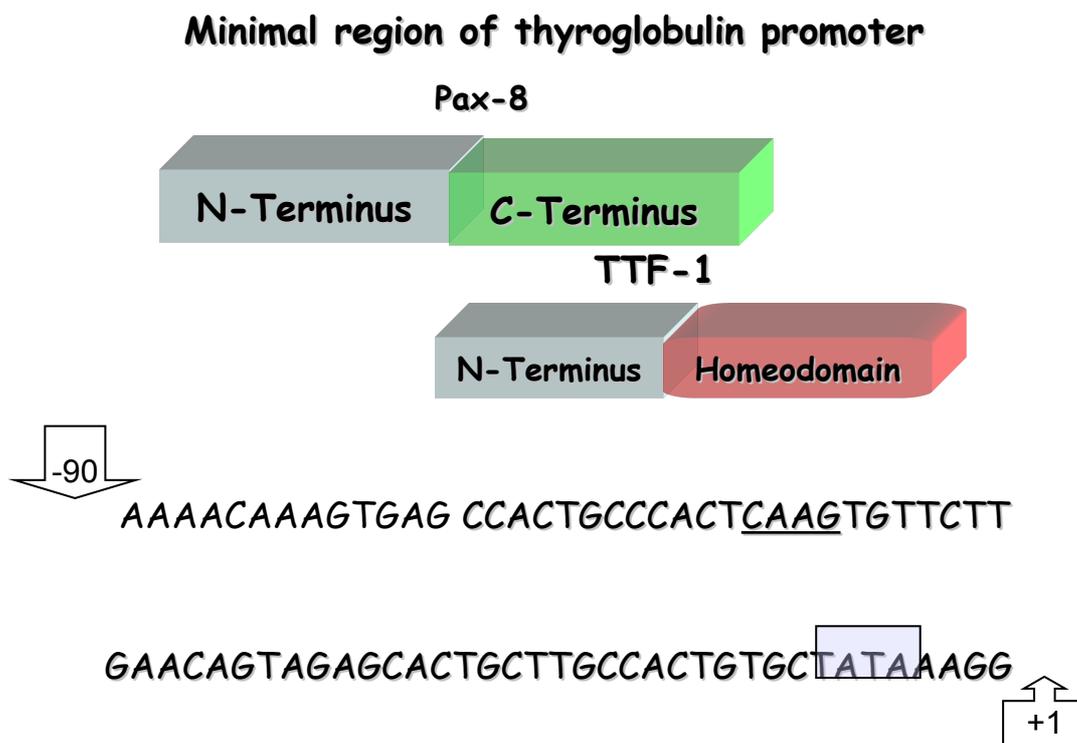


Figure 1.7 Schematic representation of Pax-8 and TTF-1 regions involved in DNA binding (Pax-8 N-terminus and TTF-1 homeodomain) and in protein-protein interaction (C-terminal region of Pax-8 C-terminal and N-terminal region of TTF-1) In the second cartoon is shown the nucleotide sequence of minimal region of thyroglobulin promoter.

In this thesis has been performed a physico-chemical characterization of the proteins TTF-1 and Pax-8. In particular has been investigated on the energetics of DNA interaction and conformational

stability of their DNA binding regions: the homeodomain of TTF-1 and the paired domain of Pax-8. The study has been performed by using spectroscopic (CD and Fluorescence) and calorimetric techniques (ITC and DSC). In addition, a structural model of the complex between TTF-1 homeodomain and its target DNA, generated by combining homology modeling and molecular dynamic simulations, has provided detailed view of interaction.

1.4 References

1. Jones S, van Heyningen P, Berman HM, Thornton JM. Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* 1999; 287:877–896.
2. Jen-Jacobson L. Protein-DNA recognition complexes: conservation of structure and binding energy in the transition state. *Biopolymers* 1997 ; 44:153–180.
3. Kerppola TK. Protein-DNA Interactions: Structure and Energetics. *Encyclopedia of Life Sciences*, 2002; Macmillan Publishers Ltd
4. Janin J. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure* 1999; 7:277–279.
5. Makhatadze GI, Privalov PL. Energetics of protein structure. *Adv. Prot. Chem.* 1995; 47:307–425.
6. Stoner CD Inquiries into the Nature of Free Energy and Entropy in Respect to Biochemical Thermodynamics. *Entropy* 2000; 2:106-141
7. Spolar RS, Record MT Jr. Coupling of local folding to an site-specific binding of proteins to DNA. *Science.* 1994; 263:777–784.
8. Record MT Jr, Zhang W, Anderson CF. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv. Prot. Chem.* 1998; 51:281–353.
9. Patikoglou G, Burley SK. Eukaryotic transcription factor–DNA complexes *Ann. Rev. Biophys. Biomol. Struct.* 1997; 26:289 – 325.
10. Love JJ, Li X, Case DA, Giese K, Grosschedl R, Wright PE. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 1995; 376:791 – 795.
11. Murphy KP, Freire E. Thermodynamics of structural stability and cooperative folding behaviour in proteins . *Adv. Protein Chem.* 1992; 43:313 – 361.
12. Ptashne M Two "What If" Experiments *Cell* 2004; 116:71-72.

13. Qian YQ, Billeter M, Otting G, Müller M, Gehring WJ, Wüthrich K. The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*. 1989; 59:573-80
14. Pabo CO, Sauer RT. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*. 1992; 61:1053-95
15. Pabo CO, Aggarwal AK, Jordan SR, Beamer LJ, Obeysekare UR, Harrison SC. Conserved residues make similar contacts in two repressor-operator complexes. *Science* 1990; 247: 1210-1213
16. Pabo CO, Nekludova L. Geometric analysis and comparison of protein-DNA interfaces: why is there no simple code for recognition? *J Mol Biol* 2000; 301:597-624
17. Tell G, Perrone L, Fabbro D, Pellizzari L, Puccillo C, De Felice M, Acquaviva R, Formisano S, Damante G. Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains. *Biochem. J.* 1998; 329:395-403
18. De Felice M, Damante G, Zannini M, Francis-Lang H, Di Lauro R. Redundant domains contribute to the transcriptional activity of the thyroid transcription factor 1. *J. Biol. Chem.* 1995; 270:26649-26656
19. Epstein JA, Glaser T, Cai J, Jepeal L, Walton DS and Maas RL. Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing. *Genes Dev.* 1994; 8:2022–2034
20. Di Palma T, Nitsch R, Mascia A, Nitsch L, Di Lauro R, Tannini M. The Paired Domain-containing Factor Pax8 and the Homeodomain-containing Factor TTF-1 Directly Interact and Synergistically Activate Transcription. *J. of biol. Chem.* 2003; 278:3395–3402
21. Zannini M, Francis-Lang H, Plachov D, Di Lauro R. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol.* 1992; 9:4230-41

CHAPTER 2

Methods

2.1 Circular Dichroism

Nearly all molecules synthesized by living organism are optically active. The optical activity of small molecules arise from their lack of symmetry, particularly, from the presence of asymmetric carbon atoms and from the effect these atoms have on any nearby chromophores. On the other hand, the optical activity of the biological samples derives, essentially, from the conformation of the macromolecules in solution. There are at least four ways that an optically active sample can alter the properties of transmitted light: optical rotation, ellipticity, circular dichroism and circular birefringence.

Circular Dichroism (CD) is the phenomenon that is recorded when an optical active sample interacts with the plane-polarised light, particularly, when the left circularly polarised light is absorbed in a different amount with respect to the right circularly polarised light. After passing through the sample each component is still circularly polarised, but the radii of the circle traced out by the electric vector of each are now different. When these two opposite circularly polarised light waves are combined together the result will be elliptically polarised light because the two components have different amplitude. Thus CD is equivalent to ellipticity. Recording the CD at different wavelength allows obtaining the CD spectrum that is strongly suggestive, for the biological samples, to the secondary structure content of the macromolecule.

For proteins, the major objective has been to deduce the average secondary structures of the peptide chain from measured CD spectra. Optical activity in the region of the spectrum between 190 nm and 230 nm (far-UV region) is dominated by the peptide backbone. Therefore, as approximation, one considers a protein simply as a linear combination of backbone regions with helices, sheets or random coil structures.

The different types of regular secondary structure found in proteins give rise to characteristic CD spectra in the far-UV (Fig.2.1).

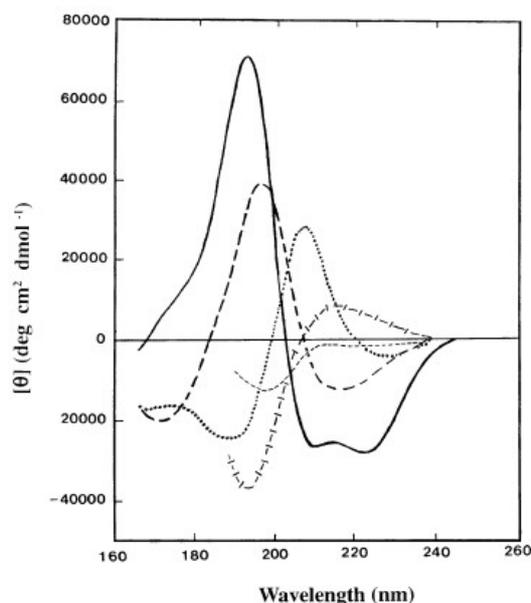


Figure 2.1 Far UV CD spectra associated with different secondary structures. Solid line, α -helix; long dashed line, anti-parallel β -sheet; dotted line, type I β -turn; short dashed line, irregular structure.

A number of algorithms exist which use the data from far UV CD spectra to provide an estimation of the secondary structure composition of proteins [1].

The spectra in the region 260–320 nm (near-UV) are a tertiary structure fingerprint and arise from the aromatic amino acids. Each of the amino acids tends to have a characteristic wavelength profile. Trp shows a peak close to 290 nm with fine structure between 290 and 305 nm; Tyr have a peak between 275 and 282 nm, with a shoulder at longer wavelengths often obscured by bands due to Trp; Phe shows weaker but sharper bands with fine structure between 255 and 270 nm. The fine structure in these bands arises from vibronic transitions in which different vibrational levels of the excited state are involved.

The actual shape and magnitude of the near-UV CD spectrum of a protein will depend on the number of each type of aromatic amino acid present, their mobility, the nature of their environment (H-bonding, polar groups and polarisability) and their spatial disposition. Although the theoretical treatment of near-UV CD spectra is not sufficiently advanced to yield significant structural insights, important progress has been made in assigning features of the spectrum to particular residues by sequential removal of aromatic side chains by site-directed mutagenesis [1].

CD spectra in the different spectral regions are invaluable for assessing the structural relationships between native and recombinant protein, and between wild-type and mutant proteins. In addition the loss of CD signals either on addition of denaturing agents (such as urea or guanidinium hydrochloride) or by an increase in temperature can be used to provide quantitative estimates of the

stability of the folded state of the native protein. CD provides an experimentally very convenient means of detecting structural changes in proteins caused by the binding of ligands, such changes which can be examined in different spectral regions. In addition CD can be used to assess the range of ligand concentrations over which structural changes take place.

A Jasco J-715 spectropolarimeter equipped with a Peltier-type temperature control system (model PTC-348WI) has been used in the present studies.

2.2 Fluorescence

Fluorescent reporter groups have served for many years as sensitive probes of macromolecular structure. Such probes can be especially useful in comparative studies such as detection of conformational changes and discrimination among structural models. Spectroscopic methods such as fluorescence are attractive because they are rapid, require small amounts of material, are nondestructive, can be carried out with commonly available equipment, and are relatively inexpensive. In addition, there is a rich library of theoretical and practical materials available to aid in data interpretation [2].

Fluorescence is defined as the emission of a photon from a singlet excited state to the ground state subsequent to absorption of a photon by a fluorophore. The fluorescence properties of a system are characterized by a number of parameters: excitation wavelength, emission wavelength, quantum yield (F), and lifetime of the excited state (t_0). The quantum yield for fluorescence is defined as the ratio of photons emitted by fluorescence to photons absorbed. Direct determination of F is generally not convenient; however, it is often sufficient to determine the relative quantum yield F_{rel} , which is the emission intensity of the system measured under defined conditions. Fluorescence is a function of the instrument (intensity of excitation source, sensitivity of the detection system, etc.) as well as the conditions of the measurement (fluorophore concentration, temperature, and solution conditions). F is most accurately obtained from integration over the wavelengths of the emission envelope; however, it may be adequate to determine F as the emission intensity at the emission wavelength maximum. Depending on the instrument, F may be given in photons per second or in arbitrary units.

The fluorescence of most part of the proteins is dominated by the tryptophan residue, and the indole nucleus of these residues is a uniquely sensitive and complicated fluorophore. The emission spectra of proteins are sensitive environment. Tryptophan residues are highly sensitive to the polarity of environment, which affects the energy levels of the first excited state with the result that the emission maximum can range from 330 nm in hydrophobic environment to 355 in water. Moreover

the tryptophan fluorescence can also be quenched by different ligand as occur in the DNA binding. The extent to which the fluorescence of a protein is quenched by DNA is proportional to the concentration of quencher. As quenching is due to the formation of a complex between the protein and DNA, the extent of quenching is proportional to the amount of bound protein. Thus by determining the extent to which the protein fluorescence is quenched when fully bound to DNA (i.e. at saturation), the fraction of bound and free protein at any point in a titration can be determined. From these data, the stoichiometry and binding constant of the interaction can often be obtained. Measurement of fluorescence has been widely used to quantify the association reactions between biological molecules. The fractional fluorescence quenching (Q_{obs}) was calculated as $(I_o - I)/I_o$, where I_o and I represent the protein fluorescence intensity at the maximum of the registered peak observed in the absence (free) and in presence (bound) of DNA, respectively. The concentration of the bound protein to DNA was calculated as:

$$[PL] = [P]_{tot} \cdot (Q_{obs}/Q_{max}) \quad (1)$$

where $[P]_{tot}$ is the total protein concentration and Q_{max} is the maximum quenching, the quenching observed when all the protein in the sample is bound to DNA, i.e. at the saturation.

An independent 1:1 binding model was used to evaluate the binding between DNA and protein:

$[P] + [L] \rightleftharpoons [PL]$. The concentration of the bound protein could be written as the difference between the total concentration of ligand in solution $[L]_{tot}$ and the fraction of ligand unbound $[L]$:

$$[PL] = [L]_{tot} - [L] \quad (2)$$

The binding constant of the reaction K_b is defined as:

$$K_b = \frac{[PL]}{[P] \cdot [L]} \quad (3)$$

The total concentrations of the protein $[P]_{tot}$ and the DNA $[L]_{tot}$ per each sample is given by:

$$[P]_{tot} = [P] \cdot (1 + K_b \cdot [L]) \quad (4)$$

$$[L]_{tot} = [L] \cdot (1 + K_b \cdot [P]) \quad (5)$$

Combining equation (4) with the equation (5) a quadratic expression is obtained which can be solved to express the concentration of the free ligand, [L], in terms of the total concentration of DNA and protein. This can be substituted into the equation (2) to give an analytical expression for [PL]:

$$[PL] = [L]_{tot} \frac{-1 - K_b([P]_{tot} - [L]_{tot}) + \sqrt{(1 + K_b([P]_{tot} - [L]_{tot}))^2 + 4K_b[L]_{tot}}}{2K_b} \quad (6)$$

Combining equations 4, 6 and 1, an expression of the intrinsic fluorescence quenching, as function of the total concentration of DNA, is obtained and that function was used to interpolate the experimental data in order to get a value for the K_b of the process.

Steady-state fluorescence measurements were made on a JASCO FP-750 spectrofluorometer equipped with a thermostated cell holders and temperature was kept constant by a circulating water bath.

2.3 Differential scanning calorimetry

Differential scanning calorimetry (DSC) is the most direct and sensitive approach for characterizing the energetics of conformational transitions in biological macromolecules. DSC is a useful technique because it leads to determine, in a single measurement, the temperature, the enthalpy and the heat capacity changes associated to the transition process. In particular, the enthalpy, being a thermodynamic potential, contains information on the states to which the system can belong within the investigated temperature range.

The principle of scanning microcalorimetry is to measure enthalpy changes of a solution of the macromolecules as a function of temperature change. This is obtained by measuring the power required to keep a sample at the same temperature as that of a reference solution as the temperature of both is increased in a linear manner.

Biological samples are usually studied in diluted solutions. Consequently, the associated thermal effects are very low, so that for detecting these phenomena highly sensitive apparatus, called micro DSC, can be used. The Setaram micro DSC has been used in the present studies. The calorimetric block of the micro DSC consists of a gold-plated metal cylinder with high thermal conductivity (Fig. 2.2). This assembly ensures excellent temperature uniformity with accurate control, which in turn determines the very good baseline stability in the calorimeter. Two hollows are machined in the block to accommodate the experimental vessel. The micro DSC is equipped with two closed steel

vessels (800 μL) that are removable and can be easily cleaned. The "measurement" vessel holds the sample to be analyzed. The "reference" vessel contains buffer to compensate for the thermal effect associated with heating of the samples. Various types of experimental vessels can be used depending on the applications. The micro DSC features a three-dimensional sensor with Joule effect calibration for highly sensitive and precise calorimetric measurements. Each vessel is surrounded by very high sensitivity Peltier elements ensuring the thermal link with the calorimetric block. These detectors are good thermal conductors that keep the temperature in the vessels identical to that in the calorimetric block. The heat-flow transducer enables the micro DSC to reach a very high sensitivity detection limit of 0.2 mW with very good measurement accuracy. The temperature of the calorimeter is set by thermostatic control of a circulating liquid which is heated or cooled by the controller from -20°C to $+120^{\circ}\text{C}$. External water circulation removes the heat. The transducer zone is protected by thermal buffers that prevent outside interference but provide direct access to the vessels. Using a liquid thermostat and precisely controlling the temperature reduces the effect of outside temperature fluctuations and allows the accurate measurement of very small thermal effects. The calorimeter was interfaced to an IBM PC computer for automatic data collection and analysis using previously described software [3].

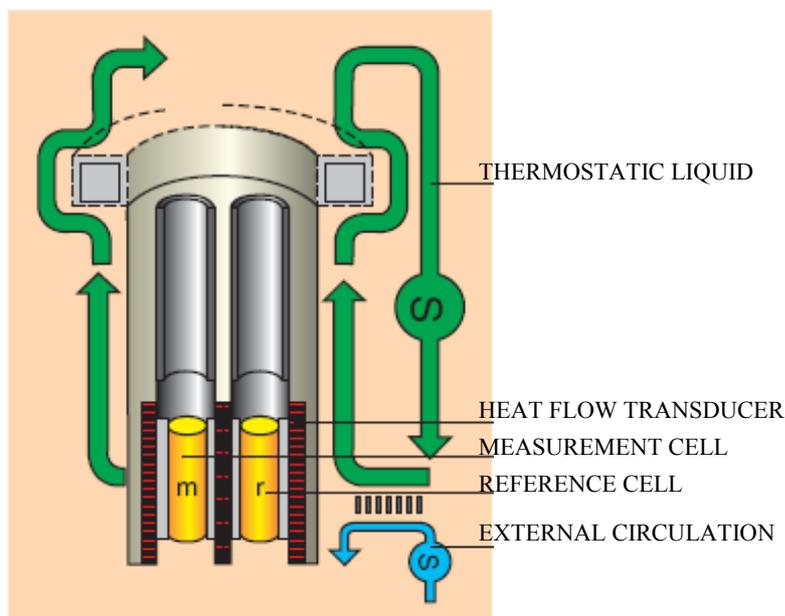


Figure 2.2 Calorimetric block.

For biological solutions, the sample normally contains the macromolecule in buffer and the reference is the buffer solution. As the temperature changes, the macromolecule undergoes a cooperative transition. This transition (protein denaturation or DNA dissociation) arises from the destruction of the numerous small forces that stabilize the native structure. Such disruption changes

the enthalpy of the system giving rise to a drop in temperature, because the process is usually endothermic. The calorimeter will provide energy to the sample to maintain its temperature at the same value as that of the reference solution. The energy is measured as a power input and the raw output is a data set of power versus temperature. Power is easily converted to the apparent molar excess heat capacity using the following equation:

$$\langle \Delta_d C_p \rangle = \frac{P}{\sigma m}$$

where $\langle \Delta_d C_p \rangle$ is the apparent excess heat capacity in $\text{J mol}^{-1}\text{K}^{-1}$, P is power in J s^{-1} , σ is the scan rate in K s^{-1} and m is the number of moles of the protein in the sample. Subtraction of the buffer baseline corrects the data for the partial molar heat capacity of the solvent, allowing the partial molar heat capacity of the macromolecule to be determined. Thus, a typical calorimetric curve reports the excess heat capacity function, $\langle \Delta_d C_p \rangle$, versus temperature. The experimental transition enthalpy, $\Delta_d H(T_d)$, is obtained by integrating the area under the curve. The transition midpoint (T_d , or the melting temperature) is the temperature corresponding to the maximum of DSC peak. $\Delta_d C_p$ is the overall denaturation heat capacity, i.e. the difference between the denatured and native macromolecule heat capacity.

The $\Delta_d H(T_d)$ is actually a net value from a combination of endothermic contributions, such as the disruption of hydrogen bonds, and exothermic processes such as the disruption of hydrophobic interactions. The sharpness of the transition peak is indicative of the cooperative nature of the unfolding process. If the unfolding produces a narrow, symmetric peak, the transition is close to a two-state, reversible and highly cooperative process [4]. However, it is frequently convenient to compare the directly measured enthalpy with indirect estimates using the classical van't Hoff equation:

$$\frac{d(\ln K(T))}{dT} = \frac{\Delta_d H^{v.H.}}{RT^2}$$

The van't Hoff analysis is based on a hypothesis of model of the process involved. In a two-state thermal denaturation process, the $\Delta_d H(T_d)^{v.H.}$ is the enthalpy change per mole of cooperative units, as defined by the model. In this case, the ratio between the values of $\Delta_d H(T_d)$ and $\Delta_d H(T_d)^{v.H.}$ is close to unity. If the unfolding transition is not two-state, but involves one or more intermediates, the transition appear broader than expected and $\Delta_d H(T_d)^{v.H.}$ is less than $\Delta_d H(T_d)$. If the macromolecule unfolds cooperatively as a dimer or higher oligomers, the transition appears sharper than expected for a two-state transition of a monomer and the $\Delta_d H(T_d)^{v.H.}$ is greater than $\Delta_d H(T_d)$.

In order to obtain correct thermodynamic data, the thermal transition must be fully reversible. This is verified by using the reheating criterion: reheating the sample after cooling and obtaining a second calorimetric curve superimposable on the first.

2.4 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a valuable tool for characterizing the macromolecular biological interactions, thanks to its general applicability and precision. ITC is a high-accuracy method for measuring binding affinities and stoichiometry; moreover, it is the only technique that directly measures the binding enthalpy. ITC also allows dissecting the free energy of binding into enthalpic and entropic components to reveal the overall nature of the forces that drive the binding reaction [5,6]. In ITC, one component of a complex (such as a protein) is present in the calorimeter's sample cell, and the second component (for example, a DNA) is slowly added in an incremental, stepwise fashion. Since all binding events are accompanied by the evolution or absorption of heat (a change in enthalpy), the analysis of these extremely small thermal effects arising from the binding allows a full thermodynamic characterization of the reaction and provides fundamental information about the molecular interactions driving the process. The calorimeter used for our experiments is the CSC 4200 ITC. A schematic diagram of the ITC is shown in (Fig. 2.3).

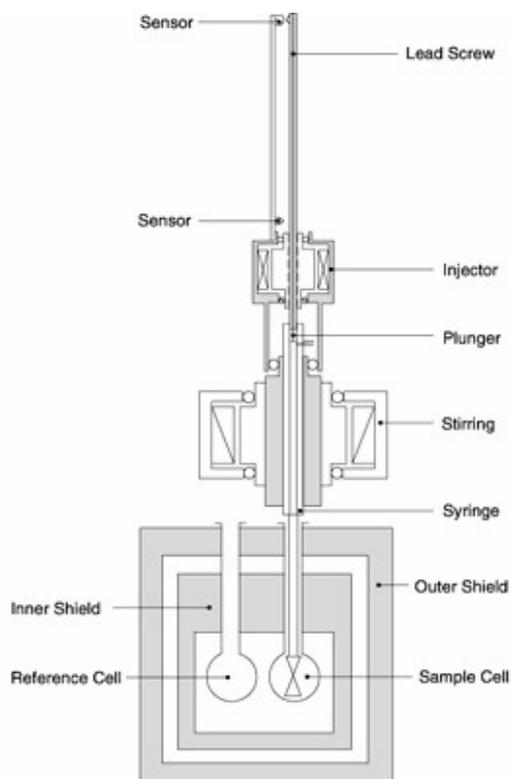


Figure. 2.3 - Schematic representation of an ITC.

The calorimeter holds two removable vessels (reference and sample cells), with a nominal volume of 1.3 ml, that are contained in an ultra-stable constant temperature bath. The reference vessel contains water or buffer solution; the sample vessel holds a solution containing one of the reactants. The two vessels are constantly kept in thermal equilibrium with the bath during the experiment. The heat flow between the reaction vessel and the isothermal block is precisely measured by thermoelectric device sensors that surround the vessel and is monitored as a function of time. The ITC measures the heat generated/absorbed when two solutions of reactants are mixed. The other reactant is contained in a precision syringe (25-250 μl) connected to the sample cell and is added in computer-controlled injections (1-20 μl). The solution in the sample vessel is stirred with a propeller to ensure rapid mixing of components. The heat change caused by any reaction occurring when the solutions are mixed, is detected by the thermoelectric device sensors, amplified and converted by an electronic output circuit to a signal corresponding to the heat change. The signal registered is a deflection peak. The CSC 4200 ITC can detect heat effects as small as 0.4 μJ , allowing titrations to be done with nmol of molecules, moreover, ITC may be run at any temperature between 0 and 110°C. In a typical ligand-macromolecule titration, the chemical reaction generated by each injection either releases or absorbs a certain amount of heat (q_i) proportional to the amount of ligand that binds to the macromolecule in a particular injection and the characteristic binding enthalpy for the reaction:

$$q_i = V \cdot \Delta_b H^\circ \cdot \Delta L_i$$

Where V is the volume of the reaction cell and is the increase in the concentration of bound ligand after the i^{th} injection [7]. The heat after each injection is therefore obtained by calculating the area under each peak. Because the amount of unbound macromolecule available progressively decreases after each successive injection, the magnitude of the peaks becomes progressively smaller until complete saturation is achieved. Once this situation is reached, subsequent injections produce similar peaks corresponding to dilution or mechanical effects that need to be subtracted from all the injection peaks before analysis.

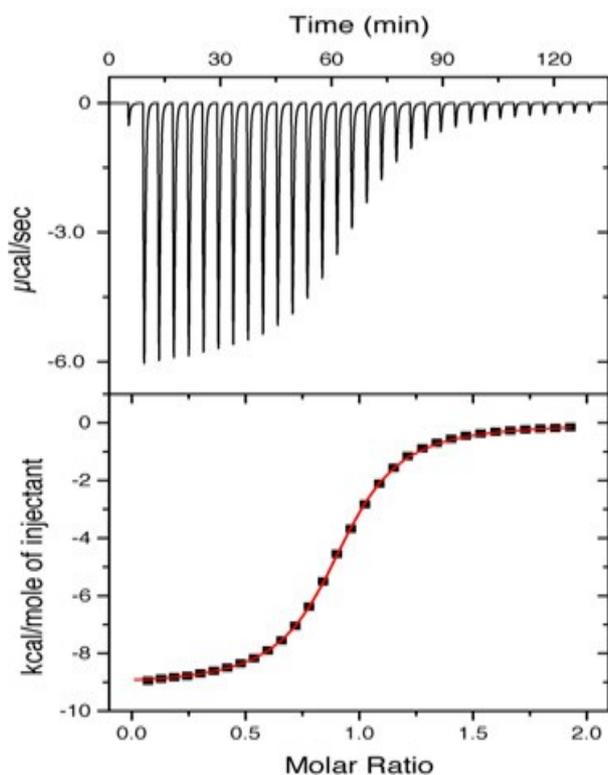


Figure 2.4 Signal produced by the sequence of injections for a typical ITC experiment (top Panel). Data after integration of each injection peak (bottom Panel).

The top panel of Figure 2.4 depicts the signal produced by the sequence of injections, and the bottom panel shows the data after integration of each injection peak. The sigmoidal shape of the bottom panel Figure 2.5, with numerous data points throughout the curved rise portion of the plot, facilitates estimation of the midpoint of the transition, and thus the stoichiometry of the binding reaction (in this case, 1:1). K_b , n (stoichiometry) and $\Delta_b H^\circ$ are calculated by iterative approximation. A value for K_b is initially estimated then the concentration of bound complex is calculated for each injection. In combination with the measured heat, these values are used to determine the average of $\Delta_b H^\circ$. The $\Delta_b H^\circ$ and the calculated concentration are then used to determine an expected heat per injection, and the error square sum between the measured and expected heat for each peak is calculated. The value of K_b is then adjusted and the process repeated until a minimum error square sum is obtained. The values of K_b , $\Delta_b H^\circ$ and n can often be calculated from a single experiment as long as the concentration of both macromolecule and ligand are accurately known [8] and chosen so that:

$$10 < K_b [M]_T < 1000$$

where $[M]_T$ is the total concentration of macromolecule in the sample cell titrated by ligand. If concentrations are not within this range, the curvature of the titration plot can be so low as to be

almost linear, or so high as to produce a step-like profile. In these cases K_b may not be estimated accurately. Typically, macromolecule concentrations in the order of 10-100 μM are used; permitting K_b values in the range $10^2 - 10^9 \text{ M}^{-1}$ to be accurately estimated. Since temperature (T) is held constant throughout the entire experiment, the free energy ($\Delta_b G^\circ$) of the binding reaction can be determined by:

$$\Delta_b G^\circ = -RT \ln K_b$$

where R is the gas constant. ITC directly measures $\Delta_b H^\circ$, so the change in entropy ($\Delta_b S^\circ$) can be determined by:

$$\Delta_b S^\circ = \frac{\Delta_b H^\circ - \Delta_b G^\circ}{T}$$

Quantification of these thermodynamic parameters reveals the physical processes involved in the binding reaction. A spontaneous binding process must have a negative $\Delta_b G^\circ$, and $\Delta_b G^\circ$ will become increasingly negative as binding becomes tighter. As seen above, free energy changes have both an enthalpic and entropic component. The enthalpic contribution to binding is primarily due to an increased number of hydrogen bonds at the ligand-target interface, and to more favourable van der Waals interactions between the two interacting molecules; the hydrophilicity of the system will determine how important electrostatic, polar and dipolar interactions will be in driving the reaction. The entropic contribution has two primary components: conformational changes, such as folding or unfolding of the macromolecules, and the release of bound solvent as hydrophobic groups interact. The large number of ordered water molecules released into the bulk solvent when the hydrophobic surfaces of the ligand and target interact provides the main driving force for hydrophobic interactions. This driving force is sufficient to compensate for the unfavourable conformational entropy of the macromolecule and ligand caused by decreased conformational and rotational freedom following binding. In addition to the entropic effect, burial of surface area also affects the heat capacity of the sample, since water molecules ordered at hydrophobic surfaces have a different heat capacity from that of water that has been released into the bulk solvent following binding.

2.5 Molecular Dynamics methods

Molecular Dynamics (MD) is a computer simulation technique in which the time evolution of a set of interacting atoms is computed by integrating their equations of motion. Therefore, MD allows us to view how a molecular system evolves through time and to derive average properties of the

system, given a simulation of sufficient length. MD is particularly useful when the system cannot be studied by experimental methods.

The equations of motion can only be solved numerically for a multi-body problem. To calculate the dynamics of the system (*i.e.* the position of each atom as a function of time), Newton's classical equations of motion are solved for each atom given an empirical force field:

$$F_i = m_i a_i$$

where F_i is the force exerted on particle i , m_i is the mass of particle i and a_i is the acceleration of particle i . The force on each atom is the negative of the derivative of the potential energy (V) with respect to the position of the atom (r_1, r_2, \dots, r_N):

$$F_i = - \frac{\partial V}{\partial r_i}$$

Once the coordinates of the atoms of a starting structure and their velocities are defined, the force acting on each atom can be calculated for each point in time $t + dt$ and a new set of coordinates can be generated. The repetition of this procedure generates a molecular trajectory corresponding to the time-dependent fluctuations of the atomic positions. The accuracy of the simulations is directly related to the potential energy function that is used to describe the interactions between particles. In MD a classical potential energy function is used that is defined as a function of the coordinates of each of the atoms.

The potential energy function is separated into terms representing covalent interactions and non covalent interactions. The covalent interactions may be described by the following terms:

$$V_{bond} = \sum_{i=1}^{N_b} \frac{1}{2} k_i^b (r_i - r_{0,i})^2$$

$$V_{angle} = \sum_{i=1}^{N_\theta} \frac{1}{2} k_i^\theta (\theta_i - \theta_{0,i})^2$$

$$V_{dihedral} = \sum_{i=1}^{N_\phi} \frac{1}{2} k_i^\phi \cos(n_i(\phi_i - \phi_{0,i}))$$

$$V_{improper} = \sum_{i=1}^{N_\xi} \frac{1}{2} k_i^\xi (\xi_i - \xi_{0,i})^2$$

the equations correspond to two, three, four and four body interactions, respectively. These interactions are represented by harmonic potentials for the bond lengths r_i , for the bond angles θ_i ,

for the improper dihedral (out of the plane) angle ξ_i and by a more complex potential for the dihedral angles Φ_i . The non-covalent (non-bonded) interactions, which correspond to interactions between particles separated by more than three covalent bonds, are usually described by Coulomb's law:

$$V_{Coulomb} = \sum_{i < j} \frac{1}{4\pi \epsilon_0 \epsilon_r} \frac{q_i q_j}{r_{ij}}$$

for the electrostatic interactions (for a pair of atoms carrying the partial charges q_i and q_j), and by a Lennard-Jones potential:

$$V_{LJ} = \sum_{i < j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

for the van der Waals interactions, where r_{ij} is the atomic distance between particles i and j . The force field parameters describe the strength of the interactions. For bonded interactions parameters are defined for bond stretching, bond bending and torsional rotation. Another set of parameters determines the strength of non-bonded electrostatic and van der Waals interactions. Electrostatic interactions are generally represented by point charges located at the center of the atom.

It should be kept in mind, however, that MD is affected by several limitations [9]. Firstly, MD is computationally very demanding and the computational load scales with the square of the system size. Simulation times are currently limited to hundreds of nanoseconds or a few microseconds at most. The phenomena that can be explored must occur with sufficient statistical significance within time scales that are encompassed by the computation.

How the system evolves through time is specified by the force field and by an integration time step that determines where the atoms will be positioned at time $t + dt$. MD requires the use of a very small time-step (1-2 fs) to achieve accurate results, because small time-steps limit the approximations that are introduced by the numerical integrator. This limits the overall scope of the simulated time and the computable properties.

According to statistical thermodynamics, physical quantities are represented by averages over configurations belonging to a certain statistical ensemble. A trajectory obtained by molecular dynamics provides such an ensemble. Therefore, a measurement of a physical quantity by simulation is simply obtained as an arithmetic average of the various instantaneous values adopted by that quantity during the MD run. Statistical thermodynamics is the link between the microscopic ensembles and the macroscopic properties. In the limit of an exact force field and very long simulation times, one could expect the phase space to be fully sampled and in that limit the

averaging process would yield exact thermodynamic properties. In practice, MD runs are always of finite length and one should exert caution when judging the sampling quality. An important constraint in deriving average properties is to extract configurations only from an ensemble at thermal equilibrium. Therefore, MD simulations start generally with an equilibration phase. Once equilibrium is reached, the simulation enters the production phase. The production run should be long enough to sample the property of interest with sufficient statistical significance.

2.6 References

1. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta*. 2005; 2:119-39.
2. Lakowicz JR (1999) *Principles of fluorescence spectroscopy*. Plenum, New York, NY
3. Barone G, Del Vecchio P, Fessas D, Giancola C, Graziano G. Theseus: A new software package for the handling and analysis of thermal denaturation data of biological macromolecules. *J. Thermal. Anal.* 1993; 38:2779.
4. Bruylants G, Wouters J, Michaux CW. Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Curr Med Chem*. 2005; 17:2011-20
5. Pierce MM, Raman CS, Nall BT. Isothermal titration calorimetry of protein-protein interactions. *Methods* 1999; 2:213-21.
6. Ababou A, Ladbury JE. Survey of the year 2005: literature on applications of isothermal titration calorimetry *J Mol Recognit*. 2007; 1:4-14.
7. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol*. 2001; 5:560-6.
8. O'Brien R, Ladbury JE, Chowdry BZ. *Protein-Ligand Interactions: hydrodynamics and calorimetry*, S. E. Harding; B. Z. Chowdry, eds.; Oxford University Press: Oxford, 2001, pp. 263-286.
9. Van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glattli A, Hunenberger PH, Kastenzholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB. Biomolecular modeling: goals, problems, perspectives *Angew Chem Int Ed Engl*. 2006; 45:4064-92

CHAPTER 3

Conformational stability and DNA binding energetics of the thyroid transcription factor 1 homeodomain (TTF-1HD)

3.1 Abstract

The conformational stability of the thyroid transcription factor 1 homeodomain, TTF-1HD, has been investigated by means of circular dichroism (CD) and differential scanning calorimetry (DSC) measurements at pH 5.0 as a function of KCl concentration. Thermal unfolding of TTF-1HD is a reversible two-state transition. The protein does not possess a great stability against temperature, showing a denaturation temperature of 32°C in the absence of salt and 51°C at 75 mM KCl. The binding energetics of TTF-1HD to its target DNA sequence has been characterized using isothermal titration calorimetry (ITC) measurements, by fluorescence measurements and complemented with circular dichroism data. At 25°C, pH 5.0 and 75 mM KCl, the binding constant amounts to $1.0 \cdot 10^8 \text{ M}^{-1}$ and the binding enthalpy change amounts to $-41.4 \text{ kJ mol}^{-1}$. The process is enthalpy driven, but also the entropy change is favourable to the complex formation. Structural models of two complexes, TTF-1HD with the target DNA sequence and TTF-1HD with a modified DNA sequence, have been constructed by using as a template the NMR structure of the complex between NK-2 HD and its target DNA, and by performing molecular dynamics simulations 3.5 ns long. Analysis of these models allows one to shed light on the origin of the DNA binding specificity characteristic of TTF-1HD.

3.2. Introduction

The regulation of gene transcription is based on specific interactions between transcription factors and their target genes. These transcription factors play central roles in cell differentiation and embryo developmental processes [1]. The sequence-specific binding of the transcriptional regulator to its target DNA sequence is a fundamental event for gene regulation. Homeodomains, HDs, are highly conserved DNA-binding domains found in numerous transcription factors. They possess a well-defined tertiary structure with three α -helix segments. The helices I and III adopt a helix-turn-helix motif suitable for DNA recognition. In particular, residues in the C-terminal part of helix III specifically recognize bases in the major groove of the double helix. The N-terminal arm of HD,

containing a stretch of basic residues, known as the nuclear localization signal, interacts with the minor groove of the double helix by means of strong but non-specific electrostatic interactions.

The homeodomain of the thyroid transcription factor 1, TTF-1HD, was first described by Di Lauro and colleagues [2]; it binds specifically the DNA sequence 5'-CAAG-3' like the HDs of the NK class, instead of the more commonly recognized 5'-TAAT-3' sequence [2-4]. The 3D solution structure of TTF-1HD has been solved by NMR [5], showing that helix I comprises residues 10-21, helix II comprises residues 28-38 and helix III comprises residues 42-59. The structure of the complex between TTF-1HD and its target DNA sequence has not yet been determined. A model of the complex was proposed more than ten years ago using available structural information on different HD-DNA complexes [6]. More recently, the topology of the TTF-1HD/DNA complex has been investigated by integrating limited proteolysis and selective chemical modification experiments with mass spectrometry methodologies [7]. The DNA binding activity has been evaluated by means of gel retardation assays and methylation-ethylation interference experiments, showing structural analogies and differences between TTF-1 and Antennapedia and Engrailed homeodomains [4]. Moreover the effects of amino acid replacements on both the stability [8,9] and DNA binding specificity [10] have been determined.

In this study we provide a more detailed physico-chemical characterization of TTF-1HD conformational stability by means of circular dichroism, CD, and differential scanning calorimetry, DSC, measurements. In addition, the binding energetics is determined by means of isothermal titration calorimetry, ITC, measurements and connected to a structural model of the complex obtained by combining molecular modelling and molecular dynamics procedures, also measurement of fluorescence has been widely utilized to quantify the association reactions between homeodomain and DNA. Also we investigated of the effect of ionic strength on the stability of the homeodomain and on the binding properties. A detailed understanding of the physical mechanism of the protein-nucleic acid recognition requires the knowledge of both 3D structure of the complex and its dynamical behaviour [11,12]. Since in the present case structural data are not available, a model of the complex has been generated and subjected to a 3.5 ns long molecular dynamics simulation in explicit water. The resulting model of the complex is shown in Figure 3.1 and analysis of the trajectory has been useful to improve the understanding of the specific recognition ability of TTF-1HD.

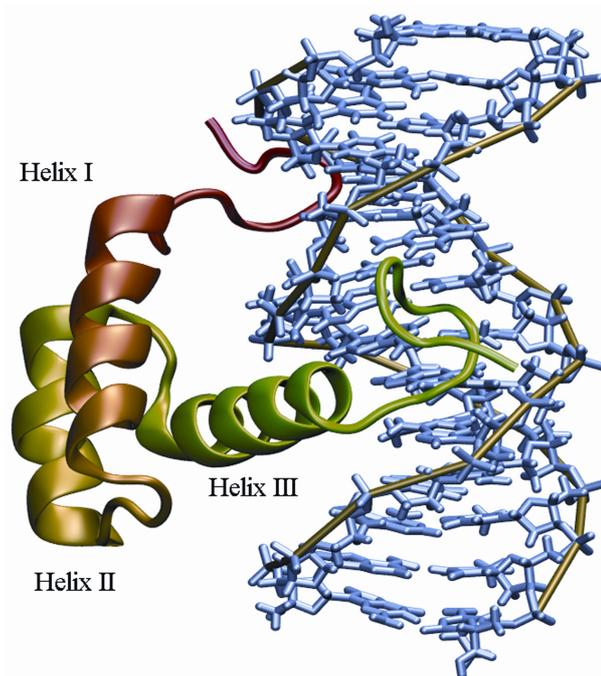


Figure 3.1 Structural model of the complex between TTF-1HD and C duplex containing the specific target sequence 5'-CAAG-3'. See text for further details.

3.3. Experimental

3.3.1. Materials.

- **Cloning and protein expression**

TTF-1 homeodomain was cloned, expressed and purified at the Department of Cellular and Molecular Biology and Pathology “L. Califano” of the University of Naples “Federico II” in collaboration with the group of Prof. Formisano. The polypeptide encompassing the TTF-1 homeodomain (68 residues corresponding to residues 160-226 of rat TTF-1 plus a methionyl residue at the N-terminus) was produced from rat cDNA [2]. The cDNA corresponding to the homeobox was amplified by PCR using appropriate oligos, the amplified DNA was cut with NdeI and PstI and cloned in pT7.7 vector. By this way an ATG and a TAA codon are positioned respectively at the 5' and 3' ends of the insert to give a translation product of 68 amino acids. Plasmids were used to transform the BL21 bacterial strain [13]. The amino acidic sequence is reported:

10	20	30	40	50
RRKRRVLF SQ	AQVYELE RRF	KQQKYL SAPE	REHLAS MIHL	TPTQVK IWFQ
60				
NHRYKM KRQA	KDKAAQQ Q			

To purifying the TTF-1 homeodomain, freshly transformed cells were grown in Luria-Bertani (LB) medium (10 g Bacto-Tryptone Difco, 5 g Bacto yeast extract Difco, 5 g NaCl in 1 liter H₂O) supplemented with 50 µg/ml ampicillin to an O.D.₆₀₀ of 0.8 at 37 °C. Cultures were induced for 2 h with 1mM Isopropyl-β-D-1-Thiogalactopiranoside (IPTG). All of the following operations were carried out at 0-4°C and were monitored by SDS gel electrophoresis. Cells were harvested by centrifugation at 9000 g in a Sorvall GS-3 rotor and washed with 50 mM phosphate buffer, pH 7.5. Cells were then resuspended in sucrose buffer (50 mM phosphate buffer, pH 7.5, 10% sucrose, 2 mM EDTA, 400 mM NaCl, 1 mM DTT, protease inhibitors) and sonicated with the intensity of 250 watt. Disrupted cells were centrifuged at 27000 g for 30 min. The crude extract was precipitated, by drop wise addition of a stock solution of 10% Polymyxin P, 50 mM phosphate buffer, pH 7.5, at a final concentration of 0.8% Polymyxin P. After stirring for 15 min, the extract was centrifuged at 27000 g for 30 min and the supernatant was collected. This fraction was directly loaded onto a BioRex A-70 (BIO-RAD cation exchange resin) column previously equilibrated in 400 mM NaCl in buffer A (50 mM phosphate buffer, pH 7.5, 10% glycerol, 1 mM DTT, protease inhibitors). The column was extensively washed with 500 mM NaCl buffer A and then eluted with a linear gradient of 500-900 mM NaCl in buffer B (50 mM phosphate buffer, pH 7.5). The peak fractions were pooled and adjusted to a final concentration of 500 mM NaCl. Minor remaining impurities present in the pooled Biorex fraction were removed by loading it onto a Pharmacia mono S 10/10-FPLC column which was then eluted with a linear gradient of 600-900 mM NaCl in buffer B [14]. Peak fractions were desalted using gel filtration (Pharmacia NAP-25 columns). Protein concentration was determined using the Bio-Rad dye reagent with bovine γ-globulin as standard, according to Bradford method (1976).

The overall final yield is 0.1-0.2 mg of pure protein per liter of culture. The purity and homogeneity of TTF-1HD were confirmed by a single band on SDS-PAGE; mass spectrometry gives a molecular weight of 8444 Da in agreement with that expected from the amino acid sequence. The purified protein solutions are exhaustively dialyzed against the appropriate buffer solution at 4 °C by using Spectra Por MW 3500 membranes; protein concentration is determined spectrophotometrically using a sequence-based extinction coefficient of 9970 M⁻¹cm⁻¹ at 280 nm [15].

- **DNA preparation**

Single stranded, 14 base oligonucleotide, 5'-CCAGTCAAGTGTTTC-3' and its complementary 3'-GGTCAGTTCACAAG-5', containing the TTF-1HD binding site 5'-CAAG-3' as part of the core, are purchased from SIGMA Genosys (Sigma Chemical Co., St.Louis, MO) and used without further purification. Lyophilized single strands are dissolved in the appropriate buffer solution and

concentrations are determined spectrophotometrically at 260 nm using extinction coefficients of $142000 \text{ M}^{-1}\text{cm}^{-1}$ and $134000 \text{ M}^{-1}\text{cm}^{-1}$ for the forward and reverse strand, respectively, calculated by the nearest neighbour model [16]. The complementary strands are hybridized to form the DNA duplex, named C, by mixing strands in equimolar ratio, heating the mixture to 90°C for 10 min and slowly cooling to room temperature. A modified DNA duplex, named Cm, obtained by hybridization of the following single stranded 14 base oligonucleotide, 5'-GAACAGATATCTGG-3' and its complementary 3'-CTTGTCTATAGACC-5', without the TTF-1HD specific binding site, is used as control. Concentrated DNA stock solutions are dialyzed against the appropriate buffer and stored at -20°C . Concentrations of diluted DNA duplex for calorimetric and spectroscopic experiments were determined spectrophotometrically at 260 nm using the extinction coefficient of $170500 \text{ M}^{-1}\text{cm}^{-1}$ [16].

Sodium acetate and phosphate buffers, KCl and other reagents of analytical grade are purchased from Sigma. Buffers for calorimetric and spectroscopic measurements are prepared with deionized and filtered through a Millipore Elix3 reagent grade system water. The pH of all solutions is determined with a Radiometer pHmeter, model PHM 93 at 25°C .

3.3.2. Differential Scanning Calorimetry. DSC measurements are carried out on a Setaram Micro-DSC III instrument, interfaced with a data translation A/D board for automatic data acquisition. A scan rate of 1.0 K min^{-1} is chosen for all experiments. Data analysis is accomplished as already described [17,18]. The van't Hoff enthalpy is calculated by the commonly used formula [19]:

$$\Delta_d H(T_d)^{\text{vH}} = nRT_d^2 \langle \Delta C_p(T_d) \rangle / \Delta_d H(T_d) \quad (1)$$

where T_d is the denaturation temperature corresponding to the maximum of DSC peak, $\langle \Delta C_p(T_d) \rangle$ is the height of the excess heat capacity at T_d , $\Delta_d H(T_d)$ is the total denaturation enthalpy change calculated by direct integration of the area of the DSC peak, R is the gas constant and n is 4 or 6 depending on the stoichiometry $N \rightleftharpoons D$ or $N_2 \rightleftharpoons 2D$. The finding that the calorimetric to van't Hoff enthalpy ratio is close to one is a necessary condition to state that the denaturation is a two-state transition [19,20]. For a two-state transition, assuming $\Delta_d C_p$ as temperature-independent, the denaturation Gibbs energy change can be calculated as a function of temperature according to the equation [19]:

$$\Delta_d G(T) = \Delta_d H(T_d) [1 - (T/T_d)] + \Delta_d C_p [T - T_d - \ln(T/T_d)] \quad (2)$$

3.3.3. Circular Dichroism. CD spectra are recorded with a Jasco J-715 spectropolarimeter equipped with a Peltier-type temperature control system (model PTC-348WI). The instrument is calibrated with an aqueous solution of D-10-(+)-camphorsulfonic acid at 290 nm [21]. The molar ellipticity per mean residue, $[\theta]$ (deg cm² dmol⁻¹), is calculated from the equation: $[\theta] = [\theta]_{\text{obs}} \cdot (\text{mrw}) / 10 \cdot l \cdot C$, where $[\theta]_{\text{obs}}$ is the ellipticity (deg), mrw is the mean residue molecular weight, 124 Da, C is the protein concentration (g mL⁻¹), and l is the optical path length of the cell (cm). Cells with 0.1 cm path length and protein concentrations of about 0.2 mg mL⁻¹ are used to record CD spectra between 190 and 250 nm, with a time constant of 16 s, a 2 nm bandwidth, and a scan rate of 5 nm min⁻¹. The spectra are signal-averaged over at least five scans and baseline corrected by subtracting a buffer spectrum, and are analyzed for secondary structure composition according to the SELCON3 method [22] using Dichroweb [23]. Thermal unfolding curves are recorded in the temperature mode at 222 nm, from 0 to 85 °C with a scan rate of 1.0 K min⁻¹. They are analyzed with the two-state N \rightleftharpoons D model whose equilibrium constant is given by [24]:

$$K_d(T) = \exp\{-\Delta_d H(T_d)^{\text{vH}}/R\} [(1/T) - (1/T_d)] \quad (3)$$

where T_d is the denaturation temperature at which $K_d = 1$ and $\Delta_d H(T_d)$ is the denaturation enthalpy change. The denaturation heat capacity change, $\Delta_d C_p$, is considered zero because it cannot reliably be determined from CD measurements [24]. The observed molar ellipticity is given by:

$$[\theta] = ([\theta]_N + [\theta]_D \cdot K_d) / (1 + K_d) \quad (4)$$

where $[\theta]_N$ and $[\theta]_D$ are the molar ellipticities of the native and denatured states, respectively, which are assumed to depend linearly on temperature. A non-linear least-squares regression was carried out to estimate the unknown parameters associated with the unfolding transition.

In order to estimate the fractional helical content from the ellipticity at 222 nm it is possible to use the following relationship provided by Richardson and Makhatadze [25]:

$$f_h = ([\theta] - [\theta]_c) / ([\theta]_h - [\theta]_c) \quad (5)$$

where f_h is the fractional helical content, $[\theta]_h$ and $[\theta]_c$ are the molar ellipticities of the full helical and full coiled states, respectively, which are given by $[\theta]_h = (-40000 + 250 \cdot T) [1 - (2.5/N_{\text{res}})]$ and $[\theta]_c = 640 - 45 \cdot T$, where T is in Celsius degrees [25].

3.3.4. Isothermal Titration Calorimetry. ITC measurements are performed on a CSC 4200 calorimeter from Calorimetry Science Corporation (CSC, Utah) with a cell volume of 1.3 mL. Protein solutions are placed in the cell, and concentrated DNA solutions in the syringe. The DNA solution is injected into the protein solution with 10 μ l aliquots at 500 s intervals. The protein concentration in the cell is about 10 μ M, and the DNA concentration in the syringe is about 100 μ M. Reverse control experiments, in which DNA solutions are titrated with TTF-1HD give identical results.

Both DNA and protein solutions are prepared with the same batch of buffer to minimize the differences in buffer composition and pH. The heat of dilution of the DNA into the solvent is measured in a separate experiment and appropriate corrections are made. The heat of dilution obtained by averaging the heats of the last 5-10 injections, and subtracted from the measured heat of binding give the same value. Raw data are integrated, corrected for non-specific heats, normalized for concentration, and analyzed assuming a single binding site by using the Bindwork program supplied with the instrument.

3.3.5 Fluorescence. The TTF-1 HD intrinsic emission fluorescence spectra were recorded by using a JASCO FP-750 spectrofluorimeter equipped with a thermostated cell holders and temperature was kept constant by a circulating water bath. The excitation wavelength was set at 290 nm using a 5nm band-width both for the emission and the excitation spectrometers. The fluorescence emission spectra were recorded between 300 and 450 nm using 1 nm steps and 1 s integration time. The experiments were performed at 5 °C by using a 1 cm sealed cell with a 500 μ L volume and corrected for background signal. Equilibrium binding measurements were performed recording fluorescence spectra at fixed concentration of protein (10 μ M) in the absence and in the presence of increasing amounts of specific DNA (0 to 20 μ M). The dissociation constant binding isotherm was obtained by using equation (6) reported in the chapter 2 incorporated into the Origin 5.0 program.

3.3.6. Construction of Complexes and Molecular Dynamics Simulations. To obtain a reasonable starting model of the complex between TTF-1HD and C duplex, the NMR structure of the vnd/NK-2 HD complexed with its target DNA sequence [26,27] (PDB code: 1NK3) is chosen as template based on the high sequence identity of both protein (74%) and DNA (57%), and on the presence of the 5'-CAAG-3' sequence in the DNA duplex (Figure 1S of Supplementary Material Chapter 3). The initial structure of TTF-1HD is generated using the coordinates of the NMR structure solved by Esposito and colleagues, PDB code: 1FTT [5]. Among the 20 NMR structures of 1FTT, we select the structure that exhibits the lowest root mean square deviation, RMSD, with respect to the 1NK3

coordinates. The DNA duplex C, containing the specific sequence and the modified DNA duplex Cm are generated with the HyperChem 7.5 software [28]. The backbone and the common residues of 1FTT are superimposed onto those of 1NK3. The backbone of each of the DNA duplexes is also superimposed onto that of 1NK3. The DNA structures are then merged with the structure of TTF-1HD. The initial structures of both complexes are then energy minimized in vacuo by 1000 steps of the steepest descent method.

All simulations are performed by means of the GROMACS package [29], employing the all-atom force field parm98 [30]. The molecules are neutralized with 14 Na⁺ ions (placed following electrostatic potential values) and solvated in a cubic box of size 70×70×70 Å³ containing about 10.700 TIP3P water molecules for each complex [31]. Initially, water molecules and ions are relaxed by a first steepest descent energy minimization with positional restraints on solute. The LINCS [32] algorithm is used to constrain the bonds and to carry out an initial simulation, 200 ps long, with the positions of the solute atoms restrained by a force constant of 3000 kJ mol⁻¹nm⁻², in order to let water molecules diffuse around the complex and to equilibrate. Simulations are carried out with periodic boundary conditions at a constant temperature of 300 K. The Berendsen algorithm is applied for temperature and pressure coupling [33], and the particle mesh Ewald method [34] is used for the calculation of electrostatic contribution to non bonded interactions (grid spacing of 0.12 nm). The trajectory length is 3.5 ns long for each complex. An MD simulation, 3 ns long, of TTF-1HD in TIP3P water has been performed in the same conditions in order to have a structural model of the free protein to be compared with the structure of the protein in the complexes.

3.4. Results

3.4.1. Conformational Stability. Far-UV CD spectra of TTF-1HD at pH 5.0, 20 mM potassium acetate and 75 mM KCl, at 0°C, 80°C and 0°C after cooling of the solution are shown in Panel A of Figure 3.2. Complete recovery of the CD signal after cooling suggests reversibility of the thermal unfolding. Analysis of CD spectra at 0 °C, by means of Dichroweb [23], indicates that the HD has a folded structure with 40% of α -helix content. The same value is obtained at 0 °C in the presence of different KCl concentrations. Previous studies on TTF-1HD reported: (a) 34% of α -helix content at 25°C from CD measurements [8]; (b) 60% of α -helix content from NMR measurements, that should be an overestimate in view of the large differences existing among the twenty conformations deposited in the PDB file 1FTT [5]; in addition, data on hydrogen/deuterium exchange of backbone peptide groups suggest that residues 51-59 in helix III are not so rigidly connected [5].

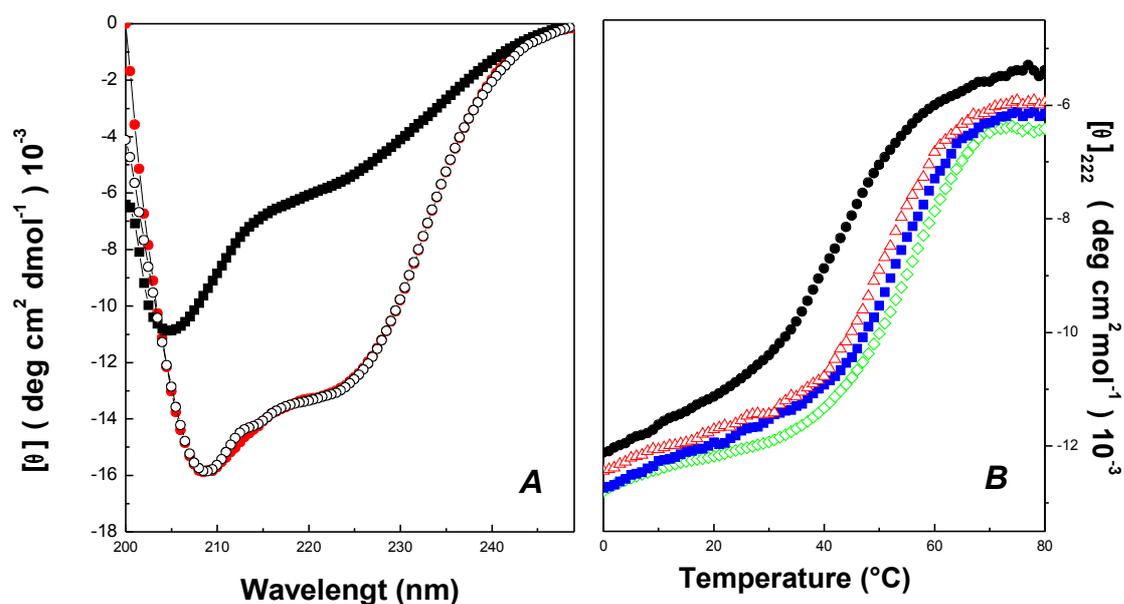


Figure 3.2 Panel A. Far-UV CD spectra of TTF-1HD at pH 5.0, 20 mM potassium acetate and 75 mM KCl, recorded at 0 °C (filled circles, red), 80 °C (filled squares, black), and at 0 °C after cooling (open circles, black). Panel B. Thermal unfolding curves of TTF-1HD, obtained by recording the molar ellipticity at 222 nm, pH 5.0, in the absence (filled circles, black) and in the presence of 75 mM KCl (open triangles red), 200 mM KCl (filled squares, blue), and 500 mM KCl (open diamonds, green).

Thermal unfolding curves are recorded in the temperature mode following the CD signal at 222 nm from 0 to 80°C with a scan rate of 1.0 K min⁻¹ in the absence and presence of increasing amounts of KCl (0-500 mM); they are shown in Panel B of Figure 3.2. Thermal denaturation curves are analyzed by means of a van't Hoff procedure, in the assumption of the two-state N \rightleftharpoons D model. Values of T_d and $\Delta_d H(T_d)^{vH}$ are listed in Table 3.1. The agreement between experimental points and calculated curves is good (not shown), indicating that the thermal unfolding of TTF-1HD is well described by the two-state N \rightleftharpoons D model. Data emphasize that the thermal stability of TTF-1HD strongly depends on KCl concentration: T_d increases from 32°C in the absence of KCl to 59°C in 500 mM KCl. This finding is in part expected because the protein has 17 positively charged (Arg + Lys) residues to strongly interact with DNA, and 5 negatively charged (Asp + Glu) residues. The thermal stability of TTF-1HD is quite independent of pH in the range 5.0 - 7.0 (data not shown), an expected result in view of the very high isoelectric point of the protein (the theoretical value is 11.1).

Table 3.1 Thermodynamic parameters for TTF-1HD unfolding determined by means of the van't Hoff analysis of the thermal transition curves obtained by recording the molar ellipticity at 222 nm in 20 mM potassium acetate, pH 5.0.

[KCl]	T _d	Δ _d H(T _d) ^{van't Hoff}
mM	°C	kJ mol ⁻¹
0	32	120
20	42	140
75	50	155
200	56	160
400	58	163
500	59	165

Note. Each values represents the value averaged over at least three measurements. The error in T_d does not exceed 0.5°C, while that in Δ_dH(T_d)^{van't Hoff} amounts to 10% of reported values.

A DSC curve of TTF-1HD recorded at pH 5.0, 20 mM potassium acetate, 75 mM KCl, is shown in Figure 3.3, curve A; a superimposable curve is obtained in the second heating of a sample previously heated up to 80°C (data not shown). Thus, according to the reheating criterion, the thermal unfolding of TTF-1HD can be regarded as a reversible process. The obtained thermodynamic values are: T_d = 51°C, Δ_dH(T_d) = 165 kJ mol⁻¹, and Δ_dC_p = 1.5 kJ K⁻¹mol⁻¹. The van't Hoff enthalpy change is Δ_dH(T_d)^{van't Hoff} = 170 kJ mol⁻¹, in agreement with Δ_dH(T_d), confirming that the thermal unfolding of TTF-1HD is well described by the two-state N ⇌ D model. A plot of Δ_dH(T_d)^{van't Hoff} versus T_d, for the different KCl concentrations investigated by means of far-UV CD, gives a good linear fit with a slope of 1.6 ± 0.1 kJ K⁻¹mol⁻¹. The latter estimate is in line with Δ_dC_p obtained from DSC curves. Thermal unfolding of C duplex is investigated by recording the CD signal at 280 nm and by means of DSC measurements at pH 5.0, 20 mM potassium acetate and 75 mM KCl. The reproducibility of both the DSC curve and the ellipticity transition curve upon cooling and reheating of the sample indicates that the thermal unfolding of C duplex is reversible. A DSC profile of C duplex, with the usual pre-melting tail, is shown in Figure 3.3, curve B; the thermodynamic parameters are: T_d = 62°C and Δ_dH(T_d) = 360 kJ mol⁻¹. They are in agreement with those derived from the van't Hoff analysis of CD transition curves, confirming that C duplex unfolding is a

bimolecular $N_2 \rightleftharpoons 2D$ two-state transition. Data indicate that at 25°C, both TTF-1HD and C duplex are in their folded conformation.

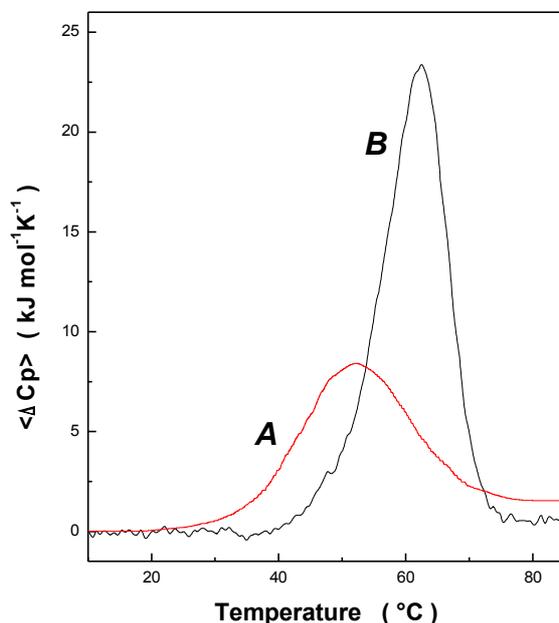


Figure 3.3 DSC profiles of TTF-1HD (curve A, red) and DNA C duplex (curve B, black) at pH 5.0, 20 mM potassium acetate and 75 mM KCl.

Thermal unfolding of TTF-1HD/C duplex (1:1) complex, at pH 5.0, 20 mM potassium acetate and 75 mM KCl, is monitored by recording the ellipticity at 222 nm and 280 nm, respectively, taking advantage of the fact that the 222 nm wavelength practically reflects the thermal transition of the protein alone, whereas the 280 nm wavelength practically reflects the thermal transition of C duplex alone. Experimental curves are shown in Figure 3.4. Comparison of the two sigmoidal curves indicates that the transition occurs always at 65 °C. Since in the same experimental conditions $T_d = 50^\circ\text{C}$ for TTF-1HD and 62°C for C duplex, the association of TTF-1HD with its target DNA duplex significantly stabilizes the protein. The sigmoidal shape of both profiles suggests that the dissociation of the complex and the unfolding of TTF-1HD and C duplex take place within a narrow temperature range in a cooperative manner. CD spectra of TTF-1HD, C duplex and their 1:1 complex, at 0 °C and pH 5.0, 20 mM potassium acetate and 75 mM KCl, are reported in Figure 3.5, together with the spectrum obtained by subtracting the spectrum of C duplex from that of the complex with the aim to single out the contribution of TTF-1HD in the complex, in the assumption that no changes occur in DNA duplex conformation upon binding. Comparison of the CD spectra of free and bound TTF-1HD indicates an increase in the protein helical content. A 10% increase in helical content has been estimated according to Eq. (5) and is in line with a previous determination [10]. It should originate from the passage of about six residues from a coil conformation to the

helical one. Similar results have been reported in the case of NK-2 and MAT α 2 HDs [35,36], in line with the observation that the length of HD recognition helix increases upon binding to target DNA sequences [26,27].

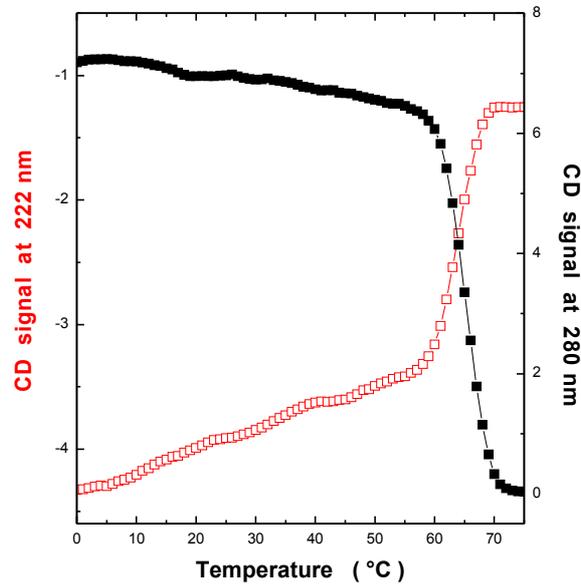


Figure 3.4 Thermal unfolding curves of the 1:1 complex between TTF-1HD and C duplex obtained by recording the CD signal at 222 nm (open squares, red) and 280 nm (filled squares, black).

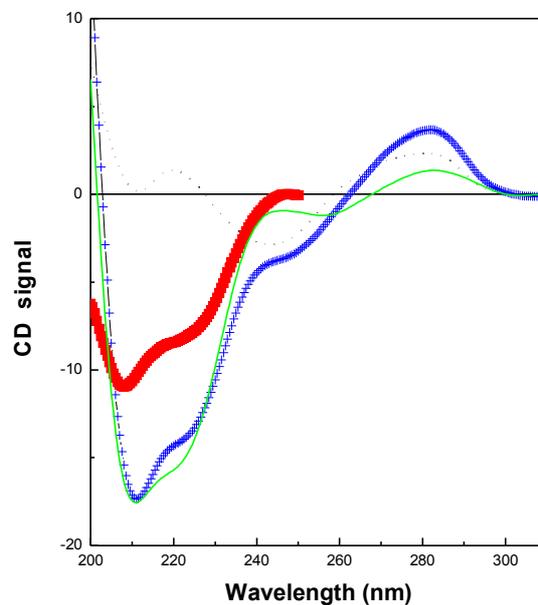


Figure 3.5 Far-UV CD spectra at pH 5.0, 20 mM potassium acetate and 75 mM KCl, of free TTF-1HD (filled squares, red), free C duplex (dot line, black), their 1:1 complex (cross line, blue); spectrum obtained by subtracting that of free C duplex from that of the complex in order to single out the contribution of the homeodomain (solid line, green).

3.4.2 Fluorescence measurement. The fluorescence spectra of TTF-1HD in the absence and presence of DNA, reported in Figure 3.6 show a decrease of fluorescence intensity in the bound state of protein. This fluorescence signal is largely determined by the Trp48, which is quenched upon its interaction with DNA. In addition, there is a blue-shift of the fluorescence maximum of the bound TTF-1HD (325 nm) in comparison with free TTF-1HD (342 nm), indicating that upon binding the Trp48 is exposed to a more hydrophobic environment. This observation is consistent with the molecular dynamic result suggesting that the side chain of Trp48 directly contact DNA by means of hydrogen bonds with phosphate oxygen atoms.

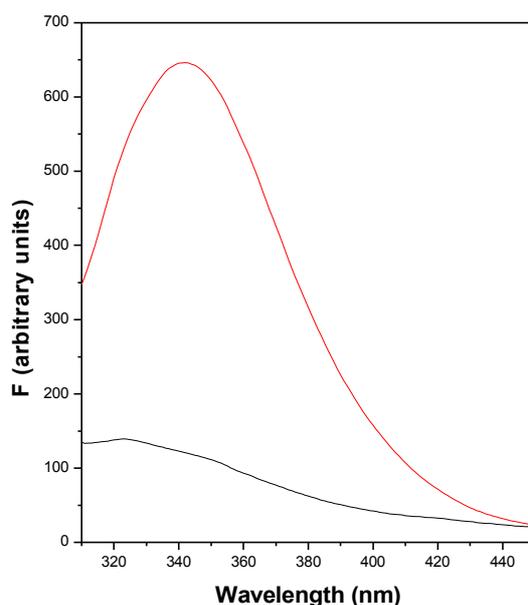


Figure 3.6: Fluorescence spectra of free (red) and bound TTF-1HD (black)

3.4.3. Binding Energetics. Binding of TTF-1HD to its target C duplex, containing the 5'-CAAG-3' specific sequence, has been measured by ITC and fluorescence measurements. Fluorescence titration conducted at 25 °C and pH 5.0, 20 mM potassium acetate and different KCl concentrations was performed to obtain the binding constant K_b . In particular, by plotting the fractional fluorescence quenching (Q_{obs}) versus the total concentration of DNA ($[L]_{tot}$) a binding curve is obtained (Fig 3.7). Interpolating the binding curve of the equation 6 (see Chapter 2), the values for the K_b are obtained (Tab 3.2). At 75 mM KCl, the value of the binding constant K_b is $1.0 \cdot 10^8 \text{ M}^{-1}$ indicated a high affinity for specific DNA and on increasing the KCl concentration, the values of K_b decrease in magnitude, thus suggesting that the binding is strongly influenced by the ionic strength of the solution.

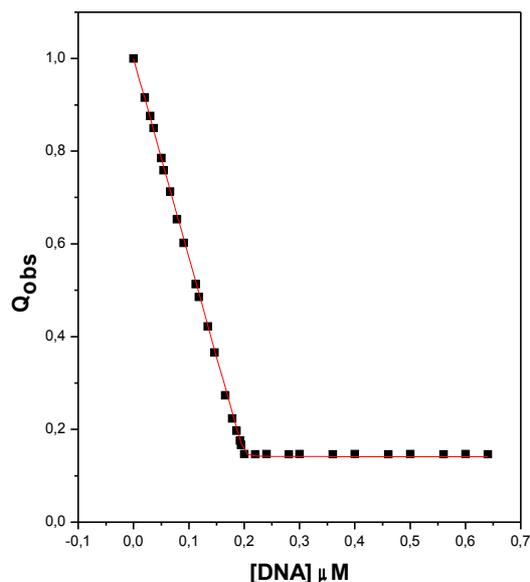


Figure 3.7 Titration of TTF-1HD with C duplex at 25 °C and pH 5.0, 20 mM potassium acetate and 75 mM KCl

Table 3.2 Thermodynamic parameters from ITC experiments for the binding of TTF-1HD with C duplex, containing its target sequence, at 25 °C and pH 5.0, 20 mM potassium acetate buffer and different KCl concentrations.

[KCl]	K_b	$\Delta_b H^\circ$	$\Delta_b G^\circ$	$T\Delta_b S^\circ$
mM	M^{-1}	kJ mol^{-1}	kJ mol^{-1}	kJ mol^{-1}
50	$6.5 (\pm 1.0) \cdot 10^8$	-44.0	-50.3	6.3
75	$1.5 (\pm 0.6) \cdot 10^8$	-41.0	-46.7	5.7
100	$7.5 (\pm 0.5) \cdot 10^7$	-39.0	-45.0	6.0
120	$3.0 (\pm 0.5) \cdot 10^7$	-37.0	-42.7	5.7
150	$8.0 (\pm 0.6) \cdot 10^6$	-34.0	-39.4	5.4
200	$4.0 (\pm 0.5) \cdot 10^6$	-31.0	-37.7	6.7
300	$3.5 (\pm 0.7) \cdot 10^5$	-24.0	-31.6	7.6
400	$2.5 (\pm 0.5) \cdot 10^5$	-21.0	-30.8	9.8

Note. The K_b values were determined by fluorescence measurements. Values are the average over four measurements in each condition; the error in the values of $\Delta_b H$, $\Delta_b G$ and $T\Delta_b S$ is always around 1.0 kJ mol^{-1} .

The ITC results are shown in Figure 3.8. Trace a in Panel A of Figure 3.8 shows a typical ITC profile for the association between TTF-1HD and C duplex. The corrected heat values are plotted as a function of the molar ratio in Panel B of Figure 3.8. Trace b in Panel A of Figure 3.8 shows the ITC profile obtained for the interaction of TTF-1HD with the modified Cm duplex, not-containing the specific target sequence. Data indicate the absence of a net exothermic effect, emphasizing a weak association between the two partners, as already determined by means of gel retardation measurements [10]. Limited proteolysis experiments pointed out that the Cm duplex provided a low protection of TTF-1HD, because complete digestion of the protein was observed when associated with Cm [7]. Raw ITC data are best-fit to a model for one binding site, providing values of K_b and $\Delta_b H^\circ$. The value of K_b are in agreement with the values of K_b obtained with the fluorescence. The binding Gibbs energy change, $\Delta_b G^\circ$, and entropy change, $T\Delta_b S^\circ$, are calculated from the equations $\Delta_b G^\circ = -RT \ln K_b$ and $T\Delta_b S^\circ = \Delta_b H^\circ - \Delta_b G^\circ$. Experimental data are collected in Table 3.2. At 25°C and 75 mM KCl, the value of the binding Gibbs energy change is $\Delta_b G^\circ = -46.7 \text{ kJ mol}^{-1}$, indicating that the association is strongly favoured from the thermodynamic point of view. The values of $\Delta_b H^\circ$ and $T\Delta_b S^\circ$ are $-41.0 \text{ kJ mol}^{-1}$ and 5.7 kJ mol^{-1} , respectively, emphasizing that the binding process is both enthalpy and entropy driven, even though the enthalpy change is largely dominant in this ionic strength condition. Similar results are obtained at 25 °C and pH 7.0, 50 mM sodium phosphate buffer, with $K_b = 4.5 \cdot 10^8 \text{ M}^{-1}$, $\Delta_b G^\circ = -49.4 \text{ kJ mol}^{-1}$, $\Delta_b H^\circ = -45.0 \text{ kJ mol}^{-1}$ and $T\Delta_b S^\circ = 4.4 \text{ kJ mol}^{-1}$.

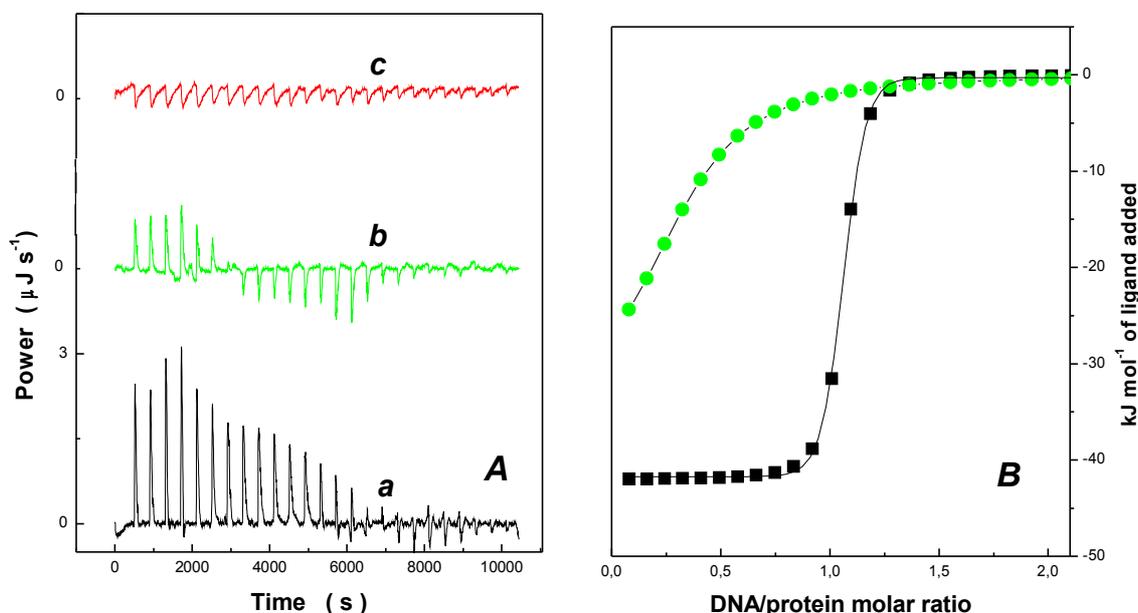


Figure 3.8 Panel A. A representative isothermal titration of TTF-1HD with C duplex (trace a); with the non-specific Cm duplex in the same buffer conditions (trace b); heat of dilution of C duplex in the buffer solution (trace c). Panel B. Binding enthalpies obtained from trace a (filled squares) and trace b (filled circles), after subtraction of the average heat of dilution from the individual integrated heats and normalization for protein concentration.

Increasing the KCl concentration, the values of both K_b and $\Delta_b H^\circ$ decrease in magnitude, as can be readily appreciated by looking at the numbers listed in Table 3.2; for instance, at 25°C and pH 5.0, 20 mM potassium acetate and 400 mM KCl, $K_b = 2.5 \cdot 10^5 \text{ M}^{-1}$ and $\Delta_b H^\circ = -21.0 \text{ kJ mol}^{-1}$. The Gibbs energy of the binding is calculated from the association constant, which is obtained both by optical method, such as fluorescence spectroscopy, and from isothermal calorimetry. Measurements of the ionic strength dependence of the association constant enables the separation of K_b (i.e. $\Delta_b G^\circ$) into its electrostatic and non-electrostatic components. Upon protein and DNA association, the formation of ion pairs between cationic residues in the protein and the DNA polyanion results in a release of counterions. The mixing of released ions with ions in the bulk solution produces significant entropy increase [37]. If the salt concentration is relatively low in aqueous solution this entropy effect is simply proportional to the number of released counterions [38]. Correspondingly, the logarithm of the association constant of protein with DNA is presented in just two terms

$$\log(K_b) = \log(K_{b-nel}) - (Z\psi + \beta) \log[NaCl] \quad (6)$$

The first term on the right-hand side of the equation (6) results from the non-electrostatic (nel) interactions between the DNA and protein, and the second results from the electrostatic effects associated with the release of counterions. The slope of this function, N is equal to $(Z\psi + \beta)$, where Z is the number of DNA phosphates interacting with the protein, ψ is the number of cations per phosphate released upon protein association, and β is the number of anions displaced from the protein on DNA binding. For short DNA duplexes ψ has been empirically determined as 0.64 [39], and the product $Z\psi$ is much greater than β . The number of ionic contacts made on forming a complex can thus be determined from the value of the slope of the linear relationship of $\log(K_b)$ with $\log[NaCl]$. A linear dependence between the log of the binding constant of a protein with DNA and the log of the salt concentration (Figure 3.9) is usually regarded as a manifestation of the electrostatic interactions in this process and the slope of this function, $\partial \log(K_b) / \partial \log[NaCl] = (Z\psi + \beta)$, actually gives the total number of counterions released upon protein binding to DNA [36]. A linear least-squares fit of data of Table 3.2 provides a slope of -4.0, with a correlation coefficient $r = -0.9933$ for eight points. The slope of the straight line gives the total number of counterions released upon association of TTF-1HD with C duplex, both cations from DNA phosphates and anions from the protein [40]. Since the number of contacts between Arg and Lys side chains of TTF-1HD helix III and the C duplex phosphates is 4-5 (see Figure 3.11),

and the empirically number of released cations per phosphate is 0.64 [39], the expected number of released cations would be around 3. The experimentally determined value of 4 suggests that one anion should be released from TTF-1HD upon association.

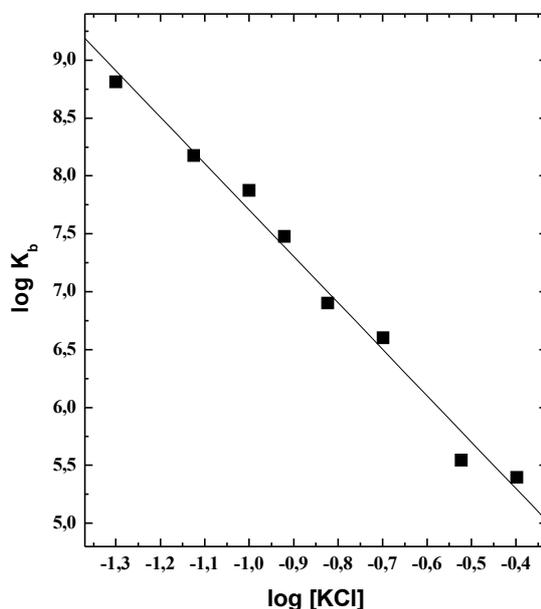


Figure 3.9 Plot of log K_b versus log [KCl] for the binding of TTF-1HD to C duplex at 25 °C and pH 5.0, 20 mM potassium acetate, investigated by means of ITC measurements.

When the salt concentration approaches to 1 M, (i.e., log [KCl] = 0) (see Figure 3.9), the electrostatic term in Eq (6) vanishes and $\Delta_b G^\circ$ approaches to the nonelectrostatic term of the Gibbs energy of association, $\Delta_b G^\circ = \Delta_b G^\circ_{\text{nel}} = RT \ln(K_b)$ [38]. This permits splitting the observed Gibbs association energy into two components: the nonelectrostatic and the electrostatic, $\Delta_b G^\circ = \Delta_b G^\circ_{\text{nel}} + \Delta_b G^\circ_{\text{el}}$. The non electrostatic component is of special interest because of the widely accepted assumption that it is this component of the binding energy that determine the specificity of DNA recognition [41].

In general, it results that, even though the binding of TTF-1HD to C duplex is strongly influenced by the ionic strength of the solution, the large and negative values of $\Delta_b H^\circ$ indicate that the association of TTF-1HD to its target DNA sequence, at room temperature, is always favoured by enthalpy, as generally found in the case of proteins that bind in a sequence-specific manner into the DNA major groove [36]. The positive values of $T\Delta_b S^\circ$ indicate that the association of TTF-1HD to its target DNA sequence, at room temperature, is always favoured by entropy, as generally found in the case of proteins that bind in a non-sequence-specific manner into the DNA minor groove

[37,41]. This finding seems to be reliable because HDs interact with both the major and minor grooves of DNA double helix [36,42].

An important thermodynamic parameter to characterize protein/DNA molecular recognition is $\Delta_b C_p$, the heat capacity change associated with binding [43]. ITC measurements, at pH 5.0 and low ionic strength, over the 5-35 °C temperature range, have not allowed a reliable determination of $\Delta_b C_p$ for the complex of TTF-1HD with C duplex, in all probability because this quantity is small. This is not really surprising if one recognizes that: (a) Privalov and colleagues reported $\Delta_b C_p \approx -1.2 \text{ kJ K}^{-1}\text{mol}^{-1}$ for several HDs [36]; (b) Ginsburg and colleagues found, at pH 7.4, 50 mM sodium phosphate, $\Delta_b C_p = 1.0 \text{ kJ K}^{-1}\text{mol}^{-1}$ for vnd/NK-2 HD, $-0.7 \text{ kJ K}^{-1}\text{mol}^{-1}$ for the mutant His52Arg of vnd/NK-2 HD, and $-0.4 \text{ kJ K}^{-1}\text{mol}^{-1}$ for the mutant His52Arg/Thr56Trp of vnd/NK-2 HD [44]. An accurate determination of $\Delta_b C_p$ for the association of TTF-1HD with C duplex will be one of the goals of a next study.

3.4.4. Molecular Dynamics Simulations. MD simulations of TTF-1HD with its target C and with modified Cm duplex complexes produce stable trajectories as shown by macroscopic properties of the systems such as potential and total energies, density (data not shown). The time evolution of the RMSD values with respect to starting structures indicates that the molecular components of the complexes have individually reached stable structures during the 3.5 ns long MD simulations (see Figure 2S of Supplementary Material Chapter 3). For the complex of TTF-1HD with C duplex, the RMSD values increase in the first 1 ns, and then fluctuate around an average value of 2.2 Å for the DNA and 3.6 Å for the protein during the last 2.5 ns. For the complex of TTF-1HD with Cm duplex, the RMSD values behave in a similar manner (see Figure 2S of Supplementary Material Chapter 3). The final structure of the complex between TTF-1HD with C duplex is reported in Figure 3.1. The RMSD between the structures of TTF-1HD in the two complexes amounts to 2.5 Å considering only backbone atoms, and 4.1 Å considering both backbone and side chain atoms.

The movements of TTF-1HD side-chains relative to the two DNA duplexes are visualized in Figure 3.10. The top view allows a direct comparison of the different movements of protein side chains. In the simulation with the C duplex, the protein side chains rearrange their positions and conformations in order to establish closer contacts with the target DNA sequence; in the simulation with the Cm duplex, instead, the protein side chains have more contacts with DNA in the starting model than at the end of the simulation (i.e., there is no gain of contacts with DNA).

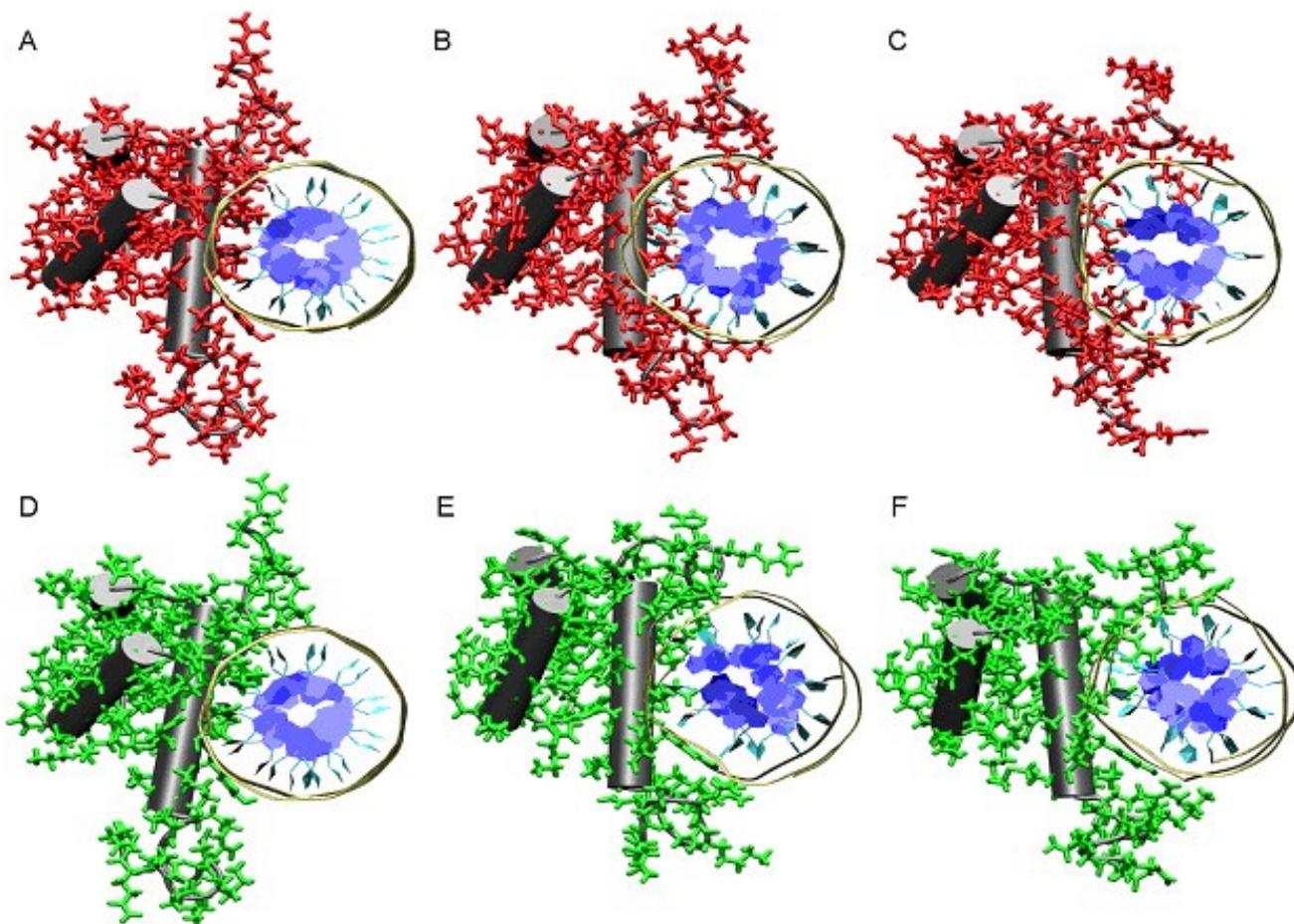


Figure 3.10 Comparison between TTF-1HD/C duplex complex (red protein) and TTF-1HD/Cm duplex complex (green protein) during the MD simulations. Starting structures (A, D), snapshots at 1.75 ns (B, E) and final snapshots after 3.5 ns (C, F). The proteins are represented by licorice and the α -helix segments are highlighted as cartoons, whereas the DNA duplexes are shown as ribbons.

In order to obtain a quantitative measure of the TTF-1HD contacts with the two duplexes, the relative reduction in the solvent accessible surface area (SASA) of individual residues upon complex formation with respect to the free protein has been calculated [45], and is shown in Figure 3.11. The pattern obtained for the complex of TTF-1HD with C duplex is similar to that reported by Dragan and colleagues for NK-2 HD [36]. In any case, changes in SASA upon complex formation cannot be considered a perfect measure of the role played by single residues in molecular recognition. This is especially true when there are both freely diffusing and slowly diffusing water molecules at the protein/DNA interface, such as in the case of the complexes between HDs and target DNA sequence [46,47]. Analysis of the last nanosecond of the MD trajectory of the specific complex indicates the presence of six water molecules that practically remain fixed in the same position, bridging TTF-1HD helix III residues and DNA bases (see Figure 3S of Supplementary Material Chapter 3). Specifically: a water is hydrogen bonded to both Gln50, A17 and C18; a

second water is hydrogen bonded to both Gln50 and phosphate oxygen of A17; a third water is hydrogen bonded to both Asn51 and A8; a fourth water is hydrogen bonded to both Arg53 and phosphate oxygen of A17; a pair of waters bridge Gln50 and Asn51 to A19. It is worth noting that: (a) these water molecules exactly correspond to those detected by Gruschus and Ferretti [46], by means of suitable 3D NMR techniques, at the interface between vnd/NK-2 HD and its target DNA sequence; (b) Wuthrich and co-workers [48] were the first to point out the importance of hydration waters for the role played by the invariant Asn51 in binding specificity. The importance of a water-dominated, wet interface for the specific recognition ability of Antp-HD emerged also from MD simulations of Gutmanas and Billeter [47].



Figure 3.11 Analysis of the contacts between the TTF-1HD residues and the base-pairs of the specific C duplex (upper panel) and the non-specific Cm duplex (down panel), determined by the relative reduction in the solvent accessible surface area, SASA, of the residues upon complex formation with respect to the free protein.

Comparison between the patterns obtained for the specific complex and the non-specific one indicates that, even though some contacts are conserved, marked differences occur in helix III and minor ones in terminal regions, with a larger number of contacts in the case of C duplex. In particular the side chains of Lys46, Gln50, Lys55 and Lys63 make similar interactions in both complexes, even though the Lys63 side chain proves to be very flexible in the complex with Cm duplex. In addition, the OH group of Tyr54 makes a good hydrogen bond (the distance is 2.9 Å) with the phosphate oxygen of A19 (notation as in Figure 1S of Supplementary Material Chapter 3) in both complexes. In the complex of vnd/NK-2 HD with its target DNA sequence (PDB code: 1NK3), the OH group of Tyr54 is hydrogen bonded to the phosphate oxygen of C20. It seems that

the position of helix III in the DNA major groove is slightly shifted on passing from the structure of the complex of vnd/NK-2 HD to the model of the complex of TTF-1HD.

The finding that the OH group of Tyr54 is hydrogen bonded to phosphate oxygen may appear surprising because it has been reported that Tyr54 plays an important role for the binding specificity of HDs of the NK class (i.e., Tyr54 is highly conserved in this class) [5,6,9,10]. However, it has to be noted that other HDs, such as Antp-HD and Engr-HD possess Met and Ala, respectively, at position 54 that cannot form a hydrogen bond with the phosphate oxygen, leading to a marked differences in the recognition ability with respect to HDs belonging to the NK class. The earlier structural model of the complex between TTF-1HD and the target DNA sequence, constructed by homology modelling, suggested that the OH group of Tyr54 forms a hydrogen bond to N7 of A19, and the aromatic ring of Tyr54 accepts a hydrogen bond from the N4 proton of C20 [6]. These predictions are not confirmed by the present model.

On the other hand, analysis of the complex between TTF-1HD and C duplex indicates that: (a) the NH₂ group of Gln44 forms a good hydrogen bond with the phosphate oxygen of A8 (the distance is 2.8 Å); (b) the side chain of the invariant Trp48 makes two hydrogen bonds with the phosphate oxygen atoms of A7 (the distances are 2.7 Å and 3.8 Å, respectively), and a third hydrogen bond with the oxygen of the phosphodiester bond of A7 (the distance is 3 Å); (c) the NH₂ group of the invariant Asn51 forms a hydrogen bond with a nitrogen atom of A8 (the distance is 3.8 Å). All these interactions are lost in the complex of TTF-1HD with Cm duplex. It is worth noting that Gln44, Trp48 and Asn51 are reported to play a fundamental role for the recognition ability of HDs [27,36].

3.5. Discussion

The overall data obtained indicate that TTF-1HD is a small globular domain not very resistant to temperature. At pH 5.0, far-UV CD measurements indicate that $T_d = 32^\circ\text{C}$ and $\Delta_d H(T_d)^{\text{vH}} = 120 \text{ kJ mol}^{-1}$ in the absence of salt, while $T_d = 50^\circ\text{C}$ and $\Delta_d H(T_d)^{\text{vH}} = 155 \text{ kJ mol}^{-1}$ in the presence of 75 mM KCl. In the latter conditions the parabola-like profile of $\Delta_d G$ versus temperature reaches the maximum of 25 kJ mol^{-1} at 231 K, while $\Delta_d G^\circ = 11.6 \text{ kJ mol}^{-1}$ at 25°C . This value corresponds to $170 \text{ J mol res}^{-1}$, a number that emphasizes the low conformational stability of TTF-1HD [49,50].

It may be useful to compare the conformational stability of TTF-1HD with that of other HDs of the NK class, which recognize the same target sequence 5'-CAAG-3'. At pH 5.0, 20 mM sodium acetate and 100 mM NaCl, NK-2 HD shows $T_d = 42.1^\circ\text{C}$, $\Delta_d H(T_d) = 154 \text{ kJ mol}^{-1}$, $\Delta_d C_p = 1.6 \text{ kJ K}^{-1} \text{ mol}^{-1}$, and $\Delta_d G^\circ = 9.4 \text{ kJ mol}^{-1}$ at 20°C [36]; at pH 7.4, 20 mM MOPS, Nkx2.5(C56S) HD shows T_d

= 44.4°C and $\Delta_d H(T_d) = 142 \text{ kJ mol}^{-1}$ in the absence of NaCl, and $T_d = 54.4^\circ\text{C}$ and $\Delta_d H(T_d) = 184 \text{ kJ mol}^{-1}$ in the presence of 100 mM NaCl [51]. The latter values indicate that TTF-1HD is more stable than NK-2 HD, but less stable than Nkx2.5(C56S) HD with respect to temperature. According to Ginsburg and colleagues [51], the higher thermal stability of Nkx2.5(C56S) HD is due to the presence of a salt bridge between Glu17 and Arg52, that, by coupling helix I to helix III, should confer rigidity to the entire structure. Such a salt bridge does not exist in the other two HDs because Arg52 is mutated to His52, whose side chain is about 5 Å away from that of Glu17 in the structure of TTF-1HD. It is worth noting that a salt bridge between Arg19 and Glu30 occurs in all the three HDs, coupling helix I to helix II [5,26,27,51]. This salt bridge is conserved during the MD simulations of TTF-1HD, since the distance between the side chains of Arg19 and Glu30 is always around 3.0-3.5 Å. The higher thermal stability of TTF-1HD with respect to NK-2 HD was suggested to be related to the polarity of residues 54 and 56 in helix III: Tyr54 and Met56 in TTF-1HD, Tyr54 and Thr56 in NK-2 HD [9]. This suggestion merited further investigation, reported in the next chapter. Taking also into account that the N-capping interaction stabilizing helix III in NK-2 HD, a strong backbone-backbone hydrogen bond between Thr41 and Gln44 [26,27], exists in both free and DNA-bound TTF-1HD, even though the hydrogen bond is somewhat longer, 3.0 Å versus 2.6 Å.

The value of $\Delta_d C_p = 1.5 \text{ kJ K}^{-1}\text{mol}^{-1}$ obtained for TTF-1HD is significantly smaller than that expected for a globular protein of the same size on the basis of the well established correlation between $\Delta_d C_p$ and the buried SASA [49,52]. This finding is in line with that of Privalov and co-workers for several HDs [36], and suggests that the native structure of TTF-1HD such as those of other HDs do not have a well-developed hydrophobic core, probably because they are not designed to be extremely stable, but to specifically interact with DNA.

ITC measurements demonstrate that the association of TTF-1HD with DNA is very specific: at 25°C and pH 5.0, 20 mM potassium acetate and 75 mM KCl, $K_b = 1.5 \cdot 10^8 \text{ M}^{-1}$ for C duplex; at 25°C and pH 7.0, 50 mM sodium phosphate, $K_b = 4.5 \cdot 10^8 \text{ M}^{-1}$ for C duplex. These values are in line with those determined by Ginsburg and colleagues [51]: at 25°C and pH 7.4, 20 mM MOPS and 100 mM NaCl, $K_b = 1.2 \cdot 10^8 \text{ M}^{-1}$ for NK-2 HD and $4.9 \cdot 10^8 \text{ M}^{-1}$ for Nkx2.5(C56S) HD, on the respective target DNA sequences. The binding specificity is confirmed and rationalized by the results of MD simulations: in the complex of TTF-1HD with C duplex there are several interactions between the side chains of residues Gln44, Trp48 and Asn51 in helix III and target DNA sequence that are entirely lost in the complex with Cm duplex (see Figures 3.10 and 3.11). This is in line with the absence of a net exothermic effect associated with the interaction of TTF-1HD with Cm duplex (see Panel B of Figure 3.8). On the other hand, at 25 °C and pH 5.0, 20 mM potassium acetate and

75 mM KCl, $\Delta_b H^\circ = -41.0 \text{ kJ mol}^{-1}$ for the association of TTF-1HD with C duplex; this large and negative enthalpy change is in line with the values obtained by Ginsburg and colleagues for NK-2 HD and Nkx2.5(C56S) HD [51].

The large and negative $\Delta_b H^\circ$ values at low ionic strength are due not only to the purely association of TTF-1HD with C duplex, but also to partial refolding of the protein induced by the interaction with its target sequence. According to CD measurements at 75 mM KCl, we suggest that six residues adopt a helical conformation upon binding (see Figure 3.5). A similar result emerges by comparing the structure of free TTF-1HD obtained by means of MD simulations in TIP3P water, 3 ns long, and the structure of the protein in the complex with C duplex (data not shown). Since the enthalpy change associated with helix-to-coil transition amounts to 3-4 kJ mol^{-1} per residue at room temperature [25], the corrected binding enthalpy change (i.e., accounting for the purely association event) would be around -20 kJ mol^{-1} . The latter value agrees with the binding enthalpy determined at high ionic strength (i.e., 400 mM KCl), where the partial refolding does not seem to occur [44]. In addition, the present estimate of the enthalpy change accounting for the purely association event, is in line with the value determined by Privalov and co-workers in the case of NK-2 HD [36], and should be due to the formation of several hydrogen bonds between TTF-1HD side chains and target DNA sequence.

It is worth noting that several slowly diffusing water molecules are localized at the TTF-1HD/C duplex interface on the basis of MD simulations; these waters correspond to the ones detected by Gruschus and Ferretti at NK-2 HD/DNA interface [46], and confirm the wet nature of the HD/DNA recognition surfaces [46-48]. These slowly diffusing waters are involved in water-mediated hydrogen bond bridges between TTF-1HD residues and DNA phosphates and/or bases. These water-mediated hydrogen bond bridges may be important for DNA binding affinity and specificity, taking into account also the entropy penalty associated with blocking a water molecule in a hydrogen bond bridge.

In conclusion, the conformational stability of TTF-1HD and the binding energetics to its target DNA sequence have carefully been investigated experimentally. A structural model of the complex, generated by combining homology modelling and molecular dynamics simulations, has provided detailed views of the interactions occurring between specific TTF-1HD side chains and the target DNA sequence. This structural information will be the basis for further investigations on mutant forms of TTF-1HD reported in the next chapter.

3.6. References

1. Duboule D. Guidebook to the homeodomain genes. 1994; Oxford University Press.
2. Guazzi S, Price M, De Felice M, Damante G, Mattei MG, Di Lauro R. Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J* 1990; 9:3691-3699.
3. Damante G, Fabbro D, Pellizzari L, Civitareale D, Guazzi S, Polycarpou-Schwartz M, Cauci S, Quadrifoglio F, Formisano S, Di Lauro R. Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. *Nucleic Acid Res* 1994; 22:3075-3083.
4. Damante G, Pellizzari L, Esposito G, Fogolari F, Viglino P, Fabbro D, Tell G, Formisano S, Di Lauro R. A molecular code dictates sequence-specific DNA recognition by homeodomains. *EMBO J* 1996; 15:4992-5000.
5. Esposito G, Fogolari F, Damante G, Formisano S, Tell G, Leonardi A, Di Lauro R, Viglino P. Analysis of the solution structure of the homeodomain of rat thyroid transcription factor 1 by 1H-NMR spectroscopy and restrained molecular mechanics. *Eur J Biochem* 1996; 241: 101-113.
6. Fogolari F, Esposito G, Viglino P, Damante G, Pastore A. Homology model building of the thyroid transcription factor 1 homeodomain. *Prot Eng* 1993; 6:513-519.
7. Scaloni A, Monti M, Acquaviva R, Tell G, Damante G, Formisano S, Pucci P. Topology of the thyroid transcription factor 1 homeodomain-DNA complex. *Biochemistry* 1999; 38:64-72.
8. Damante G, Tell G, Leonardi A, Fogolari F, Bortolotti N, Di Lauro R, Formisano S. Analysis of the conformation and stability of rat TTF-1 homeodomain by circular dichroism. *FEBS Lett* 1994; 345:293-296.
9. Tell G, Acquaviva R, Formisano S, Fogolari F, Pucillo C, Damante G. Comparative stability analysis of the thyroid transcription factor 1 and Antennapedia homeodomains: evidence for residue 54 in controlling the structural stability of the recognition helix. *Int J Biochem & Cell Biol* 1999; 31:1339-1353.
10. Fabbro D, Tell G, Leonardi A, Pellizzari L, Pucillo C, Lonigro R, Formisano S, Damante G. In the TTF-1 homeodomain the contribution of several amino acids to DNA recognition depends on the bound sequence. *Nucleic Acid Res* 1996; 24:3283-3288.
11. Tang Y, Nilsson L. Effect of G40R mutation on the binding of human SRY protein to DNA: A molecular dynamics view. *Proteins: Struct Funct Genet* 1999; 35:101-113.

12. Suenaga A, Yatsu C, Komeiji Y, Uebayasi M, Meguro T, Yamato I. Molecular dynamics simulation of trp-repressor/operator complex: Analysis of hydrogen bond patterns of protein-DNA interaction. *J Mol Struct* 2000; 526:209-218.
13. Studier FW and Moffatt BA. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes *J Mol Biol* 1986; 189:113-130.
14. Muller M, Affolter M, Leupin W, Otting G, Wuthrich K, Gehring WJ. Isolation and sequence-specific DNA binding of the Antennapedia homeodomain. *EMBO J* 1988; 7:4299-4304.
15. Pace CN, Vajdos F, Free L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 1995; 4:2411-2423.
16. Cantor CR, Warshaw MM, Shapiro H. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* 1970; 9:1059-1077.
17. Freire E, Biltonen RL. Statistical mechanical deconvolution of thermal transitions in macromolecules I. Theory and application to homogeneous systems. *Biopolymers* 1978;17:463-479.
18. Catanzano F, Giancola C, Graziano G, Barone G. Temperature-induced denaturation of ribonuclease S: a thermodynamic study. *Biochemistry* 1996; 35:13378-13385
19. Privalov PL. Stability of proteins: small globular proteins. *Adv Protein Chem* 1979; 33:167-241.
20. Zhou Y, Hall CK, Karplus M. The calorimetric criterion for a two-state process revisited. *Protein Sci* 1999; 5:1064-1074.
21. Venyaminov SY, Yang JT. Determination of protein secondary structure. In *Circular dichroism and the conformational analysis of biomolecules*. ed. G.D. Fasman Plenum Press, New York, 1996; 69-107.
22. Sreema N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 1993; 209:32-44.
23. Lobley A, Whitmore AL, Wallace BA. Dichroweb: An interactive website for the analysis of secondary structure from circular dichroism data. *Bioinformatics* 2002;18:211-212.
24. Del Vecchio P, Graziano G, Granata V, Barone G, Mandrich L, Manco G, Rossi M. Temperature- and denaturant-induced unfolding of two thermophilic esterases. *Biochemistry* 2002; 41:1364-71.
25. Richardson JM, Makhatadze GI. Temperature dependence of thermodynamics of helix-coil transition. *J Mol Biol* 2004; 335:1029-1037.

26. Gruschus JM, Tsao DH, Wang LH, Nirenberg M, Ferretti JA. Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry* 1997; 36:5372-5380.
27. Gruschus JM, Tsao DH, Wang LH, Nirenberg M, Ferretti JA. The three-dimensional structure of the vnd/NK-2 homeodomain- DNA complex by NMR spectroscopy. *J Mol Biol* 1999; 289:529-545.
28. HyperChem 7.5, Hypercube, Inc., Gainesville, Florida, USA.
29. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 1995; 91:43-56.
30. Cheatham TE, Cieplak P, Kollman PA. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* 1999; 16:845-862.
31. Jorgensen W, Chandrasekhar J, Madura J, Impey R, Klein M. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983; 79:926-935.
32. Hess B, Bekker H, Berendsen HJC, Fraaije J. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 1997; 18:1463-1472.
33. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984; 81:3684-3690.
34. Darden T, Perera L, Li L, Pedersen L. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure* 1999; 7:55-60.
35. Carra JH, Privalov PL. Energetics of folding and DNA binding of the MAT alpha 2 homeodomain. *Biochemistry* 1997; 36:526-535.
36. Dragan AI, Li Z, Makeyeva EN, Milgotina EI, Liu Y, Crane-Robinson C, Privalov PL. Forces driving the binding of homeodomains to DNA. *Biochemistry* 2006; 45:141-151.
37. Manning GS. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* 1978; 11:179-246.
38. Ha JH, Capp MW, Hohenwalter MD, Baskerville M. and Record JT jr. Thermodynamic stoichiometries of vparticipation of water, cations and anions in specific and non-specific binding of lac repressor to DNA. Possible thermodynamic origins of the “glutamate effect” on protein.DNA interactions. *J. Mol. Biol.* 1992; 228: 252-264
39. Olmsted MC, Bond JP, Anderson CF, Record MT. Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L8+) to the center of DNA oligomers. *Biophys J* 1995; 68:634-647.

40. Record MT, Zhang W, Anderson CF. Analysis of effects of salts and uncharged solutes on protein and nucleic acid equilibria and processes: a practical guide to recognizing and interpreting polyelectrolyte effects, Hofmeister effects, and osmotic effects of salts. *Adv Protein Chem* 1998; 51:281-353.
41. Dragan AI, Klass J, Read C, Churchill MEA, Crane-Robinson C, Privalov PL. DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. *J Mol Biol* 2003; 331:795-813.
42. Tsao DH, Gruschus JM, Wang LH, Nirenberg M, Ferretti JA. Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR. *Biochemistry* 1994; 33:15053-15060.
43. Spolar RS, Record TM. Coupling of local folding to site-specific binding of proteins to DNA. *Science* 1994; 263:777-784.
44. Gonzales M, Weiler S, Ferretti JA, Ginsburg A. The vnd/NK-2 homeodomain: thermodynamics of reversible unfolding and DNA binding for wild-type and with residue replacements H52R and H52R/T56W in helix III. *Biochemistry* 2001;40:4923-4931.
45. Kleinjung J, Fraternali F. POPSCOMP: an automated interaction analysis of biomolecular complexes. *Nucleic Acids Res* 2005; 33:342-346.
46. Gruschus JM, Ferretti JA. Quantitative measurement of water diffusion lifetimes at a protein/DNA interface by NMR. *J Biomol NMR* 2001; 20:111-26.
47. Gutmanas A, Billeter M. Specific DNA recognition by the Antp homeodomain: MD simulations of specific and nonspecific complexes. *Proteins: Structure, Function, and Bioinformatics* 2004; 57:772-782.
48. Qian YQ, Otting G, Wuthrich K. NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. *J Am.Chem Soc* 1993; 115:1189-1190.
49. Makhatadze GI, Privalov PL. Energetics of protein structure. *Adv Protein Chem* 1995; 47:307-425.
50. Rees DC, Robertson AD. Some thermodynamic implications for the thermostability of proteins. *Protein Sci* 2001; 10:1187-1194.
51. Fodor E, Mack JW, Maeng JS, Ju JH, Lee HS, Gruschus JM, Ferretti JA, Ginsburg A. Cardiac-specific Nkx2.5 homeodomain: conformational stability and specific DNA binding of Nkx2.5(C56S). *Biochemistry* 2005; 44:12480-12490.
52. Graziano G, Catanzano F, Barone G. Prediction of the heat capacity change on thermal denaturation of globular proteins. *Thermochim Acta* 1998; 321:23-31.

3.7 Supplementary Material - Figure of Chapter 3

Figure 1S. Sequence alignment of the TTF-1HD homeodomain with the selected template 1NK3 (A). Sequence alignment of the TTF-1HD binding DNA (C14) with the template DNA (B). The alignments were obtained using T-Coffee*. Numbered C14 duplex bases (C).

A)

```

                0          10          20          30          40          50
1FTT  -----MRRKRRVLFSSQAQVYELERRFKQQKYL SAPEREHLASMIHLTPTQVKIWFQNH
1NK3  ASDGLPNKKRKRRLVFTKAQTYELERRFRQQRYLSAPEREHLASLIRLPTQVKIWFQNH
                :*****:.*.*****:*.*****:*.*****:*.*****
                60
1FTT  RYKMKRQAKDKAAQQ--
1NK3  RYKTKRAQNEKGYEGHP
                *** **  :.*. :

```

B)

```

C14   CCAGTCAAGTGTTCC--
1NK3  TGTGTCAAGTGGCTGT
                *****

```

C)

```

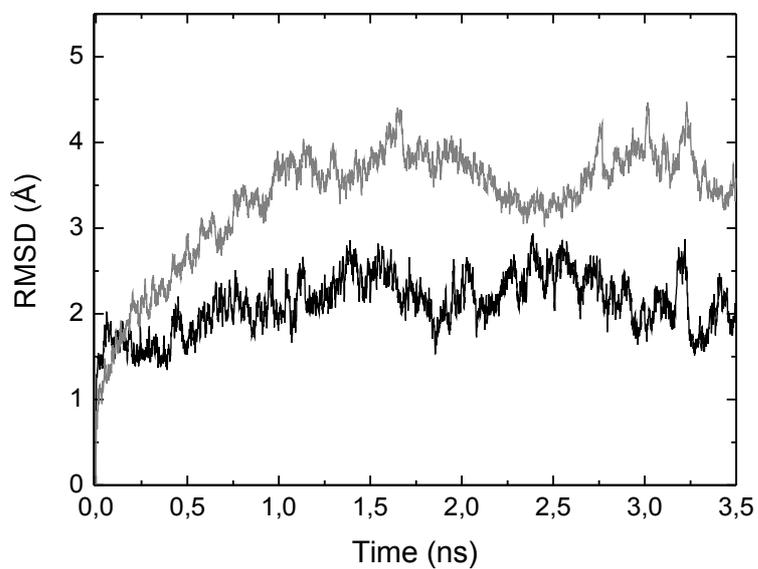
    1  2  3  4  5  6  7  8  9 10 11 12 13 14
5'C  C  A  G  T  C  A  A  G  T  G  T  T  C3'
    |  |  |  |  |  |  |  |  |  |  |  |  |
3'G  G  T  C  A  G  T  T  C  A  C  A  A  G5'
    28 27 26 25 24 23 22 21 20 19 18 17 16 15

```

*Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment, J Mol Biol 2000; 302:205-217.

Figure 2S. Time dependence of RMSD of protein backbone (grey) and DNA (black) in the TTF-1HD/C14 (A) and TTF-1HD/C14m (B) complexes during the MD simulations.

A)



B)

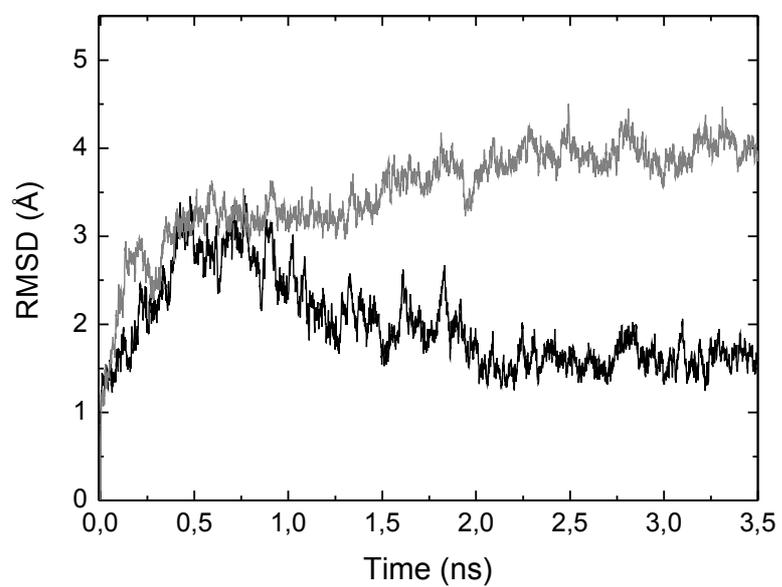
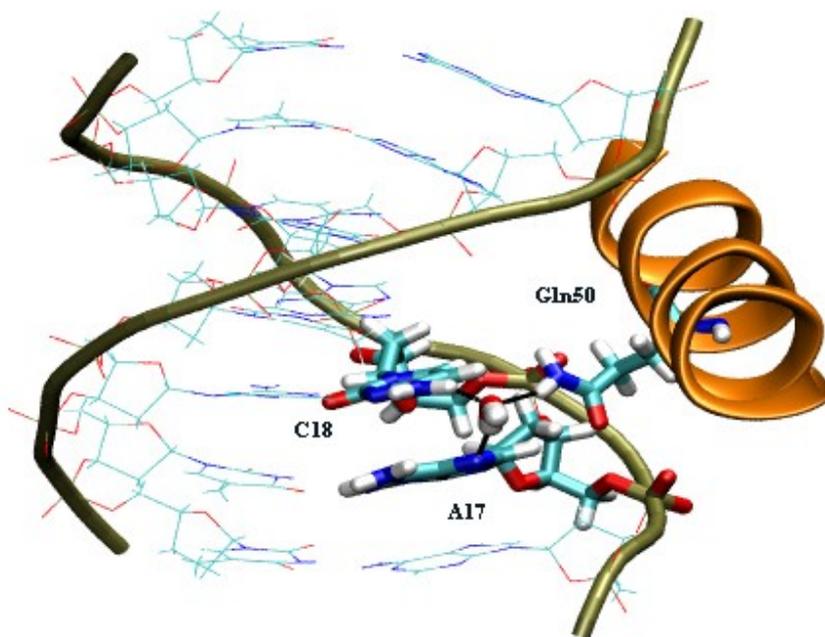
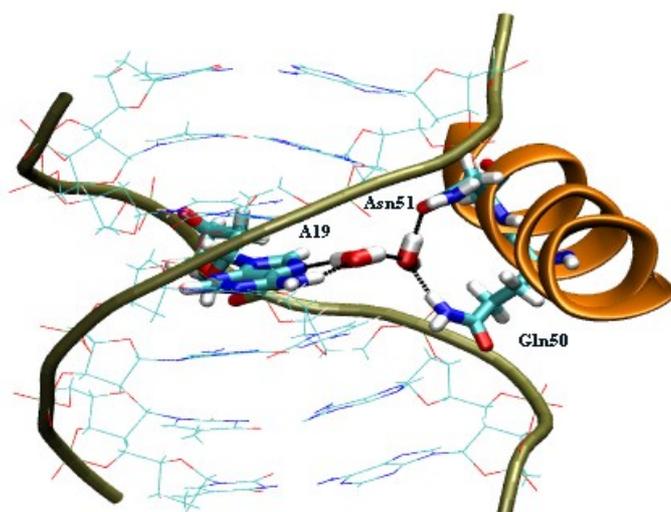


Figure 3S. Snapshots showing the fixed water molecules detected at the TTF-1HD/C duplex interface in the last ns of MD simulations.

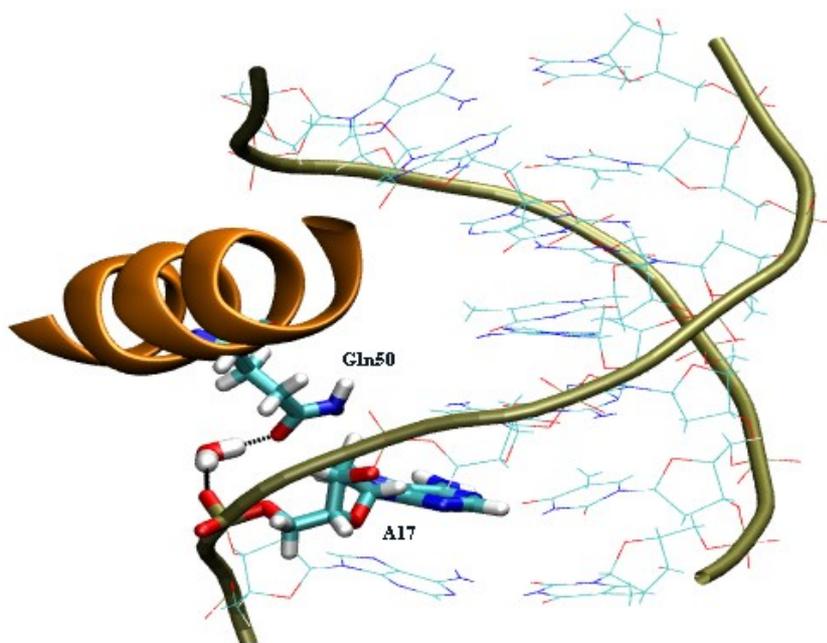
A)



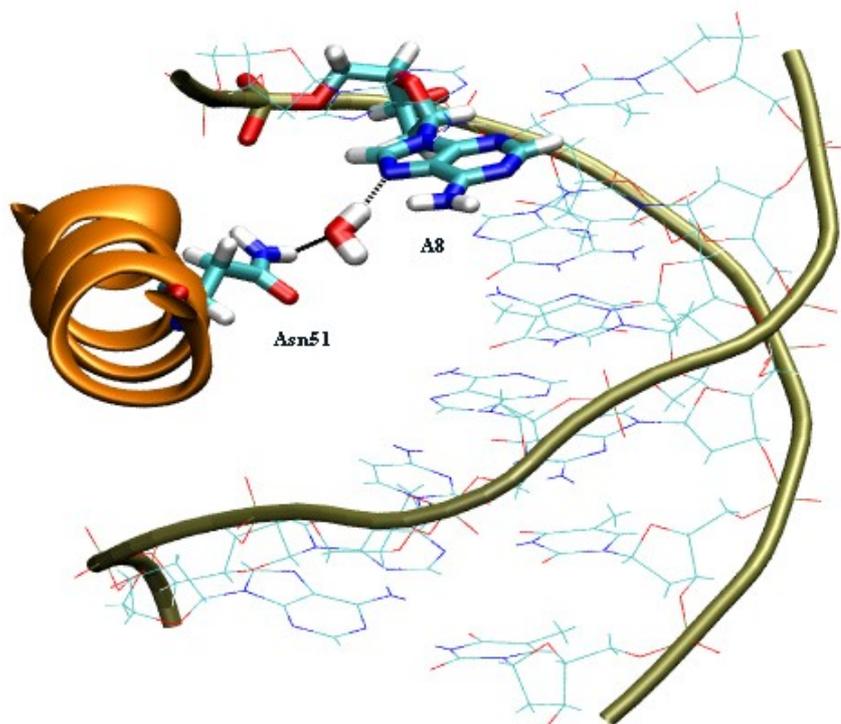
B)



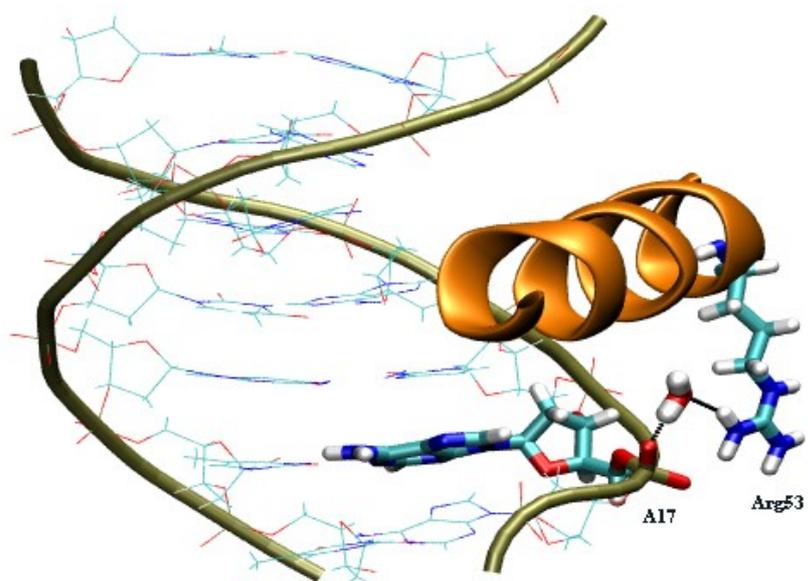
C)



D)



E)



CHAPTER 4

Characterization of TTF-1 mutant forms

4.1 Introduction

The thyroid transcription factor 1 (TTF-1) is a tissue-specific transcription factor involved in the development of thyroid and lung. It is a protein of 371 amino acids that possesses two independent transcriptional activation domains, one (159 amino acids) located in the N-terminal region of homeodomain and the other (145 amino acids) in the C-terminal region of homeodomain [1]. The homeodomain is a highly conserved DNA-binding motif that is found in numerous transcription factors throughout a large variety of species from yeasts to human. These gene-specific transcription factors play critical roles in development and adult homeostasis, and therefore, any germline mutations associated with these proteins can lead to a number of congenital abnormalities [2]. Remarkable features of structural and functional conservation have been observed within homeodomain family members. Compared with primary sequences, their three-dimensional structures are more conserved, which indicates the importance of proper architecture for correct functioning with respect to, DNA recognition and protein-protein interactions. This high degree of conservation between the sequence and structure makes the homeodomain an ideal model for studying protein-DNA interactions and gene regulation [3]. The compact homeodomain is composed by the hydrophobic core, that hold all of its three helices together, and by a flexible N-terminal arm which contains a stretch of basic residues known as the nuclear localization signal (NLS). DNA-binding specificity appears to be conferred by the helix III residues that interact with the major groove and the N-terminal arm residues that recognize the minor groove of DNA [3,4,5]. As shown in Figure 3.1 of Chapter 3, TTF-1HD behaves like the other homeodomains.

To gain a molecular-level understanding of the interactions determining the association of TTF-1HD to the target DNA sequence, we studied the conformational stability and DNA binding ability of different mutant forms of TTF-1 homeodomain. Some key point mutants have been produced by means of the site-directed mutagenesis. In particular, some residues involved in DNA recognition such as Q50, Y54 and W48 were mutated in alanine residues to give three different mutants. Moreover, a deleted mutant (desTTF-1HD), lacking five residues (RRKRR named N-tails) from the N-terminal region was produced in order to investigate the role of the N-tail in the protein stabilization and in the DNA binding.

The TTF-1 transcription factor possesses two regions one located at N-terminal and the other at C-terminal position (see Fig 1.5) that probably interact with a common intermediary factor or co-activator in order to activate the thyroid development [6]. In the aim to assist in the understanding of functional studies on transcriptional control in development, a fusion protein containing the N-terminal region and the homeodomain of TTF-1 (TTF-1NH₂HD) has been produced and characterized.

The conformational stability and DNA binding ability of the TTF-1HD mutants as well as of the fusion protein TTF-1NH₂HD have been investigated by means of circular dichroism (CD) and isothermal titration calorimetry measurements (ITC).

4.2 Experimental

4.2.1 Materials

- **Cloning and protein expression**

TTF-1 mutant forms were cloned, expressed and purified in collaboration with the group of Prof. Formisano of the Department of Cellular and Molecular Biology and Pathology “L. Califano” of the University of Naples “Federico II”.

The cDNA coding for the amino acids 160-226 of rat TTF-1 was amplified by PCR using appropriate oligos, the amplified DNA was digested with BamHI and HindIII and cloned into the vector pQE30 (Qiagen) in-frame with the coding region of six histidine residues. By this way an ATG and a TAA codon are positioned respectively at the 5' and 3' ends of the insert and an extra histidine sequence is present at the N-terminus to give a translation product of 82 amino acids. In the box is reported the amino acids sequence.

	10	20	30	40	50	60
MRGSHHHHHH	<u>GSRRKRR</u> VLF	SQAQVYELER	RFKQQKYL	SA PEREHLASMI	HLTPTQVKIW	
	70	80				
FQNHRYKMKR	QAKDKAAQ	<u>QK</u>	LN			

Mutations were incorporated into the wild-type 6HysTTF-1HD located on plasmid pQE30 using the QuikChange mutagenesis kit (Stratagene, La Jolla, CA) and the appropriate mutagenic primers.

A procedure analogous to that described above was followed to obtain a truncated form, lacking five N-terminal residue (desTTF-1HD), using as a template the cDNA coding for the amino acids 166-226. The deleted residues are underlined in the box.

The cDNA coding for the amino acids 1-222 of rat TTF-1 was cloned into the vector pQE12 (Qiagen) in-frame with the coding region of six histidine residues. By this way an ATG and a TAA codon are positioned respectively at the 5' and 3' sides of the insert and an extra histidine sequence is inserted at the N-terminus to give a translation product of 237 amino acids. In the box is reported the amino acids sequence.

10	20	30	40	50	60
MRGSHHHHHH	GS SMSPKHTT	PFSVSDILSP	LEESYKKVGM	EGGGLGAPLA	AYRQGQAAPP
70	80	90	100	110	120
AAAMQQHAVG	HHGAVTAAYH	MTAAGVPQLS	HSAVGGYCNG	NLGNMSELPP	YQDTMRNSAS
130	140	150	160	170	180
GPGWYGANPD	PRFPAISRFM	GPASGMNMSG	MGGLGSLGDV	SKNMAPLPSA	PRRKRRVLFS
190	200	210	220	230	240
QAQVYELERR	FKQQKYL SAP	EREHLASMIH	LTPTQVKIWF	QNHRYKMKRQ	AKDKAAQQ KLN

The correct frame of all constructs were confirmed by DNA sequencing. The extra histidine sequence at the N-terminus allowed protein purification by nickel/nitrilotriacetic acid (Ni-NTA) affinity chromatography. The fusion protein, was expressed in ORIGAMI Escherichia coli cells. Overnight cultures were inoculated into LB medium supplemented with 50 µg/ml ampicillin and grown to an O.D.₆₀₀ of 0.8 at 37°C. Induction was achieved by adding IPTG to a concentration of 1 mM and incubating at 37°C for an additional 3 h. Cells were harvested by centrifugation at 600 g in a Sorvall GS-3 rotor. The bacterial pellet was resuspended in lysis buffer A (6 M guanidinium chloride, 0.1 M NaH₃PO₄, 0.01 M Tris, pH 8.0) and centrifuged at 10000 g for 20 min at 10 °C. The supernatant was loaded on to an Ni-NTA column, equilibrated with buffer A, and washed with 10 vol. of buffer B (8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris, pH 8.0). The protein was eluted with buffer B (8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris, pH 4.5).

Proteins concentration were determined using the Bio-Rad dye reagent with bovine γ-globulin as standard, according to the Bradford method (1976).

The overall final yield is 0.2-0.5 mg of pure protein per liter of culture. The purity and homogeneity of proteins are confirmed by a single band on SDS-PAGE; mass spectrometry gives a molecular weight of 10066.5 Da for wild type, 10009.5 for Q50A, 9974.5 for Y54A, 9314.5 for desTTF-1HD and 26217.8 for TTF-1NH₂HD in agreement with that expected from the amino acid sequence. The purified protein solutions are exhaustively dialyzed against the appropriate buffer solution at 4 °C by using Spectra Por MW 3500 membranes; protein concentration is determined

spectrophotometrically using a sequence-based extinction coefficient of 9970 M⁻¹cm⁻¹ for wild type, Q50A and desTTF-1HD, 8480 for Y54A, 4470 for W48A and 24410 for TTF-1NH₂HD at 280 nm [7]. Single strands of 14 base oligonucleotide, containing the TTF-1HD binding site 5'-CAAG-3', named C, and a modified single strands of 14 base without the TTF-1HD specific binding site, named Cm, are purchased from SIGMA Genosys (Sigma Chemical Co., St.Louis, MO) and used without further purification. The complementary strands are hybridized to form the DNA duplex, by mixing strands in equimolar ratio, heating the mixture to 90°C for 10 min and slowly cooling to room temperature.

Sodium acetate and phosphate buffers, KCl and other reagents of analytical grade are purchased from Sigma. Buffers for calorimetric and spectroscopic measurements are prepared with deionized and filtered through a Millipore Elix3 reagent grade system water. The pH of all solutions is determined with a Radiometer pHmeter, model PHM 93 at 25°C.

4.2.2 Circular Dichroism.

CD spectra are recorded with a cells with 0.1 cm path length and protein concentrations of about 0.2 mg mL⁻¹ are used to record CD spectra between 190 and 250 nm, with a time constant of 16 s, a 2 nm bandwidth, and a scan rate of 5 nm min⁻¹. The spectra are signal-averaged over at least five scans and baseline corrected by subtracting a buffer spectrum, and are analyzed for secondary structure composition according to the SELCON3 method [8] using Dichroweb [9]. Thermal unfolding curves are recorded in the temperature mode at 222 nm, from 0 to 85 °C with a scan rate of 1.0 K min⁻¹.

4.2.3 Isothermal Titration Calorimetry.

ITC experiments were carried out at 25°C using a high-sensitivity CSC-5300 Nano-ITC calorimeter from Calorimetry Science Inc. (Lindon, Utah), with a cell volume of 1.0 mL. Proteins solutions were placed in the cell at the concentration of 10 μM, and concentrated DNA solutions (70 μM) in the syringe. The DNA solution was injected into the proteins solution with 10 μl aliquots at 400-800 second intervals, prior to each injection, the heat evolution returns to zero, ensuring that the binding process equilibrates on the time scale of the injection. Both DNA and protein solutions were prepared with the same batch of buffer to minimize the differences in buffer composition and pH. The heat of dilution of DNA into the buffer solution was measured in a separate experiment and appropriate corrections were made. The heat of dilution obtained by averaging the heats of the last 5-10 injections, and subtracted from the measured heat of binding

give the same value. Raw data were integrated, corrected for non-specific heats, normalized for concentration, and analyzed by using the Bindworks program supplied with the instrument.

4.3 Results

4.3.1 Single mutants: Q50A, Y54A and W48A

Even though a large variety of homeodomain proteins have arisen in the course of evolution, their amino acid sequences have been conserved to a high degree. Several positions in the homeodomain sequence are occupied by the same amino acid in more than 95% of the 346 known homeodomains [10], including specific position within helix I (Leu16, Phe20) and helix III (Trp48, Phe49, Asn51 and Arg53). The high conservation of residues in the helix III has great implication on DNA binding and on the overall stability of the homeodomain tertiary structure. The invariant hydrophobic residues (Trp48 and Phe49) maintain the hydrophobic core by establishing favorable interactions with the hydrophobic amino acids in the helix I and II. The residues at position 50, 47 and 54 are much less conserved and their variability presumably contributes to the DNA-binding specificity [11]. To understand the role played by a single amino acid in the thermal stability and the DNA binding specificity, we undertook a study on the mutant forms Q50A Y54A and W48A. The mutant W48A was cloned, expressed and purified, but the resulting protein did not retain the folded structure, probably because the mutation of the invariant Trp48, in the centre of the hydrophobic core, prevent the correct folding of the protein.

Far-UV CD spectra at pH 5.0, 20 mM potassium acetate and 75 mM KCl of the mutant proteins are shown in Figure 4.1 The experimental data show that Q50A possess a far-UV CD spectrum very similar to that of wild-type, suggesting that no evident structural changes in the secondary structure occurred upon mutation. Conversely, the spectrum of Y54A shows an increase of secondary structure content. Indeed, the analysis of CD spectra at 0°C, by means of Dichroweb [9], shows that the mutant Y54A has a 52% of α -helix content, while TTF-1HD and Q50A have 42% of α -helix content. A similar difference in α -helix content was also found in the Y54M mutant form of NK-2 homeodomain. In fact, Weiler and coworkers [12] reported that the substitution of a hydrophobic amino acid in this position favors the formation of a longer helix III.

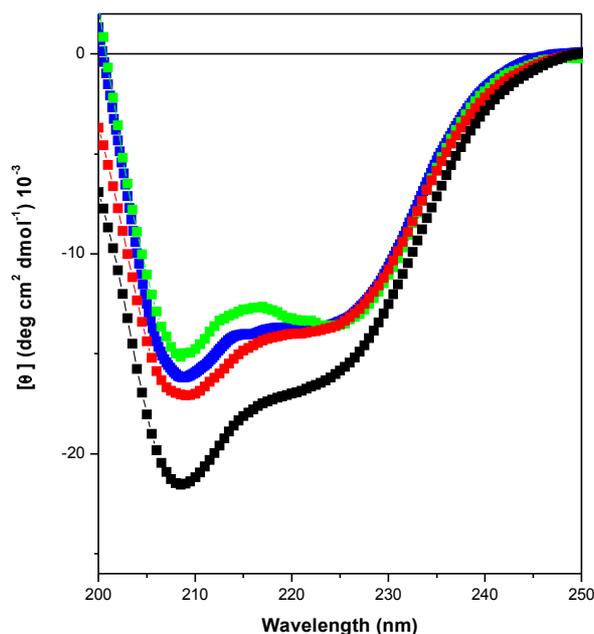


Figure 4.1. Far-UV CD spectra of TTF-1HD (Red), Y54A (black), Q50A (green) and desTTF-1HD (blue) recorded at 25°C, 20 mM potassium acetate and 75 mM KCl, pH 5.0.

Thermal unfolding curves are recorded in the temperature mode following the CD signal at 222 nm from 0 to 80°C with a scan rate of 1.0 K min⁻¹ (data not shown). From the analysis of thermal denaturation curves the values of denaturation temperature (T_d) for all the mutants were obtained and collected in Table 4.1. The TTF-1HD and Q50A mutant have the same T_d of 49.8°C, while a higher thermal stability is observed for Y54A that show a T_d of 52.1°C. This higher thermal stability is in agreement with the literature results on similar homeodomains systems for which was shown that the residues at position 54 and 56 are important in controlling homeodomains thermal stability [13,14]. In particular, if both residues at position 54 and 56 are hydrophobic, as in the case of Antp (Met and Trp, respectively) or hydrophilic, as in the case of Mat- α 2 (Arg and Gln, respectively) the thermal stability is higher (55.5°C for Ant and 52.5°C for Mat- α 2) than as in the case of TTF-1 (Tyr and Met, respectively) or NK-2 (Tyr and Thr, respectively), where a combination of hydrophobic/hydrophilic residues is present [15]. These observation are consistent with the results obtained: the thermal stability and the α -helix content is higher in the Y54A mutant respect to the TTF-1HD wild type, due to the replacement of a hydrophilic amino acid (Tyr) with an hydrophobic (Ala) (Y54A).

Table 4.1. Denaturation temperature of TTF-1HD and mutant forms unfolding determined by means the analysis of the thermal denaturation curves obtained by recording the molar ellipticity at 222 nm in 20 mM potassium acetate buffer, pH 5.0, 75 mM KCl.

	T_d (°C)
TTF-1HD	49.8
Y54A	52.1
Q50A	49.8
desTTF-1HD	53.3
TTF-1NH ₂ HD	48.5

Note. Each T_d represents the value averaged over at least three measurements. The error for the T_d does not exceed 0.5 °C.

Near-UV CD spectra of the mutant proteins compared with TTF-1HD are shown in Figure 4.2. There are some differences between the mutant proteins and the wild-type: a) the mutant Y54A shows a reduction of the maximum at 260 nm according with the substitution of a tyrosine in position 54 with an alanine; b) even if no differences in both secondary structure and thermal stability were observed for Q50A mutant, a noticeable difference in the near-UV CD spectrum is observed. Particularly a higher positive band at 260 nm is detected, possibly due to changes in the chemical environment of Try 25.

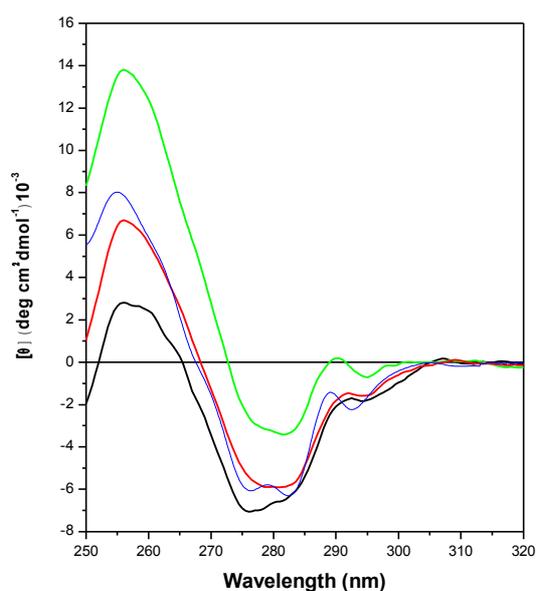


Figure 4.2 Near-UV CD spectra of of TTF-1HD (Red), Y54A (black) Q50A(green) and desTTF-1HD(blue) recorded at 25°C, 20 mM potassium acetate and 75 mM KCl, pH 5.0.

In fact, as shown in Figure 4.3 the Q50 screens the Tyr25 in the loop (Y25), while the replacement of the Gln with an Ala (Q50A) could induce changes in the chemical environment of Tyr 25, making it more exposed to the solvent. This could be the reason of increasing of CD signal at 260 nm.

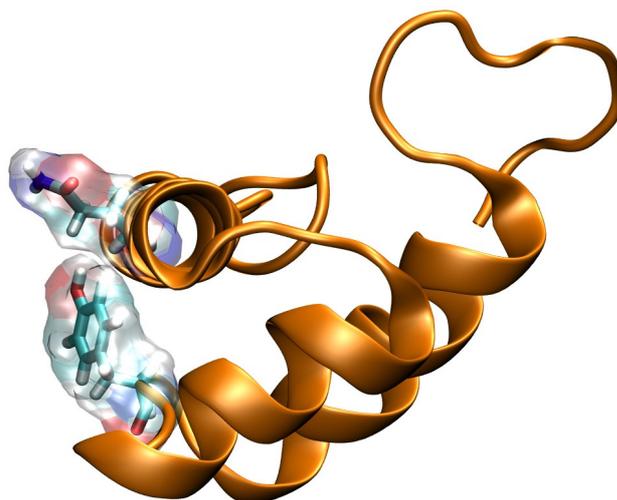


Figure 4.3 Ribbon representation of the TTF-1HD. The Gln 50 residue mutated in the experiment and the Tyr 25 are shown in licorice-like representation through the transparent surface.

The binding of TTF-1HD and its mutant forms with a 14-mer DNA duplex containing the 5'-CAAG-3' specific sequence, named C, and with another 14-mer DNA duplex, named Cm, not-containing the specific target sequence, has been measured by ITC at 25 °C. The Panel A in Figure 4.4 shows a representative ITC profile for the association between the TTF-1HD and its target DNA; Panel B and C show a profile for the association between the Y54A (Panel B) and Q50A (Panel C) and their target DNA. An exothermic heat pulse is observed after each injection of DNA into the protein solution. Instead a small exothermic heat is observed when is injected an aspecific DNA (Cm), suggesting a weaker thermal effect upon association. The area of each exothermic peak was integrated, and the heat of dilution of protein was subtracted from the integrated values. The corrected heat was normalized by the moles of DNA injected, and the resulting values were plotted as a function of the molar ratio, as shown in Panel D (TTF-1HD), E (Y54A) and F (Q50A) in Figure 4.4. The integrated ITC data were modeled to an equation based on an equivalent and independent binding site model, by using a non-linear least-squares method. The interpolation procedure provides an evaluation of the binding constant, K_b , the stoichiometry and the enthalpy change, $\Delta_b H^\circ$ in the binding process. The binding Gibbs energy change, $\Delta_b G^\circ$, and entropy change,

$T\Delta_bS^\circ$, are calculated from the equations $\Delta_bG^\circ = -RT \ln K_b$ and $T\Delta_bS^\circ = \Delta_bH^\circ - \Delta_bG^\circ$. The experimental data are collected in Table 4.2.

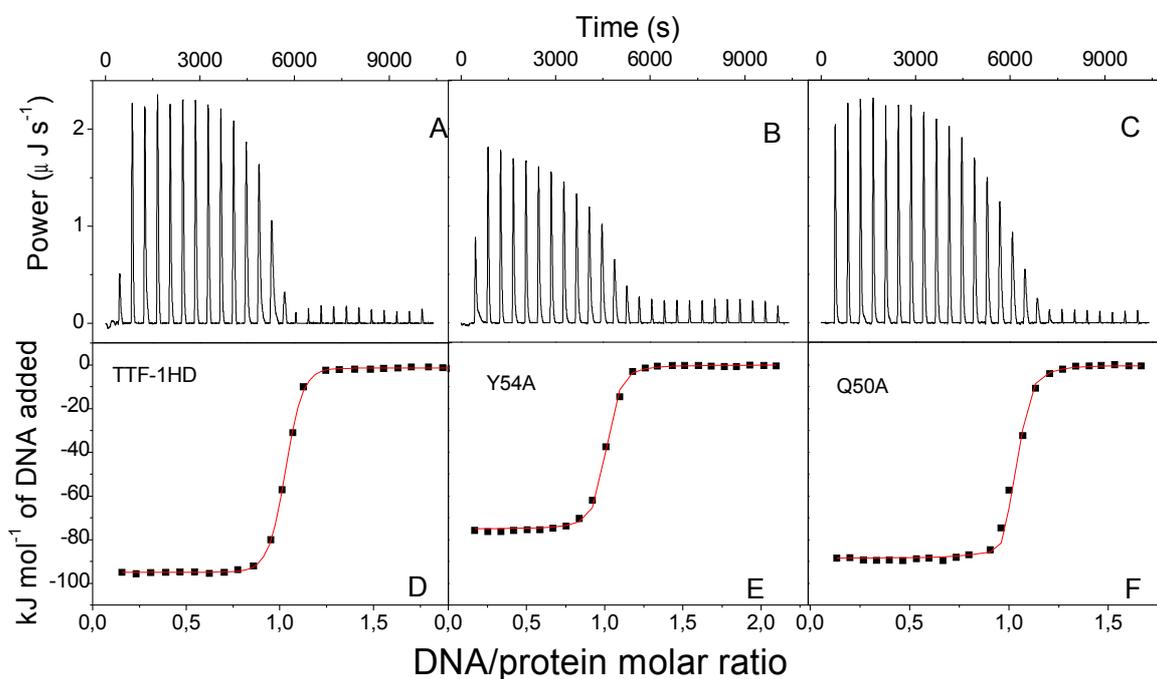


Figure 4.4 Isothermal titration of TTF-1HD with specific DNA (Panel A). ITC profile of Y54A and Q50A Panel B and Panel C respectively. Panel D,E and F show the binding isotherm obtained from trace A , B and C, respectively.

Table 4.2. Thermodynamic parameters from ITC experiments for the binding of TTF-1HD and mutant forms to C duplex, containing its target sequence, at 25 °C and pH 5.0, 20 mM potassium acetate 75 mM KCl. The stoichiometry is always $n = 1$.

	K_b (M^{-1})	Δ_bH° ($kJ mol^{-1}$)	Δ_bG° ($kJ mol^{-1}$)	$T\Delta_bS^\circ$ ($kJ mol^{-1}$)
TTF-1HD	$2.1 (\pm 0.6) 10^8$	-95.3 ± 1.3	-47.5 ± 1.0	-47.8 ± 1.2
Y54A	$4.9 (\pm 0.7) 10^7$	-76.7 ± 1.1	-43.8 ± 1.0	-32.9 ± 1.3
Q50A	$5.2 (\pm 0.4) 10^7$	-90.8 ± 1.0	-40.6 ± 1.0	-50.2 ± 1.2
desTTF-1HD	$5.9 (\pm 0.6) 10^6$	-75.2 ± 1.1	-38.6 ± 1.0	-36.6 ± 1.1
TTF-1NH ₂ HD	$1.5 (\pm 0.2) 10^7$	-65.5 ± 1.2	-40.9 ± 1.0	-24.6 ± 1.2

Note. The K_b values were determined by fluorescence measurements. Values are the average over four measurements in each condition; the error in the values of Δ_bH , Δ_bG and $T\Delta_bS$ is always around $1.0 kJ mol^{-1}$

The value of the binding constant K_b is $2.1 \cdot 10^8 \text{ M}^{-1}$ so that $\Delta_b G^\circ$ is $-47.5 \text{ kJ mol}^{-1}$, indicating that the association between the wild type to DNA is strongly favored from a thermodynamic point of view and enthalpy driven. All the mutants of the homeodomain show an unchanged 1:1 stoichiometry but lower binding constants of $4.9 \cdot 10^7 \text{ M}^{-1}$ and $5.2 \cdot 10^7 \text{ M}^{-1}$, respectively.

The ITC measurements of Y54A show a decrease of both the binding constant and enthalpy. In fact, K_b passes from $2.1 \cdot 10^8 \text{ M}^{-1}$ to $4.9 \cdot 10^7 \text{ M}^{-1}$ and $\Delta_b H^\circ$ from $-95.3 \text{ kJ mol}^{-1}$ to $-76.7 \text{ kJ mol}^{-1}$. To rationalize the role of this single amino acid in the binding thermodynamics we used the results of the molecular dynamics simulations. We observe that during the MD simulation, the OH group of Tyr54 makes a good hydrogen bond (the distance is 2.9 \AA) with the phosphate oxygen of A19 (Fig 4.5), so the replacement with an alanine, that cannot form any hydrogen bond with the phosphate oxygen may determine both the decrease in binding constant and enthalpy of the mutant compared to wild type.

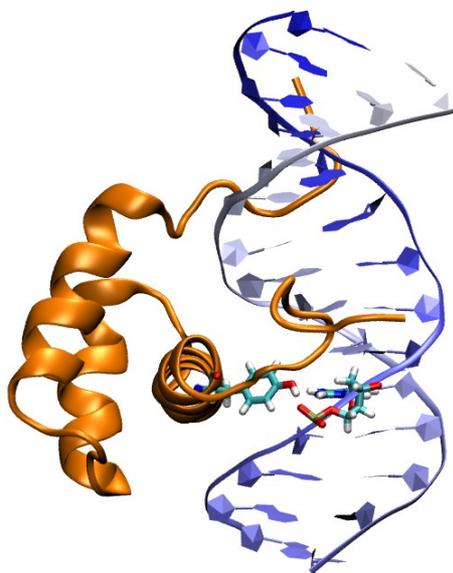


Figure 4.5 Snapshots showing the hydrogen bond detected at Y54/DNA interface in the last ns of MD simulations

Respect to Q50A from MD simulation results (see Figure 3S of Supplementary Material Chapter 3) there are four water molecules that remain fixed between Gln50 and DNA bases. Specifically: a water molecule is hydrogen bonded between Q50 and both A17 and C18; a second water is hydrogen bonded to both Q50 and phosphate oxygen of A17; a third and fourth waters bound Q50 and N51 to A19. Thanks to the side chain of Q50 an optimal fluctuating network of hydrogen bonds occur. When an Ala is present at position 50 this network of hydrogen bonds could be disrupted. Experimental data show that the binding constant decrease from $2.1 \cdot 10^8 \text{ M}^{-1}$ to $5.2 \cdot 10^7 \text{ M}^{-1}$ but the $\Delta_b H^\circ$ values are very close. This result suggests that the total numbers of hydrogen bonds does not

change but as the binding constant changes and then $\Delta_b G^\circ$ the difference is related to an entropic contribution. This result is in agreement with NMR experiments on NK-2 that indicate that the side chain of Q50 experience a disordered environment [16], and similar broadening was observed for the Ant-DNA complex [17]. Then the entropy associated with the disordered Q50/DNA interaction could have important thermodynamic consequences favouring homeodomain-DNA complex formation.

The overall results indicate that the single mutations of Y54A and Q50A does not prevent the binding because the substitution of a single amino acid is not sufficient to abolish the DNA binding specificity, that can be determined by a combinatorial molecular code. The thermodynamic signature of binding is then unchanged and it is enthalpy driven.

4.3.2 Deleted mutant: desTTF-1HD

The homeodomain structure possesses a highly charged N-terminal segment (N-tail) protruding from the hydrophobic core in free state, and having a role in the DNA bound state. The interaction of this N-terminal segment with the DNA is characterized by strong electrostatic interactions between the positively charged side chains of amino acids residues and the negatively charged phosphate groups. In addition, several hydrogen bonds are also formed with the bases into the minor groove of DNA [18].

The far- and near-UV CD spectra of desTTF-1HD mutant (Fig. 4.1 and Fig 4.2 in green) suggest that no evident structural changes occur in the secondary and tertiary structure upon deleted. The analysis of denaturation curves shows an increase of denaturation temperature which rises from 49.8 °C to 53.5 °C (Tab 4.1). This results is in agreement with the data reported by Dragan et al. on NK-2 and Ant and their truncated forms (desNK-2 and desAnt) [19]. The authors shown that the N-terminal arms are unfolded in the free homeodomains and that the stability of truncated forms is higher by about 2 kJ/mol than that of the complete homeodomains, with a corresponding increase in T_d of about 3.5°C. This suggested that the charged N-terminal extension destabilized the folded domain and then its deletion makes the hydrophobic core more stable [18].

Figure 4.6 shows an ITC profile of the binding of the desTTF-1HD with its target DNA (Panel A) and with DNA not-containing the specific target sequence (Panel B). In Panel A is observed an higher exothermic heat pulse respect to that of Panel B emphasizing the specific binding of the first and the weaker association between the two partners in the second case.

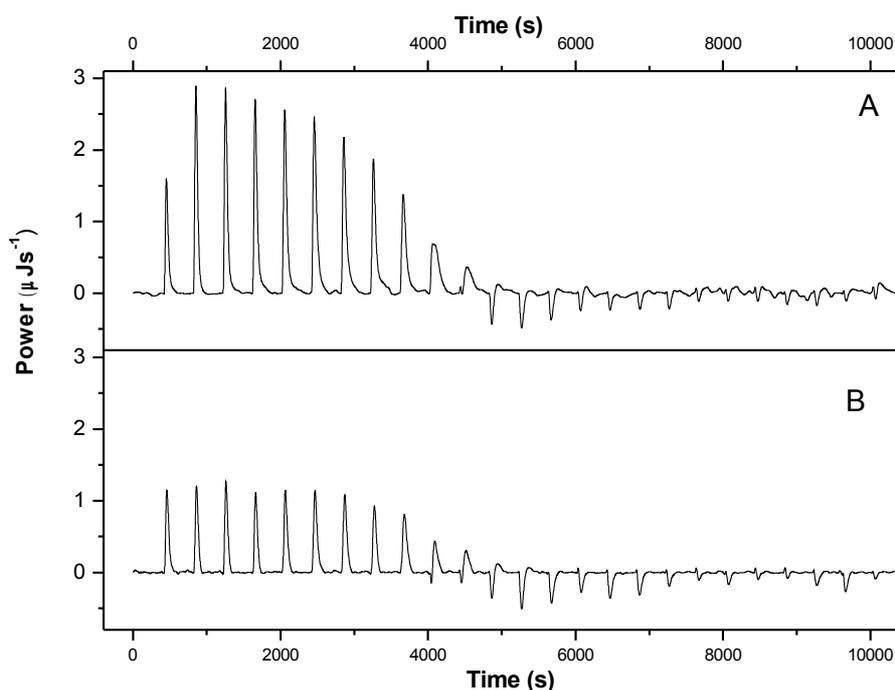


Figure 4.6 Isothermal titration of desTTF-1HD with specific (Panel A) and non specific (Panel B) DNA

The experimental data obtained by the interpolation procedure are collected in Table 4.2. The results show that the removal of the N-terminal residues results in a decrease of both binding constant (from $2.1 \cdot 10^8 \text{ M}^{-1}$ to $5.9 \cdot 10^6 \text{ M}^{-1}$) and enthalpy change (from $-95.3 \text{ kJ mol}^{-1}$ to $-75.2 \text{ kJ mol}^{-1}$) suggesting that N-terminal extension gives a contribution to the DNA specificity. Very similar results are reported for NK-2 and Ant where the removal the N-terminal residues reduces both binding constant and enthalpy [17]. Thus, the N-terminal arm concur to the complex formation through an enthalpic negative contribution by increasing the number of specific contacts with the minor groove of DNA.

4.3.3 Recombinant fusion protein : TTF-1NH₂HD

As mentioned TTF-1 is a modular protein that has two independent domain located to N- and C-terminal with respect to homeodomain. In particular it has been reported that the TTF-1 N-terminal region interacts with the C-terminal region of the Pax-8 transcription factor that is also involved in the development of thyroid and lung [4].

The analysis of far-UV CD spectra of TTF-1NH₂HD (Fig. 4.7) shows only 22 % α -helix content and over 60% disordered structure, in fact a small negative band at 222 nm and a very deep band to 202 nm are observed. The extent of the α -helix content in TTF-1NH₂HD suggest that the

homeodomain is the only folded region while the N-terminal region is probably fully disordered in solution. This is according to the analysis of thermal denaturation curve, in fact denaturation temperature of TTF-1NH₂HD is decreased respect the homeodomain (see Tab 4.1), suggesting that the presence of the additional unordered region makes the protein less stable.

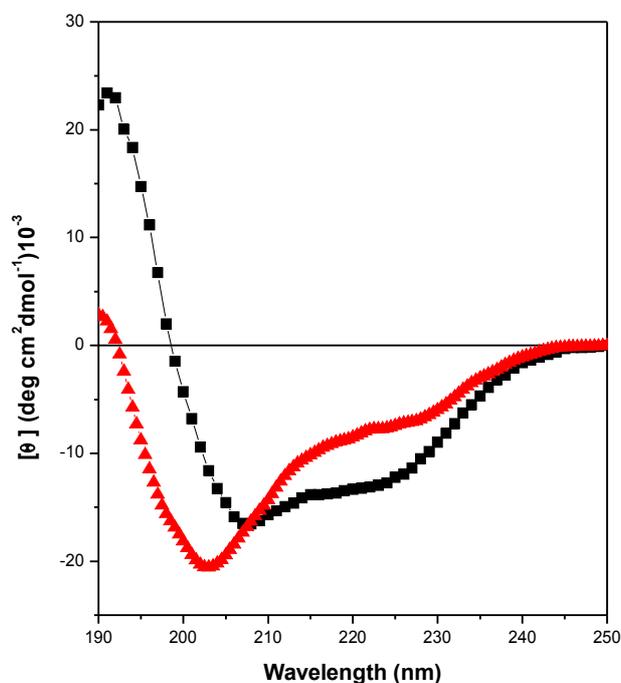


Figure 4.7. Far-UV CD spectra of TTF-1HD (black) and TTF-1NH₂HD (red)

In Figure 4.8 are reported the ITC profile for the association between the TTF-1NH₂HD and its target DNA (Panel A) and the corrected heat normalized by the moles of DNA injected plotted as a function of the molar ratio, (Panel B). The analysis of the ITC data (Tab 4.2) shows that the N-terminal region decreases the binding affinity of TTF-1 homeodomain, in fact a reduction of both constant (from $2.1 \cdot 10^8 \text{ M}^{-1}$ to $1.5 \cdot 10^7 \text{ M}^{-1}$) and binding enthalpy (from $-95.3 \text{ kJ mol}^{-1}$ to $-65.5 \text{ kJ mol}^{-1}$) is observed. These results are consistent with the hypothesis that the N-terminal region of TTF-1 exists in a random coil conformation and that does not interact with DNA. In fact, it is well known that the biological function of TTF-1 in the activation of thyroid genes depends on the DNA binding recognition of the homeodomain, while N-terminal region might interact with others factors for activating the transcription [6].

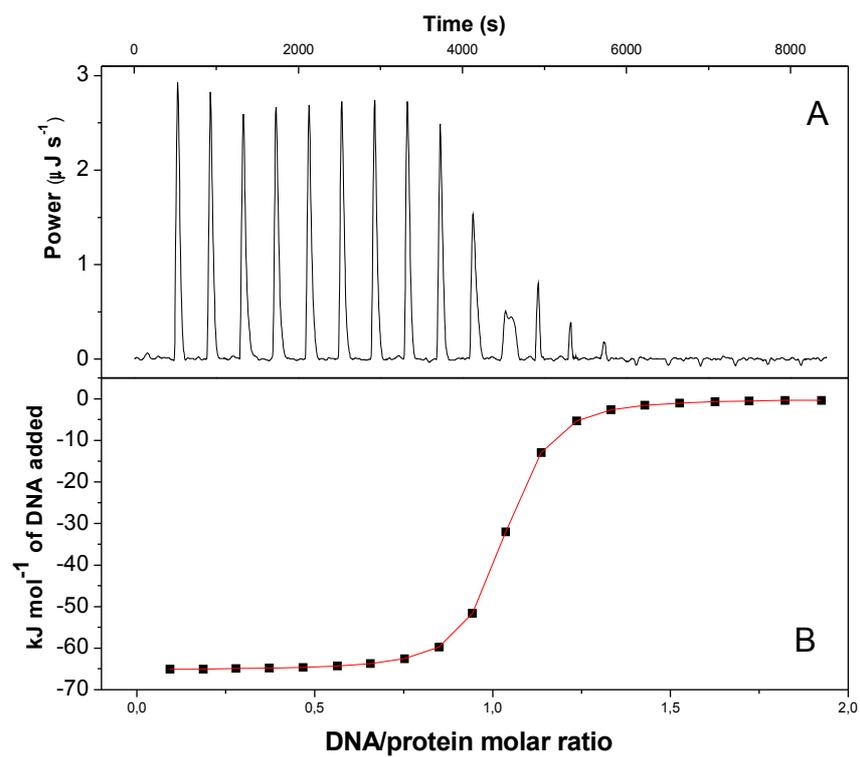


Figure 4.8 Isothermal titration of TTF-1NH2HD with specific DNA (Panel A), and binding isotherm (Panel B).

4.4. References

1. Tell G, Perrone L, Fabbro D, Pellizzari L, Puccillo C, De Felice M, Acquaviva R, Formisano S, Damante G. Structural and functional properties of the N transcriptional activation domain of thyroid transcription factor-1: similarities with the acidic activation domains *Biochem. J.* 1998; 329:395-403
2. D'Elia AV, Tell G, Paron I, Pellizzari L, Lonigro R, Damante G. Missense Mutations of Human Homeoboxes: A Review. *Human mutation* 2001; 18: 361-374
3. Young-In Chi Homeodomain Revisited: a Lesson from Disease-causing Mutations *Hum Genet.* 2005; 116:433–444
4. Di Palma T, Nitsch R, Mascia A, Nitsch L, Di Lauro R, Tannini M. The Paired Domain-containing Factor Pax8 and the Homeodomain-containing Factor TTF-1 Directly Interact and Synergistically Activate Transcription. *J. of biol. Chem.* 2003; 278:3395–3402
5. Guazzi S, Price M, De Felice M, Damante G, Mattei MG, Di Lauro R. Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. *EMBO J* 1990;9:3691-3699.
6. De Felice M, Damante G, Zannini M, Francis-Lang H and Di Lauro R Redundant domains contribute to the transcriptional activity of the thyroid transcription factor 1. *J. Biol. Chem.* 1995; 270:26649-26656
7. Pace CN, Vajdos F, Free L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 1995;4:2411-2423.
8. Venyaminov SY, Yang JT. Determination of protein secondary structure. In *Circular dichroism and the conformational analysis of biomolecules.* ed. G.D. Fasman Plenum Press, New York, 1996;69-107
9. Lobley A, Whitmore AL, Wallace BA. Dichroweb: An interactive website for the analysis of secondary structure from circular dichroism data. *Bioinformatics* 2002;18:211-212.
10. Gehring WJ, Affolter M, Burglin T. Homeodomain proteins *Annu RevBiochem.* 1994; 63: 487-526.
11. Banerjee-Basu S, Baxevanis AD. Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acid research* 2001; 29:3258-3269
12. Weiler S, Gruschus JM, Tsao DH, Yu L, Wang LH, Nirenberg M, Ferretti JA. Site-directed mutations in the vnd/NK-2 Homeodomain *J. of biol. Chem.* 1998; 273:10994–11000
13. Tsao DH, Gruschus JM, Wang LH, Nirenberg M, Ferretti JA. Elongation of helix III of the NK-2 homeodomain upon binding to DNA: a secondary structure study by NMR. *Biochemistry* 1994; 33: 15053–15060

14. Qian YQ, Furukubo-Tokunaga K, Resendez-Perez D, Muller M, Gehring WJ, Wüthrich K. Nuclear magnetic resonance solution structure of the fushi tarazu homeodomain from *Drosophila* and comparison with the Antennapedia homeodomain. *J. Mol. Biol.* 1994; 238: 333–345
15. Tell G, Acquaviva A, Formisano S, Fogolari F, Puccillo C, Damante G. Comparative stability analysis of the thyroid transcription factor 1 and Antennapedia in controlling the structural stability of the recognition helix. *Int. J. Biochem. Cell. Biol.* 1999;31:1339-1353.
16. Gruschus JM, Tsao DH, Wang LH, Nirenberg M, Ferretti JA. Interactions of the vnd/NK-2 homeodomain with DNA by nuclear magnetic resonance spectroscopy: basis of binding specificity. *Biochemistry.* 1997; 36:5372-80
17. Billeter M, Qian YQ, Otting G, Müller M, Gehring W, Wüthrich K. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J Mol Biol.* 1993 234:1084-93.
18. Tóth-Petróczy A, Simon I, Fuxreiter M, Levy Y. Disordered tails of homeodomains facilitate DNA recognition by providing a trade-off between folding and specific binding. *Am Chem Soc.* 2009; 42:15084-5.
19. Dragan AI, Li Z, Makeyeva EN, Milgotina EI, Liu Y, Crane-Robinson C, Privalov PL. Forces driving the binding of homeodomains to DNA. *Biochemistry* 2006;45:141-151.

CHAPTER 5

Thermal stability of Pax8 Prd domain and DNA binding energetics

5.1 Introduction

The Paired-box (PAX) gene family is a developmentally crucial small gene family, that encodes a set of well-characterized transcription factors. They are named for the paired box DNA-binding domain that is common and conserved in all the family members [1]. PAX genes are regulators of tissue development and cellular differentiation in embryos, in fact the expression of PAX genes is observed primarily during fetal development, and typically it is down-regulated in the adult organism. However, in a few tissues, the expression of PAX genes persists into adult life [2].

Pax-8, the only member of the family expressed in the thyroid tissue, is involved in the morphogenesis of the gland and in the transcriptional regulation of thyroid-specific genes. The molecular mechanisms involved in Pax-8 control of thyroid cell differentiation have been investigated in details, both in vitro and in vivo. In mature thyroid follicular cells, Pax-8 regulates the expression of thyroglobulin (Tg) and thyroperoxidase (TPO) [3]. It has been demonstrated that Pax-8 synergizes with TTF-1 in the regulation of thyroid-specific gene expression and that such a synergism correlates with a physical interaction between the two factors [4].

The Pax-8 protein shares a bipartite functionality consisting of a N-terminal DNA-binding region and a C-terminal transactivation region. The Pax-8 protein contains in its C-terminal region a transactivation domain whose deletion leads to a strong reduction or abrogation of Pax-8 ability to activate transcription from its target gene promoters. Deletion experiments of Pax-8 indicate that C-terminal domain is involved in a binding with the N-terminal activation domain of TTF-1 [4].

The Pax-8 N-terminal region contains three domains: a Paired Box (Prd) domain, a conserved octapeptide, and a further homeodomain. Pax-8 Prd domain consists of a well conserved 128-amino acids region, formed by two distinct subdomains known as PAI (N-terminal) and RED (C-terminal), each containing a helix-turn-helix motif joined by a linker region. DNA binding studies have demonstrated that the two subdomains of the Pax-8 Prd domain bind DNA independently, and that binding is regulated via a redox based mechanism centered on the glutathionylation of specific cysteines in the N-terminal region (Cys45 and Cys57) [5].

The thermal stability of Pax-8 Prd domain and the binding energetics to its target DNA have been characterized by means of isothermal titration calorimetry (ITC), complemented with circular dichroism (CD) experiments. Our study has shown that the binding process is enthalpy driven and

the unfavourable entropy reflects a remarkable protein conformation change upon binding. The binding-induced conformational change has been clearly highlighted by using CD spectroscopy and an evaluation of the secondary structure content change, upon binding, has been derived. The thermodynamic signature of interaction, together with the protein refolding upon binding, suggest that Pax-8 Prd interacts with the DNA through the major groove; thus showing an interaction mode that is typical of the close homologous proteins Pax-6 Prd and Pax-5 Prd.

5.2 Experimental

5.2.1 Materials.

- **Protein solutions preparation.**

Pax-8 paired domain was provided by the group of Prof. Esposito of University of Udine Department of Science and Biomedical Technology.

Pax-8 Prd domain was a 128-residue segment which encompasses residues 17-145 of the human Pax-8 protein. Here we study a 159-residue fusion protein containing a 13-residue C-terminal linker ending with a His₆ affinity tag. The fusion protein was expressed and purified using an optimized ad hoc protocol [6]. The purity and homogeneity of protein were confirmed by means of ESI mass spectroscopy which revealing a main m/z peak with mass equal to 17288.7 Da.

For this study two different buffers, with different ionic strength, have been chosen. The first condition, (buffer A) contains 50 mM sodium phosphate buffer, pH 6.5 while the second, (buffer B) contains, together with the same amount of sodium phosphate at the same pH, 75 mM of sodium chloride. In order to keep Cys45 and Cys57 in the active reduced state, we added 1mM TCEP in buffer solution for the spectroscopic and calorimetric measurements. Before each experiment the purified protein solutions were dialyzed against the appropriate buffer solution at 4 °C by using Spectra Por MW 10000 membranes. Protein concentration was determined by means UV measurement using a sequence-based extinction coefficient of 10095 M⁻¹cm⁻¹ at 280 nm [7].

- **DNA Preparation**

Single stranded, 24 base oligonucleotide, 5'-TGATGCCCACTCAAGCTTAGACAG-3' and its complementary strand (Thyropoxidase promoter) were purchased from SIGMA Genosys (Sigma Chemical Co., St.Louis, MO) and used without further purification. Lyophilised single strands were dissolved in the appropriate buffer solution, and concentrations were determined measuring the UV absorbance at 260 nm using an extinction coefficients of 230400 M⁻¹cm⁻¹ and 228800 M⁻¹cm⁻¹ for the forward and reverse strand, respectively calculated by the nearest neighbour model [8]. The

complementary strands were hybridised to form the DNA duplex, by mixing strands in a 1:1 molar ratio, dialyzed against the buffer and annealed by heating the solution to 90°C for 10 min and slowly cooling it down to the room temperature. The concentration of duplex was evaluated by UV measurement at 90°C, using as molar extinction coefficient the sum of the coefficients of the single strands. Sodium phosphate buffer, NaCl and other reagents of analytical grade were purchased from Sigma. Buffers for calorimetric and spectroscopic measurements were prepared with deionised and filtered water through a Millipore Elix3 reagent grade system. The pH of all solutions was determined with a Radiometer pH-meter, model PHM 93 at 25 °C.

5.2.2 Circular Dichroism. CD spectra were recorded with a Jasco J-715 spectropolarimeter equipped with a Peltier-type temperature control system (model PTC-348WI). The instrument was calibrated with an aqueous solution of D-10-(+)-camphorsulfonic acid at 290 nm [9]. The molar ellipticity per mean residue, $[\theta]$ (deg cm² dmol⁻¹), was calculated from the equation: $[\theta] = [\theta]_{\text{obs}} \cdot (\text{mrw}) / 10 \cdot l \cdot C$, where $[\theta]_{\text{obs}}$ is the ellipticity (mdeg), mrw is the mean residue molecular weight, 109 Da, C is the protein concentration (g L⁻¹), and l is the optical path length of the cell (cm). Cells with 0.1 cm path length and protein concentrations of about 0.1-0.3 mg mL⁻¹ were used to record CD spectra between 190 and 250 nm, with a time constant of 16 s, a 2 nm bandwidth, and a scan rate of 5 nm min⁻¹. The spectra were signal-averaged over at least five scans and baseline corrected by subtracting a buffer spectrum, and were analyzed for secondary structure composition according to the CDSSTR method [10] and reference database 4 [11] using Dichroweb [12]. Thermal unfolding curves were recorded in the temperature mode following the signals at 222 and 280 nm for protein and DNA, respectively from 0 to 90°C with an heating rate of 1.0 K min⁻¹.

The evaluation of the helical content in the protein samples was derived from the ellipticity at 222 nm using the relationship provided by Richardson and Makhatadze [13]: $f_h = ([\theta] - [\theta]_c) / ([\theta]_h - [\theta]_c)$ where f_h is the fractional helical content, $[\theta]_h$ and $[\theta]_c$ are the molar ellipticities of the full helical and full coiled protein with a certain number of residues (N_{res}) respectively, which are given by $[\theta]_h = (-40000 + 250 \cdot T) [1 - (2.5/N_{\text{res}})]$ and $[\theta]_c = 640 - 45 \cdot T$, where T is the temperature in Celsius degrees.

5.2.3 Isothermal Titration Calorimetry. ITC experiments were carried out at 25°C using a high-sensitivity CSC-5300 Nano-ITC calorimeter from Calorimetry Science Inc. (Lindon, Utah), with a cell volume of 1.0 mL. DNA solutions were placed in the cell, and concentrated protein solutions in the syringe. The protein solution was injected into the DNA solution with 10 µl aliquots at 400-800 second intervals, prior to each injection, the heat evolution returns to zero, ensuring that the binding

process equilibrates on the time scale of the injection. The DNA concentration in the cell was about 10 μM , and the protein concentration in the syringe was about 70 μM . Both DNA and protein solutions were prepared with the same batch of buffer to minimize the differences in buffer composition and pH. The heat of dilution of the protein into the buffer solution was measured in a separate experiment and appropriate corrections were made. The heat of dilution obtained by averaging the heats of the last 5-10 injections, and subtracted from the measured heat of binding give the same value. Raw data were integrated, corrected for non-specific heats, normalized for concentration, and analyzed by using the Bindworks program supplied with the instrument. Thermodynamic parameters were determined by modelling the experimental data to a theoretical equation based on a simple independent and equivalent binding sites model. The thermodynamic data directly evaluated with this procedure were the binding constant (K_b), the apparent molar ratio between the amount of titrant and the amount of analyte (n) and the change in the standard enthalpy $\Delta_b H^\circ$ per amount of titrant in the binding reaction. The binding enthalpy may also include several other contributions like the heat of protonation of the buffer (if proton transfer occurs in the binding reaction) and/or heat from conformational changes occurring in the analyte and/or titrant upon binding. The change in the standard Gibbs energy and the entropic contribution were calculated using the fundamental relationships $\Delta_b G^\circ = -RT \ln K_b$ ($R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$, $T = 298 \text{ K}$) and $-T\Delta_b S^\circ = \Delta_b G^\circ - \Delta_b H^\circ$.

5.3 Results

5.3.1 Thermal stability of the protein, the DNA and the complex in different buffer conditions.

Far-UV CD spectra of Pax-8 Prd in buffer A, at 0°C and 90°C are reported in Figure 5.1. The deconvolution analysis of CD spectra at 0°C, by means of Dichroweb [12], indicates that the protein has a folded structure with 27% of α -helix content. The same value was obtained at 0°C in buffer B indicating that a higher ionic strength of the buffer solution does not interfere with the folding of the protein. In order to investigate the stability of the protein, the DNA and complex, thermal-induced unfolding curves were recorded following the change of the CD signal between 0 to 90°C, with an heating rate of 1.0 K min^{-1} . The temperature-induced unfolding curves of the protein alone were recorded following the CD signal at 222 nm. After heating, a complete recovery of this signal was observed thus suggesting that the thermal unfolding is a reversible process. Unfolding temperature values of 55.5°C and 58.5°C were obtained in the two different buffer conditions reported in Table 5.1.

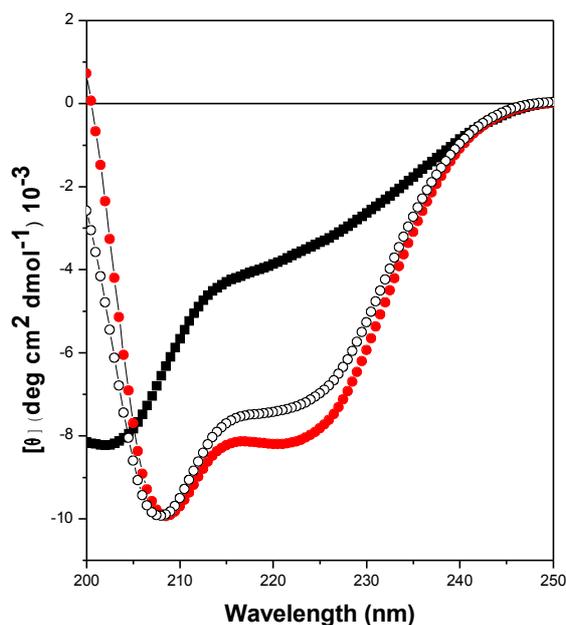


Figure 5.1. Far-UV CD spectra of Pax-8 Prd in the buffer A, recorded at 0°C (filled circles, red), 90°C (filled squares, black), and at 0°C after heating (open circles, black).

Table 5.1. Transition temperatures for the protein, DNA and the (1:1) complex obtained by recording the molar ellipticity at 222 nm or 280 nm in two different buffer conditions (A) 50 mM sodium phosphate, 1mM TCEP pH 6.5 and (B) 50 mM sodium phosphate, 1mM TCEP, 75 mM NaCl, pH 6.5

sample	T_d (°C)	
	Buffer (A)	Buffer (B)
Pax-8 Prd	55.5	58.5
24 mer DNA duplex	67.5	72.0
Complex 1:1 (at 222 nm)	69.5	69.5
Complex 1:1 (at 280 nm)	69.5	72.0

Note : The error in T_d does not exceed 0.2 °C.

The thermal-induced dissociation of 24-mer DNA duplex were investigated by following the CD signal at 280 nm and transition temperatures of 65.0°C and 72.0°C were obtained in the two buffer conditions (see Table 5.1). Moreover, the thermal-induced unfolding curve of the 1:1 protein and DNA complex were acquired by following the CD signal, as function of the temperature, at both 222 nm and 280 nm. The latter experiment is advantageous because of the signal at 222 nm reflects the thermal transition of the protein in the complex, whereas the 280 nm is, practically, indicative of the thermal-induced dissociation of the DNA in the complex. The experimental transition curves of the complex, in buffer A, are shown in Figure 5.2.

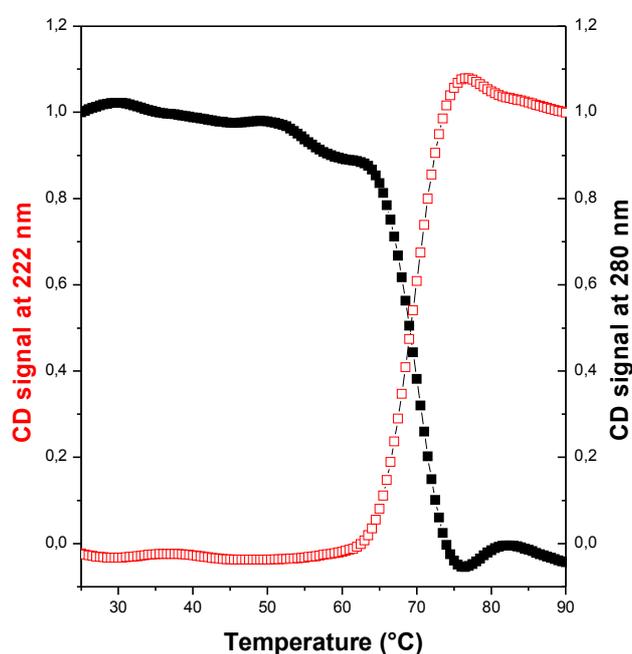


Figure 5.2. Thermal unfolding curves of the 1:1 complex between Pax-8 Prd and the DNA duplex obtained by recording the CD signal at 222 nm (open squares, red) and 280 nm (filled squares, black) in buffer A.

The comparison of the two curves clearly indicates that both the protein and the DNA, in the complex, synergic unfolds/dissociates at the same temperature, 69.5°C, which is higher of the transition temperatures of the two free molecules. The last experiment were performed in the buffer B as well, and the results are collected in Table 5.1; in that case the DNA dissociation temperature does not change upon the complex formation but the unfolding temperature of the protein increases from 58.5 °C up to 69.5 °C. Intriguingly, thus report in Figure 5.3, reheating the complex solution, both in buffer A or buffer B, the same transition curves were obtained, thus suggesting, that the complex formation is completely reversible. Clearly, the binding process strongly stabilises the protein structure while little or negligible effects have been observed for the DNA duplex structure.

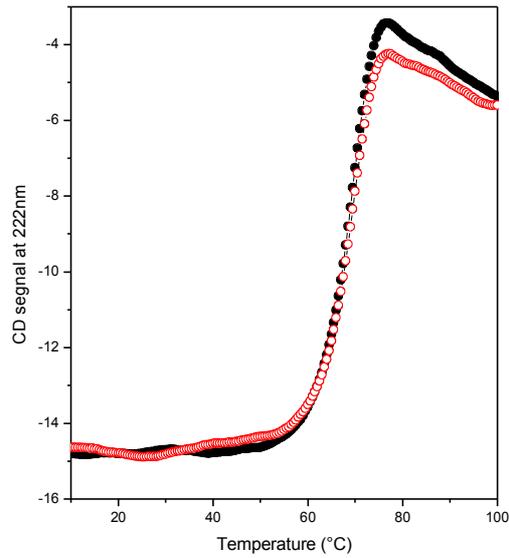


Figure 5.3 . Thermal-induced denaturation curves of the first (open circles, red) and second heating (filled circles, black) of a 1:1 complex of Pax-8 Prd with 24-mer DNA duplex by following changes of CD signal at 222 nm. The two curves are super-imposable suggesting that the complex formation is a fully reversible event.

It was possible to appreciate the protein stabilization by comparing the unfolding curves of the free protein and the one obtained following the CD signal of the protein in the complex as show in Figure 5.4. The two curves are centred on two different temperatures and, interestingly, they have two different slopes indicating that the unfolding enthalpy of the protein increases when it is bound to DNA.

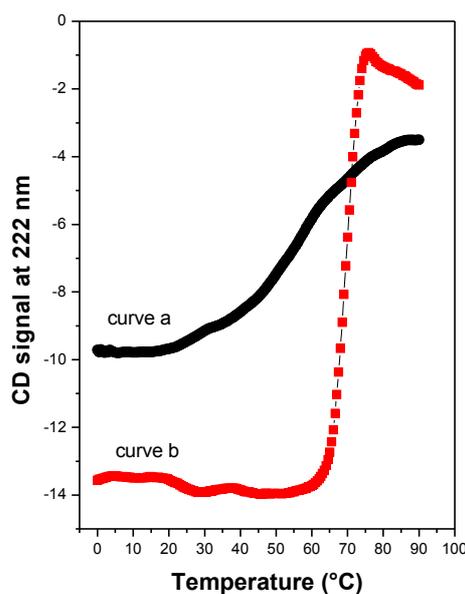


Figure 5.4 CD signal at 222 nm as function of temperature of free Pax-8 Prd (curve a, black) and in a 1:1 complex with the 24-mer duplex (curve b, red) recorded in buffer A.

5.3.2 Protein structural changes upon DNA interaction. The CD spectra of Pax-8 Prd, 24-mer DNA duplex and their 1:1 complex, at 0°C in buffer A, are reported in Figure 5.5, together with the spectrum obtained by subtracting the spectrum of DNA from that of the complex with the aim to single out the contribution of protein in the complex, in the assumption that no changes occur in DNA duplex conformation upon binding. Comparison of the CD spectra of free and bound Pax-8 Prd indicates a remarkable increase in the protein secondary structure. Particularly, about a 13% increase in the helical content was estimated according to Richardson and Makhatadze equation (see Experimental). It could be possibly originate from the passage of some residues from a coil to a helical conformation.

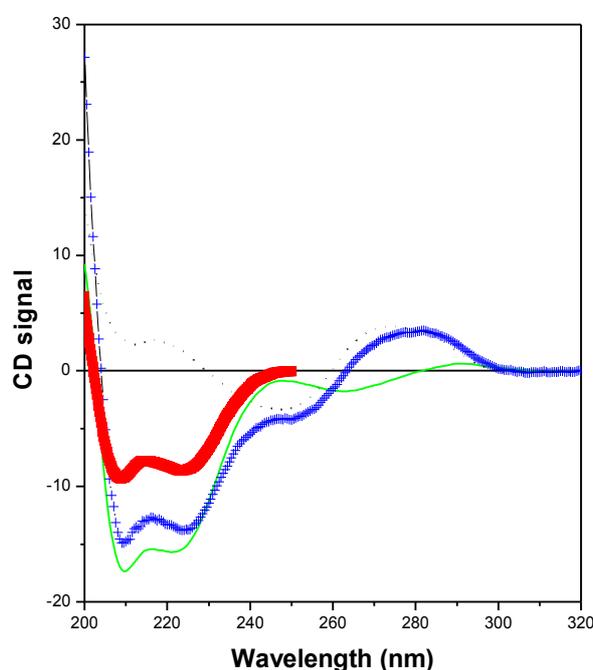


Figure 5.5. Far-UV CD spectra of free Pax-8 Prd (filled squares, red), free DNA duplex (dot line, black), their 1:1 complex (cross line, blue); spectrum obtained by subtracting that of free 24-mer duplex from that of the complex (solid line, green). The spectra are recorded at 25°C in buffer A.

5.3.3 Energetics of the protein/DNA interaction. The binding of Pax-8 Prd with 24-mer DNA duplex, containing a specific sequence of the tyroperoxidase promoter, has been measured at 25 °C in both buffer conditions. Figure 5.6, Panel A shows a representative ITC profile for the association between the protein and its target DNA. An exothermic heat pulse is observed after each injection of protein into the DNA solution. The area of each exothermic peak is integrated, and the heat of dilution of protein is subtracted from the integrated values. The corrected heat is normalized by the moles of protein injected, and the resulting values are plotted as a function of the molar ratio, as shown in Figure 5.6 Panel B.

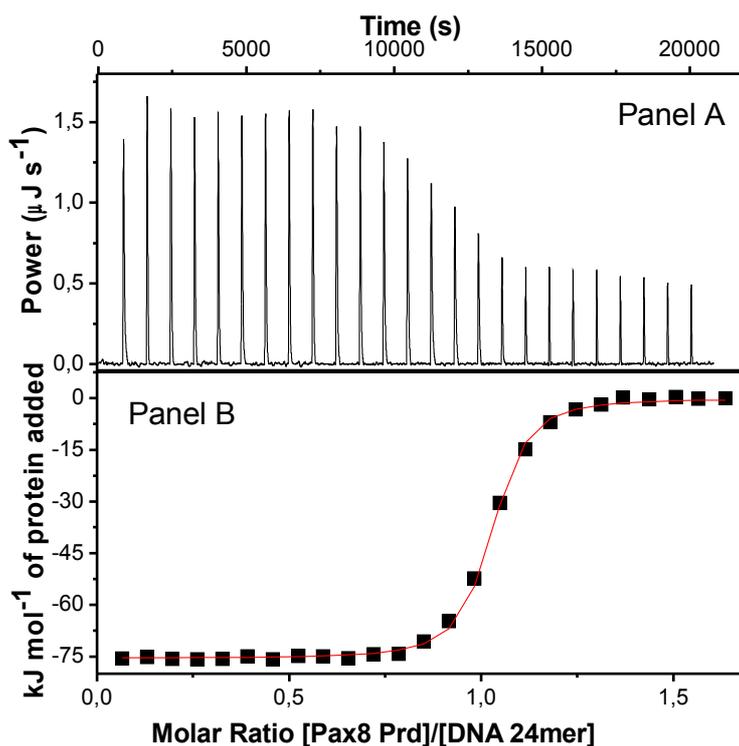


Figure 5.6. Panel A. Representative isothermal titration experiment of Pax-8 Prd with the 24-mer DNA duplex in buffer A. Panel B. Binding enthalpies obtained integrating the raw data from the Panel A, after subtraction of the average heat of dilution from the individual integrated heats and normalization for protein concentration; the solid curves represent the best-fit obtained with the equivalent and independent binding sites model.

The raw ITC data are modelled to a theoretical equation based on an equivalent and independent binding site model, by using a non-linear least-squares method. The interpolation procedure provides an evaluation of the binding constant, K_b , the stoichiometry and the change of enthalpy, $\Delta_b H^\circ$ in the binding process. The experimental data are collected in Table 5.2. The value of the binding constant K_b is $4.0 \cdot 10^7 \text{ M}^{-1}$ so that $\Delta_b G^\circ$ is $-43.3 \text{ kJ mol}^{-1}$, indicating that the association is strongly favoured from the thermodynamic point of view. In buffer A the values of $\Delta_b H^\circ$ and $T\Delta_b S^\circ$ are $-74.6 \text{ kJ mol}^{-1}$ and $-31.3 \text{ kJ mol}^{-1}$, respectively, emphasizing that the binding process is enthalpy driven. The negative entropic contribution to the binding could be mainly due to the protein induced folding upon binding resulting in an unfavourable reduction of conformational entropy. In buffer B constant K_b is $7.0 \cdot 10^6 \text{ M}^{-1}$ so that $\Delta_b G^\circ$ is $-39.2 \text{ kJ mol}^{-1}$ indicating an expected decrease of the binding constant, due to the reduced electrostatic interactions because of the increased ionic strength of the buffer. The values of $\Delta_b H^\circ$ and $T\Delta_b S^\circ$ are $-70.0 \text{ kJ mol}^{-1}$ and -30.0 kJ

mol⁻¹, respectively, indicating that the global thermodynamic signature of the event is conserved and the binding process remains enthalpy driven.

Table 5.2. Thermodynamic parameters from ITC experiments for the binding, at 25 °C, of Pax8 Prd to the 24-mer DNA duplex in both buffer (A) and buffer (B). The values of the thermodynamic properties and the binding constant have been obtained from a fitting procedure based on an independent and equivalent binding sites model as described in the experimental section, and for the experiments in buffer A the interpolation has been repeated with a competitive specific-non-specific binding model.

Buffer	K _b (M ⁻¹)	Δ _b H (kJ mol ⁻¹)	Δ _b G (kJ mol ⁻¹)	TΔ _b S (kJ mol ⁻¹)
A	4.0 (± 0.5) · 10 ⁷	-75.0 ± 1	-43.3 ± 0.3	-31.0 ± 1.0
B	7.0 (± 1) · 10 ⁶	-70.0 ± 1	-39.2 ± 0.3	-30.0 ± 1

Note : Each values represents the value averaged over at least three measurements.

5.4 Discussion

Combining together our spectroscopic and calorimetric data we could delineate a general picture illustrating the molecular mechanism of recognition between Pax-8 Prd and its target DNA. The interaction results in the formation of a specific 1:1 complex. Upon binding, significant conformational changes of the protein occur (helix formation, domains reorientation, global refolding). This is not surprising in the light of its structural features; the protein contains two small domains connected between each other by a very long and flexible linker. Both, the two domains and the linker region, could host refolding events upon DNA interaction. In particular, in order to contact the DNA target the two domains have to work synergistically choosing the best orientation to better perform the binding activity. This means that the two domains, which were completely independent in solution, in the DNA-free conformation, strongly reduce the number of accessible orientation with respect to each other in presence of DNA. The DNA-induced orientation effect of the two domains could be one of the reasons of the high negative value of the entropic contribution to the interaction. Moreover, the choice of a fixed orientation of the two domains in the presence of the DNA could be the driving force that obligates the linker region to choose a fixed conformation. An analysis of both the enthalpic and the entropic contributions could be due using as reference

model the structures of the complexes of two homologous proteins Pax-5 and Pax-6 with their DNA targets [14,15].

5.4.1 Analysis of the enthalpic and the entropic contributions to the binding. The large negative enthalpy change in binding reactions is mainly due to the formation of the recognition contacts, such as hydrogen-bonds, ion-pairs and non polar contacts between DNA and protein surfaces and to the protein refolding heat generated upon interaction. To get a rational interpretation of the thermodynamic parameters found in our characterization, structural information are essential. Pax-8 Prd shares a high sequence and structural homology with other two protein, Pax-5 Prd and Pax-6 Prd and, interestingly, for both of them the X-ray structures in complex with their target DNA are available [9,10]. Since the high homology of sequence between those proteins and Pax-8 Prd, it is tempting to speculate that they could represent good models for the complex of Pax-8 Prd and DNA target. Comparing together the solution structure of the unbound Pax-8 Prd with the Pax-5 Prd and Pax-6 Prd complexes, it is clear that the free form of Pax-8 Prd already contains all the secondary structure elements that are detectable in the bound form of the other two proteins; suggesting that the refolding events that take place during the DNA interaction are not dramatic. Intriguingly, the helix III of the N-terminal domain of Pax-8 Prd, which could be involved, together with the helix VI of its C-terminal domain, in the DNA recognition, appears to be shorter than expected, probably, this could be a possible refolding site upon DNA interaction. In particular, 5-6 residues belonging to the helix III could refold upon DNA binding, generating an enthalpic contribution of about -18 kJ mol^{-1} , since the evaluation of the enthalpy change associated with helix-to-coil transition account for $3-4 \text{ kJ mol}^{-1}$ per residues, at room temperature [13]. According to our CD data, upon DNA interaction, the protein secondary structure content increases of about the 10-13% thus suggesting that up to 10 residues (about -30 kJ mol^{-1}) could be involved in the refolding process. This finding suggests that the helix III is not the only region affected by the conformational change. In fact, the model represented by the Pax-5 Prd and/or Pax-6 Prd /DNA complex suggests that another region could host refolding events; in particular, the N-terminal region could fold into a short β -hairpin contacting DNA through the minor groove. In the large negative enthalpy change are contained possible positive contributions. A potential source of unfavourable enthalpy is the DNA distortion but this is not that the case since our CD analysis did not highlight any conformational change for the DNA [11]. Positive enthalpy could also derive from the disruption of the water molecules shell of the interacting interfaces, but this is usually a small contribution both in the case of desolvation of non-polar and polar surfaces. On the other hand, at the latter small enthalpic effect is associated a very large entropy change which is, usually, the dominant source of favourable entropy. The DNA

interaction promotes a global stabilization of the entire protein structure, probably reducing the flexibility of the system. In the hypothesis that the structure of Pax-8 Prd with its DNA closely resembles the one of Pax-5 Prd and/or Pax-6, it is possible to exclude the refolding of the linker region, too. In particular, in the bound form of Pax-6 Prd the linker region adopts an extended conformation laying partially in the minor groove of the DNA. If this interaction mode is conserved for Pax-8 Prd, it is possible to predict that the linker region will not contribute to the enthalpy change but its effect will be counted in the entropy change value. The linker, spanning the minor groove of the DNA, could displace ordered water molecules. It is well known that the water molecules soaking a double helix of DNA are not structured in the major groove but in the minor groove, because of the hindrance, they are very well structured between them all and the DNA backbone; a number of previous studies have been devoted to understand the thermodynamics of such a system. Privalov et al. found that a seven base pairs double helix DNA possesses a spine of eleven water molecules structured in the minor groove in two hydration shells [16] and successively, Jana et al. evaluated, theoretically, an entropic contribution of 0.86 kcal/mol, at 300 K, to displace one molecule of water from the minor groove [17]. According to the structure of the complex between Pax-6 Prd and its DNA the inter-domains linker spans a segment which is eight base pairs long displacing about 11 water molecules, if this contribution is unchanged in the Pax-8 Prd interaction the related entropic contribution will be about 39 kJ mol⁻¹. Our results suggested that this large positive value is counterbalanced and over taken resulting in a net negative (unfavourable) entropy change. The net entropic contributions ($\Delta_b S^\circ$) upon DNA interaction are widely considered resulting from the interplay between three major entropic contributions. The solvation favourable entropy change ΔS_{solv} , due to enhancement in the degrees of freedom of solvent molecules as a result of their restructuring and displacement, particularly around non polar groups, upon molecular associations. In this contribution is counted the releasing of counter ions from DNA upon interaction, as well. The conformational entropy change, ΔS_{conf} , is the unfavourable entropic change that arises from the restriction of conformational degrees of freedom of the backbone and side chain atoms upon molecular associations. Finally, the mixing entropy change, ΔS_{mix} is the unfavourable entropic change due to the restriction in the translational, rotational and vibrational degrees of freedom of molecules upon binding [18]. Several lines of evidence suggest that ΔS_{mix} typically contributes no more than about $-10 \text{ cal mol}^{-1}\text{K}^{-1}$ of entropy penalty to the overall entropic change upon binding [19,20]. On the basis of these considerations, we attribute the global unfavourable entropic change obtained by experimental ITC data as resulting from a large loss of conformational degrees of freedom of backbone and side chain atoms in both the protein and DNA.

We can then conclude that the interaction between Pax-8 Prd and the target DNA is a strongly thermodynamically favoured process. The remarkable energy Gibbs change of the system upon interaction results from the combination of a large and negative enthalpic contribution ($\Delta_b H^\circ < 0$) that drives an unfavourable entropy change ($-T\Delta_b S^\circ > 0$). According to the structure of Pax-6 Prd /DNA complex, the two domain of the Pax-8 Prd could contact the DNA through the major groove, while the linker region could contact the minor groove together with the N-terminal region of the protein. The enthalpic contribution accounts for specific recognition contacts formed between the protein and the DNA, and also for a large refolding of the protein possibly localised in the region of the helix III and in the N-terminal portion of the protein upon binding. The net unfavourable entropic contribution, $\Delta_b S^\circ$, is derived from the greater immobilization of the interface and from coupling between DNA binding and partial protein folding. This contribution is sufficiently high to neutralize and overtake the positive entropic contribution by the displaced water molecules from the interacting surfaces and in particular from the minor groove.

5.5 References

1. Xu W, Rould MA, Jun S, Desplan C, Pabo CO. Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell*. 1995; 80:639-50
2. Wang Q, Fang WH, Krupinski J, Kumar S, Slevin M, Kumar P. Pax genes in embryogenesis and oncogenesis. *J Cell Mol Med*. 2008; 12:2281-94
3. Espinoza CR, Schmitt TL, Loos U. Thyroid transcription factor 1 and Pax8 synergistically activate the promoter of the human thyroglobulin gene. *J Mol Endocrinol*. 2001; 27:59-67
4. Zannini M, Francis-Lang H, Plachov D, Di Lauro R. Pax-8, a paired domain-containing protein, binds to a sequence overlapping the recognition site of a homeodomain and activates transcription from two thyroid-specific promoters. *Mol Cell Biol*. 1992; 9:4230-41
5. Codutti L, van Ingen H, Vascotto C, Fogolari F, Corazza A, Tell G, Quadrifoglio F, Viglino P, Boelens R, Esposito G. The solution structure of DNA-free Pax-8 paired box domain accounts for redox regulation of transcriptional activity in the pax protein family. *J. Biol. Chem*. 2008; 283:33321-8
6. Cao X, Kambe F, Lu X, Kobayashi N, Ohmori S, Seo H. Glutathionylation of two cysteine residues in paired domain regulates DNA binding activity of Pax-8. *J Biol Chem*. 2005; 280: 25901-6

7. Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 1995;11: 2411-2423.
8. Cantor CR, Warshaw MM, Shapiro H. Oligonucleotide interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers* 1970; 9: 1059-1077.
9. Venyaminov SY, Yang JT. Determination of protein secondary structure. In *Circular dichroism and the conformational analysis of biomolecules.* ed. G.D. Fasman Plenum Press, New York, 1996;69-107.
10. Compton LA, Johnson WC Jr. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* 1986;155:155-167
11. Sreema N, Woody RW. A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* 1993;209:32-44.
12. Whitmore AL, Wallace BA. Dichroweb, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Reserch* 2004; 32:668-73.
13. Richardson JM, Makhatazde GI. Temperature dependence of thermodynamics of helix-coil transition. *J Mol Biol* 2004; 335:1029-1037.
14. Garvie CW, Hagman J, Wolberger C. Structural Studies of Ets-1/Pax5 Complex Formation on DNA. *Molecular Cell*, 2001; 8:1267-1276.
15. Xu HE, Rould MA, Xu W, Epstein JA, Maas RL, Pabo CO. Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.* 1999; 13:1263–1275.
16. Privalov PL, Dragan AI, Crane-Robinson C, Breslauer KJ, Remeta DP, Minetti CASA. What Drives Proteins into the Major or Minor Grooves of DNA? *J Mol Biol.* 2007; 365:1–9.
17. Jana B, Pal S, Maiti PK, Shiang-Tai L, Hynes JT, Bagchi B. Entropy of Water in the Hydration Layer of Major and Minor Grooves of DNA. *The Journal of Physical Chemistry* 2006; 110:19611-19618 .
18. Jen-Jacobson L, Engler LE, Jacobson LA, Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure.* 2000, 8:1015-23.
19. Tamura A, Privalov PL. The entropy cost of protein association. *J. Mol. Biol.* 1997; 273: 1048–1060.
20. Siebert X, Amzel LM. Loss of translational entropy in molecular associations. *Proteins* 2004; 54:104–115.

CONCLUSIONS AND PERSPECTIVES

In the recent years the structure and function of site-specific DNA binding proteins are matters of considerable interest because of the crucial roles of these proteins in the expression and regulation of genetic information. From a purely structural viewpoint, it is now evident that nature has devised a wide range of strategies to build proteins that recognize a specific DNA base sequence. The homeodomain is the highly conserved DNA-binding domain of a class of proteins that function as transcriptional regulators, specifying positional and temporal information in the commitment of embryonic cells to specific developmental pathways. An early step in the cascade of events associated with development is the sequence-specific binding of the transcriptional regulator. Various genetic diseases and developmental abnormalities have been mapped to base changes in a homeobox, the gene that encodes the homeodomain protein. Such mutations often alter structural stability, DNA recognition specificity, or binding affinity to the DNA of the homeodomain. The quantitative description of the forces that govern the formation of biomolecular complexes is of primary importance to understand what forces drive the molecules to interact with each other. These forces modulate transcriptional factors activity and their alteration caused by mutations can lead to human diseases.

During my research years I investigated on the homeodomain of TTF-1 (TTF-1HD) and the paired domain of Pax-8. These domain are part of modular proteins and they are of fundamental importance in controlling the development of thyroid binding the thyroglobulin (Tg) and thyroperoxidase (TPO) promoters. In particular I investigated on their thermodynamic stability as well as on the energetics of the interaction with DNA.

The results obtained for TTF-1 homeodomain indicated that this small globular protein is poorly resistant to temperature and not very stable. In fact at pH 5.0 its denaturation temperature, T_d is 50 °C and $\Delta_d G^\circ$ evaluate at 25°C is 11.6 kJ mol⁻¹. This value emphasizes the low conformational stability of the protein. The value of $\Delta_d C_p$ is 1.5 kJ K⁻¹mol⁻¹, suggesting that the homeodomain has not got a well-developed hydrophobic core, probably because it is not designed to be extremely stable, but to specifically interact with DNA. In addition the thermal stability TTF-1HD strongly depends on KCl concentration T_d increases from 32 °C in the absence of KCl to 59 °C in 500 mM KCl. This finding is in part expected because the protein has 17 positively charged (Arg + Lys) residues that strongly interact with DNA, and 5 negatively charged (Asp + Glu) residues.

Binding measurements demonstrate that the association of TTF-1HD with DNA is very specific in our experimental condition in fact K_b is $1.5 \cdot 10^8$ M⁻¹. The large and negative value of $\Delta_b H^\circ$ indicate that the association of TTF-1HD to its target DNA sequence, is favoured by enthalpy, as generally

found in the case of proteins that bind in a sequence-specific manner into the DNA major groove. The positive value of $T\Delta_b S^\circ$ indicate that the association of TTF-1HD to its target DNA sequence, is favoured by entropy, as generally found in the case of proteins that bind in a non-sequence-specific manner into the DNA minor groove. This finding seems to be reliable because HDs interact with both the major and minor grooves of DNA double helix. Both enthalpy and entropy.

Binding demonstrate that the association of TTF-1HD with DNA is very specific in fact K_b is $1.5 \cdot 10^8 \text{ M}^{-1}$ in our experimental condition.

A more detailed information has been obtained from molecular dynamics simulation suggesting a direct and water-mediated interaction between homeodomain and DNA. This structural information has been the basis for further investigations on mutant forms of TTF-1HD. To understand the role played by a single amino acid in the thermal stability and the DNA binding specificity, we reported the characterization of single mutant forms Q50A and Y54A and a deleted form (desTTF-1HD).

The overall results indicate that these point mutations do not prevent the binding because the substitution of a single amino acid is not sufficient to abolish the DNA binding ability even if it result considerably weakened. The thermodynamic signature of binding is unchanged and it is enthalpy driven. Furthermore, the results of the N-terminal arm indicated that the charged N-terminal extension destabilized the folded domain and concur to the complex formation through an enthalpic negative contribution by increasing the number of specific contacts with the minor groove of DNA.

As biological essays have shown that the N-terminal region of TTF-1 is crucial for transcription control of thyroid development the fusion protein containing the N-terminal region and the homeodomain of TTF-1 has been produced (TTF-1NH₂HD) and characterized. The obtained results suggest that this N-terminal region exists in a random coil conformation and that does not interact with DNA. In fact, it is well known that the biological function of TTF-1 in the activation of thyroid genes depends on the DNA binding recognition of the homeodomain, while N-terminal region interact with others factors. Among them Pax-8 transcription factor is one of the most important, in fact it known that there is a direct interaction between Pax-8 and TTF-1, which occurs via protein-protein interaction in the absence of DNA.

This modular protein consist of a N-terminal DNA-binding region and a C-terminal transactivation region that interact with N-terminal region of TTF-1. The study of DNA-binding region of Pax-8 is an important step to understand the activation mechanism of thyroid gene. In particular the thermal stability of Pax-8 Prd domain and energetics of interaction to its target DNA have been characterized. The remarkable value of the Gibbs energy change of interaction results from the combination of a large and negative enthalpic contribution that drives an unfavourable entropy

change. The enthalpic contribution accounts for specific recognition contacts formed between the protein and the DNA, and also for a large refolding of the protein upon binding. The net unfavourable entropic contribution, $\Delta_b S^\circ$, is derived from the greater immobilization of the interface and from coupling between DNA binding and partial protein folding. To support and complement our thermodynamic results further structural investigation on Pax-8 prd/DNA complexes, are needed.

The molecular mechanisms leading to the synergistic activity of Pax-8 and TTF-1 implicate a region of the Tg and TPO promoter, containing 5'-CAAG-3' sequence, which is where the binding sites of the two factors overlap. It is, however, still unclear if both the two proteins bind to the DNA or if only one of them is bound and thus recruits the other one to the promoter. Thus it would be interesting to study the interaction of these two full-length proteins in absence and in presence of both the Tg and TPO promoters. Furthermore, to support biological studies, the full-length TTF-1 and Pax-8 could be characterized in order to study the influence of the additional domains in DNA-binding.

PUBLICATIONS

Most of the results presented in this Ph.D thesis have been published in the following paper:

Pompea Del Vecchio **Paola Carullo** Guido Barone Bruno Pagano Giuseppe Graziano Alessio Iannetti Renato Acquaviva Antonio Leonardi and Silvestro Formisano. Conformational stability and DNA binding energetics of the rat thyroid transcription factor 1 homeodomain. *Proteins* 2008 70 748-60

Abstract

The conformational stability of the rat thyroid transcription factor 1 homeodomain, TTF-1HD, has been investigated by means of circular dichroism (CD) and differential scanning calorimetry (DSC) measurements at pH 5.0 as a function of KCl concentration. Thermal unfolding of TTF-1HD is a reversible two-state transition. The protein is not stable against temperature, showing a denaturation temperature of 32°C in the absence of salt and 50°C at 75 mM KCl. The binding energetics of TTF-1HD to its target DNA sequence has been characterized by means of isothermal titration calorimetry (ITC) measurements, complemented with CD data. At 25°C, pH 5.0 and 75 mM KCl, the binding constant amounts to $1.5 \cdot 10^8 \text{ M}^{-1}$ and the binding enthalpy change amounts to 241 kJ mol⁻¹. The process is enthalpy driven, but also the entropy change is favorable to complex formation. To gain a molecular level understanding of the interactions determining the association of TTF-1HD to the target DNA sequence structural information would be requested, but it is not yet available. Therefore, structural models of two complexes, TTF-1HD with the target DNA sequence and TTF-1HD with a modified DNA sequence, have been constructed by using as a template the NMR structure of the complex between NK-2 HD and its target DNA, and by performing molecular dynamics simulations 3.5 ns long. Analysis of these models allows one to shed light on the origin of the DNA binding specificity characteristic of TTF-1HD.

During my Ph.D time, I have also published other three papers produced in collaborations with other research group on conformational stability of other proteins. The following are the references and their abstracts.

Fabrizia Foglia, **Paola Carullo** and Pompea Del Vecchio. The effect of trimethylamine N-oxide on RNase A stability: a DSC study. *Journal of Thermal Analysis and Calorimetry* 2008 91 67-72

Abstract

The thermal stability of bovine pancreatic ribonuclease (RNase A) has been investigated in the presence of trimethylamine N-oxide (TMAO), a naturally occurring osmolyte, by means of differential scanning calorimetry (DSC) and circular dichroism (CD) measurements at neutral and acid pH conditions. It is well known that compatible osmolytes such as TMAO are effective in stabilizing protein structure and counteracting the denaturing effect of urea and guanidinium hydrochloride (GuHCl). Calorimetric results show that TMAO stabilizes RNase A at pH 7.0 and does not stabilize the protein at pH 4.0. RNase A thermal denaturation in the presence of TMAO is

a reversible two-state N \rightleftharpoons D process. We also show that TMAO counteracts the urea and GuHCl denaturing effect at neutral pH, whereas the counteracting ability is lost at acid pH.

Pompea Del Vecchio, Mikael Elias, Luigia Merone, Giuseppe Graziano, Jérôme Dupuy, Luigi Mandrich, **Paola Carullo**, Bertrand Fournier, Daniel Rochu, Mosè Rossi, Patrick Masson, Eric Chabriere and Giuseppe Manco. Structural determinants of the high thermal stability of SsoPox from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 2009 13 461-470

Abstract

Organophosphates (OPs) constitute the largest class of insecticides used worldwide and certain of them are potent nerve agents. Consequently, enzymes degrading Ops are of paramount interest, as they could be used as bioscavengers and biodecontaminants. Looking for a stable Ops catalyst, able to support industrial process constraints, a hyperthermophilic phosphotriesterase (PTE) (SsoPox) was isolated from the archaeon *Sulfolobus solfataricus* and was found to be highly thermostable. The solved 3D structure revealed that SsoPox is a noncovalent dimer, with lactonase activity against “quorum sensing signals”, and therefore could represent also a potential weapon against certain pathogens. The structural basis of the high thermostability of SsoPox has been investigated by performing a careful comparison between its structure and that of two mesophilic PTEs from *Pseudomonas diminuta* and *Agrobacterium radiobacter*. In addition, the conformational stability of SsoPox against the denaturing action of temperature and GuHCl has been determined by means of circular dichroism and fluorescence measurements. The data suggest that the two fundamental differences between SsoPox and the mesophilic counterparts are: (a) a larger number of surface salt bridges, also involved in complex networks; (b) a tighter quaternary structure due to an optimization of the interactions at the interface between the two monomers.

Carla Esposito, **Paola Carullo**, Emilia Pedone, Giuseppe Graziano, Pompea Del Vecchio and Rita Berisio. Dimerisation and structural integrity of Heparin Binding Hemagglutinin A from *Mycobacterium tuberculosis*: Implications for bacterial agglutination. *FEBS Letters* 2010 584 1091-1096

Abstract

Heparin Binding Hemagglutinin A (HBHA) is hitherto the sole virulence factor associated with tuberculosis dissemination from the lungs, the site of primary infection, to epithelial cells. We have previously reported the solution structure of HBHA, a dimeric and elongated molecule. Since oligomerisation of HBHA is associated with its ability to induce bacterial agglutination, we investigated this process using experimental and modelling techniques. We here identified a short segment of HBHA whose presence is mandatory for the stability of folded conformation, whose denaturation is a reversible two-state process. Our data suggest that agglutination-driven cell-cell interactions do not occur via association of HBHA monomers, nor via association of HBHA dimers and open the scenario to a possible trans-dimerisation process.