

**UNIVERSITA' DEGLI STUDI DI NAPOLI
FEDERICO II**



FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E NATURALI

DOTTORATO IN SCIENZE CHIMICHE XVIII CICLO

**STUDIO DELLE INTERAZIONI FRA SISTEMI PROTEICI E
ACIDI NUCLEICI MEDIANTE TECNICHE DI RISONANZA
MAGNETICA NUCLEARE**

**STUDIES OF THE INTERACTIONS PROTEINS/NUCLEIC ACIDS BY
NUCLEAR MAGNETIC RESONANCE**

Candidata

Caterina ALFANO

Tutore

Prof. Carlo Pedone

Relatore

Prof.ssa Delia Picone

Coordinatore

Prof.ssa Rosa Lanzetta

2002-2006

INDEX

CHAPTER I INTRODUCTION

1.1 Human La protein: a RNA binding protein	1
1.1.1 Rules and localisation of La proteins	1
1.1.2 Structural organisation of hLa and its interaction with RNA	2
1.1.3 Our aims on the human La project	7
1.2 DNA binding proteins	8
1.2.1 Calorimetric evidences of protein/DNA interactions	9
1.2.2 The systems chosen for study	11

CHAPTER II MATERIALS AND METHODS

2.1 DNA synthesis	14
2.1.1 PCR-ESRA Synthesis of 12-mer Sox-DNA	14
2.1.2 PCR-ESRA Synthesis of 16-mer Antp-DNA	15
2.2 Protein preparations	15
2.2.1 Antennapedia C39S homeodomain	15
2.2.2 La proteins	17
2.3 NMR measurements	19
2.3.1 NMR measurements on Antennapedia project	19
2.3.2 NMR measurements on La protein project	19
2.4 NMR Gel shift assays on La protein fragments	21

CHAPTER III RESULTS AND DISCUSSION: ANTENNAPEdia HOMEODOMAIN

3.1 The Antennapedia homeodomain in a wider contest	22
3.2 DNA synthesis: the PCR-ESRA method	23

3.2.1 Sox-5 DNA synthesis	26
3.2.2 Antennapedia DNA synthesis	37
3.2.3 PCR cost of preparing doubly-labelled DNA	41
3.3 Antennapedia C39S homeodomain	42
3.3.1 Protein preparation	42
3.4 NMR analysis of the Antennapedia homeodomain	47
3.4.1 Resonance assignment	49
3.4.2 NMR relaxation analysis	54

CHAPTER IV RESULTS AND DISCUSSION: HUMAN LA PROTEIN

4.1 Protein preparation	60
4.1.1 La1-194 and LaWT proteins purification	61
4.2.1 La105-202, La105-334, La1-334 and La1-344 proteins purification	63
4.2 Gel shift assays	67
4.2.1 The EMSA technique	67
4.2.2 Gel shift assays with 10 nt oligo(U)	69
4.3 Structural characterization of La protein	71
4.3.1 NMR in protein structure determination	72
4.3.2 Sequence-specific assignment of La105-202	74
4.3.3 Secondary structure analysis of La105-202	74
4.3.4 Restraints analysis for structure calculation of La105-202	77
4.3.5 Structural calculation of La105-202	81
4.3.6 Chemical shift mapping of the hLa N-terminal domain	83
4.4 Binding of La with the Hepatitis C Virus	86

APPENDIX I ANALYSIS OF DNA SEQUENCES BY COMPUTER PACKAGE	90
---	----

APPENDIX II NMR DATA OF ANTENNAPEDIA HOMEODOMAIN	118
---	-----

APPENDIX III	NMR DATA OF HUMAN LA PROTEIN	126
APPENDIX IV	PROTEIN SEQUENCES AND PROTPARAM OUTPUTS	131
REFERENCES		139

PREFACE

The proteins we observe in nature have evolved, through selection pressure, to perform specific functions. The functional properties of proteins depend upon their three-dimensional structures which arise because particular sequences of amino acids in polypeptide chains fold to generate, from linear chain, compact domains with specific three-dimensional structures. The folded domains either can serve as modules for building up large assemblies such as virus particles or muscle fibres or can carry oxygen or that regulate the function of DNA or that regulate the “life” of RNA. To understand the biological function of proteins we would therefore like to be able to deduce or predict the three-dimensional structure from the amino acid sequence: this is not currently possible, however. We cannot do it. In spite of considerable efforts over the last 25 years, this folding problem is still unsolved and remains one of the most basic intellectual challenges in molecular biology. Since the three-dimensional structures of individual proteins cannot be predicted, they have instead to be determinate experimentally by x-ray crystallography or NMR techniques. Over the past 30 years the structures of thousands proteins have been solved by these two techniques and this has generate a body of information from which a set of basic principles of protein structure has emerged. These principles make it easier for us to understand how protein structure is generated, to identify common structural themes, to relate structure to function and to see fundamental relationship between different proteins. However our current knowledge of the relation between structure and function of proteins is insufficient to deduce the function of a protein from its structure alone, although structural homology with proteins of known function can sometimes allow this. It is necessary then to combine biochemical studies with structural information.

In this project we investigated the relationship structure/function of protein/nucleic acids systems by nuclear magnetic resonance and biochemical assays showing as both approaches are necessary. In particular we focussed our studies on two different interactions: protein/DNA and protein/RNA interactions.

CHAPTER I

INTRODUCTION

1.1 Human La protein: a RNA binding protein

1.1.1 Rules and localisation of La proteins

The La protein is an ubiquitous, abundant, monomeric phosphoprotein found in the nucleus of eukaryotic cells where it associates predominantly with the a short polyuridylylate sequence (UUU_{OH}) at the 3' end of almost all nascent RNA polymerase III (pol III) transcripts (Wolin *et al.* 2002, Maraia *et al.* 2001). The La protein (also called SS-B) was first described as an autoantigen in patients suffering from the rheumatic diseases systemic lupus erythematosus and Sjogren's syndrome (Mattioli *et al.* 1974, Alspaugh *et al.* 1975). The name La derives from the name of the patient in which the antibody was first detected, and SS-B refers to the prevalence of the antibodies in Sjogren's syndrome.

In yeast the La protein (Lh1p1) associates with RNAs transcribed by pol II that contain a 3' end. The specific binding of La to the 3' termini in these precursor molecules protects them from exonuclease digestion and thereby regulates downstream processing (Fan *et al.* 1998, Yoo *et al.* 1997). In the case of pre-tRNA, La helps to ensure the correct endonucleolytic digestion of the 5' and the 3' extensions and has recently been shown to act as a chaperone in promoting the correct folding of these molecules (Fan *et al.* 1998, Yoo *et al.* 1997, Chakshumathi *et al.* 2003). Another facet of the role of La in regulating RNA processing is evident from observations that the protein serves to retain precursor RNA molecules in the nucleus (Simons *et al.* 2003, Intine *et al.* 2002, Raats *et al.* 1996). Moreover, La has been reported to assist the assembly of functional ribonucleoprotein particles, an activity that may be promoted by its association with RNA helicases (Pannone *et al.* 2001, Fouraux *et al.* 2002). In addition to this rather complex functional profile, La also has a role in translation (Wolin *et al.* 2002, Maraia *et al.* 2001). For example La

can bind to the internal ribosome entry sites (IRES) of hepatitis C virus (HCV) and the X-linked inhibitor of apoptosis protein, serving in both cases to stimulate translation initiation. In the case of the HCV IRES, which lacks a 3'-UUU_{OH}, La appears to bind specifically to an internal sequence in the vicinity of the initiator AUG codon (Ali *et al.* 2000, Holcik *et al.* 2000, Pudi *et al.* 2003).

1.1.2 Structural organisation of hLa and its interaction with RNA

La protein belongs to the RNA Recognition Motif (RRM) family of protein: structural data for several RRM proteins have shown that the RRM domains exhibit in most cases a common fold, a $\beta 1-\alpha 1-\beta 2-\beta 3-\alpha 2-\beta 4$ topology with the four antiparallel β -strands creating a platform for RNA binding. The binding specificity of these RRM domains is largely determined by an extensive network of hydrogen bonds between the RNA and amino acids side chains from regions surrounding the central β -sheet motif and sequences flanking the RRM domain (Allain, *et al.* 1997, Deo, *et al.* 1999, Ito, *et al.* 1999, Price, *et al.* 1998, Xu, *et al.* 1997, Shamoo, *et al.* 1997, Maris *et al.* 2005).

Human La (hLa) is a 47 kDa monomeric modular protein of 408 amino acids residues and contains two main regions: a N-terminal domain (NTD), sequence of approximately 200 amino acids residues highly conserved across species, including yeast, and a C-terminal domain (CTD), which varies greatly in length between different eukaryotes (Figure 1).

The structural details of RNA recognition remain uncertain, but it is known that the most important determinants for 3' poly(U) RNA binding are located in the NTD (Wolin *et al.* 2002, Maraia *et al.* 2001, Jacks *et al.* 2003, Alfano *et al.* 2004).

In Dr Conte laboratories, where the La protein's part of this PhD project has been done, relevant investigations on human La protein and its interactions with RNA targets have been carried out. To study the RNA binding properties of the full-length La and its deletion mutants construct, gel mobility shift assays with a 10 nt oligo(U) RNA have been performed (Jacks *et al.* 2003). The results of these experiments are shown in Figure 2.

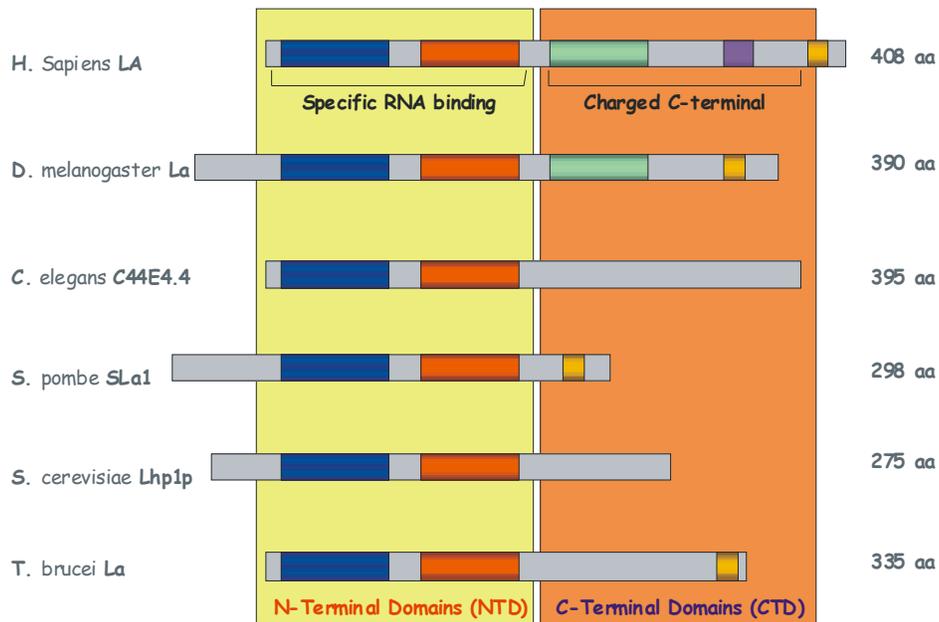


Figure 1: Domains organization of La proteins in various species. For each protein, the length in amino acids is indicated on the right. Conserved regions across different species are reported in the same colour (for more details see Figures 3 and 5). Each colour identified a different functional region: (i) La motif in blue; (ii) Central RRM in orange; (iii) C-terminal RRM in green; (iv) Nuclear localization signal (NLS) in yellow. Domains are labelled.

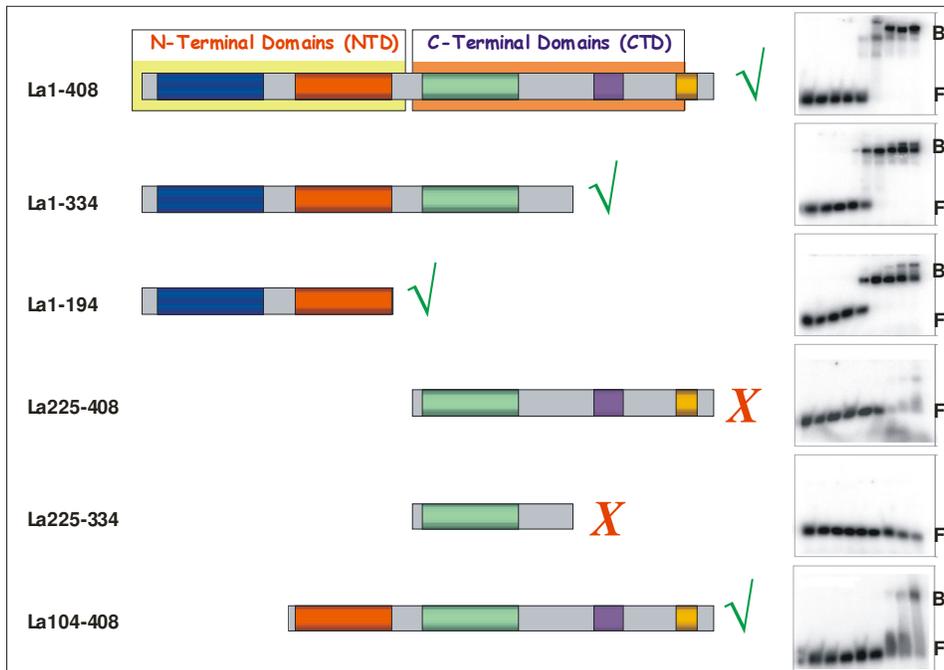


Figure 2: Deletions Mutants of hLa. Gels Shift assay shown that different constructs of La have different properties in binding poly(U) RNA (B=Bound protein, F=Free protein) (Jacks *et al.* 2003). ✓ indicate the constructs that preserve the binding affinity and specificity for poly(U), ✗ indicate the ones that lack poly(U) binding ability.

From the picture below is evident that the first two domains are essential for the binding specificity and that the truncation mutant La104–408 reduces the binding activity to a weak, non-specific interaction (Kenan, *et al.* 1995, Goodier, *et al.* 1997).

The C-terminal domain

The C-terminal domain (CTD) of La proteins varies greatly in length between different eukaryotes and its role is not fully understood (Figure 1); it is involved in nuclear retention and modulation of 5' pre-tRNA processing, but it also appears to have a distinct role in RNA recognition of large structured cellular and viral mRNA (Wolin *et al.* 2002, Maraia *et al.* 2001, Fan *et al.* 1998, Ali *et al.* 2000, Pudi *et al.* 2003, Intine *et al.* 2000 and 2003, Horke *et al.* 2000, Crosio *et al.* 2000).

In hLa protein the C-terminal domain contains an RRM domain (called the C-terminal RRM), encompassing residues 229-326, followed by a long, flexible polypeptide that contains a short basic motif (SBM) (Goodier, *et al.* 1997), a regulatory phosphorylation site on Ser366 and a nuclear localisation signal (NLS) (Marchetti, *et al.* 2000, Simons, *et al.* 1996, Rosenblum, *et al.* 1998) (Figures 1 and 3).

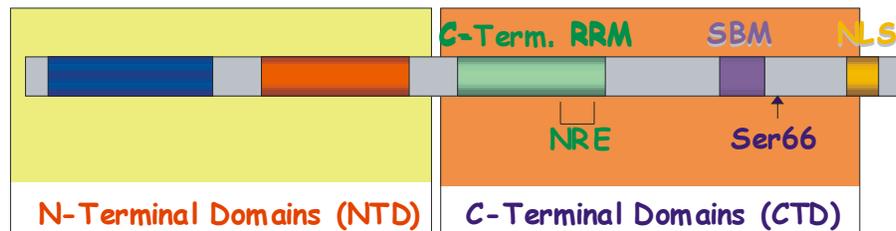


Figure 3: The C-terminal domain of hLa protein: the position of the short basic motif (SBM), nuclear localization sequences (NLS), residue Ser366 and of the Nuclear Retention Element (NRE) (see above for details) are indicated.

The Conte laboratories have provided the first structural characterisation of the C-terminal domain of La protein and provided fresh insights into its structure, RNA binding activity and oligomerisation state (Jacks *et al.* 2003). Their structural work has revealed that the C-terminal RRM of hLa protein is an unusual RRM. The novelty derives from the fact that it possesses a β -sheet comprising five rather than four strands and contains a long C-terminal helix that binds across the putative RNA

binding site, potentially impeding interactions with nucleic acids (Figure 4). This helix also incorporates a functional nuclear retention element (NRE) (Figure 3), which helps to ensure appropriate localization and processing of pre-tRNA molecules (Intine *et al.* 2002, Jacks *et al.* 2003). The C-terminal RRM is not required for specific, high-affinity interactions with poly(U) RNA, although it remains possible that this RRM, in conjunction with the SBM downstream, contributes to La interactions with non-poly(U) RNA targets such as viral RNAs, TOP mRNAs and 5'-triphosphate groups (Wolin *et al.* 2002, Maraia *et al.* 2001, Fan *et al.* 1998, Ali *et al.* 2000, Pudi *et al.* 2003, Intine *et al.* 2000 and 2003, Horke *et al.* 2000, Crosio *et al.* 2000).

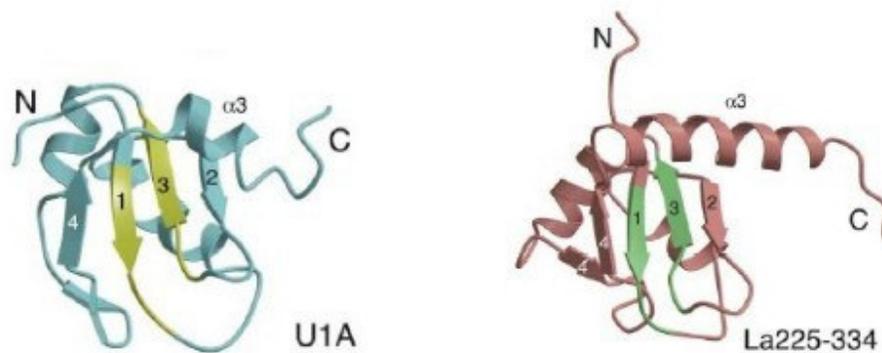


Figure 4: RRM structures comparison: Structural comparison of the RRM from La225–334 (Jacks *et al.* 2003) with U1A (Avis *et al.* 1996). The beta stands are numbered and the alpha helix 3 is labelled

The N-terminal domain

Specific recognition of UUU_{OH} sequences is a function of the N-terminal domain (NTD) of the La protein, the most conserved portion of the protein, which is comprised of two sub-domains, the La motif and an RNA recognition motif (called central-RRM) (Figure 5).

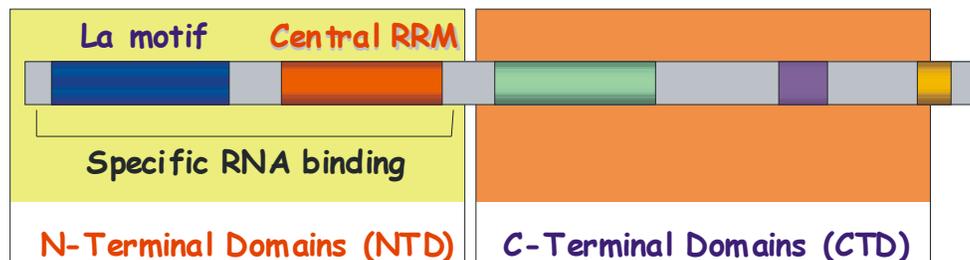


Figure 5: Sub-domain organisation in the NTD of hLa.

Both the La motif and the adjacent RRM appear to be required for high-affinity RNA binding (Kenan 1995, Chang *et al.* 1994, Goodier *et al.* 1997), and it has been suggested that the particular role of the La motif is to provide specific recognition for UUU_{OH} sequences (Wolin *et al.* 2002, Maraia *et al.* 2001, Kenan 1995). Interestingly, the coupling of a La motif to an RRM is a feature observed in a subset of La motif-containing proteins that are otherwise unrelated to La (Wolin *et al.* 2002).

The structure of the La motif, a domain of around 80 residues, has been the subject of some debate: although predicted by some to fold as a canonical RRM, others contend that it adopts a predominantly helical configuration (Wolin *et al.* 2002, Maraia *et al.* 2001, Kenan 1995, Ohndorf *et al.* 2001). Recently the tertiary structure of La motif has been solved simultaneously by NMR, in the Dr Conte laboratories (Alfano, *et al.* 2004), and by X-Ray (Dong, *et al.* 2004). Both structures reveal that the La motif is not a common RRM domain. In fact the structure comprises six α -helices and a three-stranded anti-parallel β -sheet with a topology $\alpha 1-\alpha 1'-\alpha 2-\beta 1-\alpha 3-\alpha 4-\alpha 5-\beta 2-\beta 3$ (Figure 6).

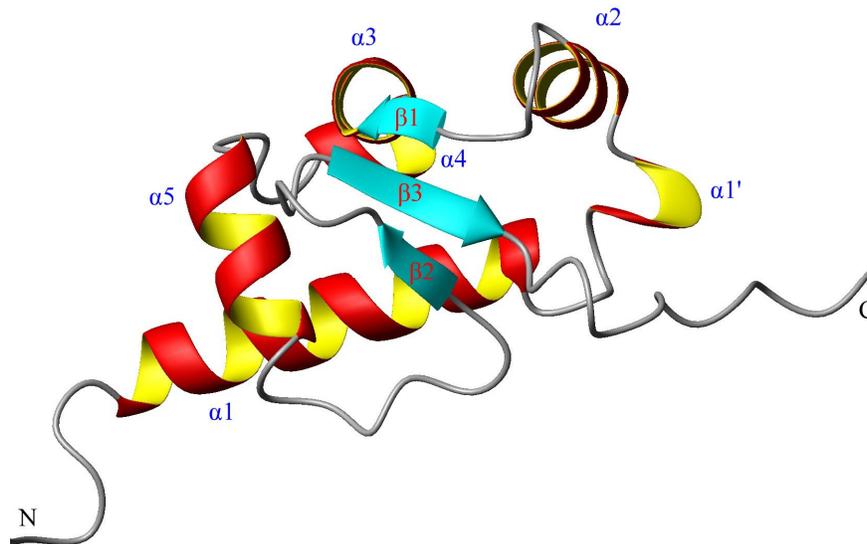


Figure 6: A representative structure of La motif (Alfano *et al.* 2004, Dong *et al.* 2004). The secondary structure elements are indicated.

As a research on databank has shown (Alfano, *et al.* 2004), the La motif is an elaborated version of a winged-helix protein, a small module found in many DNA-binding proteins that folds with $\alpha\beta\alpha\alpha\beta\beta$ topology to give a three-strand β -sheet packed against three helices (Gajiwala *et al.* 2000). Intriguingly members of this family exhibit variable modes of interaction with DNA. Many winged-helix proteins bind in the major groove of B-DNA largely via contacts with the so-called “recognition” helix α_3 (corresponding to α_5 in the La motif), with wings 1 and 2 making variable contributions to DNA recognition (Gajiwala *et al.* 2000, Clark *et al.* 1993). Moreover, in other cases the recognition helix plays only a minor role in DNA binding, while the wings or other surface features dominate the interactions with nucleic acid (Gajiwala *et al.* 2000, Okuda *et al.* 2000). The La motif represents an interesting new structural variant on this theme, being a winged-helix protein that binds RNA.

1.1.3 Our aims on the human La project

The main aim of this research work is to understand which are the features that give La the ability to bind various targets like the 3' poly(U) of polIII transcripts as well as viral RNA which do not contain poly(U). To understand the mechanisms of action of this protein it is necessary to obtain high-resolution structures for La/RNA complexes. As a general approach in structural biology, and in NMR in particular, it helps to define the boundaries of the system under investigation and also in some cases to reduce the size. Often it is easier to solve the structure of individual domains (specially by NMR) and then trying to assemble the domains together to understand the structure of larger domains and/or complexes. Since almost all the determinants of hLa for RNA recognition and high affinity binding appear to be located in its NTD, I focused my studies on this region. In particular, for the La part of my PhD project, I solved the solution structure of the Central-RRM by NMR and mapped the location and extent of the poly(U) binding site on the surface of the NTD of hLa by chemical shift mapping experiments. Furthermore, to examine the RNA binding properties of La, I performed gel mobility shift assays with a 10 nt oligo(U) RNA and a 27mer Hepatitis C RNA and several deletion mutants construct of hLa (Alfano *et al.* 2004).

1.2 DNA binding proteins

An important class of proteins in living cells are those that control the “read-out” of genetic information from the DNA. They do this by recognising and then binding to particular precise sequences of the bases in DNA that are located at points close to the start of genes. The majority of this specific DNA binding domains recognise sequence in the major groove, frequently using an α -helix, and do not give rise to large distortions of the DNA (figure 7A, Billeter *et al.* 1993). However, there are several contrasting examples in which protein binding results in a substantial DNA bend (Figure 7B, Murphy *et al.* 2001): TATA box binding protein (TBP), the HMG box, integration host factor (IHF) and the purine repressor (PurR) are the best known examples (Werner *et al.* 1996). Common to all five is the binding of protein elements to the minor groove with intercalation of protein sidechains, although the protein fold and the sidechains used vary widely. In each case, however, a substantial protein/DNA interface area is achieved by unwinding the DNA somewhat and bending it towards the major groove, generating a broad and open minor groove surface which is fairly hydrophobic.

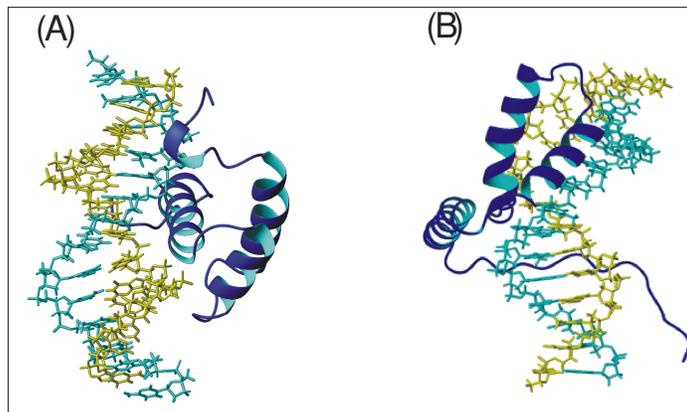


Figure 7: (A) Antennapedia homeodomain complex (Billeter *et al.* 1993): binding in the major groove does not rise to distortion of the DNA. (B) Sry complex (Murphy *et al.* 2001): When a binding on the minor groove occurs, a big distortion of the DNA towards the major groove is present.

The energetics of the interaction on the minor groove are expected to be different from the cases in which α -helices are inserted into the major groove with little DNA distortion. Although much effort has been directed to the determination of the thermodynamic parameters (enthalpy and entropy) controlling the free energy of

protein binding into the major and minor groove of DNA, the understanding of these macroscopic quantities in microscopic terms requires further study.

1.2.1 Calorimetric evidences of protein/DNA interactions

A point of particular interest is that for minor groove DBD/DNA interactions (typically at AT-rich sequences and where DNA bending occurs), calorimetric studies have shown that at lower temperatures the binding is entropy driven, i.e. there is a positive entropic contribution and similar observations have been made for drugs binding to the minor groove (Marky *et al.* 1987, Haq *et al.* 1997, Privalov *et al.* 1999, Muller *et al.* 2001) (Figure 8A). Furthermore, recent calorimetric measurements on the binding of a 10-residues peptide (the so-called AT-hook from HMA1) to the minor groove of DNA show that even at 40° the association is endothermic, i.e. is entropically driven (Dragan *et al.* 2003). In contrast, major groove binders are enthalpy driven (exothermic) and the binding entropy is more negative (Takeda *et al.* 1992, Merabet *et al.* 1995, Lundback *et al.* 1996, Dragan *et al.* 2006) (Figure 8B).

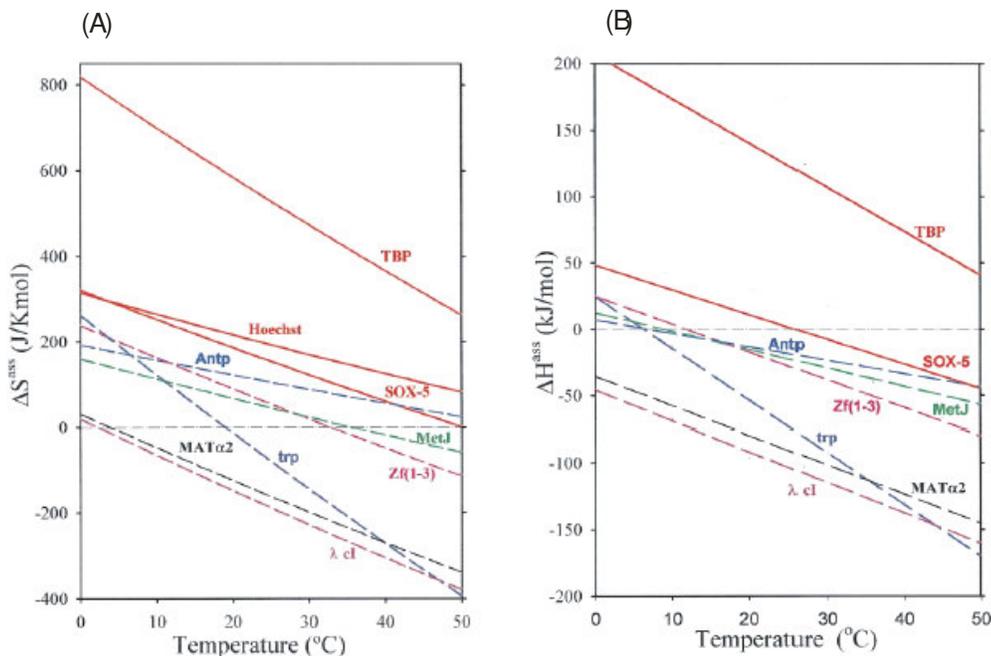


Figure 8: Temperature dependence of the entropy (A) and enthalpy (B) of association of DNA-binding proteins and domains with the corresponding target duplexes (Privalov *et al.* 1999). In both (A) and (B) thick red lines are used for minor groove binders and broken lines for major groove binders

There are not substantial differences between major and minor groove binding as regards the change in accessible surface area on complex formation, i.e. the area that becomes dehydrated on complex formation, and although the minor groove/protein interface is generally more hydrophobic, with less H-bond, than the major groove interface, measurements of the reduction in heat capacity (ΔC_p) on complex formation demonstrate effective dehydration of the interfaces in each case. It is important to recall that in the critical temperature range of 25-37 °C, the positive entropy of dehydration of apolar and polar surface is expected to be of similar magnitude (from measurement on proteins, Makhatadze *et al.* 1995), so the more hydrophobic interface of minor groove interactions does not immediately provide an explanation for their more positive entropy of association. Neither can the observed differences be explained by reductions in translational/rotational freedom that occurs on interaction (which are anyway very small, Tamura *et al.* 1997, Zidek *et al.* 1999). Thus a microscopic explanation is needed for the differences in the thermodynamics of the two categories, in particular for the large positive entropy associated with the formation of minor groove complexes containing highly distorted DNA.

Two quite different explanations have been proposed: the first is that there is a usually high degree of mobility in minor groove complexes (Jen-Jacobsen *et al.* 2000). This might be in the internal modes of either the protein or the DNA components, or at the interface. Although one might expect DNA to become stiffer on binding of a DBD, no NMR studies have been published to date documenting changes in mobility within DNA on binding to ligands or proteins (largely due to the difficulties of isotopically labelling DNA).

An alternative explanation, which is supported by Prof Crane-Robinson laboratories (where this part of the PhD project has been done), is that since the minor groove of AT-rich DNA is characterised by the presence of an extensive and regular zigzag spine of first and second-shell hydration along the floor of the groove (Drew *et al.* 1981, Liepinsh *et al.* 1992, Johannesson *et al.* 1998), when this is shed to the surrounding bulk water a substantial entropy increase would be generated. This could be one of the reasons why minor groove binding to AT-rich sequences has evolved as the best way to counteract the enthalpy input required to generate considerable DNA distortion with the minimum of protein.

Knowledge of the enthalpies and entropies of the DNA/protein interactions provides only a macroscopic picture of the driving forces that control the recognition process, but what is happening at the atomic level? The real essence of protein/DNA binding specificity and selectivity is incompletely revealed by measuring the thermodynamic parameters: a fully understanding of the problem requires definition of the microscopic bases of this thermodynamic quantities. This is the final object of the proposed project.

1.2.2 The systems chosen for study

In order to obtain explanations at the microscopic level for the observed macroscopic energetics of the interactions between sequence specific DNA binding domains (DBDs) and their DNA recognition sites, this project will compare a major groove complex (Antennapedia homeodomain) and a minor groove complex (HMG box from m-Sox5). In the Prof Crane-Robinson laboratories (where this part of my PhD thesis has been done) the free solution structure of Sox-5 has been already determined (Cary *et al.* 2001), and its structure in complex is in progress. An NMR-derived solution structure has been obtained for the drosophila Antennapedia homeodomain complexed with 14 bp DNA and this has been used in the present study as start point (Qian *et al.* 1994, Qian *et al.* 1993, Billeter *et al.* 1993). Furthermore, the thermodynamics of the interaction of this two systems have already been studied in considerable detail (Privalov *et al.* 1999, Crane-Robinson *et al.* 1998, Jelesarov *et al.* 1999, Takeda *et al.* 1992, Merabet *et al.* 1995, Lundback *et al.* 1996, Dragan *et al.* 2006).

To obtain the microscopic information on these systems, I used NMR. Using isotopic labelling techniques I examined differences in backbone mobility between the free and complexed components, in particular backbone of the two proteins.

Antennapedia homeodomain

The homeobox is a highly conserved DNA segment that was first found in homeotic genes of *Drosophila* and subsequently in many other eukaryotic organisms, including vertebrates (McGinnis *et al.* 1984a, McGinnis *et al.* 1984b, Scott *et al.* 1984,

Carrasco *et al.* 1984, Levine *et al.* 1984). The homeobox encodes the homeodomain, which represents a conserved protein domain that is shared by the various homeotic proteins. The homeodomain encoded by the Antennapedia gene of *Drosophila* consists of about 60 residues located close to the C-terminus of the Antennapedia protein. Its structure has been solved by Wuthrich group in both free and complexed state (Qian *et al.* 1994, Qian *et al.* 1993, Billeter *et al.* 1993). As many transcription factors which bind DNA, Antennapedia homeodomain is an helix-turn-helix domain that binds its specific sequence on the DNA by intercalation of an helix into the major groove (Figure 9).

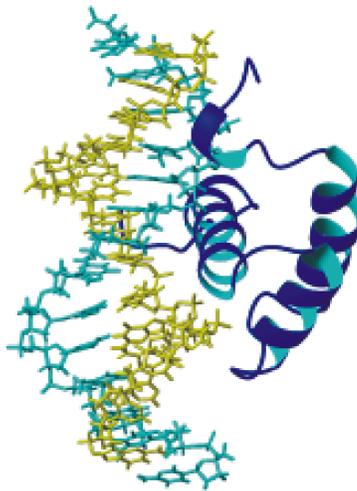


Figure 9: NMR Structure of Antennapedia homeodomain in complex with a 14mer DNA (Billeter *et al.* 1993).

HMG from mouse Sox-5

In mammalian embryogenesis the male determining factor is the SRY protein encoded on the Y chromosome, which switches development from the default (female) pathway to male development (Sinclair *et al.* 1990). The characteristic feature of the SRY protein is the presence of a sequence-specific DNA-binding High Mobility Group box (HMG box), and a number of other proteins have been described containing HMG boxes with sequences closely homologous to that in SRY (called Sox proteins, where **Sox** stands for **SRY-related HMG box**). Alternative splicing of the mouse Sox-5 (mSox-5)

gene product results in a short form of the protein, found in the nuclei of postmeiotic spermatids (Denny *et al.* 1992; Connor *et al.* 1994) and a long form (L-mSox-5) expressed in chondrocytes (Lefebvre *et al.* 1998). Both forms of the protein contain the single HMG box. This HMG box recognizes the 6-bp sequence AACAAAT (Denny *et al.* 1992) with an association constant K_a of $6.2 \times 10^7 \text{ M}^{-1}$ (Privalov *et al.* 1999). The fold of the murine Sox-5 (mSox-5) HMG box in free solution has been determined by multidimensional NMR (Cary *et al.* 2000) and has been found to adopt the characteristic twisted L-shape made up of two wings: the major wing comprising helix 1 (F10–F25) and helix 2 (N32–A43), the minor wing comprising helix 3 (P51–Y67) in weak antiparallel association with the N-terminal extended segment (Figure 10).

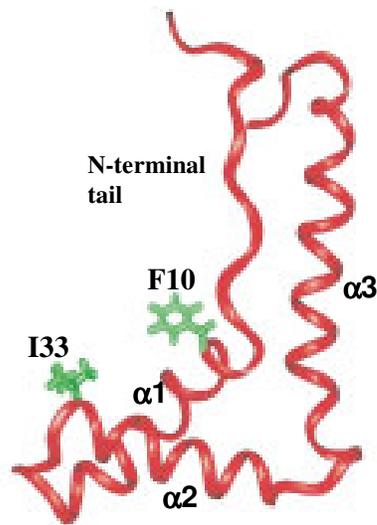


Figure 10: Characteristic L-shape motif present in HMG box. In particular in the figure it is reported HMG box of Sox-5 protein (Cary *et al.* 2000). Helices 1 and 2 form the minor wing, instead the helix 3 and the N-terminal tail form the major wing. The binding of protein elements to the minor groove occurs with intercalation of protein hydrophobic side-chains (in green).

CHAPTER II

MATERIALS AND METHODS

2.1 DNA synthesis

2.1.1 PCR-ESRA Synthesis of 12-mer Sox-DNA

The Sox double strand DNA $5' - \text{CCAAACAATAGG} - 3'$
 $3' - \text{GGTTTGTATCC} - 5'$, both unlabelled and $^{13}\text{C}/^{15}\text{N}$ labelled, was synthesised by the PCR-ESRA method, using *HaeIII* digestion, followed by purification using chromatography techniques.

60ng of the single-stranded oligonucleotide $5' - \text{CCAAACAATAGGCCAAACAATAGG} - 3'$ (termed Sox24G) (INVITROGEN) and 60ng of its complementary strand (termed Sox24H) were used in an amplification reaction using the Expand High Fidelity PCR kit (Roche-Diagnostics) in the presence of 0.2 mM dNTPs (unlabelled or $^{13}\text{C}/^{15}\text{N}$ labelled) and 1.5 mM MgCl_2 , and cycled for 8 rounds at 94°C for 1 min, 56°C for 2 min, and 72°C for 6 min in a GeneAmp 9600 cycler (Perkin Elmer). 600 μl of this amplified mixture were used as template for a further amplification of 9.6 ml of reaction mixture cycled 40 times under the same conditions in 100 μl aliquots. Amplified aliquots were pooled and warmed to 80°C for 5 min and then cooled on ice for 3 hrs. The mixture was then dialysed overnight at 4°C against 50 mM NaCl, 10 mM Tris-HCl pH 7.25, 10 mM MgCl_2 , followed by digestion with *HaeIII* restriction enzyme (New England Biolabs) at 37°C for 16 hrs. The double-stranded product obtained was purified on a MONO Q column (Pharmacia) using a linear NaCl gradient from 30% to 75% of Buffer B (Buffer A: 10mM Sodium Phosphate pH 6.8 and 0.2 mM EDTA; Buffer B: 10 mM Sodium Phosphate pH 6.8, 0.2 mM EDTA and 1 M NaCl). Peak fractions were dialyzed extensively against 10 mM potassium phosphate pH 6.0, 100 mM KCl and 0.1 mM EDTA, then concentrated using VIVASPIN tubes in order to prepare the

NMR sample. All PCR conditions were optimised on a small scale (100µl reaction mixture) in order to improve the yield.

2.1.2 PCR-ESRA Synthesis of 16-mer Antp-DNA

The Antennapedia double stranded DNA $5' - \text{ACAAAGCCATTAGAGT} - 3'$ / $3' - \text{TGTTTCGGTAATCTCA} - 5'$, both unlabelled and $^{13}\text{C}/^{15}\text{N}$ labelled, was also synthesised by PCR-ESRA method, using *RsaI* digestion (New England Biolabs), followed by purification using chromatography techniques.

The two single-stranded oligonucleotides (respectively termed Antp32A and Antp32B) used as templates were:

$5' - \text{ACAAAGCCATTAGAGTACAAAGCCATTAGAGT} - 3'$ **Antp32A**

$5' - \text{ACTCTAATGGCTTTGTACTCTAATGGCTTTGT} - 3'$ **Antp32B**

The optimised annealing temperature was 50°C. All steps were performed as described for Sox double stranded DNA in section 2.1.1. With respect to Sox DNA synthesis, the dialysis step the dialysis step, before restriction enzyme digestion, was not carried out.

2.2 Protein preparation

2.2.1 Antennapedia C39S homeodomain

The plasmid encoding the Antennapedia C39S homeodomain from *Drosophila* was provided by the Privalov laboratory (Johns Hopkins University). The protein was over-expressed in *Escherichia Coli* strain BL21(DE3)-pLysS grown on either rich media or minimal media supplemented with 0.8 g/l [^{15}N]ammonium chloride and 2 g/l [^{13}C]glucose. Cell pellets were resuspended in 100 mM Tris-HCl pH 8.0, 200 mM

KCl, 1 mM EDTA, 2 mM CaCl₂, 10 mM MgCl₂, 2 mM DTT and 1 mM PMSF, then disrupted by sonication.

The cleared supernatant was depleted of nucleic acids by polyethylene-imine precipitation (final concentration 0.5%). The white flocculent precipitate was removed by centrifugation at 12000 rpm for 40 min at 4°C. The Antennapedia C39S homeodomain was precipitated by addition of ammonium sulphate (final concentration 85% w/v saturation) at 4°C. The pellet was resuspended in 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 mM DTT and 100 mM NaCl, then dialyzed extensively against the same buffer.

The protein was purified in two steps of ion exchange chromatography. First a DEAE step using a linear NaCl gradient (Buffer A: 50 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 mM DTT and 100 mM NaCl; Buffer B as buffer A plus 1 M NaCl). Secondly, a SP step again using a linear NaCl gradient (Buffer A: 20 mM sodium phosphate pH 6.0, 0.1 mM EDTA, 1 mM DTT and 100 mM NaCl; Buffer B as buffer A plus 1 M NaCl). Peak fractions were dialyzed extensively against 10 mM Potassium Phosphate pH 6.0, 100 mM KCl and 0.1 mM EDTA.

To prepare the NMR samples the dialyzed solutions were concentrated to 1 mM using VIVASPIN tube cut-off 3000 MW and 10% of D₂O was added into each sample.

¹³C/¹⁵N Antennapedia/unlabelled DNA complex

0.546 μmol of ¹³C/¹⁵N-labelled Antennapedia protein were mixed with 0.63 μmol of unlabelled 16-mer Antp-DNA (DNA 10% in excess) synthesised by PCR-ESRA method. Both protein and DNA were in Potassium Phosphate pH 6.0, 100 mM KCl and 0.1 mM EDTA. 525 μl of protein was added in 21 times (aliquots of 25 μl) to a very diluted DNA solution (0.63 μmol in ~5 ml solution) mixing slowly. The solution was incubated 15 minutes at room temperature after each protein addition. At the end the solution was incubated overnight at 4°C.

In order to prepare the NMR sample, the formed complex was concentrated to 0.8 mM using VIVASPIN tube cut-off 3000 MW and 10% of D₂O was added.

2.2.2 *La proteins*

La-fragments 105-202, 1-103 and 105-334:

The fragment 105-202 of human La protein had been previously sub-cloned into the expression vector JM28 (a modified form of the pET-28a vector - Novagen) using BamH1 and Xho1 restriction sites. The La-fragments 1-103 and 105–334 had been sub-cloned into the pET-30 vector using the LIC methodology (Novagen).

The His-tagged recombinant proteins were over-expressed in *Escherichia Coli* strain BL21(DE3)-pLysS grown on either rich media or minimal media supplemented with 0.8 g/l [¹⁵N]ammonium chloride and 2 g/l [¹³C]glucose. Cell pellets were resuspended in 20 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8, 1 mM PMSF, then disrupted by sonication. Proteins were purified by affinity chromatography using a Ni-NTA resin (Qiagen), following the manufacturer's protocol. The eluted proteins were dialysed into 50 mM Tris-HCl, 0.2 mM EDTA, 1.5 mM MgCl₂, 10% (v/v) glycerol pH 7.25 (buffer A), loaded onto a 5 ml Hi-Trap Heparin column (Amersham-Pharmacia Biotech) and eluted with a linear 0–2.0 M KCl gradient in buffer A.

Peak fractions of La105-202 from the heparin column were then dialyzed extensively against 20 mM sodium phosphate pH 6.0, 100 mM KCl, while peak fractions of La105-334 were dialyzed extensively against 20 mM Tris-HCl pH 7.0, 100 mM KCl, 1 mM DTT. To prepare the La105-202 NMR sample, the dialyzed solutions was concentrated to 0.6 mM using VIVASPIN tube cut-off 5000 MW and 10% of D₂O was added.

Peak fractions of La1-103 from the heparin column were dialyzed extensively against 50 mM Tris-HC pH 8.0, 100 mM KCl, 5 mM CaCl₂ and cleavage with factor Xa enzyme (Novagen) to remove the N-terminal His-tag. The sample was then re-loaded onto the Ni-NTA column and the tag-free protein was collected in the flow-through and was again dialyzed extensively against 20 mM Tris-HCl pH 7.0, 100 mM KCl, 1 mM DTT.

La full length and La-fragment 1-194:

Plasmids encoding the full-length human La (La1-408) and the N-terminal domain La1-194 were originally provided by J. Keene and D. Kenan. Both proteins were over-expressed in *Escherichia Coli* strain BL21(DE3)-pLysS grown on rich media. Cell pellets were resuspended in buffer A (50 mM Tris-HCl pH 7.25, 0.2 mM EDTA, 1.5 mM MgCl₂, 10% (v/v) glycerol) plus 1 mM PMSF, then disrupted by sonication. Proteins were purified by a 5 ml Hi-Trap Heparin column (Amersham-Pharmacia Biotech) and eluted with a linear 0–2.0 M KCl gradient in buffer A. The eluted proteins were dialyzed extensively against buffer A and then the proteins were loaded onto a poly(U) Sepharose affinity chromatography column (Amersham-Pharmacia Biotec) and eluted with a linear 0–2.0 M KCl gradient in buffer A. Peak fractions of both La full length and La1-194 were then again dialyzed extensively against 20 mM Tris-HCl pH 7.0, 100 mM KCl, 1 mM DTT.

La-fragment 225-334:

The fragment 225–334 of human La protein was obtained by “natural” proteolysis of La225-408 sub-cloned into the expression vector pQE9 (Qiagen), using BamH1 and HindIII restriction sites. The His-tagged protein was over-expressed in *Escherichia Coli* strain M15 (Qiagen) grown on rich media. Cell pellet was resuspended in 20 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8, 1 mM PMSF, then disrupted by sonication. Protein was purified in a single step by affinity chromatography using a Ni-NTA resin (Qiagen), following the manufacturer’s protocol. The eluted protein was dialyzed extensively against 20 mM Tris-HCl pH 7.0, 100 mM KCl, 1 mM DTT.

2.3 NMR measurements

2.3.1 NMR measurements on Antennapedia project

All NMR spectra of both free and bounded Antennapedia protein were recorded at 293 K on Varian Inova spectrometer operating at 14.1 Tesla, located at the Biophysical Laboratories of the University of Portsmouth.

To obtain sequence specific assignments for the backbone, HNCA, CBCA(CO)NH and HNCO experiments were collected (Bax and Grzesiek, 1993). While ^{15}N -edited NOESY-HSQC and ^{13}C -edited NOESY-HSQC experiments (Fesik and Zuiderweg 1988) were acquired for the side chains resonances assignment.

For relaxation measurements, T1, T2 and $\{^1\text{H}\}$ - ^{15}N NOE experiments were acquired using the pulse sequences adapted from standard schemes (Kay *et al.* 1989).

All spectra were processed using NMRPipe/NMRDraw (Delaglio *et al.*, 1995) and analysed using XEASY (Bartels *et al.*, 1995).

2.3.2 NMR measurements on La protein project

All NMR spectra of La-fragments were recorded at 293 K on Varian Inova spectrometers operating at 14.1 and 18.8 Tesla, located respectively at the Biophysical Laboratories of the University of Portsmouth and at the Biomedical NMR Centre of the National Institute for Medical Research in London.

Resonances assignment and distance restrains of La105-202

For La105-202 fragment, HNCA, HN(CO)CA, CBCANH, CBCA(CO)NH and HNCO experiments were collected to obtain sequence specific assignments for the backbone (Bax and Grzesiek, 1993). For the side chains resonances assignment and for distance restrains, the following experiments were acquired: ^{15}N -edited TOCSY-HSQC and NOESY-HSQC (Fesik and Zuiderweg, 1988), HCCH-TOCSY and ^{13}C -edited NOESY-HSQC. For relaxation measurements, T1, T2 and $\{^1\text{H}\}$ - ^{15}N NOE experiments were acquired using the pulse sequences adapted from standard schemes

(Kay *et al.* 1989). NH residual dipolar couplings were measured in a ternary complex composed of ~5% (v/v) alkyl-poly(ethylene glycol) C8E5, ~0.8% (v/v) n-octanol and 20 mM sodium phosphate, 100 mM KCl, pH 6 (Rückert *et al.* 2000). The final concentration of the proteins in this media was ~ 0.15 mM. Precise measurements of ^1JNH splittings were obtained from ^1JNH ^1JNH modulated 2D spectra (Tjandra *et al.* 1996).

All spectra were processed using NMRPipe/NMRDraw (Delaglio *et al.*, 1995) and analysed using XEASY (Bartels *et al.*, 1995).

Structure calculation of La105-202

The solution structure of La105-202 was calculated using a combined torsion angle and Cartesian coordinates dynamics protocol executed in X-PLOR (Brunger 1992) modified to include dipolar coupling restraints (Clare *et al.* 1998). The structure was calculated from random starting coordinates on the basis of NOE distance restraints, dihedral angle restraints, hydrogen bond distance restraints and NH residual dipolar coupling restraints. Dihedral angle restraints were obtained using the backbone torsion angle prediction package TALOS (Cornilescu *et al.* 1999). For the NH residual dipolar coupling restraints, the initial estimates of the magnitude and rhombicity of the alignment tensor have been obtained using the maximum-likelihood method (Warren *et al.* 2001). The structures were displayed using MolMol (Koradi *et al.* 1996) and analyzed using PROCHECK-NMR (Laskowski *et al.* 1996).

RNA titration of La1-194

RNA titration experiments were done by adding increasing amounts of unlabeled synthetic oligonucleotides ten-nucleotide oligo(U) (5'-UUUUUUUUU-3') and five-nucleotide oligo(U) (5'-UUUUU-3') (Dharmacon) to ^{15}N -labeled La1-194 fragment. ^1H - ^{15}N HSQC spectra were recorded at RNA/protein mole ratios of 0.2:1, 0.5:1, 0.8:1, 1:1, 1.2:1 and 1.5:1 to follow the resonances perturbed by RNA binding to the protein. The titration buffer used was 20 mM Tris-HCl, 100 mM KCl, 1 mM DTT, pH 7.

2.4 Gels shift assays on La protein fragments

The binding of all the deletion mutants of La protein to radiolabeled RNA targets was assessed using gel shift mobility assays. The synthetic RNAs 10 nt oligo(U) (prepared and gel purified by Dhamarcon Research Inc.) and Hepatitis C RNA fragments (prepared and purified by Simon Pennel at the National Institute for Medical Research) were 5'-end-labeled with [γ -³²P]ATP using T4 polynucleotide kinase (NEB). Unincorporated nucleotides were removed on G-25 mini spin columns (Amersham-Pharmacia Biotech). In gel shift assays, protein-RNA binding reactions were incubated for 15 min at room temperature in 10 mM HEPES, 100 mM KCl, 3 mM MgCl₂, 5% glycerol, 1 mM DTT; in some cases 50 μ g/ml of yeast tRNA (Boehringer Mannheim) was added to the reaction mixture as competitor. After addition of 12 μ l of 30% Ficoll (Sigma-Aldrich), the samples were loaded onto a pre-electrophoresed 9% native polyacrylamide gel at 4°C in 0.5X TBE buffer. The gels were run for 1 hr at 125V and dried onto Whatman 3MM chromatography paper.

CHAPTER III

RESULTS AND DISCUSSION: ANTENNAPEDIA HOMEODOMAIN

3.1 The Antennapedia homeodomain in a wider contest

All work done on the Antennapedia homeodomain is a part of a bigger research programme carried out in the Crane-Robinson laboratories. The final aim is to obtain explanations at the microscopic level of the interactions between DNA binding domains (DBDs) and their recognition sites. The forces that drive protein/DNA recognition have been defined macroscopically by making thermodynamic measurements, i.e. obtaining the Gibbs energy from the binding constant and the enthalpy of binding by titration calorimetry, thereby calculating the entropic contribution. This does not however give microscopic explanations, i.e. indicate the detailed molecular origins of the measured quantities. This can best be derived from NMR measurements in solution, since these can define close contacts between defined atoms and indicate changes in molecular motion that can contribute to entropy changes.

The ultimate aim of such studies is to be able to “interfere with” the protein-DNA recognition step to either increase or decrease the rate at which a gene is read. This is crucial since the basic reason for the run-away activity of cancer cells is that key genes either are in overdrive or dampened down.

The people involved in this project have examined and are continuing to examine, differences in atomic mobility between the free and complexed components of the systems described in Chapter I. In particular, our efforts were and continue to be directed to relaxations measurements of the backbone and side chains of the proteins (free and complexed) and, most importantly, of the phosphodiester chains, sugars and bases of the DNAs, a subject little studied to date.

The Antennapedia backbone mobility of both free and complexed protein was investigated here using the well-established ^{15}N approach (Kay *et al.* 1989), by measuring T_1 , T_2 and heteronuclear ^1H [^{15}N] NOEs for the amide nitrogens, as will be explained below. The Sox-5 backbone mobility is similarly under study. The side chain dynamics of both systems will later be studied using the recently developed methodologies employing ^2H relaxation (Muhandiram *et al.* 1995; Mittermaier *et al.* 1999) and ^{13}C relaxation (Wand *et al.* 1996).

Another part of my PhD project concerns the assignment of all the Antennapedia resonances (free and complexed). In fact, the proton assignment of the free protein was already available, as was the complete assignment of the Antennapedia DBD in the complex but under experimental conditions, i.e. pH and temperature, very far from ours (Qian *et al.* 1994, Qian *et al.* 1993, Billeter *et al.* 1993). Therefore, it was necessary to re-assign all the resonances under our experimental conditions in order to compare the relaxation results with the published calorimetric data.

As concerns the DNA targets duplexes, their internal motions (both free and complexed) are being studied at a large number of labelled nuclear sites simultaneously: the purine C2/C8, pyrimidine C5/C6 and sugar C1', C3' and C4' atoms using ^{13}C relaxation, the imino ^{15}N relaxation and also backbone ^{31}P relaxation since the bending of the DNA results in an unusually wide dispersion of ^{31}P chemical shifts (Castagne *et al.* 2000). To obtain both unlabelled and $^{13}\text{C}/^{15}\text{N}$ labelled DNA targets, the PCR-ESRA method of synthesis was used.

3.2 DNA synthesis: the PCR-ESRA method

The aim of the first part of the project was to optimise the PCR-ESRA methodology for both the Sox-5 and Antennapedia DNA target sequences. The **PCR-ESRA** method (**P**olymerisation **C**hain **R**eaction-**E**ndonuclease **S**ensitive **R**estriction **A**mplification, Louis *et al.* 1998), with respect to normal PCR reactions, has the novelty that the primer and template are the same molecule. This is possible because the double stranded DNA used contains two tandem repeats of the target DNA sequence (i.e. the binding site for the protein of interest), separated by a restriction

site (Figure 3.1). In the annealing step this DNA can partially (as well as fully) overlap, allowing the 5' overhangs to act a template for polymerisation of a new sequence by 3' extension. After many reaction cycles long duplexes are formed, containing many target DNA sequences, separated by a restriction site and therefore cleavable with the appropriate endonuclease (Figure 3.2).

Although this methodology does not give an especially high yield, probably due to the instability of the deoxynucleotides triphosphates over extended cycles of heating and cooling, it is useful to prepare labelled duplex DNA for NMR studies. In fact, although chemical synthesis of DNA oligomers is a simple and automated procedure, its use is limited to those laboratories that have the manpower and chemical expertise to synthesise uniformly labelled phosphoramidites.

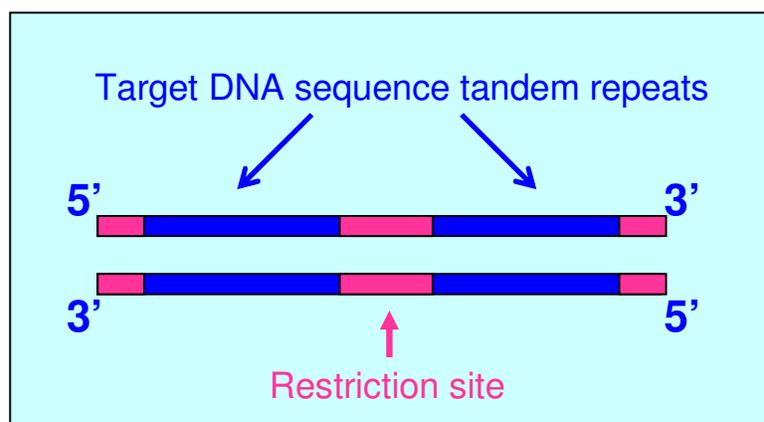


Figure 3.1: Schematic of a typical double stranded DNA used as the initial template for the PCR-ESRA method. The target DNA sequence is in blue, while sequences involved in the restriction site are in red.

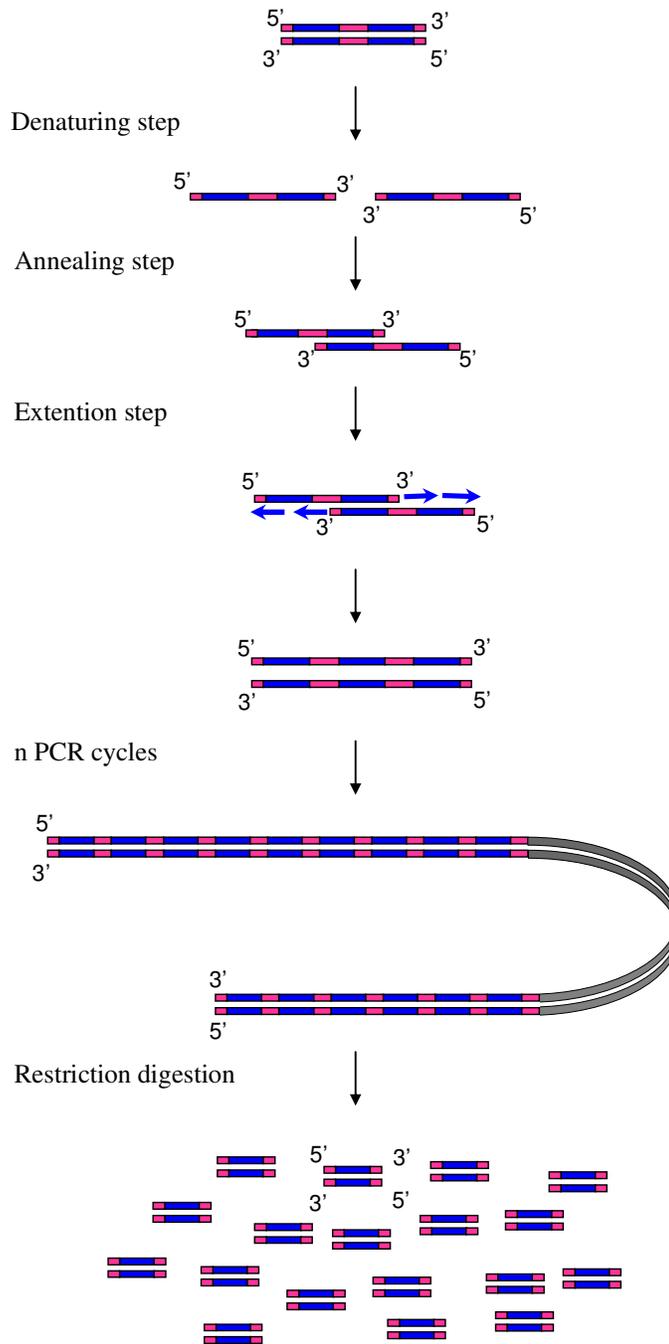


Figure 3.2: Schematic of the PCR-ESRA method. The blocks in blue represent the protein binding site whereas those in red indicate the restriction site of an appropriate blunt-ended restriction endonuclease. After denaturing and re-annealing, the DNA will be partially overlapped and the 5'-overlaps will then be used as template to extend both 3' ends by the DNA polymerase. After many reaction cycles, long duplexes are formed (the block in grey simply indicates a long extension), containing many target DNA sequences, separated by the restriction site and therefore cleavable with the appropriate endonuclease.

3.2.1 Sox-5 DNA synthesis

For the Sox-5 DNA target sequence (AACAAAT), we decided at the beginning to use the following 12 bp sequence:

5' -CCGAACAATCGG-3'

3' -GGCTTGTTAGCC-5'

This sequence was very similar to that used in measuring binding energetics (Privalov *et al.* 1999) and was suitable for the PCR-ESRA methodology because it allowed formation of the HaeIII restriction site GGCC. We then designed the following DNA duplex to use as template in the polymerization reaction, where the Sox-5 binding site is in blue, while residues involved in the Hae III restriction site are shown in red:

5' -CCGAACAATCGGCCGAACAATCGG-3' Sox 24A

3' -GGCTTGTTAGCCGGCTTGTTAGCC-5' Sox 24B

Using this duplex the PCR-ESRA method gave a very poor yield (about 300 μg , which corresponds to 12% of yield, see section 3.2.3 for more details). Analysing the sequence using a computer package which assesses all possible secondary structures and calculates their free energies of stabilisation (ΔG) and melting temperatures T_m (Zuker 2003), we realised that each single strand could form an intra-molecular stable double hairpin and thus prevent the formation of inter-molecular duplexes. The sequence was therefore seen to be unsuitable for the PCR extension reaction (Figure 3.3).

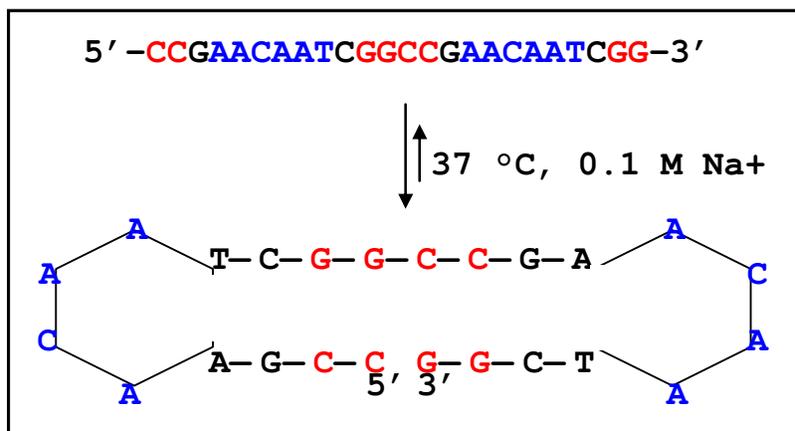


Figure 3.3: Formation of a double hairpin within the Sox 24A strand at 37°C. The T_m calculated for this structure by a computer package (Zuker 2003) is 56.2°C and the ΔG of stabilisation = -3.46 Kcal/mol.

The formation of such structures in the single strands was strongly supported by native polyacrylamide gel electrophoresis (Figure 3.4). Both the Sox 24A and Sox 24B 24nt strands migrated slightly faster than a reference 12bp duplex DNA, suggesting the presence of a compact structure.

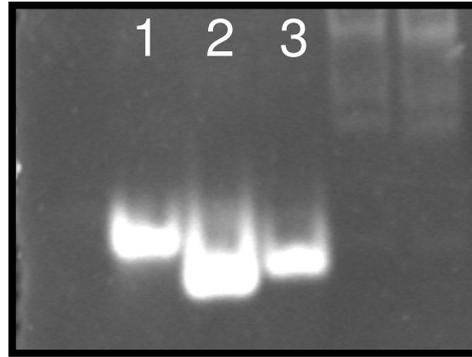


Figure 3.4: Native polyacrylamide gel electrophoresis of Sox DNA strands (16% acrylamide in TBE buffer, stained with ethidium bromide). Lane 1: 12 bp reference duplex DNA; Lane 2: Sox 24A; Lane3: Sox 24B.

The DNA template for PCR-ESRA was then changed to the following, (maintaining the same restriction site), a sequence which is not predicted to form stable hairpins by the Zuker 2003 computer package.

5' -**CCAACAATAGGCCAACAATAGG**-3' Sox 24G

3' -**GGTTTGTTATCCGGTTTGTTATCC**-5' Sox 24H

This was checked by polyacrylamide gel electrophoresis (Figure 3.5) and it was found that the 24nt single strands run significantly slower than the reference, implying a much larger structure than that of the 12 bp duplex (which has, of course, an identical charge):

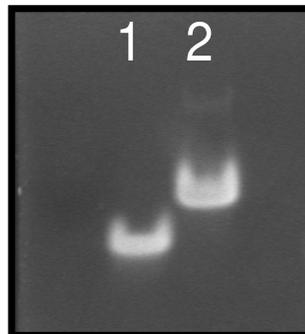


Figure 3.5: Native polyacrylamide gel electrophoresis of a Sox DNA strand (16% acrylamide in TBE buffer, stained with ethidium bromide). Lane 1: 12 bp reference duplex DNA as in Fig 14; Lane 2: Sox 24H.

Once confirmed that both Sox24G and Sox24H strands were suitable for the PCR-ESRA methodology, the optimisation of the PCR amplification process was carried out by varying the following parameters:

1. Annealing temperature: 44, 50, 56 and 60 °C.
2. Number of cycles: 20, 30, 40, 50 and 60 cycles.
3. Concentration of MgCl₂: 1.0, 1.5, 2.0 and 2.5 mM.
4. Amount of dNTPs: 0.05, 0.1, 0.2, 0.4, 0.6, 0.8 and 1.0 mM

All optimization trials were performed on 100 µl of PCR reaction mixture and the PCR products were digested with HaeIII and then quantified by 16% polyacrylamide native gel electrophoresis in TBE buffer stained with ethidium bromide. Once identified, the best temperature of annealing was used for all other optimization trials. As Figure 3.6 shows, the best results were obtained using the following parameters:

1. Annealing temperature: 56 °C.
2. Number of cycles: 40 cycles.
3. Concentration of MgCl₂: 1.5 mM.
4. Amount of dNTPs: 0.2 mM

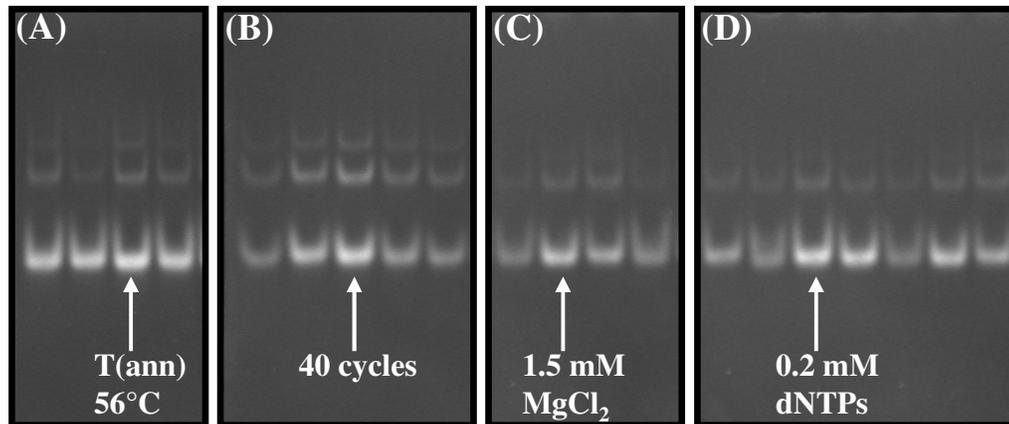


Figure 3.6: Optimization of the PCR-ESRA method for Sox DNA synthesis. The results were analysed on 16% polyacrylamide native gels in TBE buffer stained with ethidium bromide as a function of: (A) The annealing temperature (44, 50, 56 and 60 °C); (B) The number of cycles (20, 30, 40, 50 and 60 cycles); (C) The concentration of MgCl₂ (1, 1.5, 2.0 and 2.5 mM); (D) The amount of dNTPs (0.05, 0.1, 0.2, 0.4, 0.6, 0.8 and 1 mM). Arrows indicate the optimal conditions obtained for each parameter, i.e. those which gave the best yield. The upper bands (with a higher molecular weight) are probably due to incomplete digestion by the Hae III.

Another critical point was found to be the ramping speed of the temperature between sequential steps of the PCR reaction. The best yields were achieved using a Robocycler PCR machine in which the temperature ramping speed is extremely fast. Furthermore, we noticed that when the temperature cycler contained a silver block (with consequently faster temperature changes), the amplification was more efficient than if the block was of aluminium (Figure 3.7).

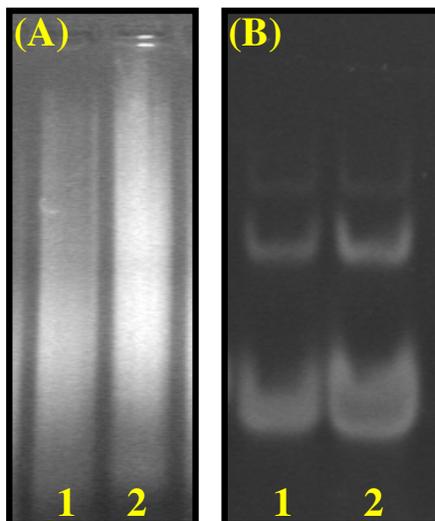


Figure 3.7 (A) 0.8% native agarose gel stained with ethidium bromide of products immediately after the polymerization reaction; (B) 16% native polyacrylamide gel stained with ethidium bromide of products after the restriction digestion. Lanes 1 are products obtained using a temperature cycler with an aluminium block and Lanes 2 are the products obtained using a temperature cycler with a silver block. A greater yield is obtained when the polymerization reaction was performed using a silver block.

After the optimisation, a large-scale PCR reaction was performed using the established conditions from the test experiments, i.e. 1.5 mM MgCl₂ and 0.2 mM dNTPs. As suggested by Louis *et al.*, the large scale PCR-ESRA was performed as two separate steps (Figure 3.8). The first amplification step is necessary to increase the amount of template, which is then distributed to multiple wells for the second amplification step. Because of the increasing length of DNA strands during the several cycles of the second step, it was decided to increase the extension time to 6 min, while it was 3 min during the first step.

After the polymerization reaction, phenol extraction was performed to remove the DNA polymerase and the DNA was then precipitated with ethanol and

resuspended in TE buffer for the HaeIII restriction digestion. Although several digestion trials were conducted to establish the best digestion conditions, i.e. the time of digestion, cleavage with the HaeIII enzyme always yielded a small amount of undigested material, which migrated in the polyacrylamide gel at the position of a dimer (Figures 3.6 and 3.7B). It was always possible however to separate this dimeric form using a linear NaCl gradient with a cation exchange monoQ column, following the large-scale amplification (Figures 3.10 and 3.11).

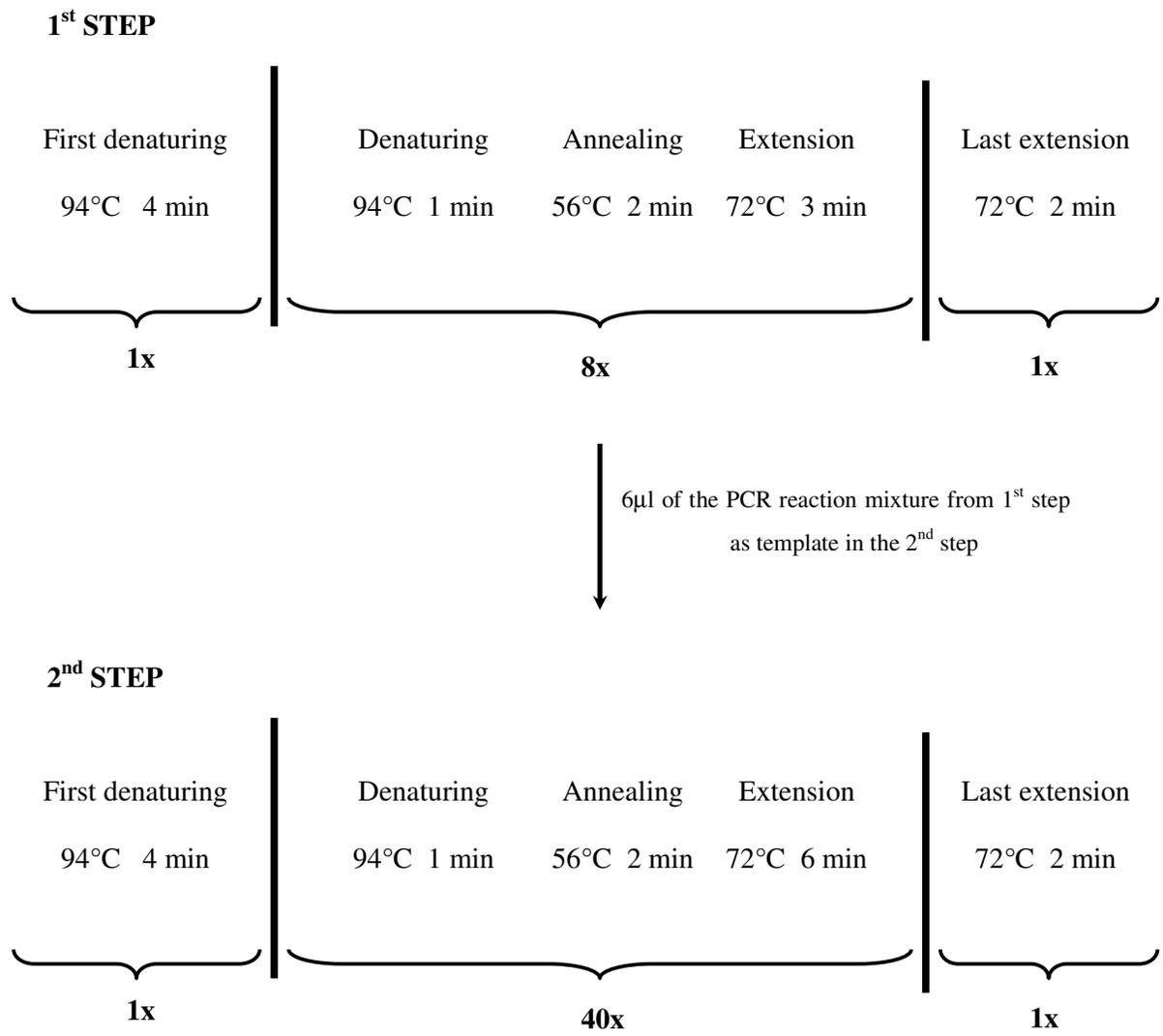


Figure 3.8: Schematic of the protocol followed in the two-step PCR ESRA reaction. The extension time of the second step (6 min) is longer than that used during the first step (3 min) because of the increasing length of the DNA strands during the 2nd step. The solution conditions were those established in the test experiments, i.e. 1.5 mM MgCl₂ and 0.2 mM dNTPs.

The final product yield of the large-scale amplification was however still quite low (about 500 µg from 2.5 mg of dNTPs) and the purification strategy was therefore modified. In particular, the presence of a white pellet at the phenol-water interface was noticed during the phenol extraction step and this could be a precipitated DNA product. It was therefore decided to perform the Hae III digestion without first purifying the DNA. A dialysis step was however included immediately after the PCR reaction since the PCR buffer was not compatible with the Hae III restriction enzyme. In addition, a new annealing step (at 80 °C for 5 min, then on ice for 3 hr) was introduced after amplification but before Hae III digestion with the aim of increasing the base pairing of the DNA strands so that the final product contains very few residual single strand overlaps. Such overlaps would be completely hydrolysed to free dNTPs by the restriction enzyme (Figure 3.9).

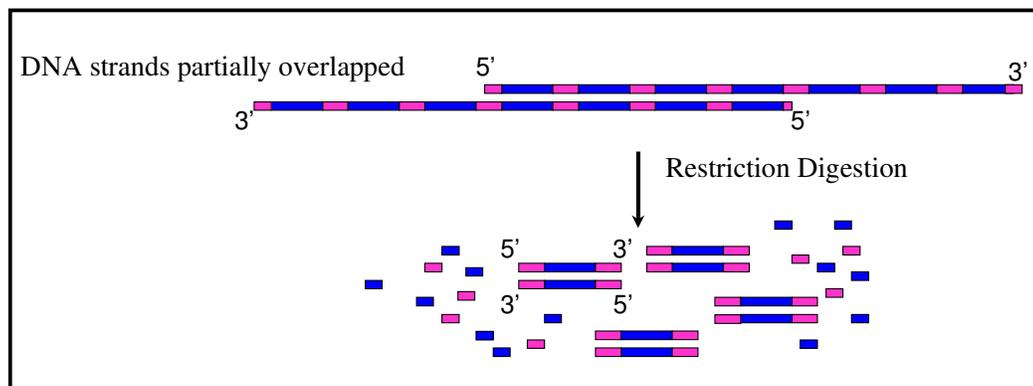


Figure 3.9: Schematic of the restriction digestion performed after the PCR reaction. Residual single strands at both 3' and 5' ends would be completely hydrolysed to free dNTPs by the restriction enzyme, resulting in a decreased yield of duplex target DNA.

This hypothesis was supported by experimental evidence. In fact, a high quantity of free nucleotides was present in the flow through of the cation exchange monoQ column when Hae III digestion was performed without a final annealing step (Figure 3.10 A), while a slow re-annealing of the final PCR products resulted in a consistent decrease of free nucleotides in the flow through (Figure 3.10 B).

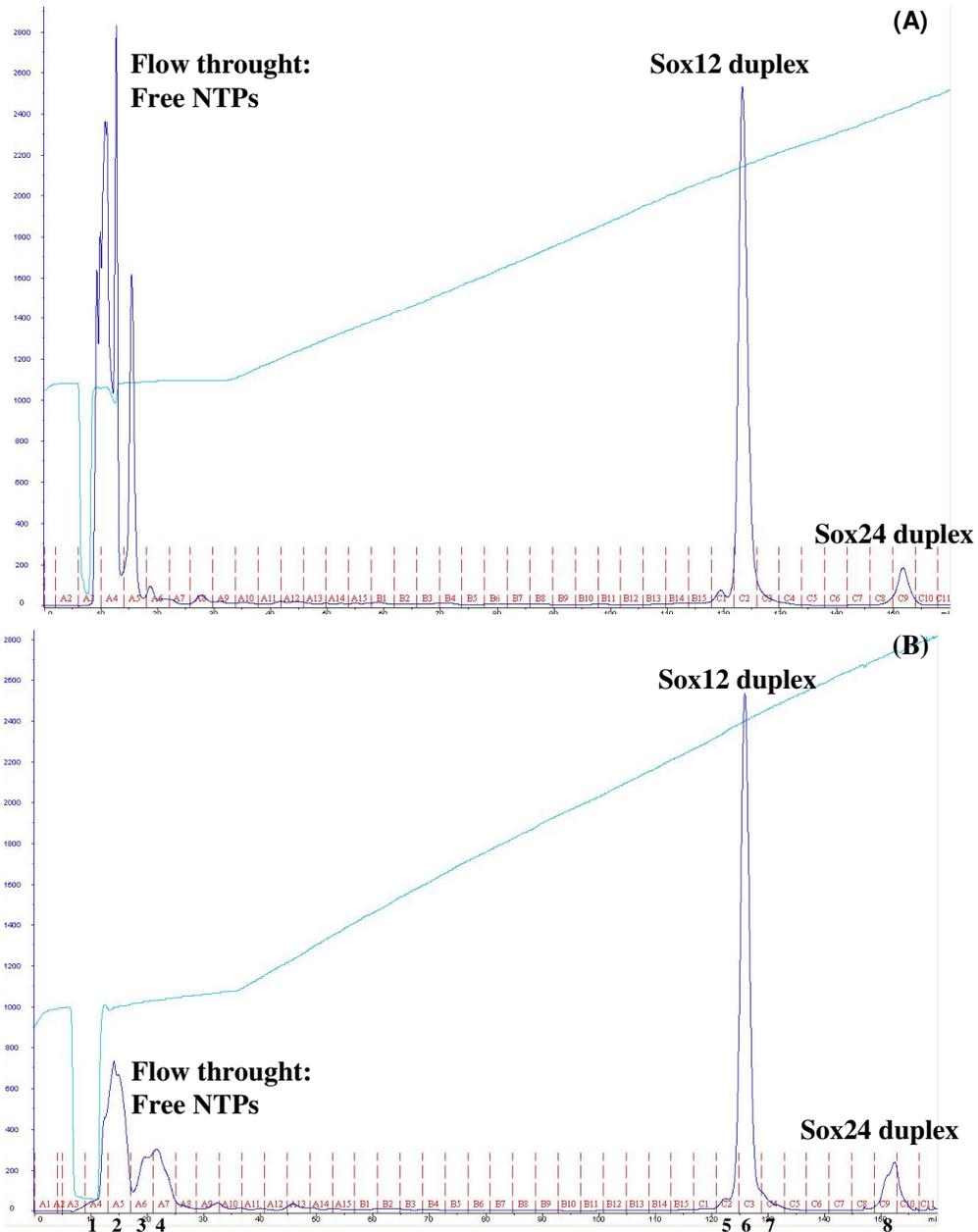


Figure 3.10: FPLC traces from monoQ column purification of the Sox-5 DNA target (Sox12 duplex). (A) The DNA was Hae III digested immediately after the PCR reaction: ~10 ml of sample loaded. (B) A final annealing step was included after the PCR reaction and the DNA then digested with Hae III: ~5 ml of sample loaded. It is clear that there is a decrease in free dNTPs in the flow through in B relative to A. Numbers (1-8) on the bottom of B correspond to the lanes of the polyacrylamide gel in Figure 3.11.

The monoQ peak fractions were checked on a 16% polyacrylamide native gel in TBE buffer stained with ethidium bromide, which confirmed the purity of the synthesised DNA (Figure 3.11).

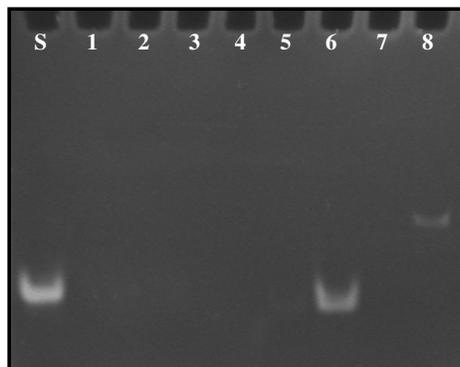


Figure 3.11: Native polyacrylamide gel electrophoresis of Sox DNA (16% acrylamide, stained with ethidium bromide). Lane S: 12 bp reference duplex DNA; Lanes 1 to 8 are labelled in accord with the column fractions of Figure 3.10 B. In particular, Lanes 1-4 correspond to the flow through which contains free nucleotides not visible on the gel. Lanes 5-7 correspond to the bigger peak, which contains the Sox12 duplex DNA. Lanes 8 corresponds to the small peak, which probably contains a dimer of Sox12 duplex as a result of incomplete Hae III digestion.

The modifications described gave rise to a 90% increase in yield, such that ultimately about 900 μg of both unlabelled and $^{15}\text{N}/^{13}\text{C}$ labelled Sox-5 DNA were obtained from 2.5 mg of dNTPs. More details concerning yield and costs are reported in Section 3.2.3.

To check the quality of the DNA samples, mono-dimensional NMR spectra were acquired using a INOVA VARIAN spectrometer operating at 14.1 Tesla located at the Biophysics Laboratories of the University of Portsmouth (Figure 3.12). In a one-dimensional NMR spectrum the signal is recorded as a function of one time-variable. Therefore, after Fourier transformation, a spectrum displaying intensities versus frequencies is obtained. If two or more spins are scalar-coupled, the fine structure of a multiplet will be observed. In the 1D spectrum, for complicated molecules like proteins, it is difficult to establish which spins are coupled. In our case, however, due to the small size of the DNA the 1D spectrum gave the required information.

In Figure 3.12 the mono-dimensional NMR spectrum of unlabelled Sox-5 DNA is reported. Around 1 ppm it is possible to discriminate clearly the signals corresponding to the methyl groups of the thymines. This indicates the good quality of the synthesised DNA sample.

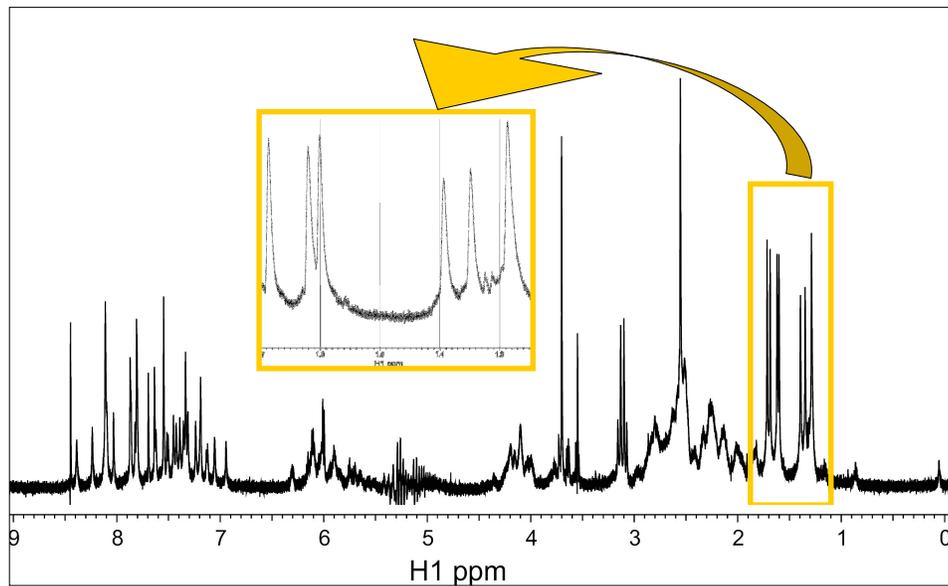


Figure 3.12: ^1H -1D NMR spectrum of the unlabelled Sox5 DNA target sequence. The spectrum was acquired in D_2O using a spectrometer operating at 600 MHz. In the zoomed yellow box is reported the region of the methyl groups of thymines. Their clear discrimination in the spectrum indicates the quality of the synthesised DNA sample.

A two-dimensional homonuclear NOESY NMR spectrum was then acquired for the unlabelled sample in order to perform the sequential assignment of proton resonances (Figure 3.13).

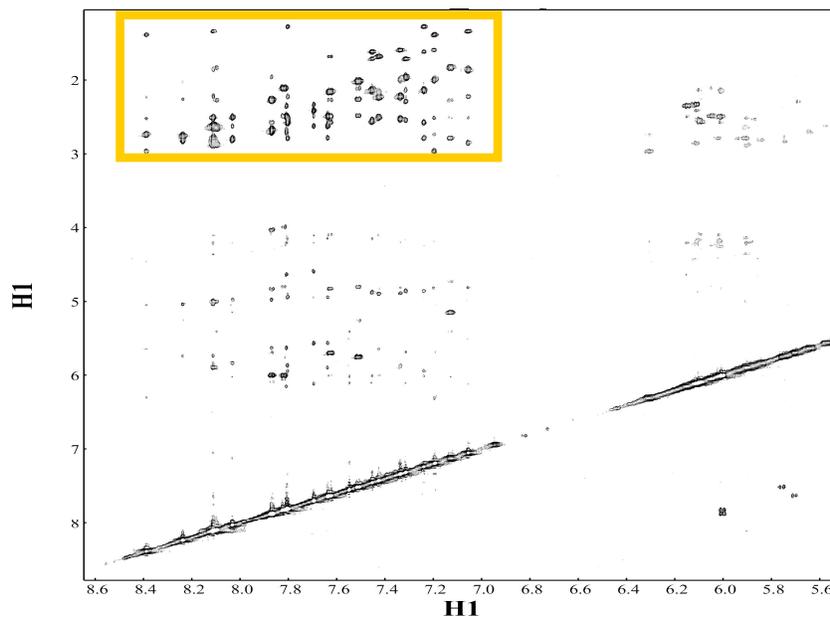


Figure 3.13: The 2D homonuclear NOESY spectrum of the Sox12 duplex. In the yellow box is reported the region of all the sugar $\text{H}2'$, $\text{H}2''$ protons and the CH_3 methyl group of thymines, for which the assignment had been made (see Figure 3.15).

With respect to a mono-dimensional spectrum, in a two-dimensional spectrum the signal is detected as a function of two time-variables t_1 and t_2 . A series of spectra is then recorded where the time t_1 , which corresponds to the delay between two RF pulses, is systematically incremented. Therefore, the recorded signal is a matrix, with axes corresponding to t_1 and t_2 (the acquisition time), on which Fourier transformation has to be performed so that each signal is a function of two frequency variables. Figure 3.14 represents the general idea behind a 2D experiment.



Figure 3.14: Schematic representation for a 2D NMR experiment. In a two-dimensional spectrum the signal is detected as a function of two time-variables t_1 and t_2 . A series of spectra is recorded increasing the time t_1 systematically. The recorded signal is a matrix, with axes corresponding to t_1 and t_2 .

Only the assignment of the sugar H2' and H2'' of the Sox12 duplex is reported here, while the complete assignment of both unlabelled and $^{15}\text{N}/^{13}\text{C}$ -labelled Sox12 duplex was performed by Dr Karim Snoussi, post-doctoral researcher on this project.

Figures 3.15 A and B report the sequential assignment of both strands of the Sox12 duplex:

Strand G: 5' -C₁ C₂ A₃ A₄ A₅ C₆ A₇ A₈ T₉ A₁₀ G₁₁ G₁₂ -3'

Strand H: 3' -G₁₃ G₁₄ T₁₅ T₁₆ T₁₇ G₁₈ T₁₉ T₂₀ A₂₁ T₂₂ C₂₃ C₂₄ -5'

For each nucleotide residue, proton H6 or H8 (H6 for pyrimidines, while H8 for purines) are correlated with their own H2' and H2'' sugar protons. Furthermore, weaker NOEs are observable for the H2' and H2'' sugar protons of the previous nucleotide in the sequence. The methyl groups of thymines give cross-peak with their own H6 proton and with the H6 or H8 proton of the previous nucleotide in the sequence. To distinguish H2' and H2'' signals, it had been considered that H2' gives stronger NOE signals with respect H2''. In fact, the H2' protons are closer to their own H6 or H8 protons than the H2''.

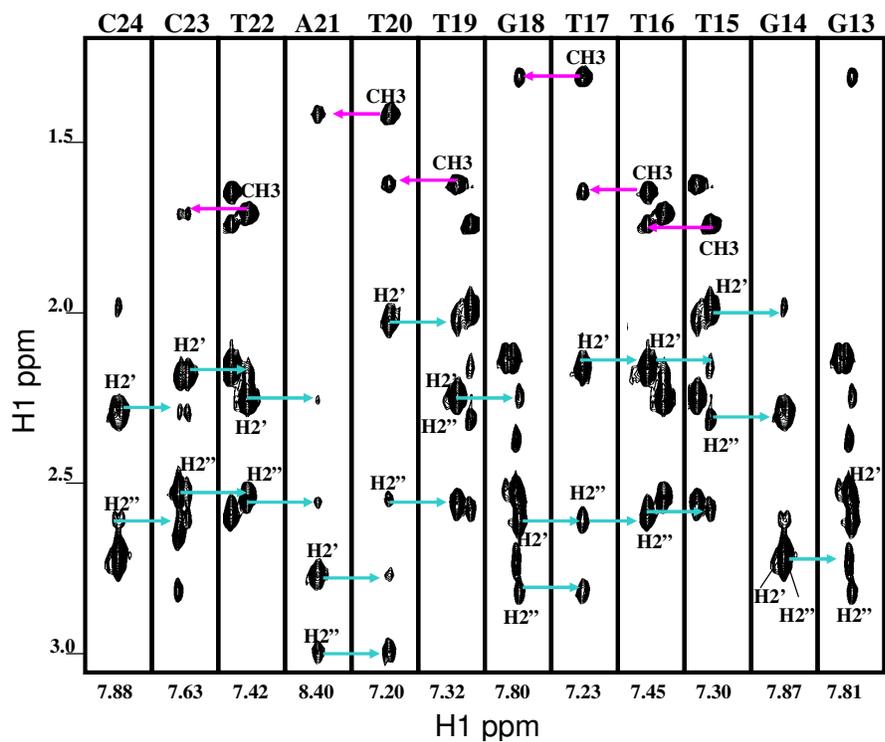
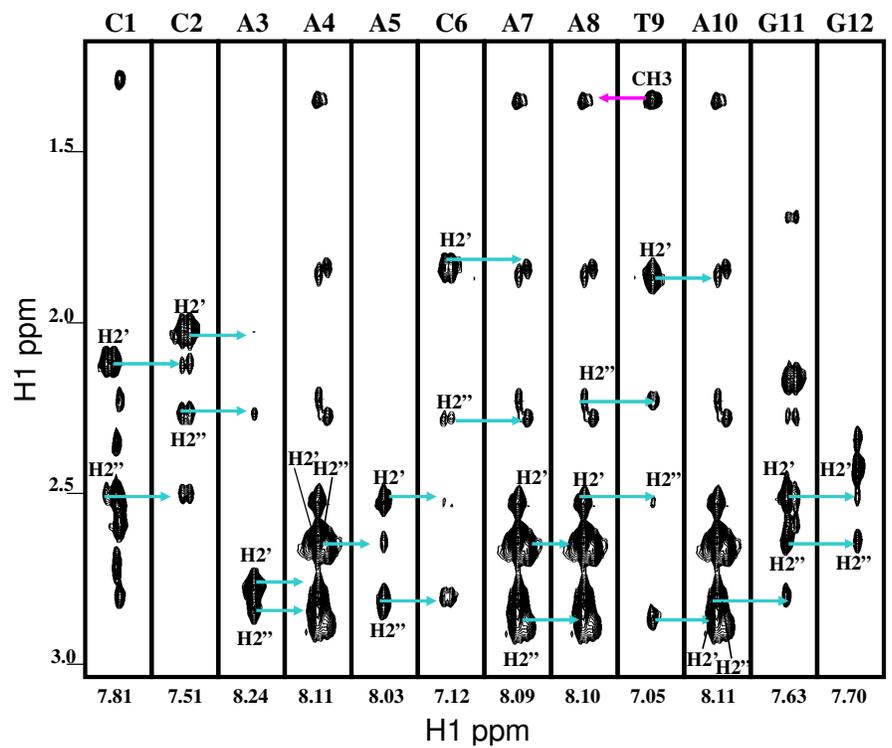


Figure 3.15: Sequential assignment of signals from the sugar H2', H2'' protons and the CH₃ methyl group of thymines in Strand G (top panel) and Strand H (bottom panel) of the Sox12 duplex.

3.2.2 *Antennapedia* DNA synthesis

After the experience with the Sox-5 DNA synthesis, optimization of the PCR-ESRA method to synthesise the *Antennapedia* DNA target started “on paper” in order to design oligonucleotide templates avoiding hairpin formation. An extensive analysis of several candidates was performed using the computer package cited above (Zuker 2003). All the designed sequences preserved as much as possible the original of *Antennapedia* DNA duplex used for the NMR structure determination of the complex (Qian *et al.* 1993, Billeter *et al.* 1993, Qian *et al.* 1994):

5' -GAAAGCCATTAGAG-3'
3' -CTTTCGGTAATCTC-5'

Starting from this 14 bp target sequence, several templates were designed changing only residues at the 3' and 5' ends in order to form the restriction site, but retaining all the residues involved in protein recognition. All these hypothetical templates and their analyses are reported in Appendix I.

The best predicted sequence, i.e. that having the lowest T_m and the least negative free energy of stabilisation of secondary structure (ΔG), was obtained (Figure 3.16) with the following strands (termed Antp32A_RsaI and Antp32B_RsaI) designed to include a restriction site for RsaI (the 4-base site is underlined in red):

5' -ACAAAGCCATTAGAGTACAAAGCCATTAGAGT-3'
3' -TGTTTCGGTAATCTCATGTTTCGGTAATCTCA-5'

With respect to the original sequence used for the structure determination, the introduction of the RsaI restriction site meant changing the first base pair (in blue) and adding one more base pair at each end (in green).

Another good candidate was that containing the HpyCH4V restriction site (see Appendix I), but since the difference in T_m and ΔG between the HpyCH4V and RsaI oligonucleotides was not significant and both enzyme have almost the same activity, the wet experiments were started using the sequence which required the cheaper restriction enzyme (RsaI).

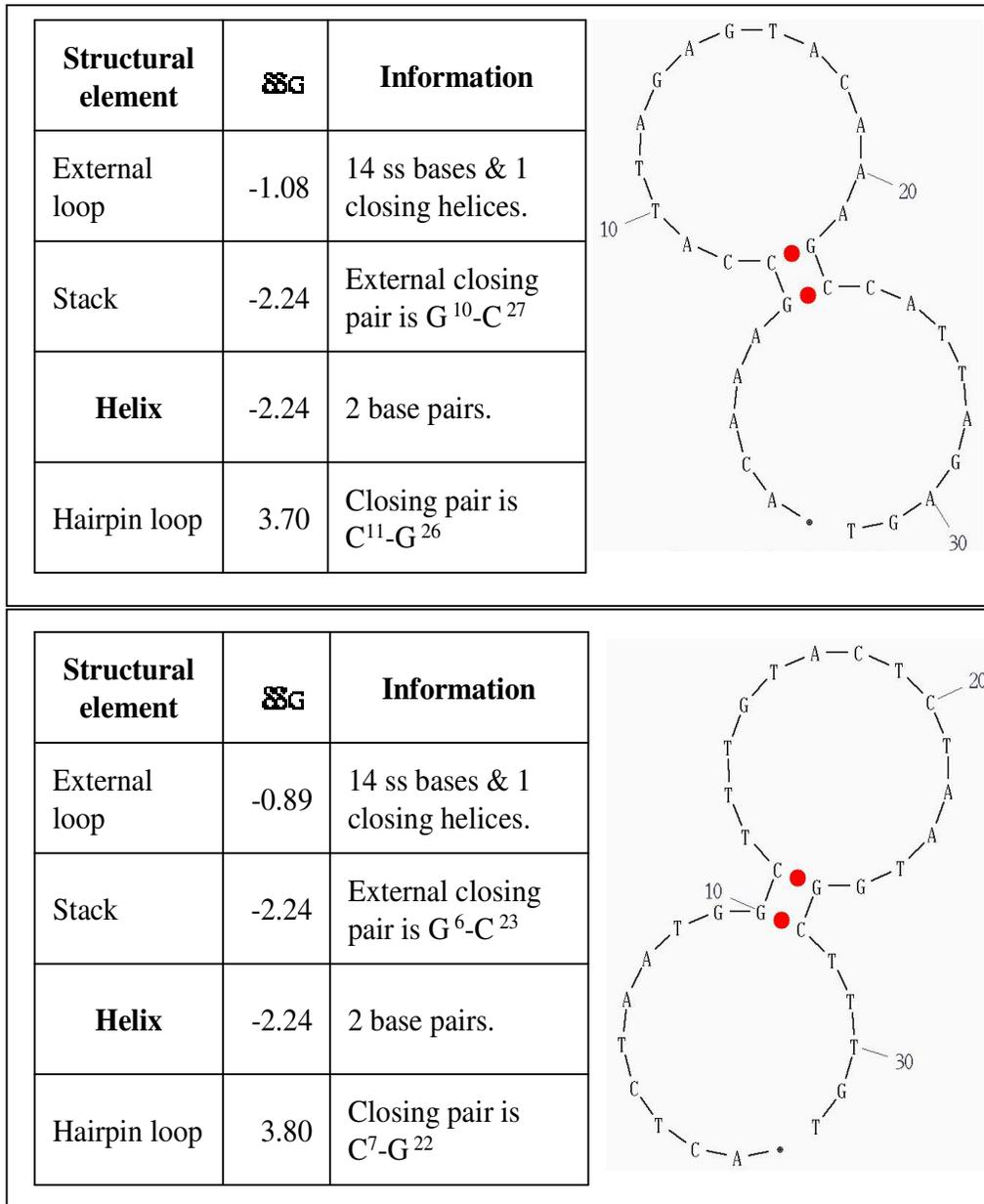


Figure 3.16: Hairpin analysis by a computer package (Zuker 2003) for oligonucleotides Antp32A_RsaI and Antp32B_RsaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C.

The optimisation of the PCR amplification process was then carried out as described above for the Sox12 DNA (section 3.2.1). The following parameters were tested:

1. Annealing temperatures: 44, 50, 56 and 60 °C.
2. Concentration of MgCl₂: 1.0, 1.5 and 2.0 mM
3. Amount of dNTPs: 0.1, 0.2 and 0.4 mM

As Figure 3.17 shows, the best results for the Rsa I oligos were obtained using the following parameters:

1. Annealing temperature: 50 °C.
2. Concentration of MgCl₂: 1.5 mM
3. Amount of dNTPs: 0.2 mM

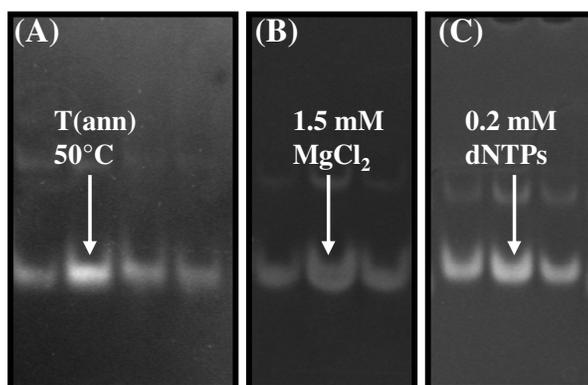
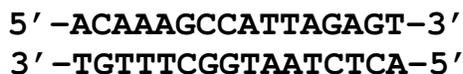


Figure 3.17: Optimization of the PCR-ESRA method for Antp DNA synthesis. The results were analysed on 16% polyacrylamide native gels in TBE buffer stained with ethidium bromide as a function of: (A) The annealing temperature (44, 50, 56 and 60 °C); (B) The concentration of MgCl₂ (1, 1.5 and 2.0 mM); (C) The amount of dNTPs (0.1, 0.2 and 0.4 mM). Arrows indicate the optimal conditions obtained for each parameter, i.e. those which gave a better yield. Upper bands (with a higher molecular weight) are probably due to incomplete restriction digestion.

The restriction digest was performed immediately after the final annealing step which followed the polymerization reaction, although for the Sox12 DNA a previous dialysis step was required. However, since the RsaI enzyme has good efficiency in the buffer used for the PCR reaction (see Materials and Methods), restriction could immediately follow polymerisation. The following duplex was finally obtained:



The total yield was ~ 44%, i.e. ~ 1.1 mg of purified DNA was obtained from each large-scale PCR-ESRA reaction. More details about yield and costs are reported in

the section below. To check the quality of the samples, mono-dimensional NMR spectra were acquired (Figure 3.18).

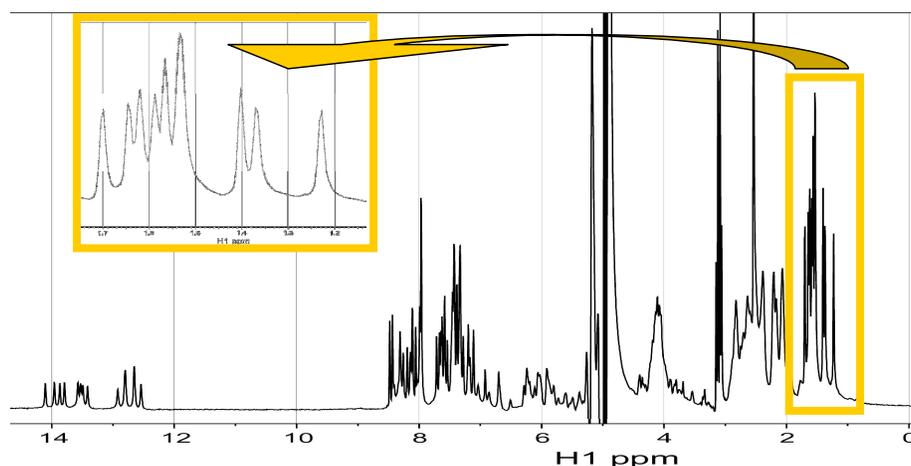


Figure 3.18: ^1H -1D NMR spectrum of the unlabelled Antp DNA target sequence. The spectrum was acquired in H_2O using a spectrometer operating at 600 MHz. In the zoomed yellow box is reported the region of the methyl groups of thymines. The clear discrimination of these signals in the spectrum indicates the quality of the synthesised DNA sample.

A 2D-NOESY spectrum was then acquired for the unlabelled Antp DNA sample (Figure 3.19).

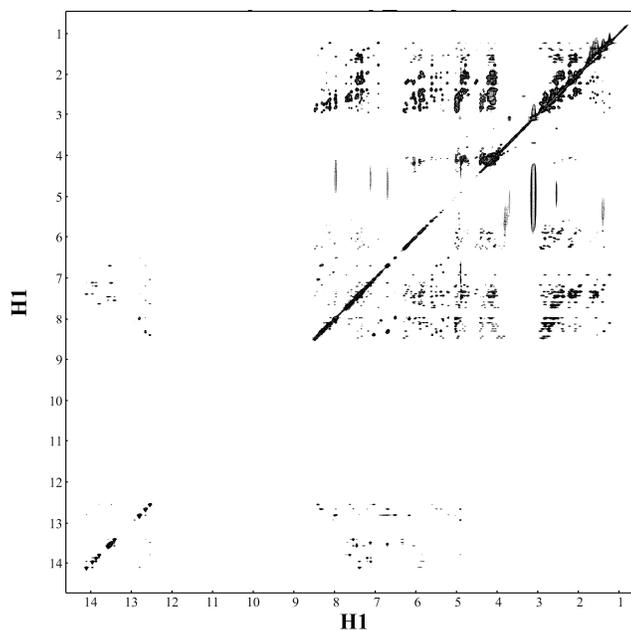


Figure 3.19: The 2D homonuclear NOESY spectrum acquired in H_2O at 600 MHz of the Antp DNA duplex.

3.2.3 Cost of preparing doubly-labelled DNA

In order to acquire good NMR datasets, the aim was to obtain 0.6 mM DNA samples in 600 μ l of both the Sox and Antp DNAs, i.e. ~2.6 mg of Sox12 duplex (MW: ~7200 Daltons) and ~3.5 mg of Antp16 duplex (MW: ~9600 Daltons) were required. From a large scale PCR-ESRA, performed using ~625 μ g of each mono-nucleotide, we obtained ~900 μ g of purified Sox12 DNA and ~1.1 mg of purified Antp16 DNA. The final yield was thus 36% for the Sox12 duplex and 44% for the Antp16 duplex.

Although these yields might seem low for an enzymatic process, the cost of preparing a good labelled NMR sample using PCR-ESRA is about 10 times less than buying it from a supplier. The analysis given below is applicable also for Antp16 DNA synthesis since the same quantity of reagents was used. In fact, although the yield obtained for Antp DNA synthesis was better than for Sox12, the final quantity in grams needed to prepare the NMR sample was higher for Antp16. Furthermore, both restriction enzymes used (i.e. HaeIII and RsaI) have a similar cost.

To prepare a DNA sample the following steps were required:

1. PCR-ESRA Reaction
2. Restriction enzyme digestion
3. Mono Q column purification

For preparing a labelled NMR DNA sample, the following reagents were purchased:

~480 μ g	Strand A template (HPLC purified)	£	~55.00
~580 μ g	Strand B template (HPLC purified)	£	~55.00
3x1000 units	Restriction enzyme (HaeIII/RsaI)	£	~150.00
2 x 500	units Taq Polymerase	£	472.80
4 x 10 mg	$^{13}\text{C}/^{15}\text{N}$ dNTP Set	£	1500.00
Total cost		£	~2232.80

With the exception of the restriction enzymes, not all of the purchased reagents were used for preparing an NMR sample and the real final cost per sample was about £825, i.e. with £1650 we prepared both the Sox12 and Antp16 DNA samples. This analysis does not of course include the cost of the monoQ column, which is considerable. Other (cheaper) cation-exchange columns or different purification strategies might alternatively be applied.

3.3 Antennapedia C39S homeodomain

Having synthesised both the Sox-5 and Antennapedia DNAs, attention was focussed on the Antennapedia protein. The aim was to make measurements of internal motions in the Antennapedia protein, both free and in a complexed state and then compare the resulting data with those obtained for the Sox-5 protein. In this way, the plan was to compare a major groove binding domain (the Antennapedia homeodomain) with a minor groove binding domain (the Sox-5 HMG box) and then obtain explanations at a microscopic level for the differences observed in the macroscopic thermodynamic quantities, the enthalpy and entropy of binding (Chapter I). For this, both unlabelled and $^{13}\text{C}/^{15}\text{N}$ labelled Antennapedia C39S homeodomain was prepared and thereby both an unlabelled-DNA/labelled-protein complex and a labelled-DNA/unlabelled-protein complex generated.

3.3.1 Protein preparation

The plasmid encoding the *Drosophila* Antennapedia C39S homeodomain (provided by the Privalov laboratory - Johns Hopkins University) was transfected into *Escherichia Coli* strain BL21(DE3)-pLysS by heat shock. The transformed cells were grown on either rich LB media or minimal media and protein expression then induced by adding IPTG. Cells were disrupted by sonication in the presence of PMSF as protease inhibitor. The purification protocol of the Antennapedia C39S homeodomain (termed AntpC39S) was based on its high pI value (theoretical pI = 11.20: <http://www.expasy.org> and Appendix IV) and optimised by the Privalov laboratory. Expressed AntpC39S is not incorporated into inclusion bodies, so the purification was carried out from the cleared supernatant obtained following sonication. Since AntpC39S is a DNA binding protein, the sample was depleted of nucleic acids by polyethylene-imine precipitation. The protein was then precipitated by addition of ammonium sulphate, the pellet resuspended in Buffer A at pH 7.5 for DEAE chromatography and then dialysed extensively against the same buffer overnight (see Materials and Methods).

DEAE column step:

The dialysed sample was loaded onto a DEAE-Sepharose anion-exchange column (Pharmacia), which had been previously equilibrated with Buffer A at pH 7.5 (see Materials and Methods). At this pH both the DEAE column and AntpC39S are positively charged, so the column retained certain contaminating proteins and nucleic acids which had not been removed by polyethylene-imine precipitation, leaving AntpC39S in the Flow Through (FT) (Figures 3.20 and 3.21). The FT also contained other bacterial proteins, which were removed by a second ion-exchange step.

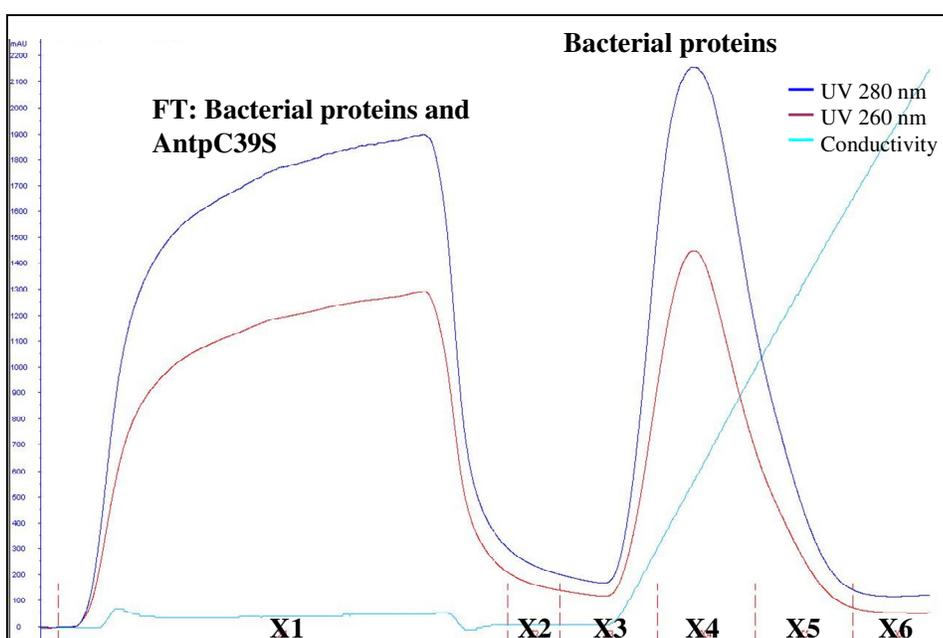


Figure 3.20: DEAE column trace of AntpC39S purification. The FT contained AntpC39S and bacterial proteins, which were removed in a second ion-exchange step. The elution was followed by UV at two wavelengths in order to observe proteins (280 nm, the blue line) and nucleic acids (260 nm, the purple line). The conductivity was monitored as a proxy for the salt gradient (the cyan line).

Peak fractions were checked by SDS-PAGE (Figure 3.21), which confirmed that AntpC39S did not bind to the DEAE, being present in the Flow Through. This was then dialysed over night against Buffer A pH 6.0 for SP cation-exchange chromatography (see Materials and Methods).

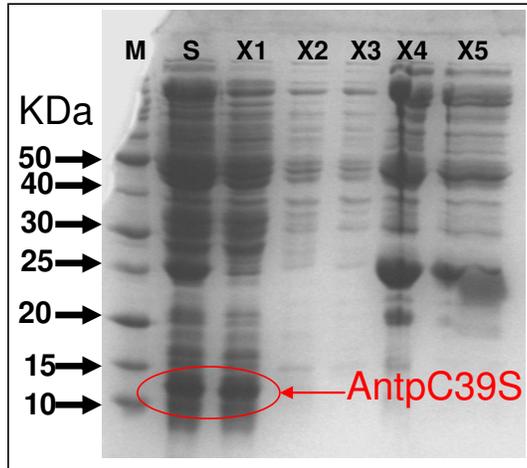


Figure 3.21: SDS PAGE analysis of DEAE column fractions. Lane M is a protein marker ladder (Invitrogen), while Lane S is the starting material. Each lane is labelled in accord with the column fractions of Figure 3.20. AntpC39S bands are circled in red. Although the presence of other bacterial proteins co-purifying with AntpC39S is evident, the DEAE step was able to remove a consistent fraction of other, in the main larger, bacterial proteins.

SP column step:

The dialysed FT from the DEAE column step was loaded onto a SP cation-exchange column which had been previously equilibrated with Buffer A at pH 6.0 (see Materials and Methods). At this pH AntpC39S is positively charged and since the column was negatively charged, the protein was retained. AntpC39S was then eluted using a linear NaCl gradient which allowed separation from residual bacterial proteins (Figure 3.22). The peak fractions were checked by SDS PAGE (Figure 3.23) which showed the effective purification of the Antp C39S protein on the SP column. Fractions with the highest purity level were pooled and dialysed overnight against the NMR buffer (see Materials and Methods) and then concentrated into a volume of ~1 ml on VIVASPIN tubes with a 5 kDa molecular cut off.

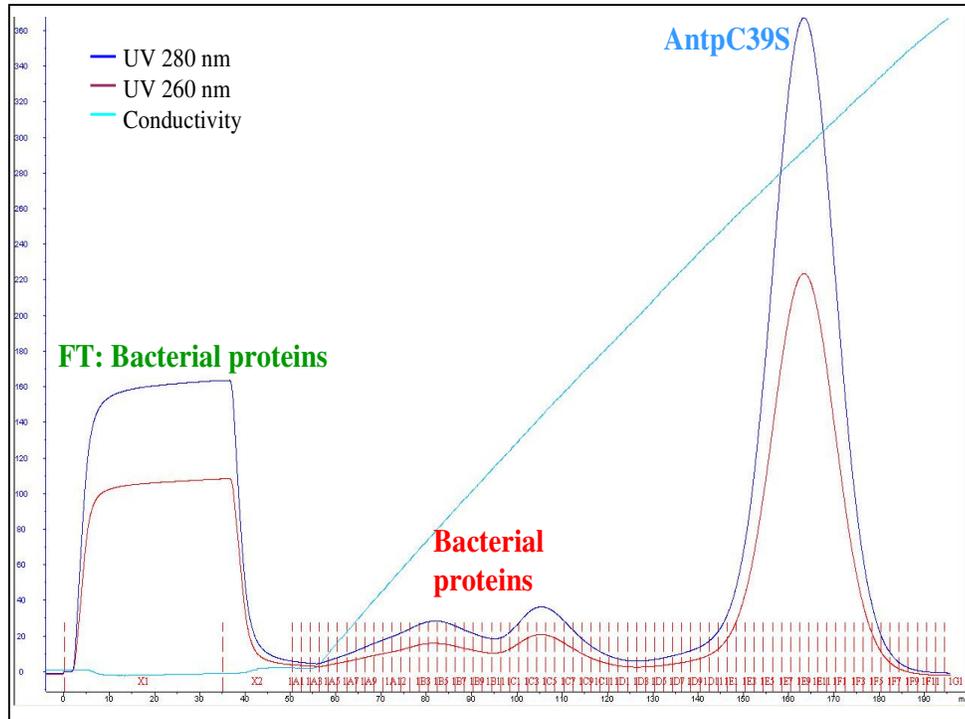


Figure 3.22: SP column trace of AntpC39S purification. The elution was followed by UV at two wavelengths in order to observe proteins (280nm, the blue line) and proteins plus nucleic acids (260nm, the purple line). The conductivity was monitored as a proxy for the salt gradient (the cyan line).

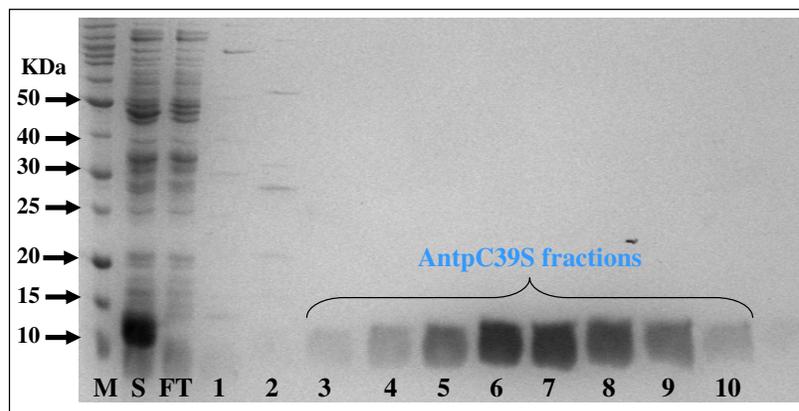


Figure 3.23: SDS PAGE analysis of SP column fractions. At the left hand side is the protein marker ladder (Invitrogen). Lane S corresponds to the starting material from the DEAE column, while Lane FT corresponds to the Flow Through. Lanes 1 and 2 represent impurity fractions of bacterial proteins and Lanes 3-10 contain AntpC39S (see Figure 3.22 for more details). The SDS PAGE analysis showed that a high level of purity was obtained. Antp bands are unusually broad since a high salt concentration remained from the NaCl gradient.

From 1 litre of rich broth, about 20 mg of purified AntpC39S was obtained, while about 12 mg of labelled protein was obtained from 1 litre of minimal medium. The final concentration was evaluated by UV measurement at 280nm using an empirical extinction coefficient of $15470 \text{ M}^{-1}\text{cm}^{-1}$ (<http://www.expasy.org> and Appendix IV). In order to acquire the various NMR datasets, just one sample of unlabelled and one of doubly-labelled AntpC39S were prepared since the purified protein domain proved to be of high stability.

Part of the $^{13}\text{C}/^{15}\text{N}$ labelled sample was used to acquire all the NMR datasets for the free protein and the remainder was used to form a complex with its target DNA (see section 3.2.2). The unlabelled AntpC39S sample was used to form a complex with doubly labelled DNA in order to investigate the mobility of the nucleic acid in the complex. All NMR experiments and analyses are reported in the sections below. The method of 1:1 complex formation is explained in Material and Methods.

3.4 NMR analysis of the *Antennapedia* homeodomain

To check the quality of the samples, a mono-dimensional NMR spectrum was acquired for the unlabelled protein (Figure 3.24), while two-dimensional [^1H - ^{15}N]-HSQC NMR spectra (Bodenhausen *et al.* 1980) were acquired for both the labelled protein and the unlabelled DNA/labelled protein complex (Figures 3.25 and 3.26).

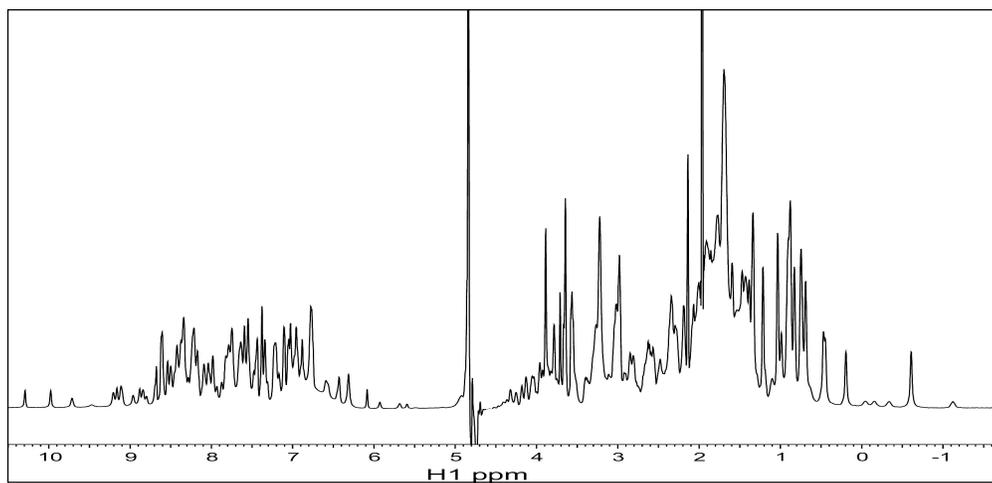


Figure 3.24: ^1H -1D NMR spectrum of the unlabelled AntpC389S protein acquired at 293 K in 10mM Potassium Phosphate pH 6.0 and 100 mM KCl with a VARIAN spectrometer operating at 600 MHz.

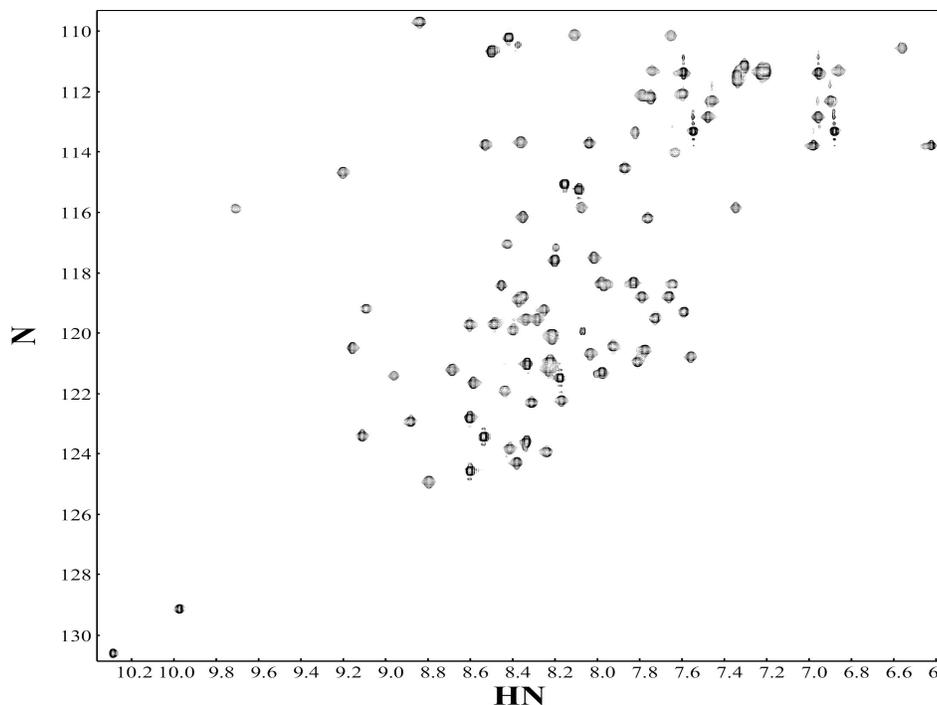


Figure 3.25: [^1H - ^{15}N]-HSQC NMR spectrum of the free AntpC39S acquired at 293 K with an INOVA VARIAN spectrometer operating at 14.1 T. The broad dispersion of the cross-peaks indicates that protein was folded and appropriate for NMR analysis.

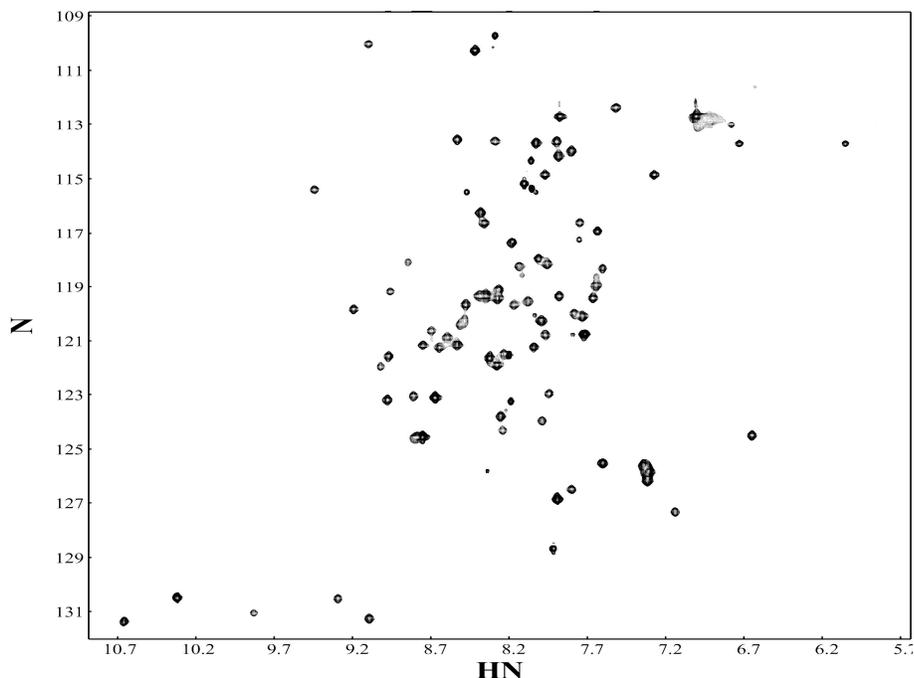


Figure 3.26: [^1H - ^{15}N]-HSQC NMR spectrum of AntpC39S in complex with its target DNA. The spectrum was acquired at 293 K with an INOVA VARIAN spectrometer operating at 14.1 T. The broad dispersion of the cross-peaks indicates that protein was folded after the complex formation and appropriate for NMR analysis.

The 2D [^1H - ^{15}N]-HSQC and [^1H - ^{13}C]-HSQC NMR spectra are the basic hetero-nuclear NMR experiments used to study bio-molecules. In particular, in this kind of spectrum the correlation is observed of a hetero-nucleus to its directly attached proton through one-bond coupling. Hence the name **HSQC** experiment: **H**etero-nuclear **S**ingle-**Q**uam coherence experiment. Since the ^1H - ^{15}N and ^1H - ^{13}C one-bond coupling constants are comparatively large (100 Hz for ^1H - ^{15}N , 140 Hz for $^1\text{H}\alpha$ - $^{13}\text{C}\alpha$ and 30-40 Hz for $^1\text{H}\beta$ - $^{13}\text{C}\beta$), ^{15}N -HSQC and ^{13}C -HSQC experiments are highly efficient. The pulse sequence for a ^{15}N -HSQC experiment is reported in Figure 3.27.

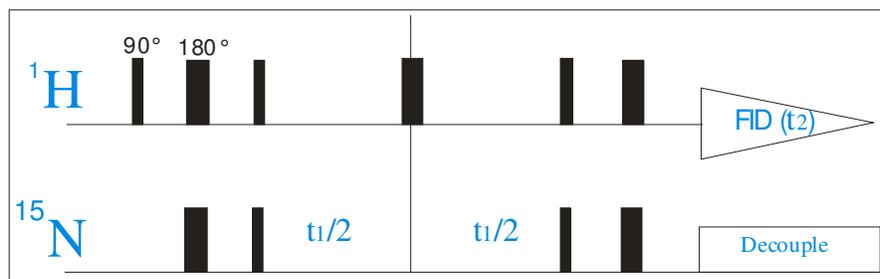


Figure 3.27: Pulse sequence scheme of ^{15}N -HSQC experiment (Bodenhausen et al. 1980). t_1 is the evolution time. The proton magnetization is recorded during the acquisition period t_2 .

In particular, the ^{15}N -HSQC spectrum is considered a “finger-print” of the protein, since the broad dispersion of the cross-peaks and the absence of minor peaks indicates whether the protein is folded and is pure, i.e. appropriate for NMR analysis or not. As Figures 3.25 and 3.26 show, these features were found for both AntpC39S samples.

All NMR datasets were obtained on a Varian INOVA spectrometer operating at 14.1 T located in the Biophysics Laboratories of the University of Portsmouth. As explained in the first section of this chapter, our efforts were directed to relaxation measurements of the backbone and side-chains of both free and complexed protein. To achieve this goal, first of all it was necessary to re-assign all the protein resonances in both states. This because the resonance assignment already available in the literature (Qian *et al.* 1993, Billeter *et al.* 1993, Qian *et al.* 1994) was obtained only for the free protein at pH 4.3 and only for protons resonances.

3.4.1 Resonances assignment

The aim of the NMR assignment stage was to associate each spectral signal with a specific nucleus in the molecule under study. This process was performed in two parts: (i) the assignment of resonances corresponding to backbone atoms; (ii) the assignment of resonances corresponding to side chains. To achieve the assignment of AntpC39S both free and in its DNA complex, the following NMR experiments were performed:

- 2D hetero-nuclear experiments: [^1H - ^{15}N]-HSQC and [^1H - ^{13}C]-HSQC
- 3D hetero-nuclear double resonance experiments: ^{15}N -edited and ^{13}C -edited NOESY-HSQC
- 3D hetero-nuclear triple resonance experiments: HNCA, CBCA(CO)NH and HNC0

The choice of three-dimensional hetero-nuclear experiments was because 2D homo-nuclear NMR spectra analyses generally present several problems with larger molecules. Both the number of proton resonances and the resonance line-widths increase with size. This leads to overlap of resonances and therefore to ambiguous assignments to specific proton pairs of 2D cross peaks. By increasing the

dimensionality of the spectrum, problems associated with spectral overlap can be solved. Another problem related to large molecules is the signal broadening associated with the increased rotational correlation time which makes it more difficult to observe small proton-proton J-coupling, and magnetization transfer in 2D spectra based on such small J-couplings becomes inefficient. This disadvantage is overcome by hetero-nuclear multidimensional NMR methods which have now become standard in studying bio-macromolecules (Fesik *et al.* 1990, Clore *et al.* 1991, Bax *et al.* 1993). The 3D NMR pulse sequences are derived from a combination of 2D experiments as shown in the Figure 3.28. The third dimension is created by introducing a second evolution period, t_2 . Therefore, in a 3D-experiment the signal, which is monitored during the acquisition time t_3 , is a function of two evolution times, t_1 and t_2 , which are incremented independently. The signal is a matrix with three axes corresponding to t_1 , t_2 and t_3 and then, after Fourier transformation, a spectrum with three different frequency dimensions ω_1 , ω_2 , ω_3 is obtained. As Figure 3.29 shows, the 3D spectrum can be imagined as a cube.

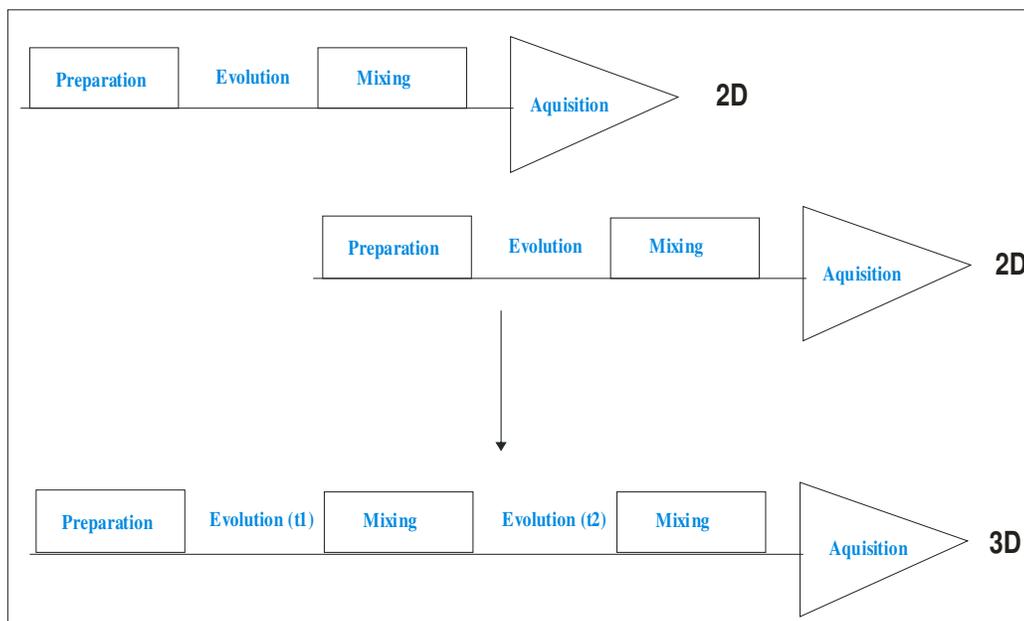


Figure 3.28: General scheme for a 3D NMR experiment. The 3D NMR pulse sequences are derived from a combination of 2D experiments. The third dimension is created by introducing a second evolution period, t_2 , the signal, which is monitored during the acquisition time t_3 , is then a function of two evolution times, t_1 and t_2 , which are incremented independently.

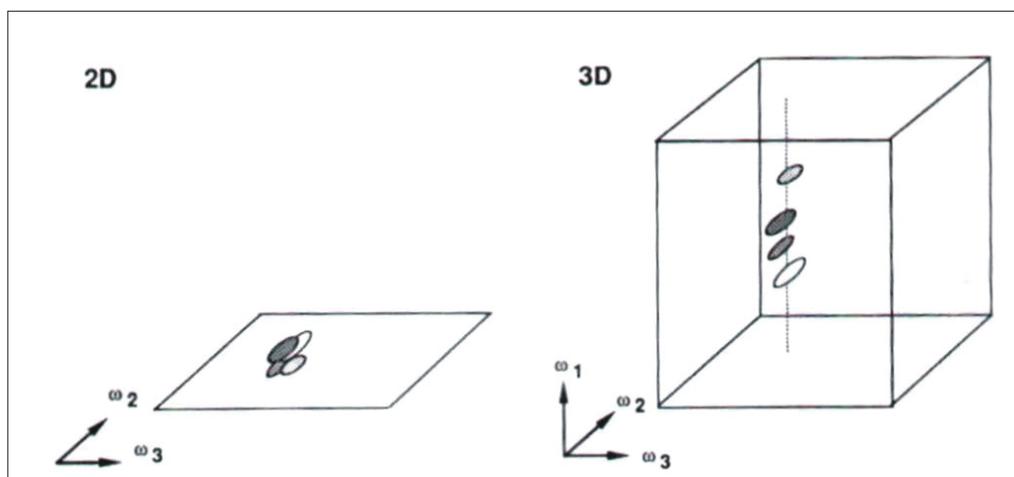
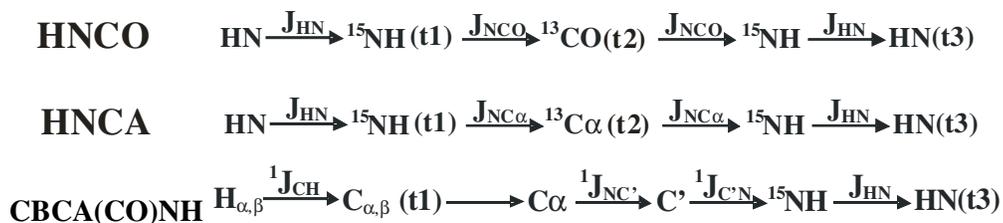


Figure 3.29: Schematic view of a 3D NMR experiment. The 3D spectrum can be imagined as a cube where signals overlapped in a 2D spectrum can be solved in the third dimension.

To obtain the assignment of AntpC39S (free and complexed), the ^{15}N and H^{N} chemical shifts were identified in the $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra on the basis of the available assignments and then the sequence specific backbone assignment was confirmed using the 3D ^{15}N -edited NOESY-HSQC spectra, where short range $\text{H}^{\text{N}}-\text{H}^{\text{N}}$, intra-residue, sequential and long range NOEs are observed. That strategy was chosen because the small dimension of the AntpC39S protein allowed a good resolution of the NMR signals and because it is a highly α -helical protein, so almost all sequential H^{N} NOEs were strongly present in the spectra. Having performed the ^{15}N -NOESY-HSQC spectra analysis, which, with the exception of a few residues, allowed the sequential assignment of both free and complexed AntpC39S, we analysed the triple resonance HNCA, CBCA(CO)NH and HNCO spectra. This kind of experiment correlates backbone H^{N} , $^{15}\text{N}^{\text{H}}$, H^{α} , $^{13}\text{C}^{\alpha}$, ^{13}CO and side-chain H^{β} and $^{13}\text{C}^{\beta}$ spins using one-bond heteronuclear coupling constants (Ikura *et al.* 1990). In particular, in the HNCO spectrum the chemical shifts of the H^{N} proton and the $^{15}\text{N}^{\text{H}}$ nitrogen of one residue (i) are correlated with the ^{13}CO of the previous residue (i-1), while in the HNCA spectrum one of the cross-peaks correlates each H^{N} with the $^{13}\text{C}^{\alpha}$ of residues i and i-1. The CBCA(CO)NH triple resonance NMR experiment generates correlations between $\text{N}_i-\text{N}_i^{\text{H}}-\text{C}_{i-1}^{\alpha}-\text{C}_{i-1}^{\beta}$. Additionally, the assignment of C^{α}

and C^β chemical shifts also helps in identifying the amino acid. The overall flow of magnetization in the three experiments is schematised below:



The analysis of the three triple resonance spectra described above, allowed the assignment of all ${}^{13}\text{C}$ backbone resonances and the identification of all ${}^{15}\text{N}^{\text{H}}$ backbone resonances which were not identified using the 3D ${}^{15}\text{N}$ -edited NOESY-HSQC spectra. As example of the sequential assignment, Figures 3.30 and 3.31 report three different strips (corresponding to residues Arg11-Tyr12-Gln13) of both HNCA and ${}^{15}\text{N}$ -NOESY-HSQC spectra of the free AntpC39S.

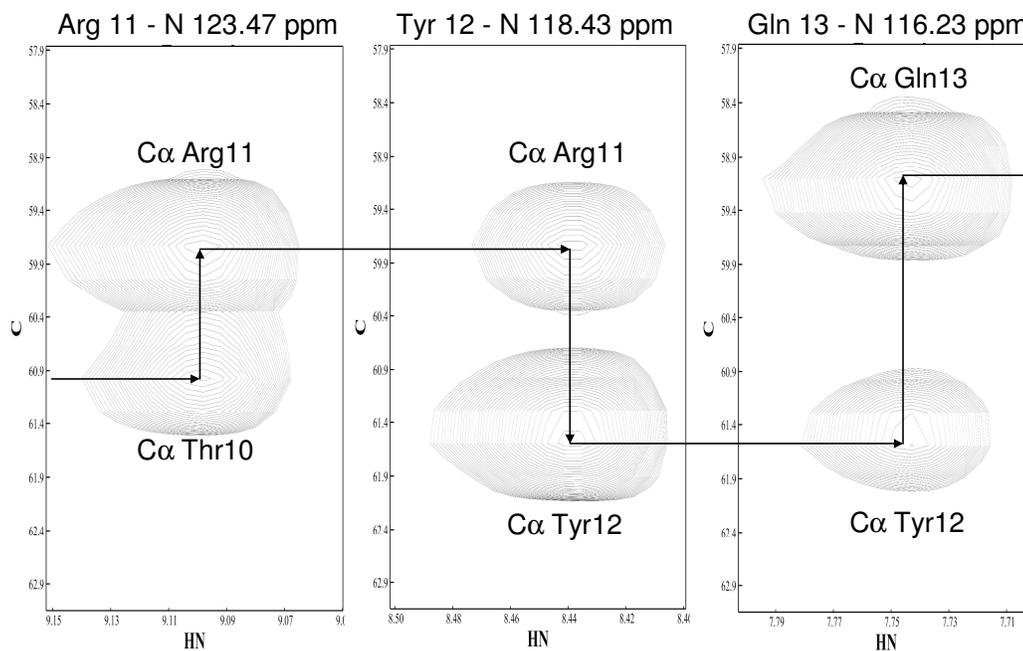


Figure 3.30: Strips extracted from the HNCA spectrum of the free AntpC39S. The figure shows the sequential assignment of residues Arg11-Tyr12-Gln13.

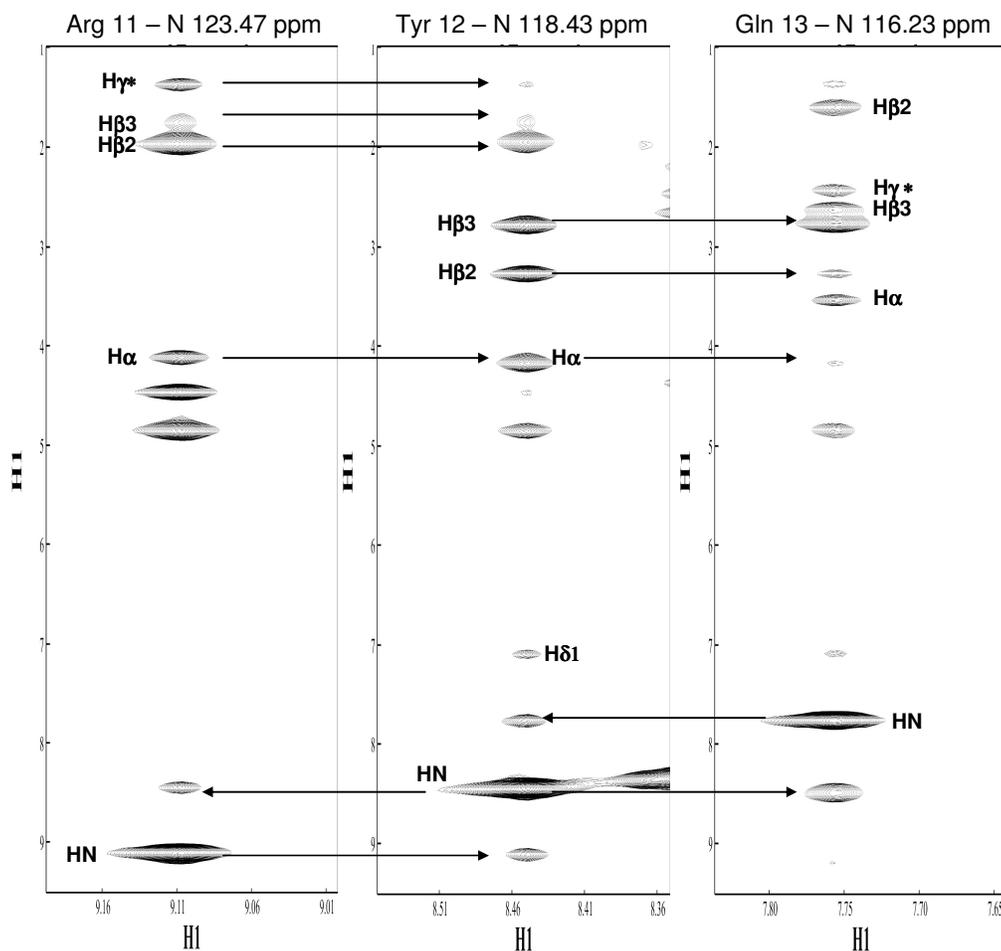


Figure 3.31: Strips extracted from the ^{15}N -NOESY-HSQC spectrum of free AntpC39S showing the sequential assignment of residues Arg11-Tyr12-Gln13.

Finally, all the sequential backbone assignments were obtained with the exception of residues M1, R2, K3 and R4 for the free protein and of residues M1, R2 and K3 for the protein in the complex. The assignment of residue R4 in the complex confirms the good quality of the data. In fact, it is reported that this residue is involved in a salt bridge when DNA binding occurs. Appendix II gives the assignments of AntpC39S in both the free and complexed states. In particular, Figures II.1 and II.2 report the assigned $[^1\text{H}-^{15}\text{N}]$ -HSQC spectra, where to each signal is assigned at least one amino acid of the protein. Concerning the proton resonances, the assignment performed here was not too different from that already available in the literature. Appendix II reports a comparison of the two assignments.

3.4.2 NMR relaxation analysis

Although globular proteins are normally well packed and adopt ordered 3D structures, they still have a variety of motions such as bond vibrations, side-chains rotations, segmental motions and domain movements. Motions are vital for protein function, which depends on alteration in 3D structure in response to specific molecular interactions. The binding of a ligand to a protein can result in changes in a protein's structure and dynamics; these changes can be local or can involve extended regions of the protein. In order to understand how proteins work, we need to know which motions are present and how they are related to the protein's biological function. Detailed knowledge of protein dynamics is therefore crucial in understanding the mechanism of protein function, including protein folding, ligand recognition, allostery and catalysis.

Multidimensional NMR methods combined with isotope labelling, can provide access to virtually every atom in a molecule, unique for protein structural studies. This not only allows characterization of the structure and interaction of proteins in their native milieu, but also provides unparalleled possibilities for obtaining a complete atomic-level resolution picture of protein dynamics from picoseconds up to seconds, the range where most motions relevant to protein function take place. Appropriate experimental data for the description of rapid internal dynamics of proteins can be obtained from relaxation rates of distinct states of magnetization. Auto-relaxation rates are the longitudinal relaxation rates, the time constant of which is called the spin-lattice relaxation time (T_1), and the transversal relaxation rates, the time constant of which is called spin-spin relaxation time (T_2). The transfer of magnetization, according to some relaxation mechanism, from one nucleus to another, is known as cross relaxation. This process is the basis of NOE spectroscopy. In general, two mechanisms of interactions with the environment are important for relaxation in proteins: the dipole-dipole interaction and the chemical shift anisotropy (CSA). For the dipolar interaction, the rotational motion of the vector connecting the two interacting nuclei relative to the external magnetic field is relevant, as well as any modulation of the length of this vector. The latter is not important if the distance between two interacting nuclei is fixed by a chemical bond. This makes ^{15}N and ^{13}C relaxation attractive for studies of mobility because these nuclei relax primarily by

interaction with the directly attached protons (Engelke *et al.* 1999). For a CH spin system the contribution of the CSA terms amounts to about 3% and can therefore be neglected. For a NH vector however, the CSA contribution is of the order of 20%. Furthermore, the influence of dipolar-heteronuclear CSA cross-correlation is significant in both cases and can potentially affect measurements of T1, T2 and NOEs. Fortunately, this contribution can be eliminated from relaxation measurements by continuous inversion of the proton magnetization during the relaxation period (Engelke *et al.* 1999).

T1, T2 and heteronuclear NOE measurements

The longitudinal relaxation time T1 describes the rate at which a disturbed spin ensemble again reaches a state of equilibrium. The transversal relaxation time T2 describes the rate with which the phase coherence between the spins of a spin-ensemble is decreasing. The steady state heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOEs are determined as the ratio of cross-peak intensities in two experiments, with and without presaturation of amide ^1H nuclear spins. The heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOE experiment (NOESY) is based on the Nuclear Overhauser Enhancement effect where a change in intensity of the NMR signal of the nucleus S is observed during the perturbation of the equilibrium magnetization of an adjacent nucleus I.

These three relaxation parameters were measured for both free and complexed AntpC39S. In both cases, eight experiments were performed for T1 measurements, using the following values of the relaxation delay (0.0089, 0.049, 0.097, 0.193, 0.499, 0.997, 1.495, 0.049 seconds). T2 data sets were obtained employing relaxation delays of: 0.0086, 0.0171, 0.0342, 0.0599, 0.0771, 0.0942, 0.120 and 0.146 seconds. Two identical $^{15}\text{N}\{^1\text{H}\}$ NOE experiments were recorded: in one, proton atoms were saturated during the recycle delay, while in the second one the recycle delay was used without proton saturation. All relaxation data were processed and analysed using NMRPipe/NMRDraw software (Delaglio *et al.*, 1995). Peak volumes were measured for the calculation of all relaxation parameters. T1 and T2 were determined by fitting the peak heights at multiple relaxation delays to the

equation $I = I_0 e^{-T/T_1}$. As an example, Figure 3.32 reports the T2 fitting for phenylalanine 23 of free AntpC39S.

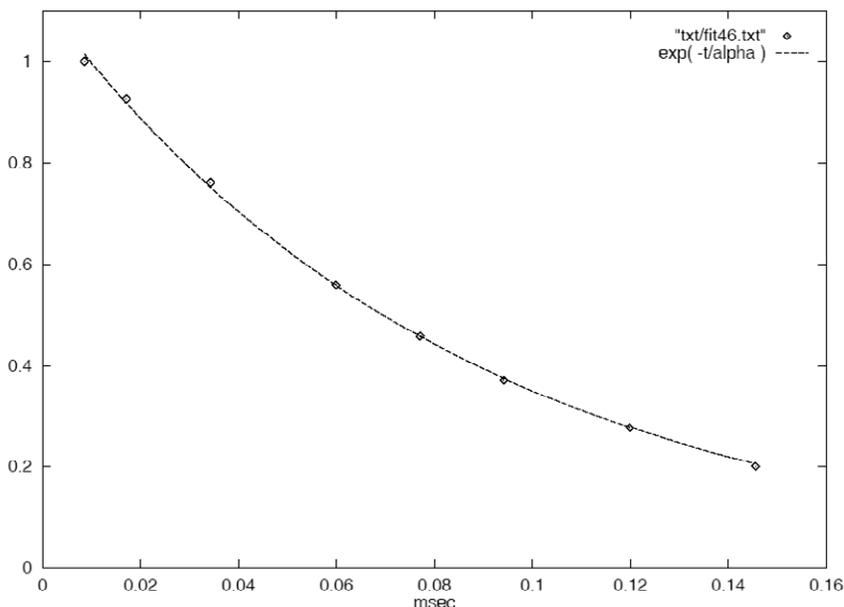
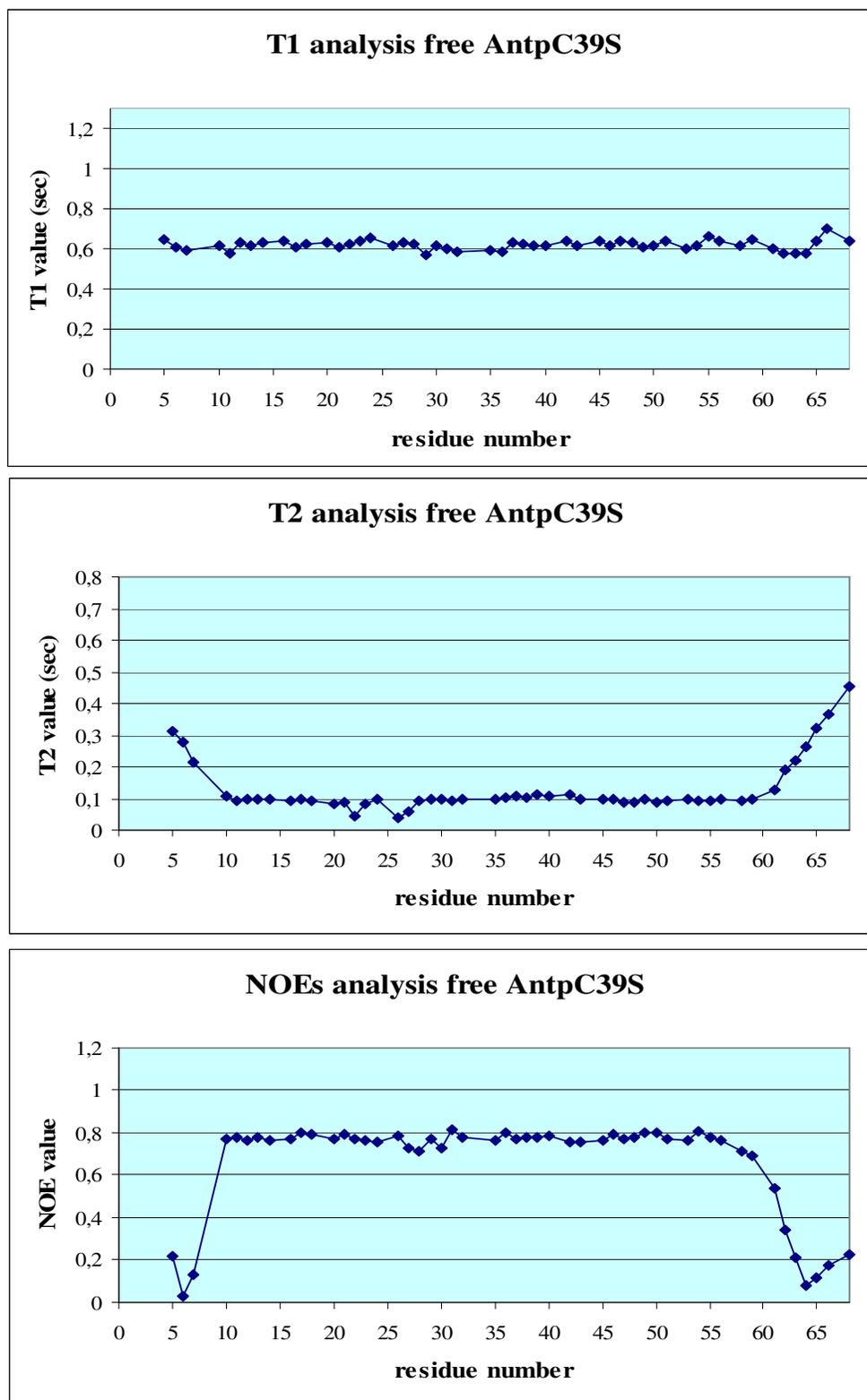


Figure 3.31: T2 fitting for residue F23 of free AntpC39S. T2 was determined by fitting the peak volume at each relaxation delay to the equation $I = I_0 e^{-T/T_1}$: the obtained value is 0.083 ± 0.001 .

The steady state heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOEs were calculated as the ratio of $^{15}\text{N}\{^1\text{H}\}$ correlation peak volumes in the spectra acquired with and without the proton saturation. The NOE data are generally more susceptible to errors than T1 and T2 because (i) the NOE experiment starts with an equilibrium ^{15}N magnetization that is about 10 fold lower than that of ^1H , hence is relatively much less sensitive, and (ii) the NOE values are derived from only two measurements, whereas T1 and T2 data are obtained from fitting multiple data points, which results in a more efficient averaging of experimental errors. The full relaxation parameters measured for both free and complexed AntpC39S are reported in Appendix II. Some residues were not fitted because of peak overlap or because the assignment was uncertain.

Figures 3.32 A and B give plots of T1, T2 and NOE values vs the residue number both free and complexed AntpC39S.



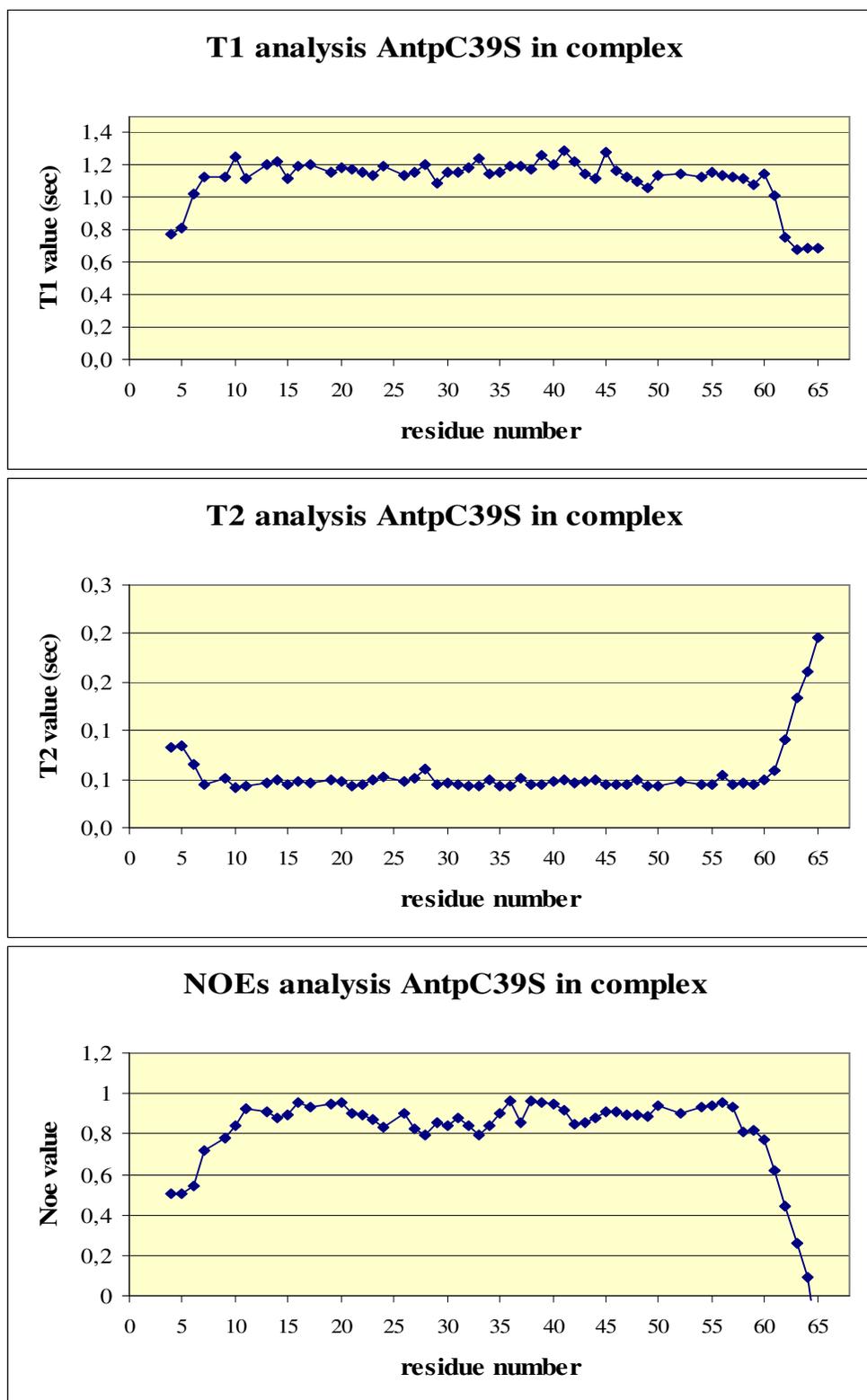


Figure 3.32B: Relaxation data analysis of AntpC39S in complex with its target DNA.

The protein appears to be in a rigid conformation both free and bound, with the exception of the N- and C-terminal regions. The data show that the N-terminal region of AntpC39 in its complex with DNA has a more rigid conformation than in the free state. In fact, as reported in the literature (Qian *et al.* 1994, Qian *et al.* 1993, Billeter *et al.* 1993), the N-terminal tail interacts with the minor groove when DNA binding occurs, while in the free state it is in a flexible conformation. Anyway, a comparison with the relaxation data of Sox-5 is necessary to understand if the differences in macroscopic parameters, i.e. ΔH and ΔS of binding, can be correlated with differences in atomic mobility.

CHAPTER IV

RESULTS AND DISCUSSIONS: LA PROTEIN

The main aim of this research work is to understand which are the features that give La the ability to bind various RNA targets like the 3' poly(U) of polymerase III transcripts as well as viral RNA, which do not contain poly(U). To achieve this, both structural and biochemical studies were carried out. In particular, to examine the RNA binding properties of La, gel mobility shift assays with several deletion mutants of hLa were performed with both a 10 nt oligo(U) RNA and a 27mer Hepatitis C virus RNA fragment. Furthermore, since almost all the determinants of hLa for oligo(U) RNA recognition and high affinity binding appear to be located in its NTD (residues 1-194), the solution structure of the CentralRRM (residues 105-202) by NMR was also determined and chemical shift mapping experiments were carried out to analyse the interaction between poly(U) and the NTD of hLa .

4.1 Proteins preparations

To explore which domains of La are involved in the RNA binding, by gel shift assays (EMSA, see section 4.2.1), the following deletion mutant proteins were prepared:

- La1-103: residues 1-103, corresponding to the La motif;
- La105-202: residues 105-202, corresponding to the Central-RRM;
- La225-334: residues 225-334, corresponding to the C-terminal RRM;
- La105-334: residues 105-334, containing both Central and C-terminal RRMs;
- La1-194: residues 1-194, corresponding to the N-terminal domain and containing both La-motif and Central-RRM;
- La1-334: residues 1-334, containing the La motif and both the Central and the C-terminal RRMs;
- La1-344: residues 1-344, containing the La motif, both the Central and the C-terminal RRMs and part of the C-terminal short basic motif;
- LaWT: residues 1-408, corresponding to the full length human La.

All protocols for La proteins preparations were already well defined in Conte laboratory. Expressed proteins were not incorporated into inclusion bodies, so all purifications were carried out from the cleared supernatants obtained following sonication. As described below, for all proteins it was used an affinity chromatography strategy.

4.1.1 La1-194 and LaWT proteins purification

Both LaWT and La1-194 were purified exploiting their affinity for nucleic acids. In particular their purifications were performed in two affinity chromatography steps: the first one using a Heparin column, while the second one using a polyU column.

Heparin column step:

The cleared supernatant obtained following sonication was loaded onto a High Trap Heparin Column pre-equilibrated with Heparin Buffer A at pH 7.25 (see Materials and Methods). This column allows the capture of both LaWT and La1-194 because the presence of sulphated glycosaminoglycan on the resin, which competed with nucleic acids in binding the protein. The protein was then eluted using a linear gradient of 2M KCl. In order to observe the separation of nucleic acids from the protein, the purification was followed by recording UV traces at two different wavelengths: 254 nm to detect nucleic acids and 280 nm to detect proteins. In Figure 4.1 it is shown the FPLC trace of LaWT obtained from the Heparin column. The same profile was obtained for La1-194. From the figure it is clear the presence of nucleic acids in flow through on the basis of the UV absorbance at 254 nm being higher than the one at 280nm. The eluted peak on the contrary gave a UV profile typical of a protein. Peak fractions were checked by SDS PAGE (Figure 4.2), which confirmed the presence of LaWT in the eluted peak. Although the Heparin column step allowed the separation of nucleic acids bounded with the protein, a second step of purification was needed because the presence of co-purified contaminants (bacterial proteins and probably degradation fragments). The fractions containing the protein of interest were then pooled together and diluted in a ratio 1:3 with Buffer A pH 7.25 (see Material and Methods) to be loaded onto the polyU column for the second step of purification.

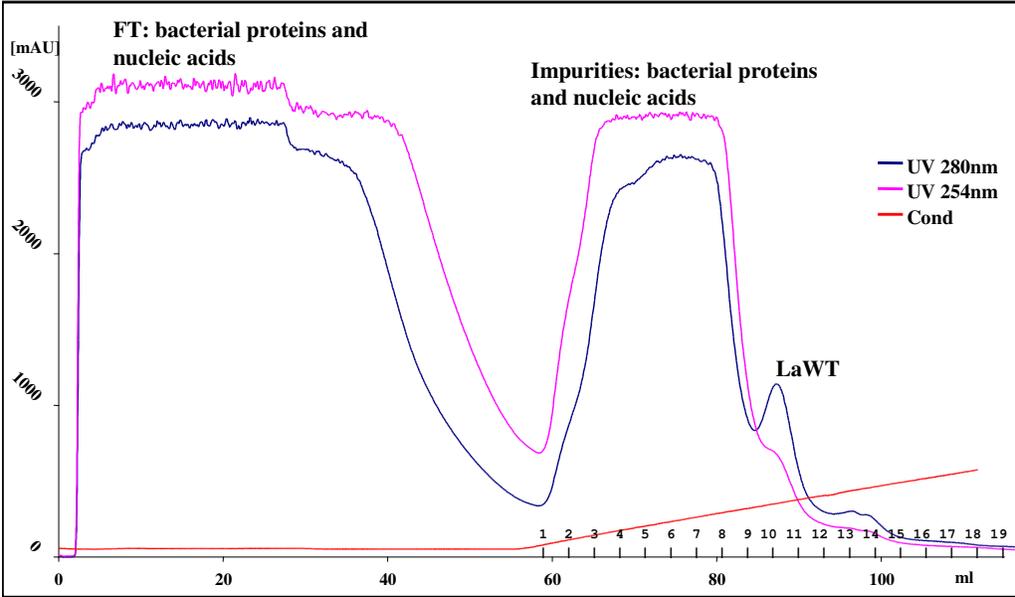


Figure 4.1: Heparin Column trace of LaWT. The elution was followed by UV at two different wavelengths: at 280 nm to observe the protein and at 254 nm to observe the nucleic acids. The FT contained almost bacterial proteins and nucleic acids, while LaWT eluted late free of nucleic acids.

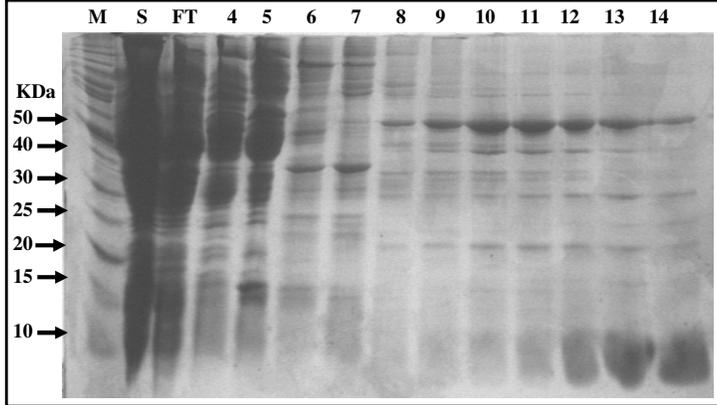


Figure 4.2: SDS PAGE analysis of Heparin column fractions. Lane M is a protein marker ladder (Invitrogen), Lane S is the starting material and Lane FT is the flow through. The other lanes are labelled in accord with the column fractions of Figure 4.1. Although the presence of other bacterial proteins copurifying with LaWT is evident, the Heparin step was able to remove a consistent fraction of other, in the main larger, bacterial proteins.

purification by Nickel affinity column. As example, below it is reported the purification of La105-202, which had been used for both gel shifts analysis (unlabelled protein) and the NMR structure determination (^{15}N and double $^{15}\text{N}/^{13}\text{C}$ labelled protein).

Nickel column step

The supernatant from cell lysis step was loaded onto a Ni-NTA column pre-equilibrated with Buffer A (see Materials and Methods). La105-202 was retained by the Nickel column, as well as all other His-tagged proteins, because the presence of the His-tag, which coordinates the Nickel ions. Elution was then performed using a linear gradient of Imydazol concentration (Figure 4.4), which competed with the protein in binding the Nickel ions. Peak fractions were checked by SDS-PAGE (Figure 4.5) and those containing La105-202 were pooled and dialysed against Heparin Buffer A overnight (see Material and Methods). As the SDS-PAGE analysis showed, the protein was not pure after the Ni-NTA column, so another purification step was performed using a High Trap Heparin column.

Eliminato: 5

Eliminato: 6

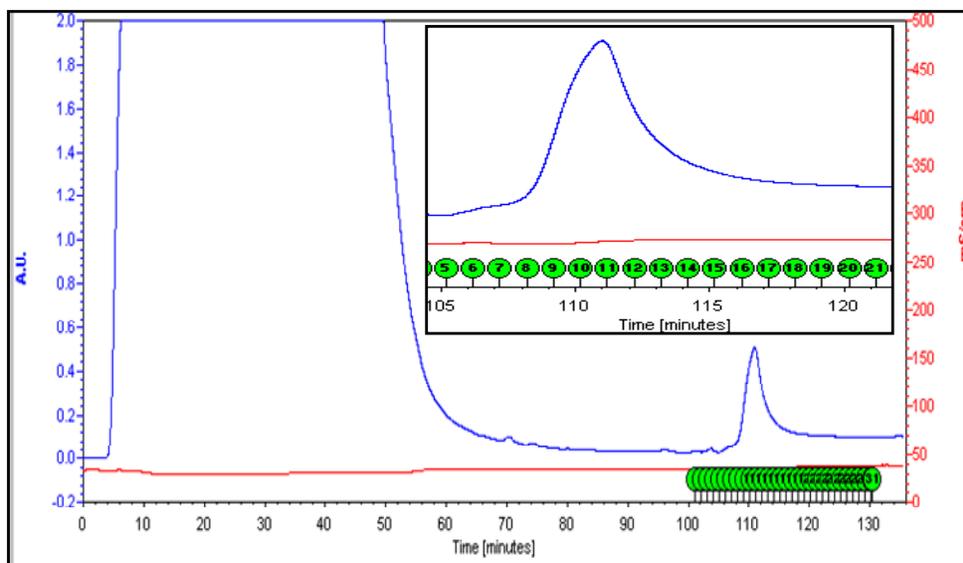


Figure 4.4: Nickel column trace of La105-202 purification. The FT contained almost all bacterial proteins, while the main peak contained the purified La105-202. The black box shows the zoomed protein peak.

Eliminato: 5

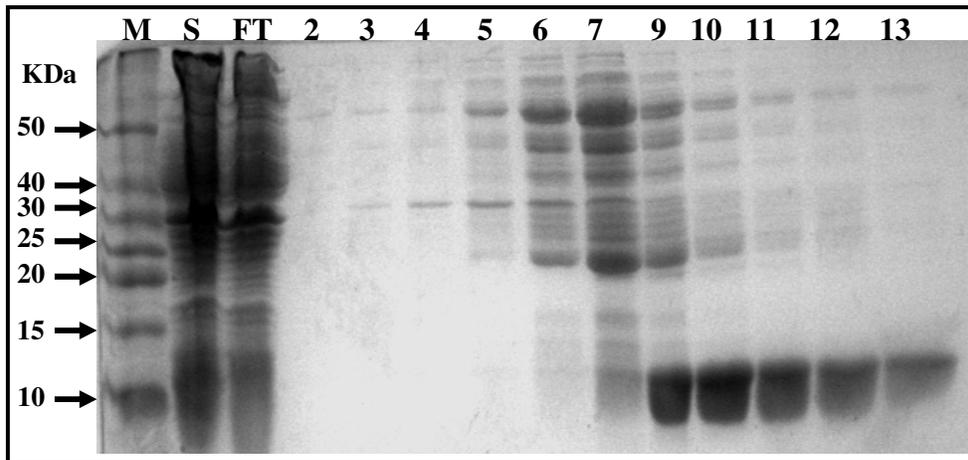


Figure 4.5: SDS PAGE analysis of Nickel column La105-202 fractions. Lane M is the protein ladder (Invitrogen). Each lane is labelled in accord with the column fractions of Figure 4.6.

Eliminato: 6

Heparin column step

The dialysed sample from the previous step was loaded onto a Heparin column equilibrated with Heparin Buffer A at pH 7.25 (see Material and Methods). This second purification step was necessary in order to remove residual bacterial proteins co-purified with La105-202 and nucleic acids, which were probably present because of the RNA binding properties of La protein.

La105-202 (theoretical pI 9.07: <http://www.expasy.org> and Appendix IV) is likely to be positively charged at pH 7.25, therefore retained by the negatively charged Heparin column. Furthermore, this column allows the separation of nucleic acid because the presence of sulphated glycosaminoglycan on the resin, which competed with nucleic acids in binding the protein. La105-202 was then eluted using a linear gradient of 2M KCl (Figure 4.6). In order to observe the separation of nucleic acids from the protein,

Eliminato: 7

the purification was followed by recording UV traces at two different wavelengths: 254 nm to detect nucleic acids and 280 nm to detect proteins. From the trace it is clear the presence of nucleic acids in flow through on the basis of the UV absorbance at 254 nm. The eluted peak on the contrary gave a UV profile typical of a protein at 280 nm. Peak fractions were checked by SDS PAGE, which confirmed the complete purification of La105-202 after the Heparin column step. Fractions with the highest purity were then

Eliminato: (Figure 4.8)

4.2 Gel shift assays

4.2.1 The EMSA technique

The **EMSA** technique (**E**lectrophoretic **M**obility **S**hift **A**ssay) is a key methodology for studying protein/nucleic acid interactions. It is based on the observation that protein/nucleic acids complexes migrate more slowly than the free nucleic acid molecules in non-denaturing polyacrylamide or agarose gel electrophoresis (Hendrickson *et. al.* 1985; Revzin *et. al.* 1989). Since the rate of nucleic acid migration is ‘shifted’ or ‘retarded’ upon protein binding, the assay is also referred to as a gel shift or gel retardation assay.

Gel shift assays can be used qualitatively to identify sequence-specific DNA or RNA binding proteins, to identify the binding region of a protein and, in conjunction with mutagenesis, to identify the important binding sequences within the upstream regulatory region of a given gene. Furthermore, gel shift assays can also be utilised quantitatively to measure thermodynamic and kinetic parameters (Fried *et. al.* 1981, Garner *et. al.* 1981, Fried *et. al.* 1984 and Fried *et. al.* 1989).

An advantage of studying interactions by an electrophoretic assay is the ability to resolve complexes of different stoichiometry or conformation. The resolution depends largely upon the stability of the complex during the brief time (approximately one minute) it migrates into the gel. Sequence-specific interactions are transient and are stabilised by the relatively low ionic strength of the buffer in which the sample is loaded. Upon entry into the gel, protein complexes are quickly resolved from free nucleic acid, in effect freezing the equilibrium between bound and free acid that existed in the loading buffer. In the gel, the complex may be stabilised by “caging” effects of the gel matrix, meaning that if the complex dissociates its localised concentration remains high, promoting prompt re-association (Fried *et. al.* 1981; Fried *et. al.* 1984). Thus even relatively labile complexes can often be resolved by this method.

Since low concentrations of oligonucleotides are typically used in order to observe variations in the amounts of protein bound, it is usually necessary to label the nucleic acids to visualise the bands on the gels. Typically, the nucleic acid is radiolabeled with ^{32}P by 5′-end labelling using $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$ and T4 polynucleotide kinase.

Eliminato: Figure 4.8: SDS PAGE analysis of Heparin column fractions. Lane M is the protein ladder (Invitrogen). Each lane is labelled in accord with the column fractions of Figure 4.8. It is possible to notice the high purity of La105-202 after the Heparin purification step. ¶

Non-denaturing TBE-polyacrylamide gels or TAE-agarose gels are used to resolve the protein/nucleic acids complexes from the free nucleic acids. The gel percentage required depends on the size of the target nucleic acid and the size, number and charge of the protein(s) that bind to it. It is important that the complex enters the gel and does not remain in the bottom of the well. Polyacrylamide gels in the range of 4-8% are typically used, although it is not uncommon for higher percentage gels to be used with lower molecular weight systems. Agarose gels (0.7-1.2%) can be used to resolve very large complexes, as for *E. coli* RNA polymerase (~460 kDa). Gels are pre-run at a constant voltage until the current becomes constant. The primary reasons for pre-electrophoresis is to remove all traces of ammonium persulfate (used to polymerise polyacrylamide gels), to distribute/equilibrate any special stabilising factors or ions added to the electrophoresis buffer and to ensure a constant gel temperature.

For RNA binding analysis of the La protein, here we performed gels shift in 9% polyacrylamide running in 0.5X TBE buffer at 125V for one hour at 4°C. To stabilise the running conditions, each gel was pre-run in the same buffer at 100V for one hour at 4°C.

All RNAs used were radiolabelled at the 5'-end with [γ -³²P]-ATP. The introduction of the radiolabelled nucleotide was performed using T4 kinase enzyme, which allowed the reaction between the hydroxyl group at the 3'-end of the radiolabelled ATP and the phosphate group at the 5'-end of the target RNAs. The reaction was stopped adding EDTA; the [γ -³²P]-ATP in excess was then removed using mini gel-filtration column G25 (Pharmacia) which retained free nucleotides and left the [γ -³²P]-RNA go to the flow trough. The total incorporation of γ -³²P was about 9%.

All gel shift assays were performed using a fixed concentration of target [γ -³²P]-RNAs (3.17 nM) and eight different concentrations (from 32 μ M to 0.014 μ M) of each protein. In particular, the protein samples were prepared as described below:

- 60 μ l of a stock solution 32 μ M of each purified protein were prepared using the following dilution buffer: 25 mM HEPES pH 7.25, 100 mM KCl, 50 μ g/ml BSA;
- from each stock solution, sequential dilution 1:3, with the dilution buffer reported above, were performed. At the end, they were obtained 40 μ l of protein

solution for each of the following concentrations: 32 μM , 10.67 μM , 3.56 μM , 1.18 μM , 0.39 μM , 0.13 μM , 0.04 μM and 0.014 μM .

To allow the binding protein/RNA, proteins were incubated with a binding solution containing the target [γ - ^{32}P]-RNA. After the incubation, Ficoll was added into each sample and 5 μl of each was then loaded into the native polyacrylamide gel.

To avoid diffusion problems which could occur during the impression period, the resolved gels were dried using Whatman 3MM chromatography paper.

4.2.2 Gel shift assays with 10 nt oligo(U)

The La protein associates predominantly with the short polyuridylylate sequence (UUU_{OH}) located at the 3' end of almost all nascent RNA polymerase III (pol III) transcripts (Wolin *et al.* 2002, Maraia *et al.* 2001). In particular, from a previous work conducted in Conte laboratories and other it is known that the determinants for poly(U) interaction are located in the N-terminal domain. La1-194 in fact exhibited poly(U) binding with the same affinity and specificity as the full-length protein (Jacks *et al.* 2003, Kenan *et al.* 1995, Maraia *et al.* 2001). Furthermore, the experiments performed previously showed that the deletion mutant La104-408, which lacks the La-motif, has a much weaker binding activity with respect the La full length and its interaction with RNA results to be not specific. The deletion mutant La105-334 also does not bind the 10 nt oligo (U) RNA (Jacks *et al.* 2003).

However an unresolved issue was to assess systematically the binding capability of the two separate domains within La1-194, therefore I performed gel mobility shift assays to test the interaction of both domains La1-103 and La105-202 with a 10 nt oligo(U) RNA.

As Figure 4.7 shows, neither of two isolated domains bind the 10 nt oligo (U) RNA at all. This result is consistent with previous data and suggests that the entire N-terminal domain 1-194 is required for poly(U) RNA interaction, and that the La motif and the Central-RRM probably work in synergy to bind to the 3' end (UUU_{OH})-sequences of the target RNAs.

Conte group therefore concentrated in understanding the molecular basis of La- poly(U) interaction. Whilst another PhD student in the laboratory, Domenico Sanfelice, which

Eliminato: 9

also provided the La1-103 sample for the gel shift assays, focussed on the structural characterisation of the La motif, I started determining the three dimensional structure of the Central RRM by NMR spectroscopy.

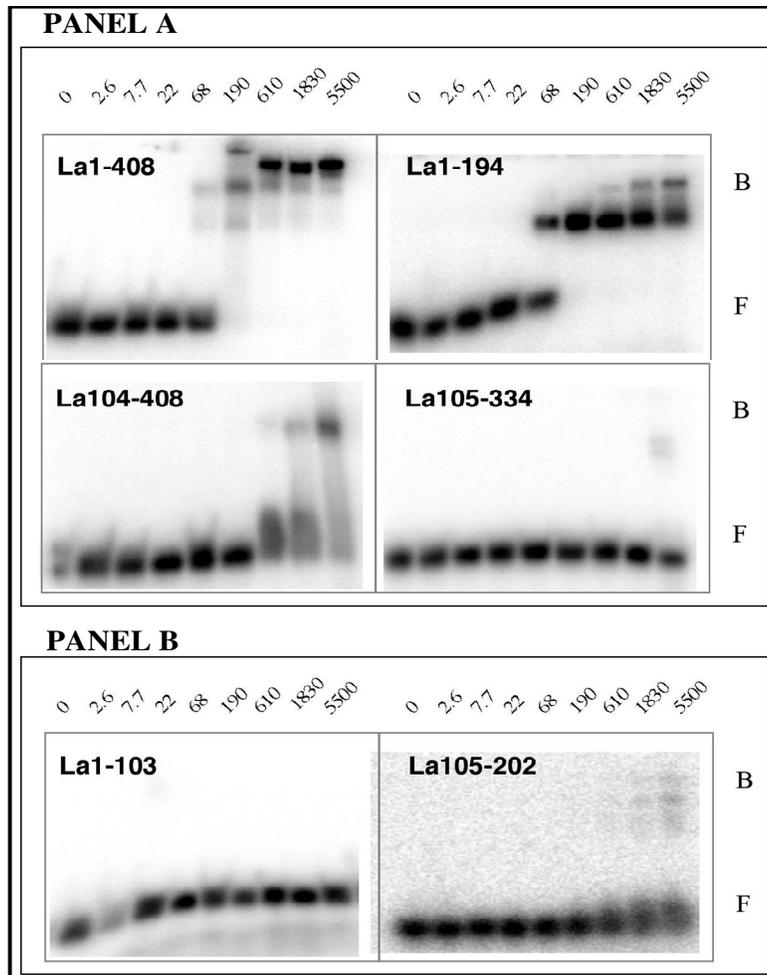


Figure 4.7: Gel mobility shift analysis of the binding affinity of La deletion mutants for 10nt oligo(U). Data shown in Panel A were published either previously (LaWT and La1-194, Jacks *et al.* 2003) or performed by other people in Conte laboratory (La104-408 and La105-334). They are here reported for comparison. In Panel B they are reported data obtained during the PhD work presented here. Numbers at the top of the figure indicate the protein concentrations (in nM) loaded in the corresponding lane. The RNA concentration was fixed at 3.17 nM. F indicates free RNA, while B indicates bound RNA. It is clear how both the La-motif (La1-103) and the Central-RRM (La105-202) are necessary for the RNA binding, even if they are not able to bind the target RNA when are isolated.

Eliminato: 9

4.3 Structural characterization of La protein

Although the gel shift analysis of La protein gave us new information regards its binding properties, for a major understanding of its mechanisms of action, it was necessary to perform structural studies, which could related with the biological function of the protein.

The techniques that allow the determination of protein structures at atomic resolution are X-Ray crystallography and NMR spectroscopy. Whereas with X-Ray crystallography it is possible to study very large biological systems, even as large as the ribosome and complete viruses, NMR can study the structure and dynamics of biomolecules in solution and requires no crystallization, but is more limited in the maximum size of systems studied. Both methods have been used to study protein-protein or protein-nucleic acid complexes, and in many cases with complementary results. NMR is particularly suited to study biomolecular interactions in weak protein-protein or protein-nucleic acid complexes, since these systems could be perturbed by crystallization.

Our collaborator at Imperial College, Dr. Stephen Curry, tried to crystallise full length La and La1-194 without success. We therefore decided to employ NMR methodologies to solve the structure of the La protein. To test the suitability of the method, we acquired a [¹H-¹⁵N]-TROSY HSQC spectrum at 293 K of LaWT using a spectrometer operating at 16.1 T located at the N.I.M.R. Biomedical NMR Centre in Mill Hill, London (Figure 4.8). Unfortunately, but expected, although the spectrum indicated a good folding of LaWT, the NMR signals were too much in overlap because the big size of the protein resulting in a not good applicability for the structure determination. Anyway, in these cases, it is a general approach in structural biology and in NMR in particular, to define the boundaries of the system under investigation in order to reduce the size. In particular, the strategy consists in solving the structure of individual domains and then trying to assemble the domains together to understand the structure of larger domains and/or complexes. This strategy had been adopted here. Conte had already solved the NMR structure of the C-terminal domain (La225-334, Jacks *et al.* 2003) and here it had been solved the structure of the Central RRM (La105-202, Alfano *et al.* 2004)). At the

Eliminato: 10

same time, in Conte laboratory had been solved the structure of the La motif (La1-103, Alfano *et al.* 2004).

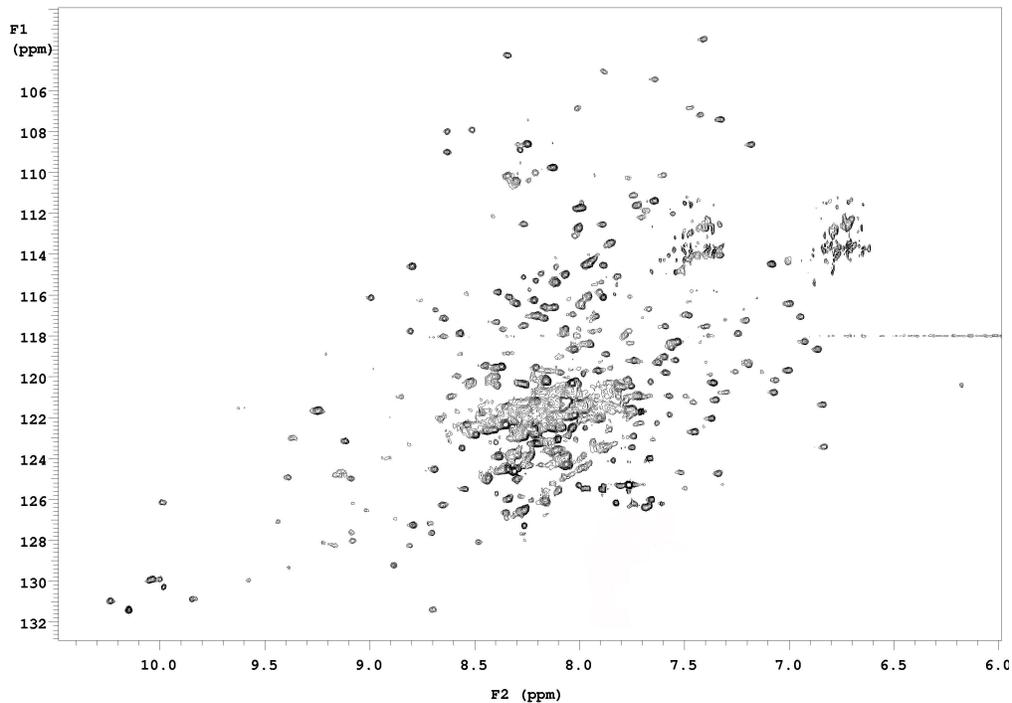


Figure 4.8: $[^1\text{H}-^{15}\text{N}]$ -TROSY HSQC spectrum of LaWT acquired at 293 K using a spectrometer operating at 16.1 Tesla. Although the spectrum shows a good folding of the protein, the NMR signals are too much in overlap because the big size of LaWT. It is then difficult to solve the structure of the protein by NMR.

Eliminato: 10

4.3.1 NMR in protein structure determination

The basic strategy for NMR structure determination can be divided in three steps:

- Sequence-specific assignment of backbone and side-chain resonances using a combination of experiments, which allow to delineate both intra-residue connectivities and through space ($< 5\text{\AA}$) sequential connectivities along the polypeptide chain.
- Identification of as many through space NOE connectivities as possible which yield a large set of approximate inter-proton distance restraints as well as determination of torsion angle restraints based on coupling constant data.

- Calculation of the three-dimensional structure on the basis of the accumulated inter-proton distances and torsion angle restrains.

Anyway, first of all it is necessary to test the suitability of the sample for NMR analysis. We then acquired an initial [^1H , ^{15}N]-Heteronuclear Single Quantum Coherence (HSQC) two dimensional spectrum (Bodenhausen *et al.* 1980) on La105-202 ^{15}N -labelled (Figure 4.9) using a VARIAN INOVA spectrometer operating at 14.1 Tesla located at the Biophysics Laboratories of the University of Portsmouth. As figure shows, the La105-202 ^{15}N -HSQC spectrum presented a good dispersion of the signals indicating that the sample was folded and suitable for the NMR studies, so we started the structure characterization of this La domain.

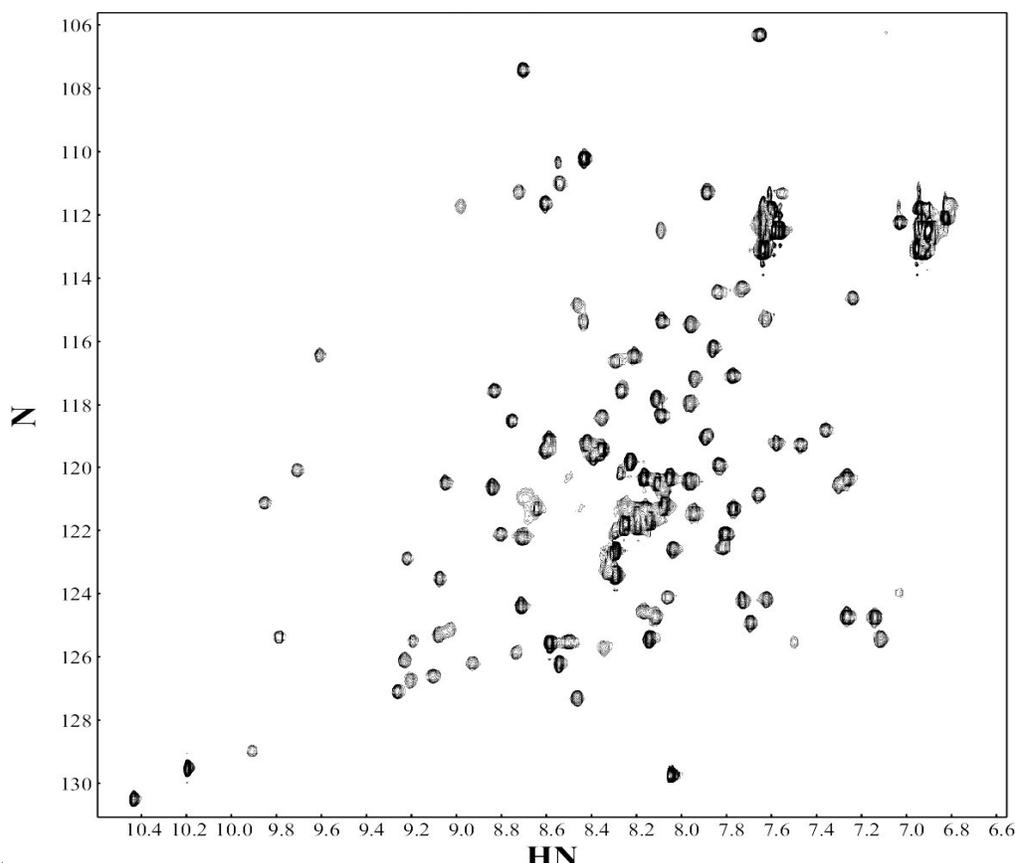


Figure 4.9: La105-202 [^1H , ^{15}N]- HSQC spectrum recorded at 14.1 T and 293 K. It is possible appreciate the dispersion of the signals, consequence of the folding of the protein.

Eliminato: 11

Formattato: Inglese (Regno Unito)

Eliminato: ¶

Eliminato: 11

4.3.2 Sequence-specific assignment of La105-202

In order to assign all NMR resonances of La105-202 atoms, which allow the extraction of all available information about inter-atomic distances and torsion angles, the following three-dimensional NMR spectra were acquired:

- 2D hetero-nuclear experiments: [^1H - ^{15}N]-HSQC and [^1H - ^{13}C]-HSQC
- 3D hetero-nuclear double resonance experiments: ^{15}N -edited and ^{13}C -edited NOESY-HSQC and ^{15}N -TOCSY-HSQC.
- 3D hetero-nuclear triple resonance experiments: HNCA, CBCA(CO)NH and HNCO

The explanation of this kind of experiments and the strategy of analysis applied are reported in Chapter III. The global assignment performed is reported in Appendix III.

The analysis of spectra listed above allowed the identification and sequential assignment for 106 residues out of the 113 of La105-202 (i.e. not including the N-terminal His-tag residues). In particular it had been possible to assign 98% of H^{N} , 97% of ^{15}N , 100% of $^{13}\text{C}\alpha$, 96% of $^{13}\text{C}\beta$, 93% of ^{13}CO .

The chemical shift values obtained from sequential assignment were used for the secondary structure determination of La105-202, which was performed by the Chemical Shift Index (CSI) method (Wishart *et al.*, 1991, Wishart *et al.* 1994). By the way, before starting the CSI analysis, it had been necessary to assign the $\text{H}\alpha$ chemical shifts. In order to do this ^{15}N -NOESY-HSQC and ^{15}N -TOCSY-HSQC three-dimensional NMR spectra were analysed, which allowed the assignment of 98% of the $\text{H}\alpha$ resonances.

4.3.3 Secondary structure analysis of La105-202

The chemical shift values obtained from sequential assignment were used for the secondary structure determination of La105-202, which was performed by the Chemical Shift Index (CSI) method (Wishart *et al.*, 1991, Wishart *et al.* 1994). This method uses the observed $^1\text{H}\alpha$, $^1\text{H}^{\text{N}}$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO and the backbone ^{15}N chemical shifts to determine regions with secondary structure. This is possible because these chemical shifts have different values in regular secondary structure elements with respect the random coil values. In particular, residues involved in a α -helix, with respect to the

random coil, present: (i) an increase of the H α and $^{13}\text{C}\beta$ chemical shifts; (ii) a decrease of the C α and CO chemical shifts. Residues involved in a β -strand rather present: (i) a decrease of the H α and $^{13}\text{C}\beta$ chemical shifts; (ii) an increase of the C α and CO chemical shifts. It is shown that the H $^{\text{N}}$ and ^{15}N chemical shifts are also influenced by the secondary structure, but the correlation is less pronounced with respect to the other nuclei. Furthermore, if the chemical shifts of the nuclei reported above not present any change with respect the random coil values, the corresponding residue is considered to be in a loop. Following this strategy, the La105-202 secondary structure was obtained. In particular, the chemical shift deviation of each residue from random coil was obtained by subtracting the experimental values from the random coil chemical shifts and then deriving the weighted consensus value $\Sigma\Delta\delta$ using the following formula:

$$\Sigma\Delta\delta = (4\Delta\delta\text{H}\alpha + \Delta\delta\text{C}\beta - \Delta\delta\text{C}\alpha - \Delta\delta\text{CO}) / \text{number of assignments}$$

The chemical shift values for random coil were downloaded from the database BioMagResBank (<http://bmr.b.wisc.edu>).

When at least four consecutive residues presented positive values of $\Sigma\Delta\delta$, the stretch was predicted as an α -helix, while if the $\Sigma\Delta\delta$ values were negative, the stretch was predicted as a β -strand. All other regions were designated as random coil.

At the end, five possible β -strands and three α -helices were predicted in the La105-202. The secondary structure prediction of La105-202 based on the CSI method is shown in

Figure 4.10.

Eliminato: 2

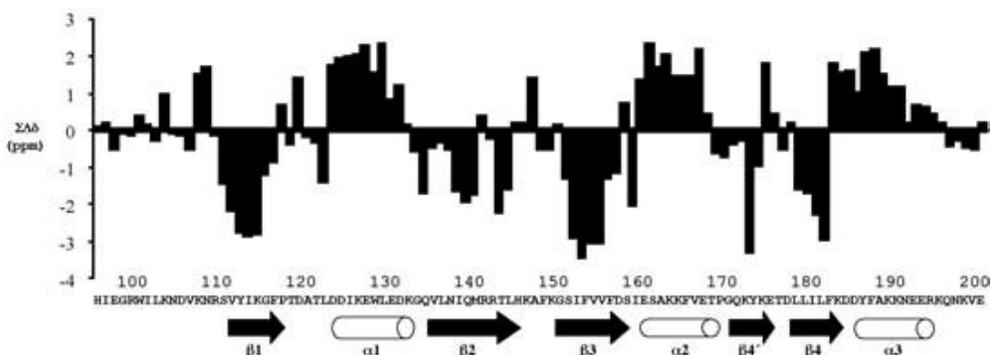


Figure 4.10: Plot of the chemical shift deviation from random coil for La105-202. The deviation was obtained by subtracting the experimental values from the random coil chemical shifts and then deriving a weighted consensus value ($\Sigma\Delta\delta$) using $(4\Delta\delta\text{H}\alpha + \Delta\delta\text{C}\beta - \Delta\delta\text{C}\alpha - \Delta\delta\text{CO}) / \text{number of assignments}$. In the bottom of the figure, the secondary structure elements are indicated.

Eliminato: 2

In parallel with the results obtained with the CSI method, the ϕ and ψ dihedral angles were calculated using the software TALOS (Cornilescu *et al.* 1999). This tool combines the chemical shifts information of $^1\text{H}\alpha$, $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO , ^{15}N with sequence information derived from a database of high-resolution NMR and X-Ray structures. In practise, TALOS analyses triplets of sequential residues of the protein under studies and then compares them with those contained in its database of protein structures. In this way, TALOS evaluates the similarity in amino acid sequence and chemical shifts of these strings. The output contains the ϕ and ψ backbone angles of the central residue of each triplet calculated as the average of those extracted from the database structures, which show the highest degree of similarity. The secondary structure prediction of La105-202 obtained using TALOS is shown in Figure 4.11. The data obtained are in agreement with those obtained with the CSI method.

Eliminato: 3

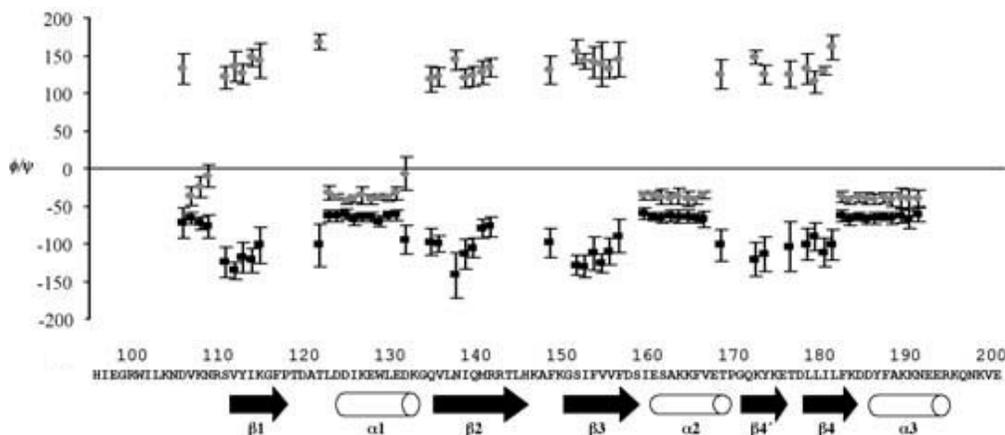


Figure 4.11: Talos derived ϕ/ψ predictions for La105-202. ϕ and ψ values are denoted by black circles and grey squares respectively, for residues where 9 or 10 database matches agree. In the bottom of the figure, the secondary structure elements are indicated.

Eliminato: 3

From both CSI and TALOS analysis, it is clear that La105-202 shows a non-canonical RRM module, with an extra β -strand ($\beta 4'$) and α -helix ($\alpha 3$) in addition to the archetypal $\beta\alpha\beta\beta\alpha\beta$ RRM fold. Surprisingly, a remarkably similar topology has already been found in the C-terminal RRM (RRM3) of hLa protein (Jacks *et al.* 2002).

4.3.4 Restraints analysis for structure calculation of La105-202

Once performed the sequence-specific assignment of La105-202 and analysed its secondary structure elements, the next step was to collect as many restraints as possible in order to calculate the tertiary structure of the protein. In particular, the following restraints were analysed and then used for the three-dimensional structure calculation:

- NOE derived restraints
- Hydrogen bond restraints
- Dihedral angles restraints
- Residual Dipolar Coupling restraints

First of all, the side chain resonances were assigned in order to determinate the NOE derived restraints. This assignment was performed using both HCCH-TOCSY and [¹⁵N]-edited TOCSY-HSQC spectra.

Side chains assignment: the HCCH-TOCSY spectrum

For large proteins it is not possible to assign the side chain resonances using ¹⁵N-NOESY-HSQC and ¹⁵N-TOCSY-HSQC. In fact, for proteins generally bigger than 10 kDa, the proton line widths are often too large and hinder efficient coherence transfers involving the relatively small ¹H-¹H J coupling, making the homonuclear or ¹⁵N experiments very inefficient. Two 3D HCCH-TOCSY experiments were introduced, one aimed at the assignment of carbon resonances and one used for assigning proton resonances of the side-chains of ¹³C- or (¹⁵N,¹³C)-labelled proteins (Bax *et al.* 1990, Oleiniczak *et al.* 1992). The HCCH-TOCSY experiments rely on magnetization transfer between ¹³C nuclei by isotropic mixing. Since the coherence transfer can be carried out in shorter periods due to the larger ¹³C-¹³C J coupling (>35Hz), the experiments are well applicable to larger molecules. Eventually, magnetization is transferred to protons using

the large heteronuclear ^1H - ^{13}C J-coupling. Therefore, the limiting factor of the HCCH J-correlation methods is the ^{13}C line width.

The flow of magnetization transfer during the (H)CCH-TOCSY experiment is given by:



The analysis of the HCCH-TOCSY spectra allowed the assignment of ^{13}C and ^1H atoms of the side chains. In this spectrum, the diagonal peak of each carbon plane corresponds to the direct attached proton, while the cross-peaks correspond to the other side chain protons. It is then clear that the peak patterns identify the amino acid type. However, in some cases, the carbon planes appeared to be crowded, and then it was necessary the combined analysis of both HCCH-TOCSY and [^{15}N]-edited TOCSY-HSQC spectra.

Eliminato: Side chains assignment: the HCCH-TOCSY analysis

NOE derived restraints

Once performed the assignment of the side chains resonances, we were able to go ahead with the identification of NOE derived restraints, which gave us information regard the proton-proton distances. In order to do this, we estimated the signal intensities of both 3D [^{15}N]-NOESY-HSQC and 3D [^{13}C]-NOESY-HSQC spectra. It is known in fact that the NOE effect between two nuclei is related to their distance by the equation:

$$\text{NOE}_{ij} \sim 1/r_{ij}^6$$

To calibrate the magnitude of NOEs it had been necessary to perform a previous analysis of the spectra based on known NOE/distance correlation. In the 3D [^{15}N]-NOESY-HSQC spectrum, this was possible because some residue strings presented strong NOEs between the H^{N} diagonal peak of residue i and the $\text{H}\alpha$ cross peak of residue $i-1$. Those kinds of NOEs are typical of a β -strand and were used to calibrate the shorter permitted distance between two atoms, which is 2.8 Å. In case of the 3D [^{13}C]-NOESY-HSQC spectrum, the calibration was instead performed using the NOEs observed for geminal $\text{H}\alpha$ glycine residues. On the based of these calibrations and of the own intensity, each NOE was assigned as strong (which corresponds a distance between

2.8 and 3.5 Å), medium (which corresponds a distance between 3.5 and 5.0 Å) and weak (which corresponds a distance upper 5.0 Å).

Exchange H^N analysis for H-bond restraints analysis

The exchange H^N analysis is very important for NMR structure calculation for two reasons: (i) is indicative of what residues are exposed to the solvent and what are inside the structure; (ii) is indicative of what residues are involved in hydrogen bonds.

In order to perform this analysis, three [1H - ^{15}N] HSQC spectra of La105-202 dissolved in D_2O were recorded at different time: immediately after the dissolution in D_2O , after 1 hour and after 12 hours. The H^N protons exchanged with the D_2O were not visible anymore on the HSQC spectra. By this analysis, all H^N were classified as: (i) fast exchanging, which where not visible immediately after the dissolution; (ii) slowly exchanging, which where still visible after 1 hour but not after 12 hours; (iii) very slowly exchanging, which where still visible after 12 hours.

We found 56 very slowly exchanging H^N all involved in hydrogen bonds as confirmed by the NOEs analysis. All these H^N were in fact of residues for which NOE patterns characteristic of secondary structures were observed.

Eliminato: ¶

Dihedral angles restraints

The dihedral angles φ and ψ calculated with the TALOS software (see above), and used for the secondary structure analysis, were also used as restraints for the ternary structure calculation.

For the first run of structure calculation, we considered only values scored 9 or 10 by TALOS, which corresponded to a high similarity degree with TALOS database. We then added values scored 8 and evaluated the restraint violations obtained from XPLOR (see below). At the final structure calculation were included 144 dihedral angles restraints.

Formattato: Rientro: Prima riga: 0 cm, Interlinea 1,5 righe

Eliminato: ¶

Residual Dipolar Coupling

Residue Dipolar Couplings (RDCs) have recently emerged as a new tool in NMR to study macromolecular structure in solution and are complementary to NOEs to provide structural information (Tolman *et al.* 1995, Tjiandra *et al.* 1997). In contrast to NOEs, which provide distance restraints between two atoms, RDCs contain distance information as well as angles formed by a vector connecting the two atoms within a tensor axis system. In fact, the r^{-6} dependence of NOEs means that NOEs can be observed only between protons which are 5 Å distant, giving short-range and local structure information. The dipolar coupling is instead defined in terms of a molecular coordinate frame and then provides information on how each dipole is related to the molecular coordinate frame and in turn to each other. As a result of the Brownian motion effects, dipolar couplings average to zero under isotropic conditions and can be observed only under anisotropic conditions. This is the reason because the RDCs are applied only recently to systems in solution, while they have been applied in solid-state NMR for long.

To obtain RDCs in solution, a co-solute is necessary to cause a partial alignment of each dipole without causing severe coupling interactions and distorted spectra. In order to obtain an anisotropic condition, the protein under study is “dissolved” in an opportune media, which mimic liquid crystalline properties, and then an NMR spectrum is acquired and compared with that acquired in isotropic conditions.

For La105-202 structure calculation, we measured $^1D_{NH}$ residual dipolar couplings. The ^{15}N labelled La105-202 sample was diluted in a ternary complex composed of ~5% (v/v) alkyl-poly(ethylene glycol) C8E5, ~0.8% (v/v) n-octanol and 20 mM sodium phosphate pH 6.0, 100 mM KCl (Rückert *et al.* 2000). A J_{NH} -modulated-HSQC spectrum was acquired at 293 K and compared with that acquired in isotropic conditions. It had been possible to define an N-H RDC values only for 21 amino acid residues. The initial values of magnitude and rhombicity of the alignment tensor were estimated -17.25 Hz and 0.25 respectively (Warren *et al.* 2001). These values together with the 21 RDCs values were introduced in XPLOR protocols for structure calculation.

Eliminato: ¶

4.3.5 Structural calculation of La105-202

All restraints obtained as described above were used to calculate the three-dimensional structure of La105-202 using the XPLOR software (Brünger *et al.* 1993). This computer package starts from the amino acid sequence, which is given as input file, in a random coil conformation. It then starts to fold the protein structure basing on all experimental restraints inserted as input files and on some empirical data of its database, which contain information about bond lengths, bond angles, van der Waals radii, etc.

It is essential to have a large number of experimental restraints to avoid the formation of wrong conformations due to the free rotation around chemical bonds.

We started the structural calculation introducing in the XPLOR protocols only unambiguous restraints, which allowed a first converged fold of La105-202. This first fold was then used to discriminate all other ambiguous restraints and those with a reasonable agreement with the obtained fold were used for a second run of structure calculation. We then started an iteration calculation monitoring at each run all restraints violations. The final run calculation included 292 short range NOEs, 140 long range, 144 dihedral angles, 56 H-bonds and 21 RDC restraints. From 100 calculated structures, a family of 20 lowest energy structures was selected and reported in Figure 4.14. The summary of structure statistic is shown in Table 4.1.

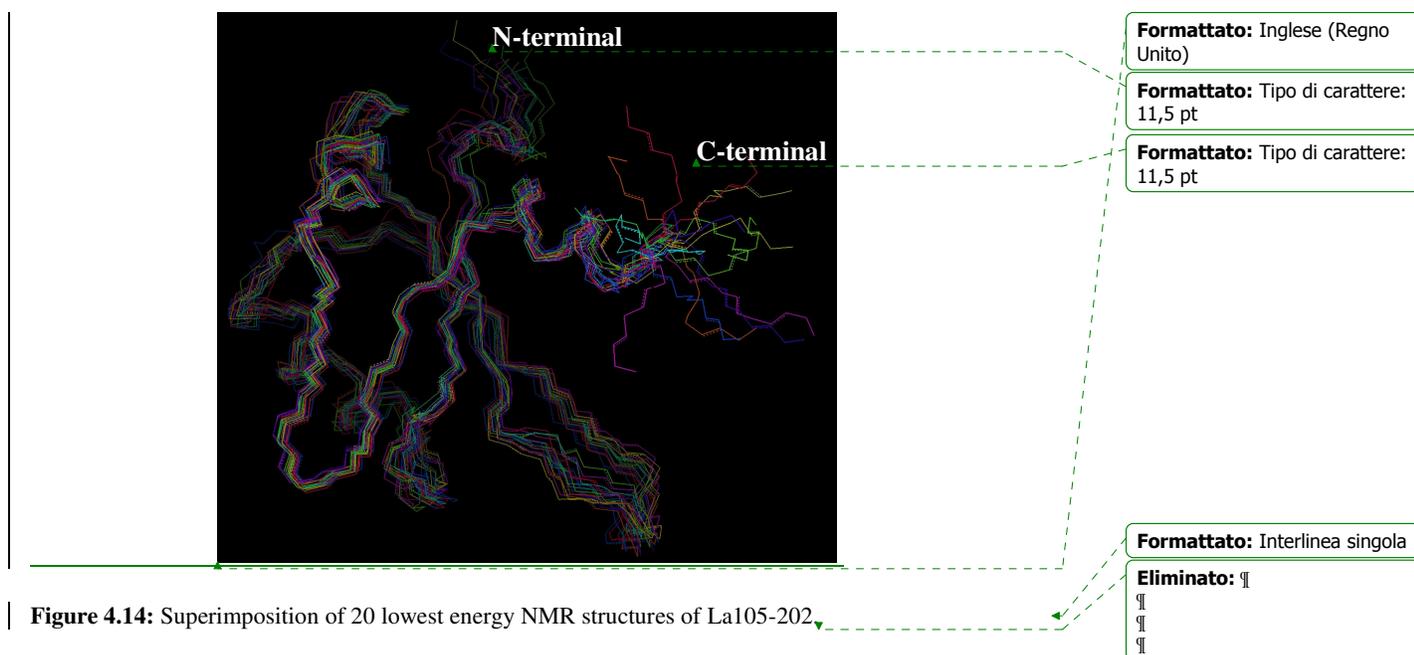


Table 4.1 Structural statistics summary for La105-202.

Total distance restraints (inter-residue)	
Short-medium range (residue i to $i + j$, $j = 1-4$)	292
Long range (residue i to $i + j$, $j > 4$)	140
Hydrogen bond	56
Total dihedral angle restraints	144
ϕ	73
ψ	71
NH residual dipolar coupling restraints	21
Restraints violations	
Distance restraint violation $> 0.3 \text{ \AA}$	none
Dihedral restraint violation $> 5^\circ$	none
NH residual dipolar coupling restraint violation $> 2 \text{ Hz}$	none
Average r.m.s.d. (\AA) among the 20 refined structures	
Backbone of structured regions ^a	0.55
Heavy atoms of structured regions	1.09
Backbone of all residues ^b	0.69
Heavy atoms of all residues	1.21
Ramachandran statistics of 20 structures ^c	
Percentage residues in	
most favoured regions	82.0%
additional allowed regions	13.9%
generously allowed regions	2.5 %
disallowed regions	1.6 %

^aResidues selected on the basis of ^{15}N backbone dynamics: 108–116, 121–134, 138–142, 154–170, 174–192.

^bResidues 108–192

^cBy PROCHECK analysis (Laskowski *et al.* 1996)

Ternary structure of La105-202

As final result of structure calculation, we obtained the three-dimensional structure of La105-202. This domain adopts a classical RRM-type fold with a topology $\beta\alpha\beta\beta\alpha\beta$ in disagreement with the secondary structure prediction (see above). This in fact gave us an extra β -strand ($\beta 4'$: residues 173-176) with respect to the canonical RRM fold. From the three-dimensional structure it is clear that the hypothetical $\beta 4'$ -strand adopts an extended structure because it lacks inter-strand interactions. Furthermore, an extra C-terminal helix is given from the secondary structure prediction. This helix ($\alpha 3$; residues 185-194), although is an unusual extension of the canonical RRM, it is well defined in the three dimensional structure. It is results in fact stabilised by hydrophobic contacts between some of its side-chains (Tyr188, Lys185, Asp186 and Asp187) and residues Tyr114, Leu183 and Phe184 on the surface of the β -sheet. The structure reveals also a short 3_{10} helical turn at the N-terminal end ($\alpha 0$; residues 108-111).

The β -sheet surface appears to be largely available for RNA interactions, in contrast with that observed in the C-terminal RRM of human La (Jacks *et al.* 2003). In the case of La105-202, in fact, the helix 3 is predominantly hydrophilic and run away from the body of the domain and towards the solvent. Of more relevance to RNA interaction is that loop $\beta 2$ - $\beta 3$ contains five basic residues which may be capable of electrostatic interactions with the sugar-phosphate backbone of RNA. Moreover in contrast to the C-terminal RRM, the β -sheet surface of the central RRM is entirely devoid of acidic residues.

4.3.6 Chemical shift mapping of the hLa N-terminal domain

To investigate which residues of the N-terminal domain of human La are involved in the RNA binding, chemical shifts mapping experiment were performed using a ^{15}N -labelled La1-194 sample. The aim of the experiment was to monitor the backbone amide chemical shift changes, using ^1H - ^{15}N HSQC NMR experiments, in La1-194 upon titration of 10 nt oligo(U) RNA. The results indicate that a substantial subset of the assigned resonances in both the La motif and the central RRM are shifted upon addition of 10 nt oligo(U) (Figure 4.15).

Eliminato: and an extra C-terminal helix ($\alpha 3$; residues 185-194)

Eliminato: .

Formattato: SpazioDopo: 15 pt, Regola lo spazio tra testo asiatico e in alfabeto latino, Regola lo spazio tra caratteri asiatici e numeri

Formattato: Tipo di carattere: Corsivo

Formattati: Elenchi puntati e numerati

Formattato: Giustificato, Interlinea 1,5 righe

Formattato: Apice

Eliminato: ¶

Formattato: Tipo di carattere: Times New Roman

Formattato: Tipo di carattere: 12 pt, Inglese (Regno Unito)

Formattato: Tipo di carattere: 12 pt, Inglese (Regno Unito)

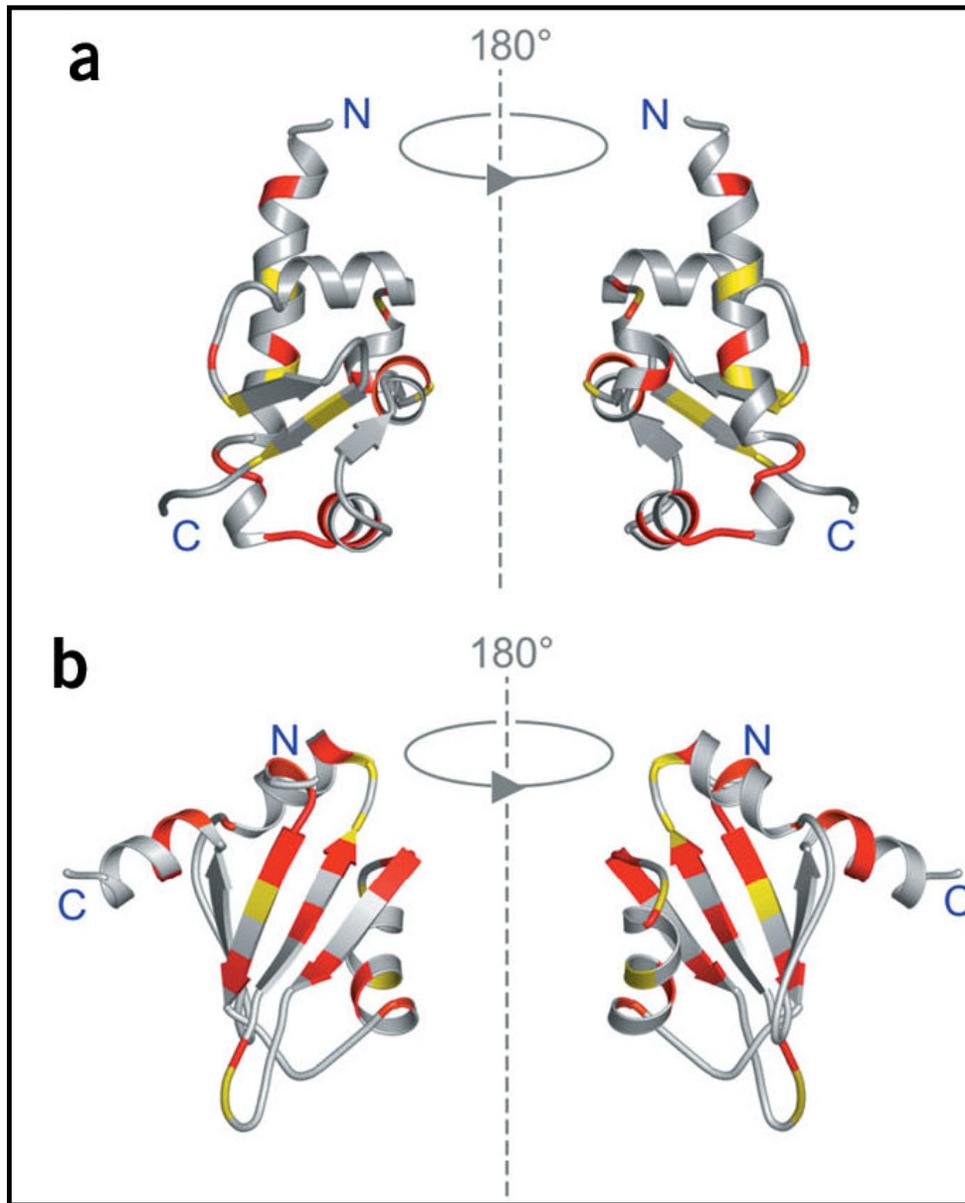


Figure 4.15: Structure mapping of the chemical shift variations obtained on RNA binding for the La motif (a) and the central RRM (b). The positions of residues that show substantial shifts are indicated on the protein secondary structures. Magnitudes of chemical shift changes are color-coded as follows: yellow, $0.05 < \Delta\delta_{AV} < 0.07$; orange, $0.07 < \Delta\delta_{AV} < 0.1$; red $\Delta\delta_{AV} > 0.1$

For the central RRM, as expected for this type of module, the residues affected by addition of RNA are mainly concentrated in the central β -sheet, the β_2 - β_3 loop, the C-

Formattato: Interlinea 1,5 righe

Formattato: Tipo di carattere: 10 pt, Grassetto

Formattato: Tipo di carattere: 10 pt

Formattato: Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)

Formattato: Tipo di carattere: 10 pt

Formattato: Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)

Formattato: Giustificato

Formattato: Tipo di carattere: 10 pt

Formattato: Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)

Formattato: Tipo di carattere: 10 pt

Formattato: Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)

Formattato: Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)

Formattato: Tipo di carattere: 10 pt

Formattato ... [1]

Formattato ... [2]

Formattato ... [3]

Formattato ... [4]

Formattato ... [5]

Formattato ... [6]

Formattato ... [7]

Formattato ... [8]

Formattato ... [9]

Formattato ... [10]

Formattato ... [11]

Formattato ... [12]

Formattato ... [13]

Formattato ... [14]

Formattato ... [15]

Formattato ... [16]

terminal α helix, and the N-terminal 3_{10} helix (Figure 4.15). Few isolated resonance shifts are located on loops α_1 - β_2 and β_3 - α_2 and on the inner surface of helices α_1 and α_2 . The RNA binding occurs then across the β -sheet surface, in agreement with previously studied RRM-RNA complexes.

On the La motif the largest chemical shifts changes associated with RNA binding map to helices α_1 , α_1' , α_2 , α_3 and α_4 and are clustered into two surface patches on one edge of the molecule. However it is striking that the structural elements (α_1' , α_2 and α_4) that are insertions in the La motif relative to the archetypal winged-helix motif are all implicated in RNA binding. This suggests that the interaction of the La motif with RNA may represent a new mode of binding for winged-helix proteins.

The presence of large areas of perturbation on both domains is consistent with the notion that the domains work together as a functional unit to bind RNA. Synergy may arise from the RNA lying in a linear fashion across a contiguous binding surface formed by the two modules, somewhat akin to the interaction observed between the polyA binding protein and its cognate RNA. Alternatively RNA binding could promote a closer association of the La motif and the RRM to form a binding pocket in which the binding surfaces from each domain contact opposite sides of the RNA. This latter model is favoured by the observation that a shorter 5 nt oligo(U) molecule which is unlikely to be able to span the full length of an open binding sites, induces a very similar pattern of chemical shift perturbations to that seen with the 10-mer oligo(U) (data not shown). However, further structural work is required to test this hypothesis.

4.4 *Binding of La with the Hepatitis C Virus*

Formattato	... [17]
Formattato	... [18]
Formattato	... [19]
Formattato	... [20]
Formattato	... [21]
Formattato	... [22]
Formattato	... [23]
Formattato	... [24]
Formattato	... [25]
Formattato	... [26]
Formattato	... [27]
Formattato	... [28]
Formattato	... [29]
Formattato	... [30]
Formattato	... [31]
Formattato	... [32]
Formattato	... [33]
Formattato	... [34]
Formattato	... [35]
Formattato	... [36]
Formattato	... [37]
Formattato	... [38]
Formattato	... [39]
Formattato	... [40]
Formattato	... [41]
Formattato	... [42]
Formattato	... [43]
Formattato	... [44]
Formattato	... [45]
Formattato	... [46]
Formattato	... [47]
Formattato	... [48]
Formattato	... [49]
Formattato	... [50]
Formattato	... [51]
Formattato	... [52]
Formattato	... [53]
Formattato	... [54]
Formattato	... [55]
Formattato	... [56]
Formattato	... [57]
Eliminato: ¶	
Eliminato: ¶	
Formattato	... [58]
Formattato	... [59]
Formattati: Elenchi puntati e numerati	... [60]

In addition to the binding with poly(U)-rich elements in RNA polymerase III transcripts, it has been shown that human La protein binds also with some viral mRNAs which not contain any uridil residue at the 3'-end and acts as a stimulator of viral translation (Maraia *et al.* 2001, Wolin *et al.* 2002) {more specific references}. In particular it has been shown that La is involved in binding the Hepatitis C Virus (HCV) IRES in the vicinity of the initiator AUG contributing to start the site selection (Ali *et al.* 1997, Ali *et al.* 2000, Pudi *et al.* 2003). Although several studies have tried to elucidate how La binds viral RNAs, its function and mode of actions are still unclear. Several investigations have probably been hampered by the use of incorrect deletion mutants of human La protein, i.e. incomplete domains which could be either unfolded or misfolded (Ali *et al.* 2000, Pudi *et al.* 2003).

. The IRES: Internal Ribosome Entry and Translation

In eukaryotic cells the translation of RNA messages involves recognition of a 7-methyl guanosine cap on the 5'-end of the mRNA by a set of initiation factors that assemble the 48S preinitiation complex (Figure 4.16) (Merrick *et al.* 1996, Pain *et al.* 1996, Sachs *et al.* 1997).

Eliminato: 33

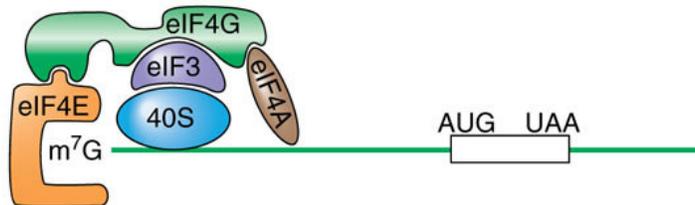


Figure 4.16: Schematic of the strategy used by eukaryotes to recruit the 43S particle to the mRNA start codon: the mRNA 7-methyl guanosine cap binds eIF4E which is in complex with eIF4G. The latter leads to 43S particle binding through the eIF3 component.

Eliminato: 3

Eliminato: 3

Alternatively, numerous viruses, such as Hepatitis C Virus, and some eukaryotic mRNAs utilise a cap-independent pathway in which an RNA element, the Internal Ribosome Entry Site (IRES, Figure 4.17), drives the complex formation by positioning the ribosome on the mRNA, either at or just upstream of the start site (Rijnbrand *et al.* 2000).

Eliminato: 34

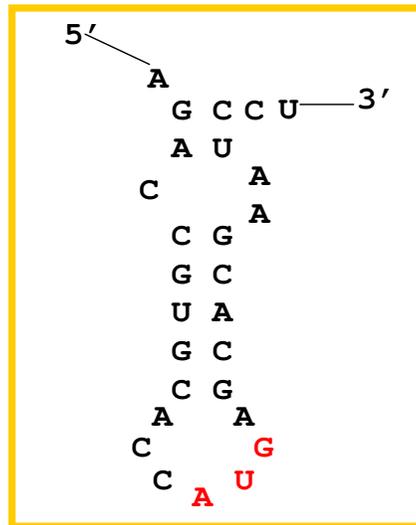


Figure 4.18: The 27 mer RNA used in our studies to investigate the binding affinity of several deletion mutants of La to Hepatitis C Virus. The chosen sequence is part of the pseudoknot region of the HCV IRES containing the AUG starting codon (in red). For a view of this region in the entire HCV IRES see Figure 4.17.

Eliminato: 35

Eliminato: 34

The deletion mutants of La used in this analysis were: LaWT (1-408), La1-344, La1-334, La105-334, La1-194, La225-334 and La105-202. Gel shift assays were performed both without and with an RNA competitor (yeast tRNA) in order to understand if the binding of La to the HCV RNA was specific or not.

As Figure 4.19 shows, from our studies resulted that La full length, La1-344, La1-334, and La1-194 bind the target viral RNA, even if an RNA competitor was present, while La105-202 presented only a weak not specific binding and La225-334 did not bind at all. All these data are in accord with those obtained with the 10 nt oligo (U) RNA, but with the difference that in the case of viral RNA probably two different complexes protein/RNA were possible since increasing the protein concentration a more shifted band was obtained. A main difference was anyway that in the case of viral RNA, La105-334 bound to the 27 mer HCV IRES even if an RNA competitors was present, while did not present any binding affinity for the poly(U) RNA. This could be the first evidence of previous hypothesis for which the C-terminal domain of La protein could be involved in RNA recognition of large structured cellular and viral mRNA (Fan *et al.* 1998, Ali *et al.* 2000, Pudi *et al.* 2003, Intine *et al.* 2000 and 2003).

Eliminato: 36

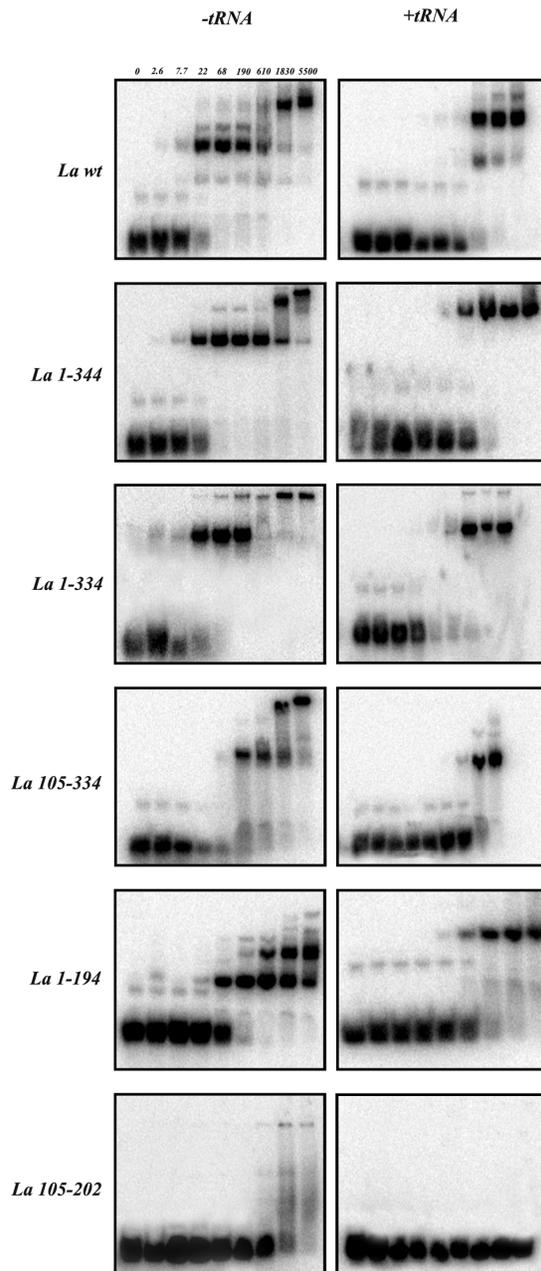


Figure 4.19: Gel mobility shift analysis of the binding affinity of La deletion mutants for 27 mer HCV IRES RNA. Numbers at the top of the Figure indicate the protein concentrations (in nM) loaded in the corresponding lane. The RNA concentration was fixed at 3.17 nM. F indicates free RNA, while B indicates bound RNA. Panel on the right is for gel shift assays performed without RNA competitor, while panel on the left is for that performed with RNA competitor.

Eliminato: 36

Pagina 84: [1] Formattato	lavinia	30/11/2006 8.01.00
Giustificato, Interlinea singola, Tabulazioni: 12,22 cm, Allineato a sinistra		
Pagina 84: [2] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt		
Pagina 84: [3] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)		
Pagina 84: [4] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt		
Pagina 84: [5] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)		
Pagina 84: [6] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt		
Pagina 84: [7] Formattato	lavinia	30/11/2006 8.01.00
Tipo di carattere: (Predefinito) Times New Roman, 10 pt, Inglese (Regno Unito)		
Pagina 84: [8] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 84: [9] Formattato	lavinia	30/11/2006 8.07.00
Giustificato, SpazioDopo: 0 pt, Interlinea 1,5 righe		
Pagina 84: [10] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 84: [11] Formattato	lavinia	30/11/2006 8.02.00
Tipo di carattere: 12 pt		
Pagina 84: [12] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 84: [13] Formattato	lavinia	30/11/2006 8.02.00
Tipo di carattere: 12 pt		
Pagina 84: [14] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 84: [15] Formattato	lavinia	30/11/2006 8.02.00
Tipo di carattere: 12 pt		
Pagina 84: [16] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [17] Formattato	lavinia	30/11/2006 8.02.00
Tipo di carattere: 12 pt		
Pagina 85: [18] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [19] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [20] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [21] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [22] Formattato	lavinia	30/11/2006 8.04.00
Tipo di carattere: 12 pt		
Pagina 85: [23] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [24] Formattato	lavinia	30/11/2006 8.04.00

Tipo di carattere: 12 pt

Pagina 85: [25] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [26] Formattato	lavinia	30/11/2006 8.04.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [27] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [28] Formattato	lavinia	30/11/2006 8.04.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [29] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [30] Formattato	lavinia	30/11/2006 8.04.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [31] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [32] Formattato	lavinia	30/11/2006 8.05.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [33] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [34] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [35] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [36] Formattato	lavinia	30/11/2006 8.05.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [37] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [38] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [39] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [40] Formattato	lavinia	30/11/2006 8.08.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [41] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [42] Formattato	lavinia	30/11/2006 8.08.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [43] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [44] Formattato	lavinia	30/11/2006 8.08.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [45] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [46] Formattato	lavinia	30/11/2006 8.08.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt

Pagina 85: [47] Formattato	lavinia	30/11/2006 7.36.00
-----------------------------------	----------------	---------------------------

Tipo di carattere: 12 pt, Inglese (Regno Unito)

Pagina 85: [48] Formattato	lavinia	30/11/2006 8.08.00
Tipo di carattere: 12 pt		
Pagina 85: [49] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [50] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [51] Formattato	lavinia	30/11/2006 8.08.00
Tipo di carattere: 12 pt		
Pagina 85: [52] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [53] Formattato	lavinia	30/11/2006 8.08.00
Tipo di carattere: 12 pt		
Pagina 85: [54] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [55] Formattato	lavinia	30/11/2006 8.08.00
Tipo di carattere: 12 pt		
Pagina 85: [56] Formattato	lavinia	30/11/2006 7.36.00
Tipo di carattere: 12 pt, Inglese (Regno Unito)		
Pagina 85: [57] Formattato	lavinia	30/11/2006 7.36.00
Giustificato, Interlinea 1,5 righe		
Pagina 85: [58] Formattato	lavinia	30/11/2006 9.26.00
Tipo di carattere: 16 pt		
Pagina 85: [59] Formattato	lavinia	30/11/2006 9.26.00
Centrato		
Pagina 85: [60] Cambia	lavinia	30/11/2006 5.55.00
Elenchi puntati e numerati formattati		

APPENDIX I

ANALYSIS OF DNA SEQUENCES BY COMPUTER PACKAGE

Antp32A_AfeI and Antp32B_AfeI strands

With respect to the original sequence used for the structure determination, the introduction of an AfeI restriction site (in red) meant changing the second base pair (in blue) and adding three more base pairs (in green):



Analysis of hairpin formation in both strands at 37°C is shown in Figures 100a and 100b.

Antp32A_BfrBI and Antp32B_BfrBI strands

With respect the original sequence used for the structure determination, the introduction of BfrBI restriction site (in red) meant changing the first two and the next to last base pairs (in blue) and adding one more base pair in each end (in green):



The hairpins formation analysis of both strands at 37°C is shown Figure 101a and 101b.

Antp36A_BsrBI and Antp36B_BsrBI strands

With respect the original sequence used for the structure determination, the introduction of BsrBI restriction site (in red) meant changing the first and the last base pairs (in blue) and adding two more base pairs in each end (in green):



The hairpins formation analysis of both strands at 37°C is shown Figures 102a and 102b.

Antp30A_BstUI and Antp30B_BstUI strands

With respect the original sequence used for the structure determination, the introduction of BstUI restriction site (in red) meant changing the next to last base pair (in blue) and adding one more base pair in the left end (in green):



The hairpins formation analysis of both strands at 37°C is shown Figures 103a and 103b.

Antp32A_FspI and Antp32B_FspI strands

With respect the original sequence used for the structure determination, the introduction of FspI restriction site (in red) meant changing the next to last base pair (in blue) and adding one more base pair between the first and the second base pairs and one more base pairs in the right end (in green):



The hairpins formation analysis of both strands at 37°C is shown in Figures 104a and 104b.

Antp28A_HaeIII and Antp28B_HaeIII strands

With respect the original sequence used for the structure determination, the introduction of HaeIII restriction site (in red) meant changing the first two base pairs and the next to last base pair (in blue):



The hairpins formation analysis of both strands at 37°C is shown in Figures 105a and 105b.

Antp28A_HpyCH4V and Antp28B_HpyCH4Vstrands

With respect the original sequence used for the structure determination, the introduction of HpyCH4V restriction site (in red) meant changing the first and the next to last base pairs (in blue):

5' -CAAAGCCATTAGTGC AAAGCCATTAGTG-3'
 3' -GTTTCGGTAATCACGTTTCGGTAATCAC-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 106a and 106b.

Antp32A_MscI and Antp32B_MscI strands

With respect the original sequence used for the structure determination, the introduction of MscI restriction site (in red) meant changing the first and the next to last base pairs (in blue) and adding one more base pair in each end (in green):

5' -CAAAGCCATTAGTGGCAAAGCCATTAGTGG-3'
 3' -GTTTCGGTAATCACCGTTTCGGTAATCAC-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 107a and 107b.

Antp32A_NruI and Antp32B_NruI strands

With respect the original sequence used for the structure determination, the introduction of NruI restriction site (in red) meant changing the last two base pairs (in blue) and adding one more base pair in each end (in green):

5' -CGAAAGCCATTAGTCGCGAAAGCCATTAGTCG-3'
 3' -GTTTCGGTAATCACGCTTTCGGTAATCACG-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 108a and 108b.

Antp32A_RsaI and Antp32B_RsaI strands

With respect the original sequence used for the structure determination, the introduction of RsaI restriction site (in red) meant changing the first base pair (in blue) and adding one more base pair in each end (in green):

5' -CAAAGCCATTAGAGTACAAAGCCATTAGAGT-3'
 3' -TGTTCGGTAATCTCATGTTTCGGTAATCTCA-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 109a and 109b.

Antp32A_ScaI and Antp32B_ScaI strands

With respect the original sequence used for the structure determination, the introduction of ScaI restriction site (in red) meant changing the first two base pairs (in blue) and adding one more base pair in each end (in green):

5' -**ACT**AAGCCATTAG**AGTACT**AAGCCATTAG**AGT**-3'
3' -**TGAT**TCGGTAAT**TCATGA**TCGGTAAT**CTCA**-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 110a and 110b.

Antp32A_SmaI and Antp32B_SmaI strands

With respect the original sequence used for the structure determination, the introduction of SmaI restriction site (in red) meant changing the second base pair and the last two base pairs (in blue) and adding one more base pair in each end (in green):

5' -**GGG**AAGCCATTAG**CCCGG**AAGCCATTAG**CCC**-3'
3' -**CC**TCGGTAAT**CGG**CC**CTTC**GGTAAT**CGG**-5'

The hairpins formation analysis of both strands at 37°C is shown in Figures 111a and 111b.

Antp32A_Afe I: 5'-GCTAAGCCATTAGAGCGCTAAGCCATTAGAGC-3'

$dG = -3.94$ $dH = -74.4$ $dS = -227.3$ $T_m = 54.2$

Structural element	dG	Information
External loop	-1.64	2 ss bases & 2 closing helices.
Stack	-2.24	External closing pair is G ¹⁵ -C ³²
Helix	-2.24	2 base pairs.
Interior loop	-0.14	External closing pair is C ¹⁶ -G ³¹
Stack	-1.28	External closing pair is C ¹⁸ -G ²⁹
Stack	-0.58	External closing pair is T ¹⁹ -A ²⁸
Stack	-1.00	External closing pair is A ²⁰ -T ²⁷
Helix	-2.86	4 base pairs.
Hairpin loop	2.90	Closing pair is A ²¹ -T ²⁶
Stack	-1.28	External closing pair is C ² -G ¹³
Stack	-0.58	External closing pair is T ³ -A ¹²
Stack	-1.00	External closing pair is A ⁴ -T ¹¹
Helix	-2.86	4 base pairs.
Hairpin loop	2.90	Closing pair is A ⁵ -T ¹⁰

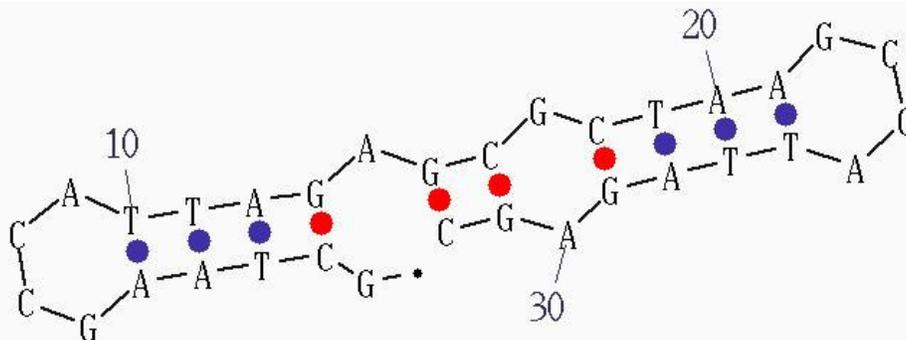


Figure 100a: Hairpins analysis by computer package of oligo Antp32A_AfeI . In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32A_BfrBI: 5'-CATAAGCCATTAGATGCATAAGCCATTAGATG-3'

dG = -0.7 dH = -20.1 dS = -62.6 T_m = 48.2

Structural element	ΔG	Information
External loop	-0.92	24 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁶ -C ²³
Helix	-2.24	2 base pairs.
Hairpin loop	2.50	Closing pair is C ¹⁷ -G ²²

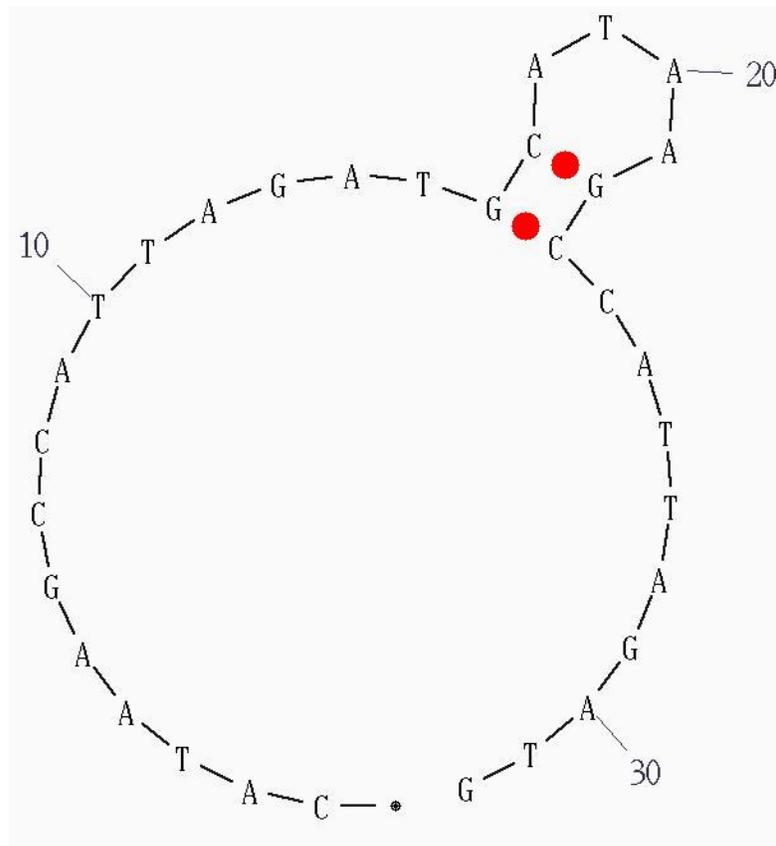


Figure 101a: Hairpins analysis by computer package of oligo Antp32A_BfrBI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32B_BfrBI: 5'-CATCTAATGGCTTATGCATCTAATGGCTTATG-3'

dG = -1.02 dH = -24.9 dS = -77.1 T_m = 50.0

Structural element	ΔG	Information
External loop	-1.38	24 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁰ -C ¹⁷
Helix	-2.24	2 base pairs.
Hairpin loop	2.60	Closing pair is C ¹¹ -G ¹⁶

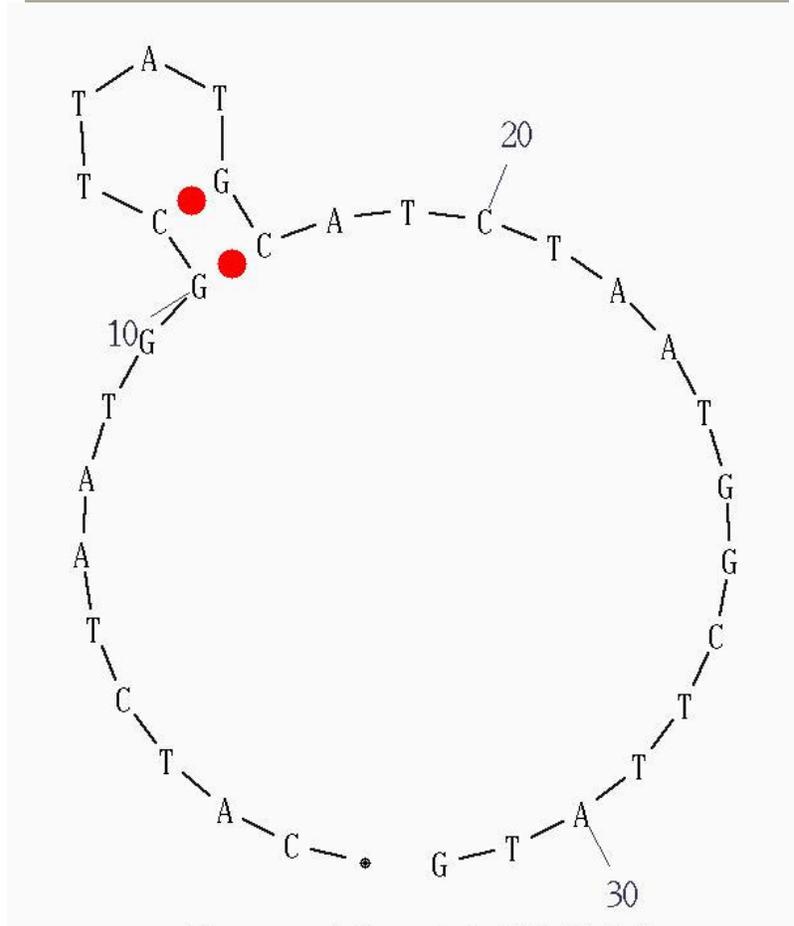


Figure 101b: Hairpins analysis by computer package of oligo Antp32B_BfrBI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp36A_BsrBI: 5'-CTCAAAGCCATTAGACCGCTCAAAGCCATTAGACCG-3'

dG = -0.62 dH = -22.0 dS = -69.0 T_m = 45.7

Structural element	$\Delta\Delta G$	Information
External loop	-0.65	27 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁸ -C ²⁶
Stack	-1.28	External closing pair is C ¹⁹ -G ²⁵
Helix	-3.52	3 base pairs.
Hairpin loop	3.55	Closing pair is T ²⁰ -A ²⁴

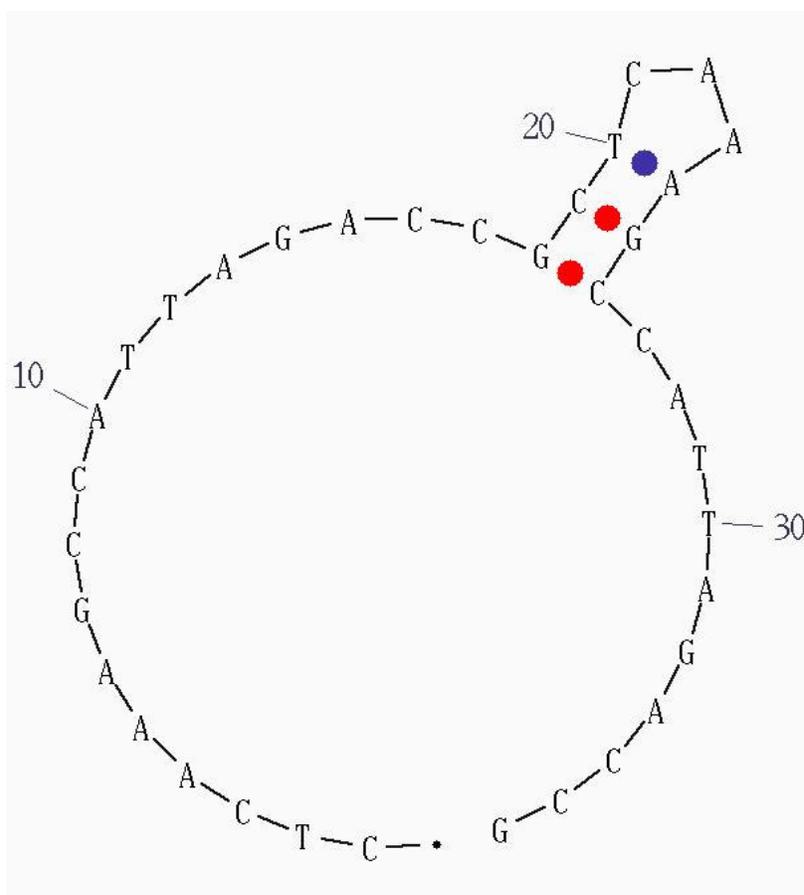


Figure 102a: Hairpins analysis by computer package of oligo Antp32A_BsrBI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp36B_BsrBI: 5'-CGGTCTAATGGCTTTGAGCGGTCTAATGGCTTTGAG-3'

$$dG = -2.39 \quad dH = -47.9 \quad dS = -146.7 \quad T_m = 53.4$$

Structural element	ΔG	Information
External loop	-0.44	16 ss bases & 1 closing helices.
Stack	-2.17	External closing pair is C ¹ -G ²⁰
Helix	-2.17	2 base pairs.
Interior loop	-1.24	External closing pair is G ² -C ¹⁹
Stack	-1.30	External closing pair is T ⁴ -A ¹⁷
Helix	-1.30	2 base pairs.
Interior loop	0.56	External closing pair is C ⁵ -G ¹⁶
Stack	-1.00	External closing pair is A ⁷ -T ¹⁴
Helix	-1.00	2 base pairs.
Hairpin loop	3.20	Closing pair is A ⁸ -T ¹³

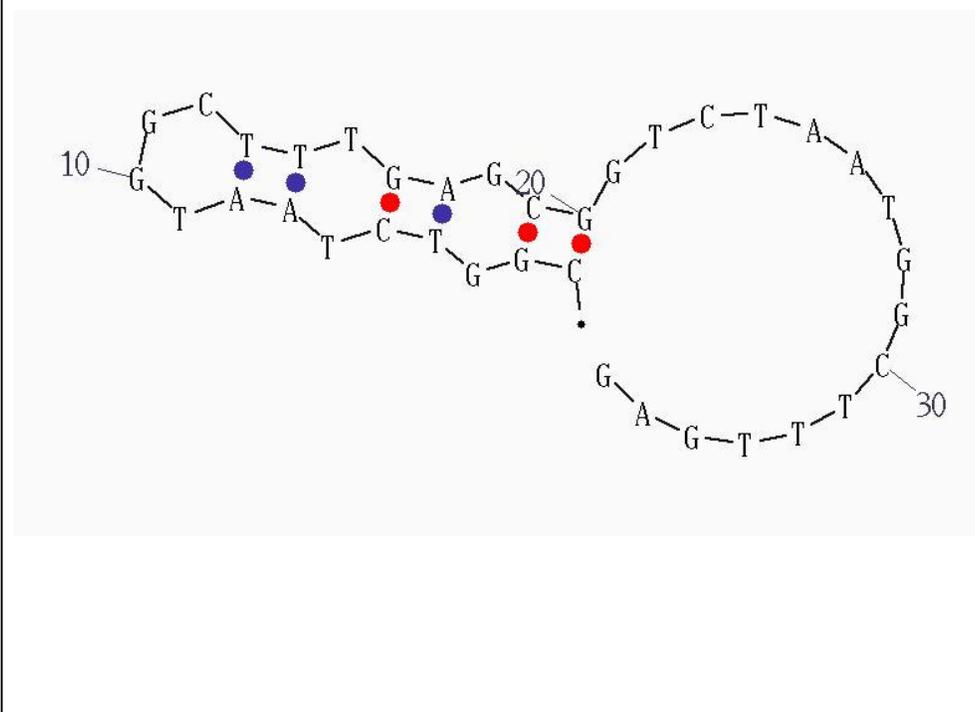


Figure 102b: Hairpins analysis by computer package of oligo Antp32B_BsrBI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp30A_BstU I: 5'-CGAAAGCCATTAGCGCGAAAGCCATTAGCG-3'

dG = -3.0 dH = -39.9 dS = -119.0 T_m = 62.2

Structural element	$\Delta\Delta G$	Information
External loop	-0.72	13 ss bases & 1 closing helices.
Stack	-2.17	External closing pair is C ¹⁴ -G ³⁰
Stack	-2.24	External closing pair is G ¹⁵ -C ²⁹
Helix	-4.41	3 base pairs.
Interior loop	0.53	External closing pair is C ¹⁶ -G ²⁸
Stack	-1.00	External closing pair is A ¹⁸ -T ²⁶
Helix	-1.00	2 base pairs.
Hairpin loop	2.60	Closing pair is A ¹⁹ -T ²⁵

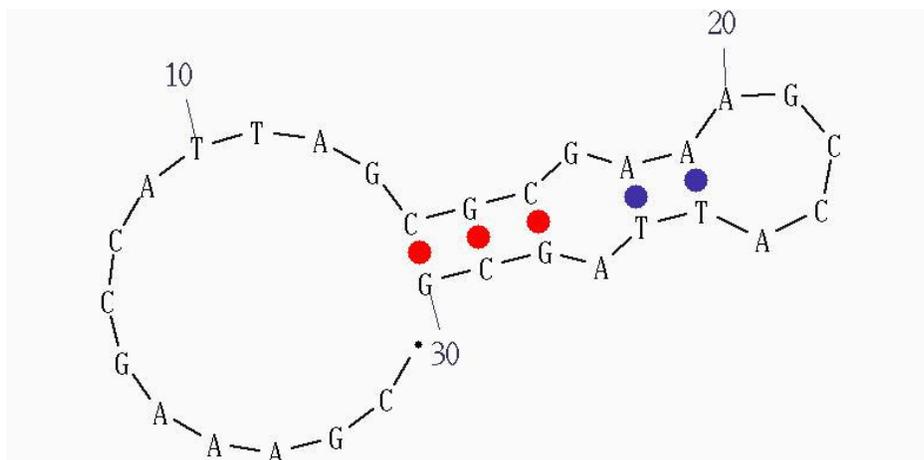


Figure 103a: Hairpins analysis by computer package of oligo Antp32A_BstUI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp30B_BstU I: 5'-CGCTAATGGCTTTCGCGCTAATGGCTTTCG-3'

dG = -1.40 dH = -49.8 dS = -156.1 T_m = 46.0

Structural element	ΔG	Information
External loop	-1.72	12 ss bases & 2 closing helices.
Stack	-2.24	External closing pair is G ¹⁷ -C ²⁵
Helix	-2.24	2 base pairs.
Hairpin loop	2.40	Closing pair is C ¹⁸ -G ²⁴
Stack	-2.24	External closing pair is G ² -C ¹⁰
Helix	-2.24	2 base pairs.
Hairpin loop	2.40	Closing pair is C ³ -G ⁹

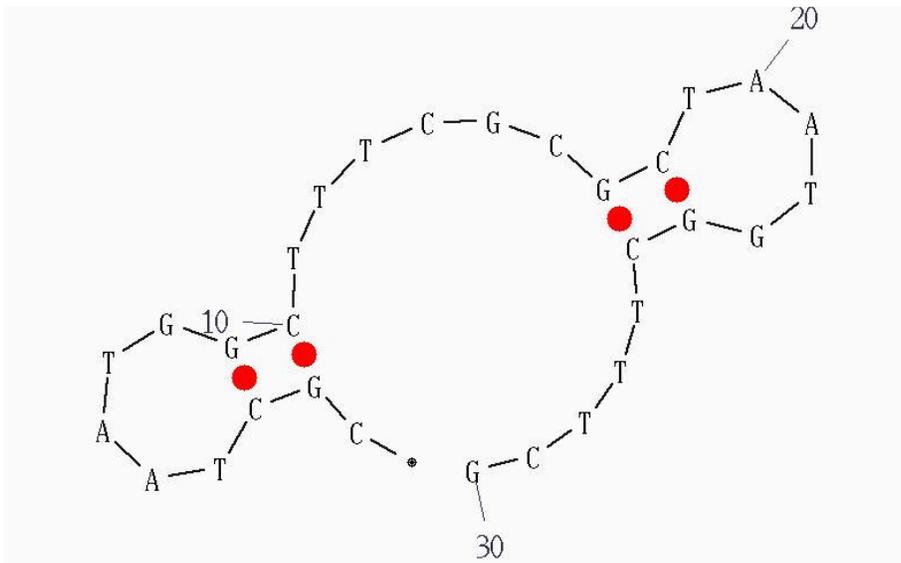


Figure 103b: Hairpins analysis by computer package of oligo Antp32B_BstUI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32A_Fsp I: 5'GCAAAGCCATTAGTGC GCAAAGCCATTAGTGC-3'

dG = -0.89 dH = -30.0 dS = -93.8 T_m = 46.6

Structural element	ΔG	Information
External loop	-0.61	14 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁵ -C ³²
Stack	-0.47	External closing pair is C ¹⁶ -G ³¹
Stack	-0.59	External closing pair is G ¹⁷ -T ³⁰
Helix	-3.30	4 base pairs.
Interior loop	1.12	External closing pair is C ¹⁸ -G ²⁹
Stack	-1.00	External closing pair is A ²⁰ -T ²⁷
Helix	-1.00	2 base pairs.
Hairpin loop	2.90	Closing pair is A ²¹ -T ²⁶

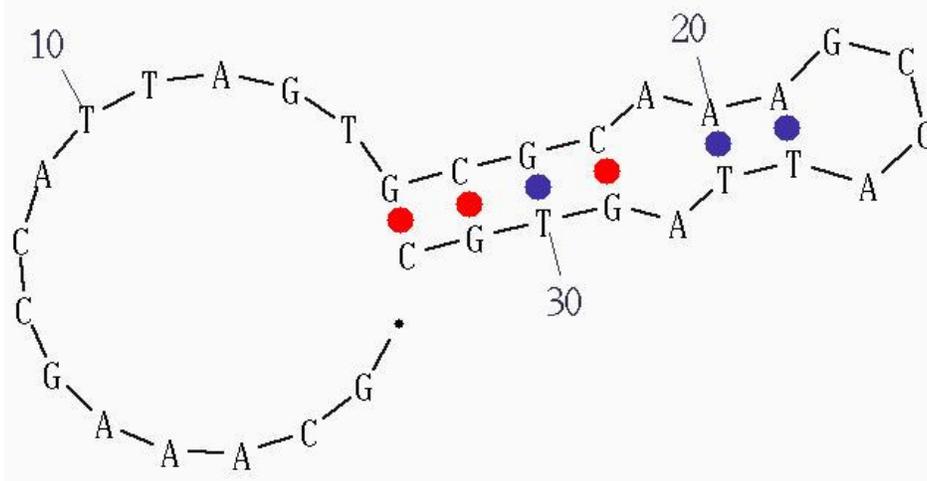


Figure 104a: Hairpins analysis by computer package of oligo Antp32A_FspI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32B_Fsp I: 5'-GCACTAATGGCTTTGCGCACTAATGGCTTTGC-3'

dG = -0.92 dH = -22.7 dS = -70.3 T_m = 49.8

Structural element	ΔG	Information
External loop	-1.38	23 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁰ -C ¹⁸
Helix	-2.24	2 base pairs.
Hairpin loop	2.70	Closing pair is C ¹¹ -G ¹⁷

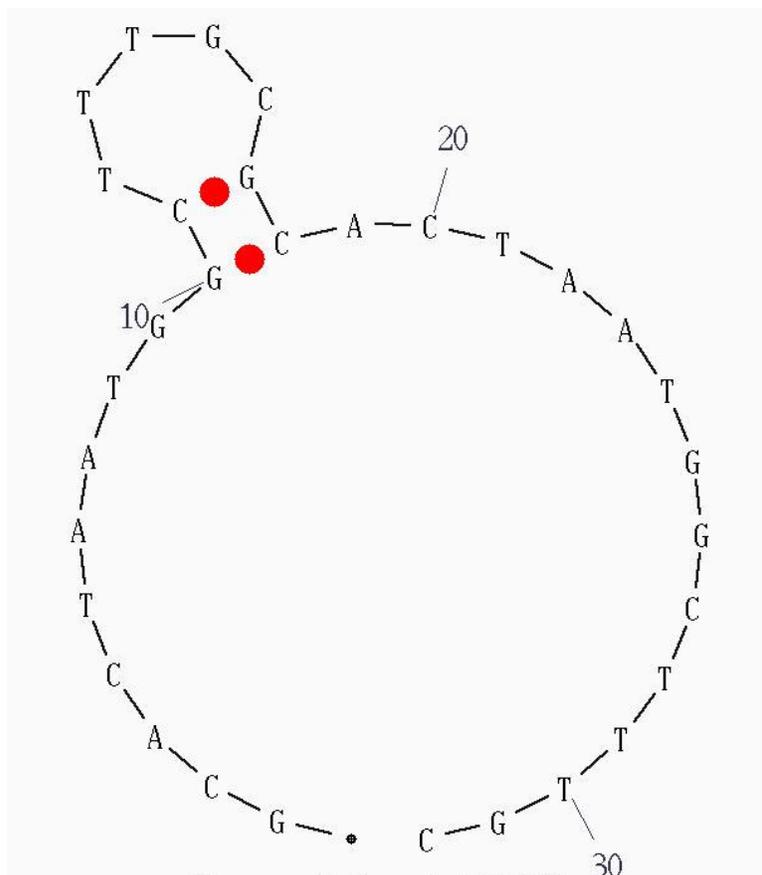


Figure 104b: Hairpins analysis by computer package of oligo Antp32B_FspI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp28A_Hae III: 5'-CCAAGCCATTAGGGCCAAGCCATTAGGG-3'

$dG = -3.07$ $dH = -46.9$ $dS = -141.2$ $T_m = 59.0$

Structural element	ΔG	Information
External loop	-2.05	9 ss bases & 2 closing helices.
Stack	-1.84	External closing pair is C ²⁰ -G ²⁷
Helix	-1.84	2 base pairs.
Hairpin loop	2.50	Closing pair is C ²¹ -G ²⁶
Stack	-2.24	External closing pair is G ⁵ -C ¹⁵
Stack	-1.84	External closing pair is C ⁶ -G ¹⁴
Helix	-4.08	3 base pairs.
Hairpin loop	2.40	Closing pair is C ⁷ -G ¹³

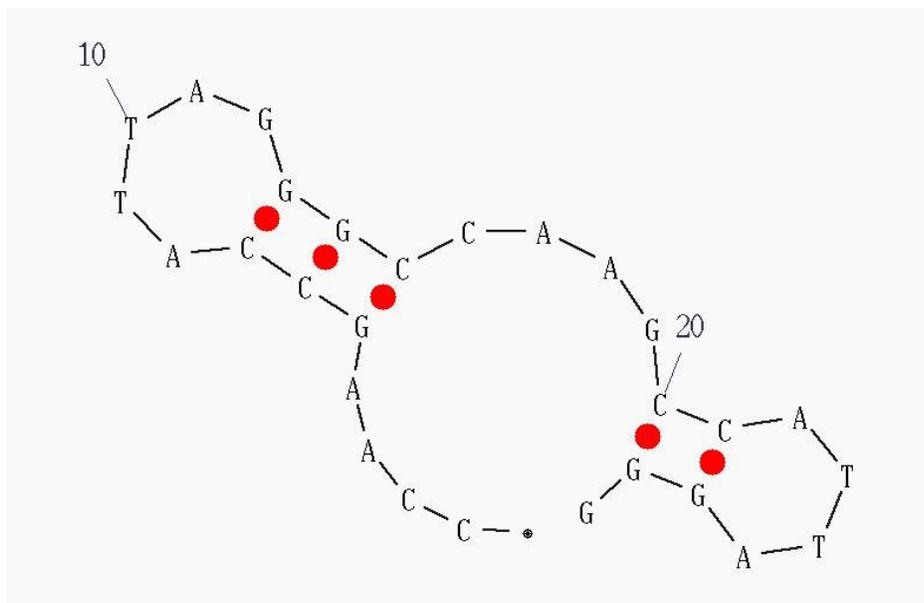


Figure 105a: Hairpins analysis by computer package of oligo Antp28A_HaeIII at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp28B_Hae III: 5'-CCCTAATGGCTTGGCCCTAATGGCTTGG-3'

$\Delta G = -2.56$ $\Delta H = -29.3$ $\Delta S = -86.1$ $T_m = 67.2$

Structural element	ΔG	Information
External loop	-1.08	17 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁴ -C ²⁴
Stack	-1.84	External closing pair is C ¹⁵ -G ²³
Helix	-4.08	3 base pairs.
Hairpin loop	2.60	Closing pair is C ¹⁶ -G ²²

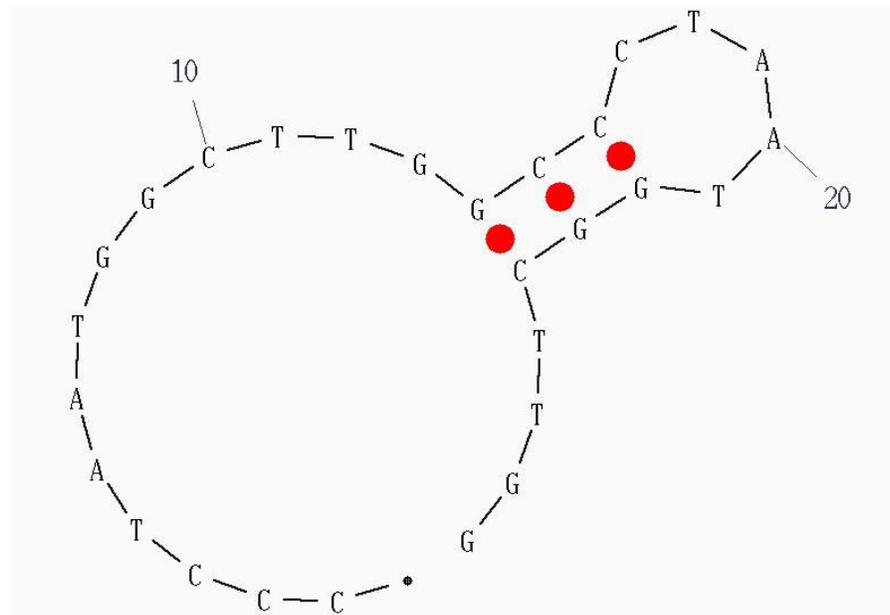


Figure 105b: Hairpins analysis by computer package of oligo Antp28B_HaeIII at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp28A_HpyCH4 V: 5'-CAAAGCCATTAGTGCAAAGCCATTAGTG-3'

$dG = -0.04$ $dH = -21.8$ $dS = -70.3$ $T_m = 37.0$

Structural element	ΔG	Information
External loop	-1.38	21 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ⁹ -C ¹⁵
Helix	-2.24	2 base pairs.
Hairpin loop	3.50	Closing pair is C ¹⁰ -G ¹⁴

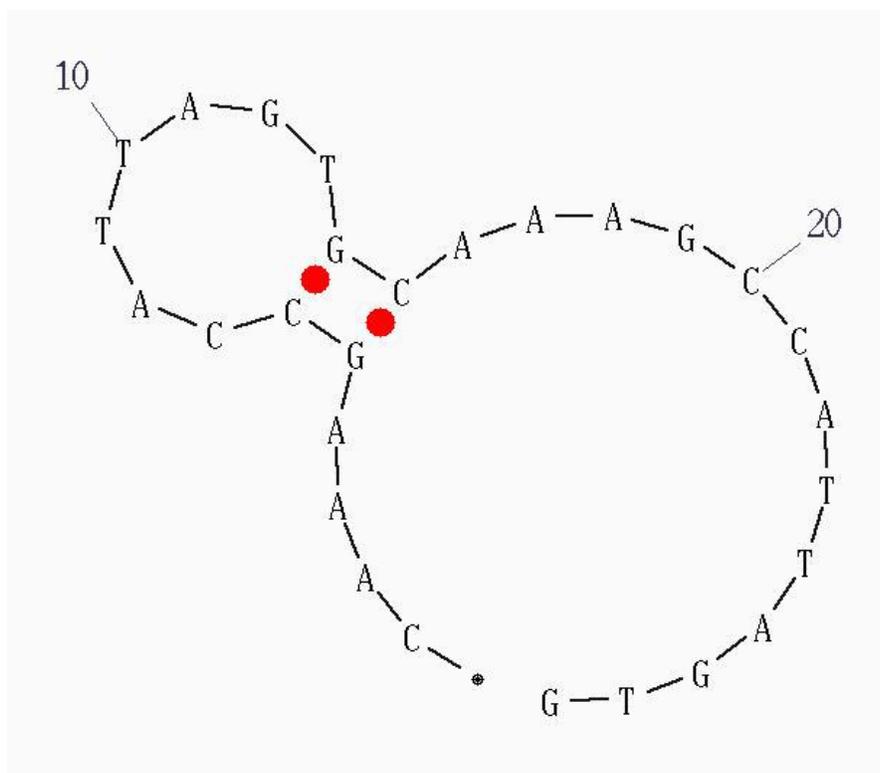


Figure 106a: Hairpins analysis by computer package of oligo Antp28A_HpyCH4V at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp28B_HpyCH4V: 5'-CACTAATGGCTTTGCACTAATGGCTTTG-3'

dG = -0.12 dH = -19.6 dS = -62.9 T_m = 38.6

Structural element	ΔG	Information
External loop	-1.40	17 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ⁵ -C ¹⁵
Helix	-2.24	2 base pairs.
Hairpin loop	3.60	Closing pair is C ⁶ -G ¹⁴

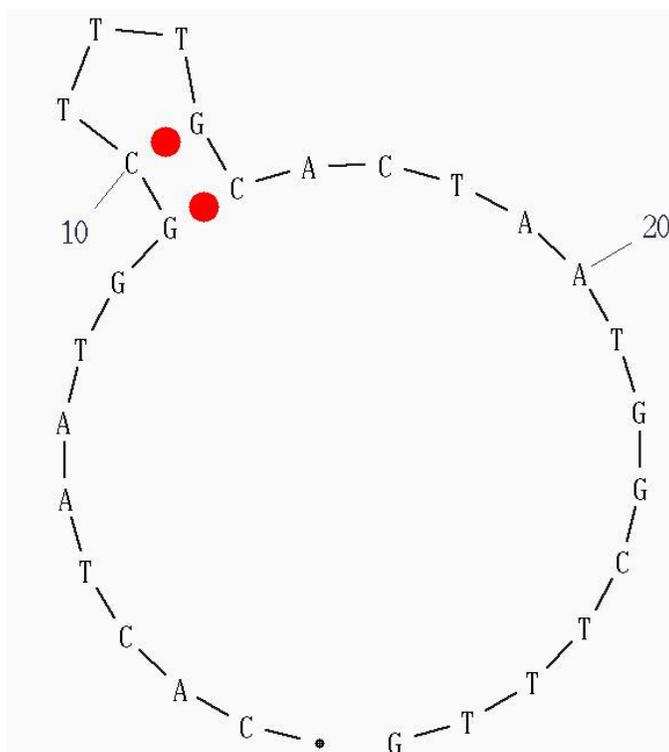


Figure 106b: Hairpins analysis by computer package of oligo Antp28B_HpyCH4V at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32B_MscI: 5'-CCACTAATGGCTTTGGCCACTAATGGCTTTGG-3'

dG = -4.33 dH = -56.5 dS = -168.3 T_m = 62.5

Structural element	ΔG	Information
External loop	-1.31	10 ss bases & 2 closing helices.
Stack	-2.24	External closing pair is G ¹⁶ -C ²⁷
Stack	-1.84	External closing pair is C ¹⁷ -G ²⁶
Stack	-1.45	External closing pair is C ¹⁸ -G ²⁵
Helix	-5.53	4 base pairs.
Hairpin loop	2.90	Closing pair is A ¹⁹ -T ²⁴
Stack	-1.84	External closing pair is C ¹ -G ¹⁰
Stack	-1.45	External closing pair is C ² -G ⁹
Helix	-3.29	3 base pairs.
Hairpin loop	2.90	Closing pair is A ³ -T ⁸

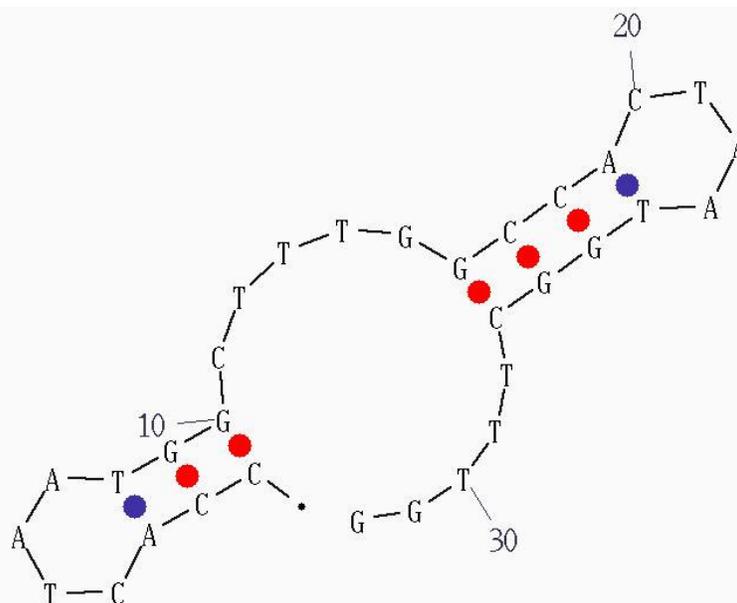


Figure 107a: Hairpins analysis by computer package of oligo Antp32B_MscI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32A_Msc I: 5'-CCAAAGCCATTAGTGGCCAAAGCCATTAGTGG-3'

dG = -5.86 dH = -60.5 dS = -176.0 T_m = 70.5

Structural element	ΔG	Information
External loop	-0.72	6 ss bases & 1 closing helices.
Stack	-1.84	External closing pair is C ⁷ -G ³²
Stack	-1.45	External closing pair is C ⁸ -G ³¹
Stack	0.07	External closing pair is A ⁹ -T ³⁰
Stack	0.71	External closing pair is T ¹⁰ -G ²⁹
Stack	-0.58	External closing pair is T ¹¹ -A ²⁸
Stack	0.71	External closing pair is A ¹² -T ²⁷
Stack	0.07	External closing pair is G ¹³ -T ²⁶
Stack	-1.45	External closing pair is T ¹⁴ -A ²⁵
Stack	-1.84	External closing pair is G ¹⁵ -C ²⁴
Stack	-2.24	External closing pair is G ¹⁶ -C ²³
Helix	-7.84	11 base pairs.
Hairpin loop	2.70	Closing pair is C ¹⁷ -G ²²

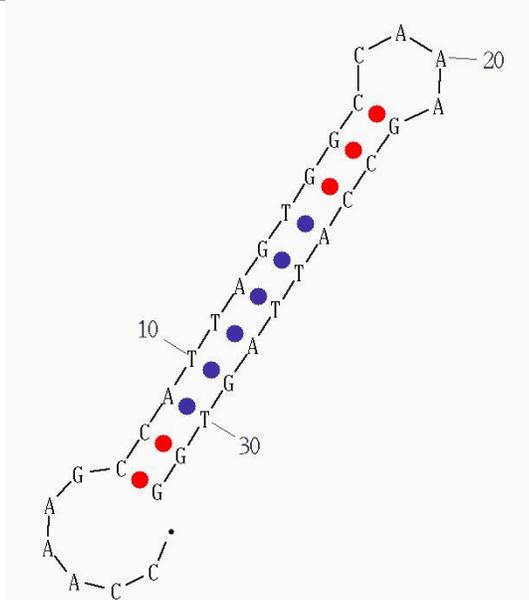


Figure 107b: Hairpins analysis by computer package of oligo Antp32A_MscI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32A_Nru I: 5'-CGAAAGCCATTAGTCGCGAAAGCCATTAGTCG-3'

dG = -1.49 dH = -22.6 dS = -68.0 T_m = 59.0

Structural element	ΔG	Information
External loop	-0.65	24 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ¹⁶ -C ²³
Helix	-2.24	2 base pairs.
Hairpin loop	1.40	Closing pair is C ¹⁷ -G ²²

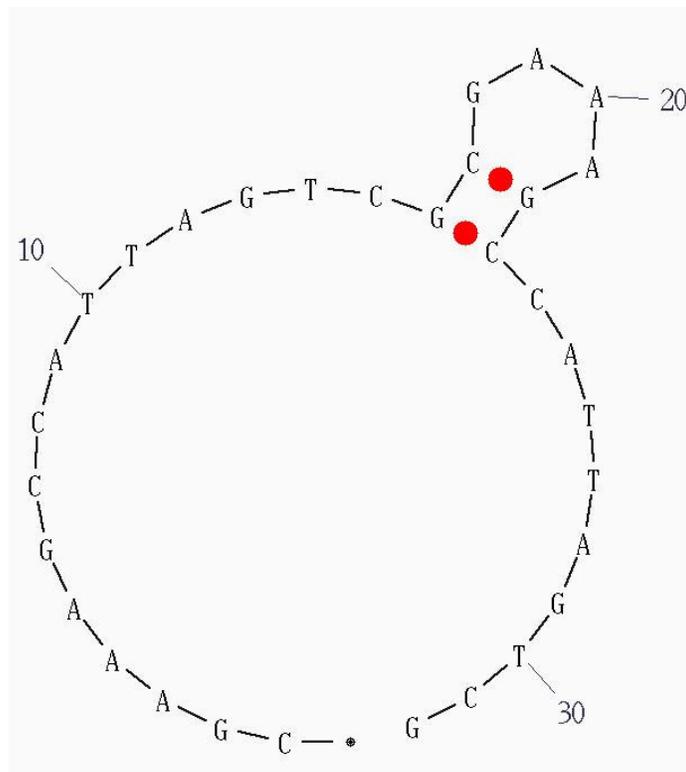


Figure 108a: Hairpins analysis by computer package of oligo Antp32A_NruI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32B_Nru I: 5'-CGACTAATGGCTTTCGCGACTAATGGCTTTCG-3'

$\Delta G = -0.19$ $\Delta H = -25.1$ $\Delta S = -80.3$ $T_m = 39.5$

Structural element	ΔG	Information
External loop	-0.72	16 ss bases & 1 closing helices.
Stack	-2.17	External closing pair is C ¹⁷ -G ³²
Stack	-1.30	External closing pair is G ¹⁸ -C ³¹
Helix	-3.47	3 base pairs.
Hairpin loop	4.00	Closing pair is A ¹⁹ -T ³⁰

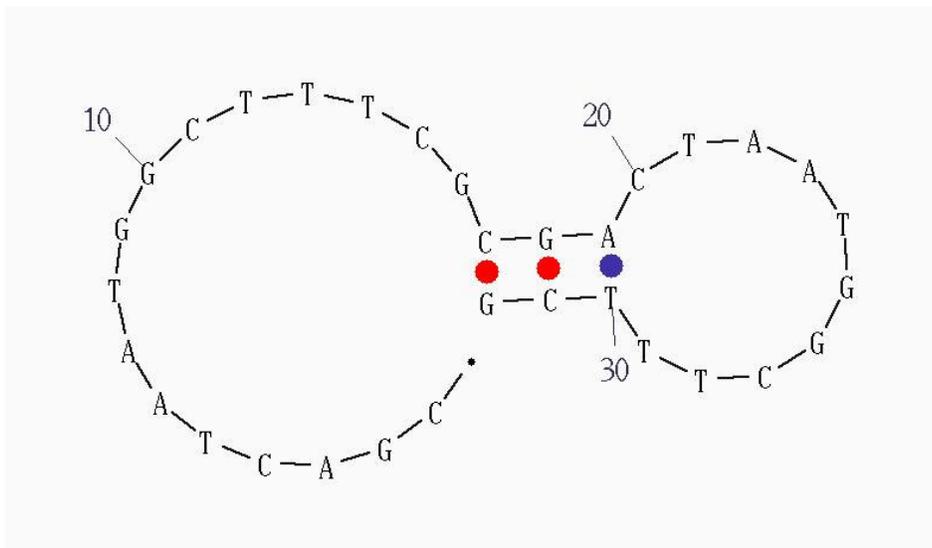


Figure 108b: Hairpins analysis by computer package of oligo Antp32A_NruI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32A_Rsa I: 5'-ACAAAGCCATTAGAGTACAAAGCCATTAGAGT-3'

$dG = 0.67$ $dH = -18.5$ $dS = -61.9$ $T_m = 25.7$

Structural element	ΔG	Information
External loop	-0.89	14 ss bases & 1 closing helices.
Stack	-2.24	External closing pair is G ⁶ -C ²³
Helix	-2.24	2 base pairs.
Hairpin loop	3.80	Closing pair is C ⁷ -G ²²

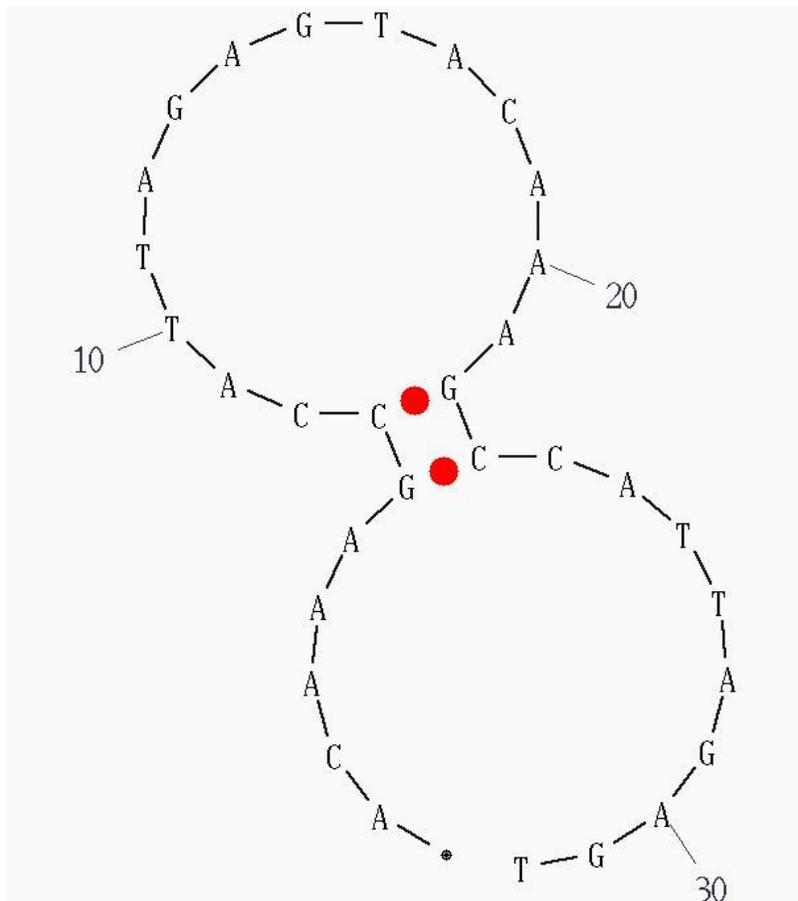


Figure 109a: Hairpins analysis by computer package of oligo Antp32A_RsaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32A_Sca I: 5'-ACTAAGCCATTAGAGTACTAAGCCATTAGAGT-3'

dG = -3.7 dH = -68.0 dS = -207.3 T_m = 54.8

Structural element	ΔG	Information
External loop	-3.76	8 ss bases & 2 closing helices.
Stack	-1.28	External closing pair is C ¹⁸ -G ²⁹
Stack	-0.58	External closing pair is T ¹⁹ -A ²⁸
Stack	-1.00	External closing pair is A ²⁰ -T ²⁷
Helix	-2.86	4 base pairs.
Hairpin loop	2.90	Closing pair is A ²¹ -T ²⁶
Stack	-1.28	External closing pair is C ² -G ¹³
Stack	-0.58	External closing pair is T ³ -A ¹²
Stack	-1.00	External closing pair is A ⁴ -T ¹¹
Helix	-2.86	4 base pairs.
Hairpin loop	2.90	Closing pair is A ⁵ -T ¹⁰

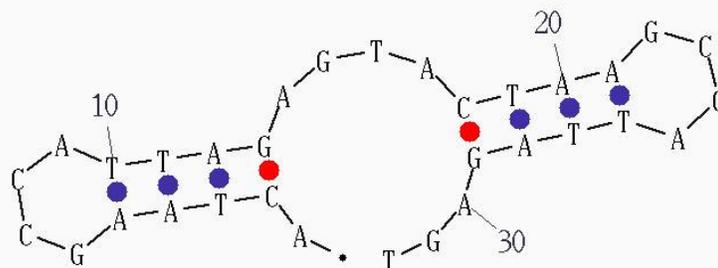


Figure 110a: Hairpins analysis by computer package of oligo Antp32A_ScaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32B_Sca I: 5'-ACTCTAATGGCTTAGTACTCTAATGGCTTAGT-3'

dG = -1.2 dH = -68.2 dS = -216.0 T_m = 42.6

Structural element	ΔG	Information
External loop	-1.86	8 ss bases & 2 closing helices.
Stack	-1.28	External closing pair is C ²⁰ -G ³¹
Stack	-0.58	External closing pair is T ²¹ -A ³⁰
Stack	-1.00	External closing pair is A ²² -T ²⁹
Helix	-2.86	4 base pairs.
Hairpin loop	3.20	Closing pair is A ²³ -T ²⁸
Stack	-1.28	External closing pair is C ⁴ -G ¹⁵
Stack	-0.58	External closing pair is T ⁵ -A ¹⁴
Stack	-1.00	External closing pair is A ⁶ -T ¹³
Helix	-2.86	4 base pairs.
Hairpin loop	3.20	Closing pair is A ⁷ -T ¹²

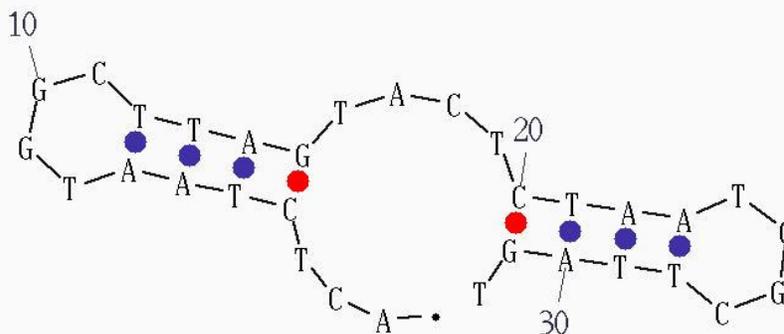


Figure 110b: Hairpins analysis by computer package of oligo Antp32B_ScaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported weak interaction are indicated in blue, while strong interaction in red.

Antp32A_SmaI: 5'-GGGAAGCCATTAGCCCGGGAAGCCATTAGCCC-3'

dG = -1.33 dH = -41.7 dS = -130.3 T_m = 47.0

Structural element	ΔG	Information
External loop	-2.05	16 ss bases & 2 closing helices.
Stack	-1.84	External closing pair is G ¹⁸ -C ²⁶
Helix	-1.84	2 base pairs.
Hairpin loop	2.30	Closing pair is C ¹⁹ -G ²⁵
Stack	-2.24	External closing pair is G ⁷ -C ¹⁵
Helix	-2.24	2 base pairs.
Hairpin loop	2.50	Closing pair is C ⁸ -G ¹⁴

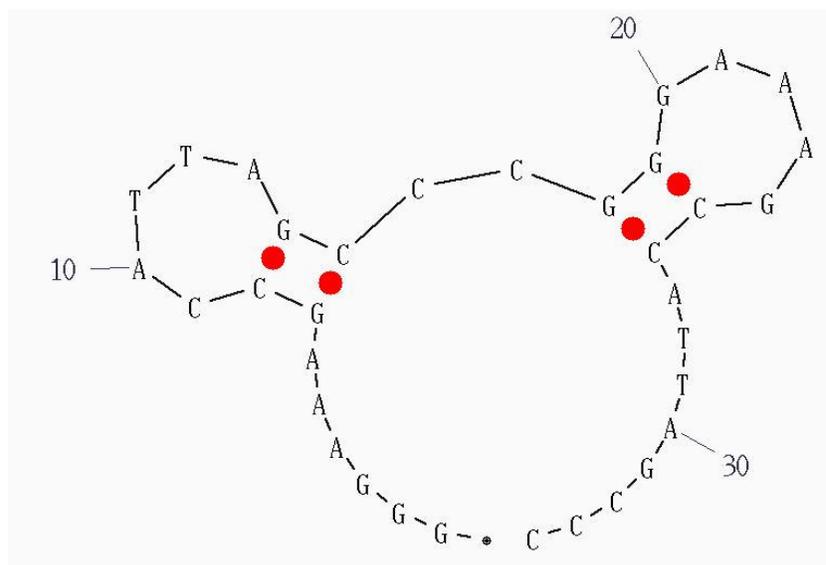


Figure 111a: Hairpins analysis by computer package of oligo Antp32A_SmaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

Antp32B_SmaI: 5'-GGGCTAATGGCTTCCCGGGCTAATGGCTTCCC -3'

$dG = -1.84$ $dH = -49.6$ $dS = -154.1$ $T_m = 48.7$

Structural element	ΔG	Information
External loop	-2.16	14 ss bases & 2 closing helices.
Stack	-2.24	External closing pair is G ¹⁹ -C ²⁷
Helix	-2.24	2 base pairs.
Hairpin loop	2.40	Closing pair is C ²⁰ -G ²⁶
Stack	-2.24	External closing pair is G ³ -C ¹¹
Helix	-2.24	2 base pairs.
Hairpin loop	2.40	Closing pair is C ⁴ -G ¹⁰

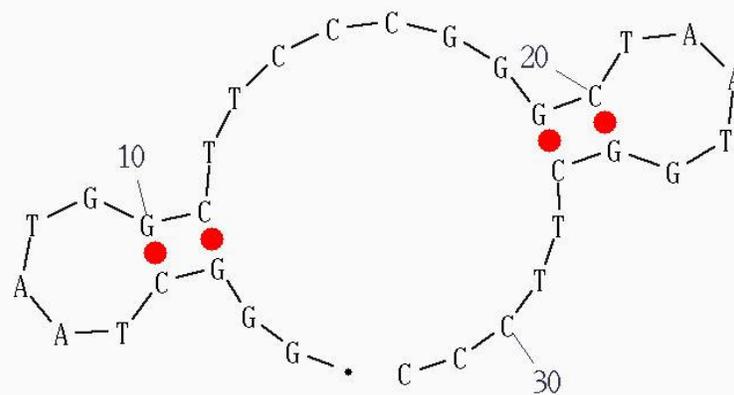


Figure 111b: Hairpins analysis by computer package of oligo Antp32B_SmaI at 37 °C. ΔG and ΔH in Kcal/mol, ΔS in cal/mol/K and T_m in °C. In the structure reported strong interaction are indicated in red.

APPENDIX II

NMR DATA OF ANTENNAPEDIA HOMEODOMAIN

APPENDIX III

NMR DATA OF LA105-202

APPENDIX IV

PROTEIN SEQUENCES AND PROTPARAM OUTPUTS

[\(http://www.expasy.org/tools/protparam.html\)](http://www.expasy.org/tools/protparam.html)

ProtParam is a computer package which computes various physico-chemical properties that can be deduced from a protein sequence such as molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

Here, ProtParam tool was used in particular to determine the theoretical pI, the MW and the extinction coefficients of all proteins under studies.

The theoretical pI value was used to perform proteins purification and to choice the stored buffers. The MW and the extinction coefficients were used to determine the proteins concentrations.

In the following pages the extracts of ProtParam online outputs are reported for all proteins used in this thesis.

Antennapedia C39S homeodomain

Sequence:

MRKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALSLTERQIKIWFQNRMKW
KKENKTKGEPG

Number of amino acids: 68

Molecular weight: 8639.0

Theoretical pI: 11.20

Amino acid composition:

Ala (A)	2	2.9%
Arg (R)	12	17.6%
Asn (N)	3	4.4%
Gln (Q)	4	5.9%
Glu (E)	7	10.3%
Gly (G)	3	4.4%
His (H)	2	2.9%
Ile (I)	4	5.9%
Leu (L)	5	7.4%
Lys (K)	8	11.8%
Met (M)	2	2.9%
Phe (F)	3	4.4%
Pro (P)	1	1.5%
Ser (S)	1	1.5%
Thr (T)	6	8.8%
Trp (W)	2	2.9%
Tyr (Y)	3	4.4%

Total number of negatively charged residues (Asp + Glu): 7

Total number of positively charged residues (Arg + Lys): 20

Atomic composition:

Carbon	C	383
Hydrogen	H	619
Nitrogen	N	125
Oxygen	O	100
Sulfur	S	2

Extinction coefficients at 280nm: 15470 M⁻¹ cm⁻¹

La full length

Sequence:

MAENGDNEKMAALEAKICHQIEYYFGDFNLPRDKFLKEQIKLDEGWVPLEIMIKFNR
LNRLTTDFNVIVEALSLSKAELMEISEDKTKIRRSRSPKPLPEVTDEYKNDVKNRSVY
IKGFPTDATLDDIKEWLEDKGOVLNIQMRRTLHKAFKGSIFVVVDSIESAKKFETPG
QKYKETDLLILFKDDYFAKKNEERKQNKVEAKLRAKQEQEAKQKLEEDAEMKSLEEK
IGCLLKSGDLDDQTCREDLHILFNHGEIKWIDFVRGAKEGIILFKEKAKEALGKAKD
ANNGNLQLRNKEVTWEVLEGEVEKEALKKIIEDQQESLNKWKSKGRRFKGKGNKA
AQP GSGKGVQFQGKTKFASDDEHDEHDENGATGPVKRAREETDKEEPASKQQKTE
NGAGDQ

Number of amino acids: 408

Molecular weight: 46837.0

Theoretical pI: 6.68

Amino acid composition:

Ala (A)	27	6.6%
Arg (R)	17	4.2%
Asn (N)	20	4.9%
Asp (D)	30	7.4%
Cys (C)	3	0.7%
Gln (Q)	19	4.7%
Glu (E)	49	12.0%
Gly (G)	27	6.6%
His (H)	6	1.5%
Ile (I)	22	5.4%
Leu (L)	32	7.8%
Lys (K)	61	15.0%
Met (M)	6	1.5%
Phe (F)	19	4.7%
Pro (P)	10	2.5%
Ser (S)	17	4.2%
Thr (T)	15	3.7%
Trp (W)	5	1.2%
Tyr (Y)	6	1.5%
Val (V)	17	4.2%

Total number of negatively charged residues (Asp + Glu): 79

Total number of positively charged residues (Arg + Lys): 78

Atomic composition:

Carbon	C	2068
Hydrogen	H	3312
Nitrogen	N	576
Oxygen	O	644
Sulfur	S	9

Extinction coefficients at 280nm: 36565 M⁻¹ cm⁻¹

La 1-103

Sequence:

AENGDNKMAALEAKICHQIEYYFGDFNLPRDKFLKEQIKLDEGWVPLEIMIKFNRL
NRLTTDFNVIVEALSLSKAELMEISEDKTKIRRSPSKPLPEVTDE

Number of amino acids: 102

Molecular weight: 11872.6

Theoretical pI: 5.12

Amino acid composition:

Ala (A)	6	5.9%
Arg (R)	5	4.9%
Asn (N)	6	5.9%
Asp (D)	7	6.9%
Cys (C)	1	1.0%
Gln (Q)	2	2.0%
Glu (E)	13	12.7%
Gly (G)	3	2.9%
His (H)	1	1.0%
Ile (I)	8	7.8%
Leu (L)	10	9.8%
Lys (K)	11	10.8%
Met (M)	3	2.9%
Phe (F)	5	4.9%
Pro (P)	5	4.9%
Ser (S)	5	4.9%
Thr (T)	4	3.9%
Trp (W)	1	1.0%
Tyr (Y)	2	2.0%
Val (V)	4	3.9%

Total number of negatively charged residues (Asp + Glu): 20

Total number of positively charged residues (Arg + Lys): 16

Atomic composition:

Carbon	C	529
Hydrogen	H	845
Nitrogen	N	139
Oxygen	O	162
Sulfur	S	4

Extinction coefficients at 280nm: 8480 M⁻¹ cm⁻¹

La 105-202

Sequence:

MGHHHHHHIEGRWILKNDVKNRSVYIKGFPTDIDLDDIKEWLEDKGQVLNIQMRRTL
HKAFKGSIFVVFDSIESAKKFVETPGQKYKETDLLILFKDDYFAKKNEERKQNKVE

Number of amino acids: 113

Molecular weight: 13424.3

Theoretical pI: 9.07

Amino acid composition:

Ala (A)	4	3.5%
Arg (R)	5	4.4%
Asn (N)	5	4.4%
Asp (D)	9	8.0%
Gln (Q)	4	3.5%
Glu (E)	9	8.0%
Gly (G)	6	5.3%
His (H)	7	6.2%
Ile (I)	8	7.1%
Leu (L)	8	7.1%
Lys (K)	16	14.2%
Met (M)	2	1.8%
Phe (F)	7	6.2%
Pro (P)	2	1.8%
Ser (S)	4	3.5%
Thr (T)	5	4.4%
Trp (W)	2	1.8%
Tyr (Y)	3	2.7%
Val (V)	7	6.2%

Total number of negatively charged residues (Asp + Glu): 18

Total number of positively charged residues (Arg + Lys): 21

Atomic composition:

Carbon	C	608
Hydrogen	H	947
Nitrogen	N	169
Oxygen	O	171
Sulfur	S	2

Extinction coefficients at 280 nm: 15470 M⁻¹ cm⁻¹

La 225-334

Sequence:

MRGSHHHHHHGSLEEKIGCLLKFSGDLDDQTCREDLHILFSNHGEIKWIDFVRGAKE
GIILFKEKAKEALGKAKDANNGNLQLRNKEVTWEVLEGEVEKEALKKI IEDQQESLN
KWKSKGR

Number of amino acids: 121

Molecular weight: 13922.7

Theoretical pI: 6.71

Amino acid composition:

Ala (A)	6	5.0%
Arg (R)	5	4.1%
Asn (N)	6	5.0%
Asp (D)	7	5.8%
Cys (C)	2	1.7%
Gln (Q)	4	3.3%
Glu (E)	15	12.4%
Gly (G)	11	9.1%
His (H)	8	6.6%
Ile (I)	8	6.6%
Leu (L)	13	10.7%
Lys (K)	16	13.2%
Met (M)	1	0.8%
Phe (F)	4	3.3%
Pro (P)	0	0.0%
Ser (S)	6	5.0%
Thr (T)	2	1.7%
Trp (W)	3	2.5%
Tyr (Y)	0	0.0%
Val (V)	4	3.3%

Total number of negatively charged residues (Asp + Glu): 22

Total number of positively charged residues (Arg + Lys): 21

Atomic composition:

Carbon	C	613
Hydrogen	H	977
Nitrogen	N	181
Oxygen	O	184
Sulfur	S	3

Extinction coefficients at 280nm: 16625 M⁻¹ cm⁻¹

La 1-94

Sequence:

MAENGDNEKMAALEAKICHQIEYYFGDFNLPRDKFLKEQIKLDEGWVPLEIMIKFNR
LNRLTTDFNVIVEALSLSKAELMEISEDKTKIRRSKPLPEVTDEYKNDVKNRSVY
IKGFPTDATTDDIKEWLEDKGQVLNIQMRRTLHKAFKGSIFVVVFDIESAKKFVETP
GQKYKETDLLILFKDDYFAKKNEE

Number of amino acids: 195

Molecular weight: 22811.1

Theoretical pI: 5.65

Amino acid composition:

Ala (A)	10	5.1%
Arg (R)	8	4.1%
Asn (N)	10	5.1%
Asp (D)	16	8.2%
Cys (C)	1	0.5%
Gln (Q)	5	2.6%
Glu (E)	20	10.3%
Gly (G)	7	3.6%
His (H)	2	1.0%
Ile (I)	14	7.2%
Leu (L)	17	8.7%
Lys (K)	25	12.8%
Met (M)	5	2.6%
Phe (F)	12	6.2%
Pro (P)	7	3.6%
Ser (S)	9	4.6%
Thr (T)	9	4.6%
Trp (W)	2	1.0%
Tyr (Y)	6	3.1%
Val (V)	10	5.1%

Total number of negatively charged residues (Asp + Glu): 36

Total number of positively charged residues (Arg + Lys): 33

Atomic composition:

Carbon	C	1029
Hydrogen	H	1623
Nitrogen	N	265
Oxygen	O	307
Sulfur	S	6

Extinction coefficients: 19940 M⁻¹ cm⁻¹

La 105-334

Sequence:

MHHHHHSSGLVPRGSGMKETAATAKFERQHMDSPDLGTGGGSGIEGRKNDVKNRSVY
IKGFPTDATLDDIKEWLEDKGOVLNIQMRRTLHKAFKGSIFVVFDSDIESAKKFVETP
GQKYKETDLLILFKDDYFAKKNEERKQNKVEAKLRAKQEQAQKLEEDAEMKSLEE
KIGCLLKFSGLDDQTCREDLHILFSNHGEIKWIDFVRGAKEGIILFKEKAKEALGK
AKDANNGNLQLRNKEVTWEVLEGEVEKEALKKI IEDQQESLNKWKSKGR

Number of amino acids: 277

Molecular weight: 31690.9

Theoretical pI: 7.75

Amino acid composition:

Ala (A)	17	6.1%
Arg (R)	12	4.3%
Asn (N)	11	4.0%
Asp (D)	19	6.9%
Cys (C)	2	0.7%
Gln (Q)	12	4.3%
Glu (E)	31	11.2%
Gly (G)	22	7.9%
His (H)	10	3.6%
Ile (I)	15	5.4%
Leu (L)	24	8.7%
Lys (K)	39	14.1%
Met (M)	5	1.8%
Phe (F)	12	4.3%
Pro (P)	4	1.4%
Ser (S)	14	5.1%
Thr (T)	9	3.2%
Trp (W)	4	1.4%
Tyr (Y)	3	1.1%
Val (V)	12	4.3%

Total number of negatively charged residues (Asp + Glu): 50

Total number of positively charged residues (Arg + Lys): 51

Atomic composition:

Carbon	C	1398
Hydrogen	H	2237
Nitrogen	N	399
Oxygen	O	427
Sulfur	S	7

Extinction coefficients at 280nm: 26595 M⁻¹ cm⁻¹

REFERENCES

- Alfano, C., Sanfelice, D., Babon, J., Kelly, G., Jacks, A., Curry, S. and Conte, M. R. **Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein.** *Nat Struct Mol Biol* **11**, 323–329 (2004).
- Ali, N., Pruijn, G. J., Kenan, D. J., Keene, J. D. and Siddiqui, A. **Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation.** *J Biol Chem* **275**, 27531–40 (2000).
- Allain, F. H., Howe, P. W., Neuhaus, D. and Varani, G. **Structural basis of the RNA-binding specificity of human U1A protein.** *EMBO J.* **18** 5764–72 (1997).
- Alsbaugh, M. A. and Tan, E.M. *J. Clin. Invest.* **55**, 1067–1073 (1975).
- Avis, J.M., Allain, F.H.T., Howe, P.W.A., Varani, G., Nagai, K., and Neuhaus, D. **Solution structure of the N-terminal RNP domain of the U1A protein: the role of the C-terminal residues in structure stability and RNA binding.** *J. Mol. Biol.* **257**, 398–411 (1996).
- Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. **The program XEASY for computer supported NMR spectral-analysis of biological macromolecules.** *J. Biomol. NMR* **6**, 1–10 (1995).
- Bax, A. and Grzesiek, S. **Methodological advances in protein NMR.** *Acc. Chem. Res.* **26**, 131–138 (1993).
- Bax, A., Clore, G. M. and Gronenborn, A. M. *J. Magn. Res.* **88**, 425–443 (1990).
- Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J. and Wüthrich, K. **Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex.** *J. Mol. Biol.* **234**, 1084–1093 (1993).
- Bodenhausen, G. and Ruben, D. J. **Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy.** *Chem. Phys. Lett.* **69**, 185 (1980).

Brunger, A.T. **X-PLOR: a System for X-ray crystallography and NMR.** *Yale University Press, New Haven* (1992).

Carrasco, A. E., McGinnis, W., Gehring, W. J. and De Robertis, E. M. **Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes.** *Cell* **37**, 409-414 (1984).

Cary, P. D., Read, C. M., Davies, B., Driscoll, P. C. and Crane-Robinson, C. **Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5.** *Protein Sci.* **10**, 83-98 (2001).

Castagne, C., Murphy, E.C., Gronenborn, A.M. and Delepierre, M. *Eur J Biochem.* **267**, 1223-1229 (2000).

Chakshusmathi, G., Kim, S. D., Rubinson, D. A. and Wolin, S. L. **A La protein requirement for efficient pre-tRNA folding.** *Embo J* **22**, 6562-6572 (2003).

Chang, Y. N., Kenan, D. J., Keene, J. D., Gatignol, A. and Jeang, K. T. **Direct interactions between autoantigen La and human immunodeficiency virus leader RNA.** *J Virol* **68**, 7008-20 (1994).

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. **Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.** *Nature* **364**, 412-20 (1993).

Clore, G. M. and Gronenborn, A. M. **Structure of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy.** *Science* **252**, 1390-1399 (1991).

Clore, G.M., Gronenborn, A.M. and Tjandra, N. **Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude.** *J. Magn. Reson.* **131**, 159–162 (1998).

Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscoll, P. C., Denny, P., Crane-Robinson, C. and Ashworth, A. **DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5.** *Nucleic Acids Res.* **22**, 3339–3346

(1994).

Cornilescu, G., Delaglio, F. and Bax, A. **Protein backbone angle restraints from searching a database for chemical shift and sequence homology.** *J. Biomol. NMR* **13**, 289–302 (1999).

Crane-Robinson, C., Read, C. M., Cary, P. D., Driscoll, P. C., Dragan, A. I. and Privalov, P. L. **The energetics of HMG box interactions with DNA. Thermodynamic description of the box from mouse Sox-5.** *J. Mol. Biol.* **281**, 705-717 (1998).

Crosio, C., Boyl, P. P., Loreni, F., Pierandrei-Amaldi, P. and Amaldi, F. **La protein has a positive effect on the translation of TOP mRNAs in vivo.** *Nucleic Acids Res* **28**, 2927-34 (2000).

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. **NMRPipe: a multidimensional spectral processing system based on UNIX pipes.** *J. Biomol. NMR* **6**, 277–293 (1995).

Denny, P., Swift, S., Connor, F. and Ashworth, A. **An Sry-related gene expressed during spermatogenesis in the mouse encodes a sequence-specific DNA-binding protein.** *EMBO J.* **11**, 3705–3712 (1992).

Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. **Recognition of polyadenylate RNA by the poly(A)-binding protein.** *Cell.* **17**, 835-45 (1999).

Dong, G., Chakshusmathi, G., Wolin, S.L. and Reinisch, K. M. **Structure of the La motif: a winged helix domain mediates RNA binding via a conserved aromatic patch.** *EMBO J.* **23**, 1000-1007 (2004).

Dragan, A. I., Liggins, J. R., Crane-Robinson, C. and Privalov, P. L. **The Energetics of Specific Binding of AT-hooks from HMGA1 to Target DNA.** *J. Mol. Biol.* **327**, 393–411 (2003).

Dragan, A.I., Li, Z., Makeyeva, E.N., Milgotina, E.I., Liu, Y., Crane-Robinson, C., Privalov, P.L. **Forces driving the binding of homeodomains to DNA.** *Biochemistry* **45**,141-151 (2006).

Drew, H. R. and Dickerson, R. E. **Structure of a B_DNA dodecamer. Geometry of hydration.** *J.Mol.Biol.* **151**, 535-556 (1981).

Engelke, J. and Rüterjans, H. *Biological Magnetic Resonance* **17**, 357-385 (1999).

Fan, H., Goodier, J. L., Chamberlain, J. R., Engelke, D. R. and Maraia, R. J. **5' processing of tRNA precursors can be modulated by the human La antigen phosphoprotein.** *Mol Cell Biol* **18**, 3201-11 (1998).

Fesik, S. W. and Zuiderweg, E. R. P. **Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules.** *Quart. Rev. Biophys.* **23**, 97-131 (1990).

Fesik, S.W. and Zuiderweg, E.R.P. **Heteronuclear 3-dimensional NMR spectroscopy. A strategy for the simplification of homo-nuclear two-dimensional NMR spectra.** *J. Magn. Res.* **78**, 588-593 (1988).

Fouraux, M. A. et al. **The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase.** *RNA* **8**, 1428-43 (2002).

Freemann, R. and Hill, H. D. W. *J. Chem. Phys.* **51**, 3140-3141 (1969).

Fried, M. and Crothers, D.M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).

Fried, M.G. and Crothers, D.M. *J. Mol. Biol.* **172**, 263-282 (1984).

Fried, M.G. *Electrophoresis* **10**, 366-376 (1989).

Gajiwala, K. S. and Burley, S. K. **Winged helix proteins.** *Curr Opin Struct Biol* **10**, 110-6 (2000).

Garner, M.M. and Revzin, A. *Nucleic Acids Res.* **9**, 3047-3059 (1981).

Goodier, J. L., Fan, H. and Maraia, R. J. **A carboxy-terminal basic region controls RNA polymerase III transcription factor activity of human La protein.** *Mol Cell*

Biol **17**, 5823-32 (1997).

Grzesiek, S., Anglister, J. and Bax, A. *J. Magn. Reson.* **101**, 114-119 (1993).

Haq, I., Ladbury, J. E., Chowdhry, B.Z., Jenkins, T. C. and Chaires, J. B. **Specific binding of Hoechst 33258 to the d(CGCAAATTTGCG)₂ duplex: calorimetric and spectroscopic studies.** *J.Mol.Biol.* **271**, 244-257 (1997).

Hendrickson, W. *BioTechniques* **3**, 346-354 (1985).

Holcik, M. and Korneluk, R. G. **Functional characterization of the X-linked inhibitor of apoptosis (XIAP) internal ribosome entry site element: role of La autoantigen in XIAP translation.** *Mol Cell Biol* **20**, 4648-57 (2000).

Horke, S., Reumann, K., Rang, A. and Heise, T. **Molecular characterization of the human La protein.hepatitis B virus RNA B interaction in vitro.** *J Biol Chem* **277**, 34949-58 (2002).

Ikura, M., Kay, L. E. and Bax, A. *Biochemistry* **29**, 4659-4667 (1990).

Intine, R. V. *et al.* **Control of transfer RNA maturation by phosphorylation of the human La antigen on serine 366.** *Mol Cell* **6**, 339-48 (2000).

Intine, R. V., Dundr, M., Misteli, T. and Maraia, R. J. **Aberrant nuclear trafficking of La protein leads to disordered processing of associated precursor tRNAs.** *Mol Cell* **9**, 1113-23 (2002).

Intine, R. V., Tenenbaum, S. A., Sakulich, A. L., Keene, J. D. and Maraia, R. J. **Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs.** *Mol Cell* **12**, 1301-7 (2003).

Ito, T., Muto, Y., Green, M. R. and Yokoyama, S. **Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF).** *EMBO J* **18**, 4523-4534 (1999).

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P. D., Curry, S. and Conte, M. R. **Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element.** *Structure (Camb)* **11**, 833-43 (2003).

Jelesarov, I., Crane-Robinson, C. and Privalov, P.L. **The energetics of HMG box interactions with DNA:thermodynamic description of the target DNA duplex.** *J. Mol. Biol.* **294**, 981-995 (1999).

Jen-Jacobsen, L., Engler, L. E. and Jacobsen, L. A. **Structural and thermodynamic strategies for site-specific DNA binding proteins.** *Structure* **8**, 1015-1023 (2000).

Johannesson, H. and Halle, B. **Minor groove hydration of DNA in solution: dependence on base composition and sequence.** *J. Amer. Chem. Soc.* **120**, 6859-6870 (1998).

Kay, L. E., Keifer, P. and Saarinen, T. *J. Am. Chem. Soc.* **114**, 10663-10665 (1992a).

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. and Torchia, D. A. *J. Magn. Reson.* **97**, 359-375 (1992b).

Kay, L.E., Torchia, D.A. and Bax, A. **Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease.** *Biochemistry* **28**, 8972–8979 (1989).

Kenan, D. J. **RNA Recognition by the Human La Protein and Its Relevance to Transcription, Translation and Viral Infectivity.** Thesis, Duke University (1995).

Koradi, R., Billeter, M. and Wüthrich, K. **MOLMOL: a program for display and analysis of macromolecular structures.** *J. Mol. Graph.* **14**, 51-55, 29–32 (1996).

Laskowski, R.A., Rullman, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. **AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.** *J. Biomol. NMR* **8**, 477–486 (1996).

Lefebvre, V., Li, P. and de Crombrughe, B. **A new long form of Sox5 (L-Sox5), Sox6**

and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO J.* **17**, 5718–5733 (1998).

Levine, R., Rubin, G. M. and Tjian, R. **Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila.** *Cell* **38**, 667-673 (1984).

Liepinsh, E., Otting, G. and Wüthrich, K. **NMR observation of individual molecules of hydration water bound to DNA duplexes-direct evidence for a spine of hydration water present in aqueous solution.** *Nucleic Acids Res.* **20**, 6549-6553 (1992).

Louis, J. M., Martin, R. G., Clore, G. M., Gronenborn, A. M. **Preparation of Uniformly Isotope-labeled DNA Oligonucleotides for NMR Spectroscopy.** *J. Biological Chemistry* **273**, 2374-2378 (1998).

Lundback, T. and Hard, T. **Sequence-specific DNA-binding dominated by dehydration.** *Proc Natl Acad Sci USA* **93**, 4754-4759 (1996).

Makhatadze, G. I. and Privalov, P. L. **Energetics of protein structure.** *Advan. Protein Chem.* **47**, 307-425 (1995)

Maraia, R. J. and Intine, R. V. **Recognition of nascent RNA by the human La antigen: conserved and divergent features of structure and function.** *Mol Cell Biol* **21**, 367-79 (2001).

Marchetti, M. A., Tschudi, C., Kwon, H., Wolin, S. L. and Ullu, E. **Import of proteins into the trypanosome nucleus and their distribution at karyokinesis.** *J. Cell Science*, **113**, 899-906 (2000).

Maris, C., Dominguez, C. and Allain, F.H.T. **The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.** *FEBS Journal* **272**, 2118–2131 (2005).

Marky, L. A. and Breslauer, K. J. **Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data.** *Proc Natl Acad Sci USA* **84**, 4359-4363 (1987)

Mattioli M. and Reichlin M. *Arthritis Rheum.* **17**, 421–29 (1974).

McGinnis, W., Hart, C. P., Gehring, W. J. and Ruddle, F. H. **Molecular cloning and chromosome mapping of a mouse DNA sequence homologous genes of *Drosophila*.** *Cell* **38**, 675-680 (1984b).

McGinnis, W., Levine, M. S., Halen, E., Kuroiwa, A. and Gehring, W. J. **A conserved DNA sequence in homeotic genes of the *Drosophila* Antennapedia and bithorax complex.** *Nature* **308**, 428-433 (1984a).

Merabet, E. and Ackers, G. K. **Calorimetric analysis of lambda cI repressor binding to DNA operator sites.** *Biochemistry* **34**, 8554-8563 (1995).

Mittermaier, A., Varani, L., Muhandiram, D.R., Kay, L.E. and Varani, G. *J.Mol.Biol.* **294**, 967-979 (1999).

Muhandiram, D.R., Yamazaki, T., Sykes B.D. and Kay, L.E. *J. Amer. Chem. Soc.* **117**, 11536-11544 (1995).

Muller, S., Bianchi, M. E. and Knapp, S. **Thermodynamics of HMGB1 interaction with duplex DNA.** *Biochemistry* **40**, 10254-10261 (2001).

Murphy, E.C., Zhurkin, V.B., Louis, J.M., Cornilescu, G., Clore, G.M. **Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation.** *J.Mol.Biol.* **312**, 481-499 (2001).

Nirmala, N. and Wagner, G. *J Am. Chem. Soc.* **110**, 7556-7558 (1988).

Ohndorf, U. M., Steegborn, C., Knijff, R. and Sondermann, P. **Contributions of the individual domains in human La protein to its RNA 3'-end binding activity.** *J Biol Chem* **276**, 27188-96 (2001).

Okuda, M. et al. **Structure of the central core domain of TFIIIEbeta with a novel double-stranded DNA-binding surface.** *EMBO J* **19**, 1346-56 (2000).

- Oleiniczak, E. T., Xu, R. X. and Fesik, S. W. *J. Biomol. NMR* **2**, 655-659 (1992).
- Pannone, B.K., Kim, S. D., Noe, D. A. and Wolin, S. L. **Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein.** *Genetics* **158**, 187-96 (2001).
- Price, S. R., Evans, P. R. and Nagai, K. **Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA.** *Nature* **394**, 645-650 (1998).
- Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I. and Crane-Robinson, C. **The energetics of HMG box interactions with DNA: thermodynamics of the DNA binding of the HMG box from mouse Sox-5.** *J.Mol.Biol.* **294**, 997-1013 (1999).
- Pudi, R., Abhiman, S., Srinivasan, N. and Das, S. **Hepatitis C virus internal ribosome entry site-mediated translation is stimulated by specific interaction of independent regions of human La autoantigen.** *J Biol Chem* **278**, 12231-40 (2003).
- Qian, Y. Q., Otting, G., Billeter, M., Muller, M., Gehring, W. J. and Wüthrich, K. **Nuclear magnetic resonance spectroscopy of a DNA complex with the uniformly C-labelled Antennapedia homeodomain, and structure determination of the DNA-bound homeodomain.** *J.Mol.Biol.* **234**, 1070-1083 (1993)
- Qian, Y. Q., Resendez-Perez, D., Gehring, W. J. and Wüthrich, K. **The des(1-6) antennapedia homeodomain : comparison of the NMR solution structure and the DNA-binding affinity with the intact Antennapedia homeodomain.** *Proc Natl Acad Sci USA* **91**, 4091-4095 (1994).
- Raats, J. M. et al. **Human recombinant anti-La (SS-B) autoantibodies demonstrate the accumulation of phosphoserine-366-containing la isoforms in nucleoplasmic speckles.** *Eur J Cell Biol* **82**, 131-41 (2003).
- Revzin, A. *BioTechniques* **7**, 346-354 (1989).
- Rosenblum, J. S., Pemberton, L. F., Bonifaci, N. and Blobel, G. **Nuclear import and the evolution of a multifunctional RNA-binding protein.** *J. Cell Biol* **143**, 887-899

(1998).

Rückert, M. and Otting, G. **Alignment of biological macromolecules in novel non-ionic liquid crystalline media for NMR experiments.** *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).

Scott, M. P. and Weiner, A. J. **Structural relationship among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila.** *Proc. Nat. Acad. Sci. USA* **81**, 4115–4119 (1984).

Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. and Steitz, T. A. **Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution.** *Nat. Struct. Biol.* **4**, 215–222 (1997).

Simons, F. H., Broers, F. J., Van Venrooij, W. J. and Pruijn, G. J. **Characterization of cis-acting signals for nuclear import and retention of the La (SS-B) autoantigen.** *Exp Cell Res* **224**, 224–36 (1996).

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkin, J. R., Griffiths, B. L., Smith, M. I., Foster, J. W., Fischauf, A. M., Lovell-Badge, R. and Goodfellow, P. N. **A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif.** *Nature* **354**, 240–244 (1990).

Takeda, Y., Ross, P. D. and Mudd, C. P. **Thermodynamics of Cro protein-DNA interactions.** *Proc Natl Acad Sci USA* **89**, 8180–8184 (1992).

Tamura, A. and Privalov, P. L. **The entropy cost of protein association.** *J. Mol. Biol.* **273**, 1048–1060 (1997).

Tjandra, N., Grzesiek, S. and Bax, A. **Magnetic field dependence of nitrogen-proton J splitting in ¹⁵N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling.** *J. Am. Chem. Soc.* **118**, 6264–6272 (1996).

Vold, R. L., Waugh, J. S., Klein, M. P. and Phelps, P. E. *J. Chem. Phys.* **48**, 3831–3832 (1968).

Wand, A.J., Urbauer, J.L., McEnvoy, R.P. and Bieber, R.J. *Biochemistry* **35**, 6116-6125 (1996).

Warren, J.J. and Moore, P.B. **A maximum likelihood method for determining D-a(PQ) and R for sets of dipolar coupling data.** *J. Magn. Reson.* **149**, 271–275 (2001).

Werner, M. H., Gronenborn, A. M. and Clore, G. M. **Intercalation, DNA kinking, and the control of transcription.** *Science* **271**, 778-784 (1996).

Wolin, S. L. and Cedervall, T. **The La protein.** *Annu Rev Biochem* **71**, 375-403 (2002).

Wüthrich, K. **NMR of Proteins and Nucleic Acids.** *Wiley, New York* (1986).

Wüthrich, K., Wider, G., Wagner, G. and Braun, W. **Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance.** *J. Mol. Biol.* **155**, 311-319 (1982).

Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A. R. **Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs.** *Structure* **5**, 559-570 (1997).

Yoo, C. J. and Wolin, S. L. **The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors.** *Cell* **89**, 393-402 (1997).

Zidek, L., Novotny, M. V. and Stone, M. J. **Increased protein backbone conformational entropy upon hydrophobic ligand binding.** *Nat. Struct. Biol.* **6**, 1118-1121 (1999).

Zuker, M. *Nucleic Acids Res.* **31**, 3406-3415 (2003).