

Open Research Online

The Open University's repository of research publications and other research outputs

Interaction of RGG and HTH motifs with nucleic acids: a study with rationally designed synthetic and recombinant polypeptides

Thesis

How to cite:

Guarnaccia, Corrado (2001). Interaction of RGG and HTH motifs with nucleic acids: a study with rationally designed synthetic and recombinant polypeptides. PhD thesis. The Open University.

For guidance on citations see [FAQs](#).

© 2001 Corrado Guarnaccia

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

**INTERACTION OF RGG AND HTH MOTIFS WITH
NUCLEIC ACIDS: A STUDY WITH RATIONALLY
DESIGNED SYNTHETIC AND RECOMBINANT
POLYPEPTIDES**

Corrado Guarnaccia

A Thesis Submitted for the Degree of Ph.D. at the Open University

Structural Biology


International Center for Genetic Engineering and Biotechnology

TRIESTE

Director of Studies: Sándor Pongor, Ph.D., D.Sc.

Second Supervisor: Daniela Rhodes, Ph.D.

June 2001


SUBMISSION DATE: 11 JUNE 2001

ProQuest Number: C808554

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest C808554

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgements

This thesis work was carried out at the International Centre for Genetic Engineering and Biotechnology, Protein Structure and Function Group. I'm very grateful to Profs. Arturo Falaschi and Francisco Baralle for giving me the opportunity to attend the Ph.D. course.

I wish to sincerely thank my Director of Studies, Prof. Sándor Pongor for his constant guidance and encouragement and my Second Supervisor, Dr. Daniela Rhodes, for her advices and helpful discussions.

I'm really indebted to my colleague Dr. Sotir Zakhariiev for introducing me to the peptide chemistry techniques and for providing me extensive instructions for the production of the amino acid building blocks. Thanks are due to Dr. Bakthisaran Raman for teaching me the fundamentals of the spectroscopy methods and for his help in the planning and analysis of experiments and results. I'm also particularly grateful to Dr. András Simoncsits for providing me the R69 plasmid constructs and useful advices in the protein expression, to Dr. Kristian Vlahovicek for his assistance in computer work and to Dr. Francesco Zanuttin for helpful discussion about mass spectrometry issues.

Last but not least I wish to thank all my friends and colleagues Cristina Acatrinei, Maristella Coglievina, Gordana Maravic, Alessandro Pintar, Marie Louise Tjörnhammar, for their constant support in these years.

Summary

The aim of this work is to provide data on the weak interactions between proteins and nucleic acids, using a sequence-specific DNA-binding protein (the recombinant 1-69 N-terminal domain of the phage 434 cI repressor, R69), and a non-specific nucleic acid binding domain (the arginine-rich RGG domain of human nucleolin produced by solid phase synthesis) as *in vitro* model systems.

Pyrene fluorescence spectroscopy and disulfide crosslinking via C-terminally attached Cys residues showed that cognate DNA acts as template for R69 dimerization even in the absence of the native C-terminal dimerization domain. Dimeric binding of R69 to DNA results in intensive pyrene excimer fluorescence resulting from the stacking of two C-terminally attached pyrene labels. The fluorescence of the pyrene monomer (383nm) is quenched in a salt/sensitive manner by both cognate and non-cognate DNA molecules, probably due to electrostatic interactions with the DNA backbone. DNA sequences as short as the half-site are sufficient to promote dimer formation in a detectable manner. On the other hand, studies with artificial hybrid operators prove that two half sites provide complete specificity to the complex assembly. These studies suggest that sequential binding of monomers is a plausible pathway for the assembly of R69 on DNA.

The RGG motif is present in many specific and non-specific nucleic acid binding proteins and its arginine residues undergo methylation by protein arginine methyltransferase type I as a part of various, not entirely known regulatory events. We used novel methods of solid phase peptide synthesis to produce methylated and non-

methylated peptides spanning sequences from residues 646-706 of the human nucleolin C-terminal domain and we used them to study nucleic acid binding in vitro.

Melting curve assays, double filter binding assays and circular dichroism studies showed that the non-sequence specific nucleic acid binding is sensitive to the geometry of the nucleic acid as it binds ssDNA>RNA>dsDNA. While binding strength is not substantially influenced by dimethylation, the ability of the peptides to modify the nucleic acid structure, detected by CD spectroscopy, is substantially decreased upon methylation.

TABLE OF CONTENTS

INTRODUCTION	1
1. DNA AND RNA RECOGNITION BY PROTEINS	3
1.1 <i>Intermolecular forces involved in the chemical recognition</i>	3
1.1.1 Electrostatic forces: salt bridges.....	4
1.1.2 Dipolar forces: hydrogen bonds	4
1.1.3 Entropic forces: hydrophobic interactions:	5
1.1.4 Dispersion forces: stacking interactions.....	5
1.2 <i>Chemical recognition at the protein-nucleic acid interface</i>	6
1.3 <i>Role of water molecules</i>	9
2. STRUCTURAL COMPLEMENTARITY OF THE INTERACTING PROTEIN AND NUCLEIC ACID	10
2.1 <i>Similarities and differences in the recognition by DNA and RNA helices</i> ..	10
2.2 <i>DNA recognition and protein folds</i>	11
2.3 <i>RNA recognition and protein folds</i>	16
3. FORMATION OF THE COMPLEX AND FOLDING EVENTS.....	18
3.1 <i>Multiple subsite recognition</i>	18
3.2 <i>The kinetics of protein-nucleic acid complex formation</i>	20
3.2.1 Non-specific and specific binding	20
3.3 <i>Induced fit folding events</i>	23
3.3.1 RNA-induced folding events.....	23
3.3.2 DNA-induced folding events.....	24
4. DNA RECOGNITION BY THE HELIX-TURN-HELIX MOTIF	27
5. INTERACTION BETWEEN PHAGE 434 REPRESSOR AND ITS OPERATOR.....	31
5.1 <i>Role in genetic switch</i>	31
5.2 <i>Structure of the repressor and its N-terminal domain (R1-69) operator complexes</i>	32
5.2.1 R1-69/OR1 complex.....	34
5.2.2 R1-69/OR2 complex	37
5.2.3 R1-69/OR3 complex	38

5.3	<i>The recognition properties of 434 repressor</i>	38
5.4	<i>The helix swap experiment</i>	40
5.5	<i>Contacts between R69 monomers</i>	41
5.6	<i>Assembly pathways of dimeric transcription factors on DNA</i>	42
6.	THE RGG MOTIF.....	44
6.1	<i>Possible involvement of the RGG box in various disorders</i>	48
7.	NUCLEOLIN.....	48
7.1	<i>The N-terminal domain</i>	49
7.2	<i>The central globular domain</i>	50
7.3	<i>The C-terminal domain</i>	51
7.3.1	<i>Arginine methylation</i>	52
8.	SPECTROSCOPIC METHODS USED IN THIS WORK TO ANALYZE PROTEIN AND NUCLEIC ACID STRUCTURE.....	56
8.1	<i>Circular dichroism spectroscopy</i>	56
8.1.1	<i>Expression of CD</i>	57
8.1.2	<i>CD determination of protein secondary structure</i>	58
8.1.3	<i>CD of protein-nucleic acid complexes</i>	59
8.2	<i>Pyrene excimer fluorescence: a probe for proximity</i>	61
	AIM OF THIS WORK	64
	MATERIALS AND METHODS	67
9.	MATERIALS.....	67
9.1	<i>Solutions</i>	67
9.2	<i>Enzymes, chemicals and purification kits</i>	68
9.3	<i>Oligonucleotides</i>	68
10.	METHODS.....	69
10.1	<i>DNA oligo labelling</i>	69
10.2	<i>RNA oligo labelling</i>	69
10.3	<i>RPHPLC purification</i>	70
10.4	<i>Electrospray Ionization Mass Spectrometry (ESI-MS)</i>	70
10.5	<i>Expression and purification of R69Cys and R*69Cys</i>	71

10.6	<i>Electrophoretic mobility shift assay (EMSA)</i>	75
10.7	<i>Pyrene Excimer Fluorescence Studies</i>	75
10.8	<i>Intermolecular Disulfide Crosslinking</i>	77
10.9	<i>Peptide synthesis</i>	78
10.10	<i>Synthesis of Fmoc-Dma(Mts)-OH</i>	80
10.11	<i>Synthesis of Fmoc-Gly(Dmb)Gly-OH</i>	81
10.12	<i>Aminoacid analysis:</i>	83
10.13	<i>Trypsin digestion of RGG and DMA-GG peptides</i>	83
10.14	<i>Double filter binding assays</i>	84
10.15	<i>DNA melting curves</i>	85
10.16	<i>Circular dichroism spectroscopy</i>	85
RESULTS:		87
SECTION I: DIMERIZATION OF THE DNA-BINDING DOMAIN OF BACTERIOPHAGE 434 REPRESSOR ON DNA		87
11.	COMPUTER MODELLING.....	87
12.	EXPRESSION AND PURIFICATION OF THE R69 MOLECULES.....	88
13.	DISULFIDE DIMERS OF R69CYS AND R*69CYS BIND <i>IN VITRO</i> TO THEIR COGNATE OPERATORS.....	91
14.	FLUORESCENCE STUDIES.....	93
14.1	<i>Labeling of R69-Cys with pyrene:</i>	93
14.2	<i>Cognate operators induce monomer fluorescence quenching and excimer fluorescence of R69-Cys-py</i>	94
14.3	<i>Monomer fluorescence is quenched upon addition of non-cognate DNA or other polyanions.</i>	97
14.4	<i>Interaction with a hybrid operator induces significant excimer fluorescence of R69-Cys -Py</i>	99
15.	INTERMOLECULAR DISULFIDE CROSSLINKING	100
15.1	<i>Cognate and hybrid DNA promote R69cys oxidative dimerization</i>	100
16.	DISCUSSION.....	105

SECTION II: N^o-METHYLATION OF ARGININE RESIDUES MODULATES THE NON-SPECIFIC INTERACTION OF THE RGG BOX WITH NUCLEIC ACIDS.....	108
17. SYNTHESIS AND CHARACTERIZATION OF THE PEPTIDES DMA-GG, RGG, KGG, RGG2	108
17.1 <i>Proteolytic cleavage resistance of RGG and DMA-GG peptides.</i>	112
17.2 <i>Circular dichroic spectra of RGG and DMA-GG peptides are indicative of β-turn structure.</i>	113
18. SYNTHESIS AND PURIFICATION OF THE POLYPEPTIDES NUC61ARG AND NUC61DMA.....	115
18.1 <i>Stepwise synthesis of Nuc61Arg and Nuc61Dma</i>	116
19. NUCLEIC ACID INTERACTION STUDIES	122
19.1 <i>Double filter binding assays</i>	122
19.1.1 Arginine dimethylation does not affect the nucleic acid binding properties of the peptide RGG.....	122
19.1.2 RGG and DMA-GG bind dsDNA with lower affinity	126
19.1.3 Nuc38Arg, Nuc61 Arg and Nuc61Dma binding to RNA and ssDNA....	127
19.2 <i>DNA melting curves</i>	129
19.3 <i>Circular dichroism studies</i>	130
19.3.1 The interaction with the RGG peptide affects the nucleic acid dichroic spectrum.	131
19.3.2 DMA-GG and KGG alter the nucleic acid conformation to a much lesser extent than RGG and RGG2.....	132
19.3.3 Influence of ionic strength on the RGG- and DMA-GG-nucleic acid interaction.....	137
20. DISCUSSION.....	139
REFERENCES	144

Figures and Tables

Figure 1 – DNA-recognition by C2-H2 zinc-fingers

Figure 2 – Molecular models of amino acid residues contacting various DNA nucleotides.

Figure 3 - Comparison of sizes of major and minor grooves in A-form and B-form double helices

Figure 4 – Ribbon models of various protein DNA-binding motifs in complex with DNA.

Figure 5 – Amino acid sequences of HTH motifs

Figure 6 - Ribbon model of two 434 repressor dimers bound to a hybrid OR1/ OR2 site.

Figure 7 – Operator sequences of the 434 (A) and P22 (B) repressors.

Figure 8 – Space-filling and ribbon representations of the R1-69 bound to OR1.

Figure 9 - Model representing the hydrogen bonds formed between adjacent base pairs in the R69/OR1 complex.

Figure 10 – Detail of the interactions between amino acid residues and nucleotides in the R1-69/ OR1 complex.

Figure 11 – Contacts between the 434 repressor and operator.

Figure 12 – Diagram of nucleolin domain organization showing the conserved sequences and structural motifs.

Figure 13 – The structures and stepwise synthesis of methylated arginine derivatives found in proteins.

Figure 14 - Origin of the CD effect.

Figure 15 – Far UV CD spectra associated with various types of protein secondary structures.

Figure 16 - CD spectra of PK5.

Figure 17 – Schematic diagram of the production of intermolecular excited dimer (excimer).

Figure 18 – Fluorescence spectra of pyrene-labeled tropomyosin in aqueous solution.

- Figure 19 - Experimental design: specific DNA should act as a template for recruiting R69 molecules.
- Figure 20 – SDS-PAGE gel of the lysate of the bacterial expression of R69Cys before and after induction and cation exchange purification of R69Cys.
- Figure 21 – Non-reducing SDS-PAGE analysis of the R69Cys fractions after cation exchange.
- Figure 22 – ESI-MS of R69Cys disulfide dimer
- Figure 23 – ESI-MS of R*69Cys disulfide dimer.
- Figure 24 – RPHPLC analysis of R69Cys monomer.
- Figure 25 – The Hmb- and Dmb- amide protecting groups
- Figure 26 - Compared chromatograms from the amino acid analysis of peptides RGG and DMA-GG.
- Figure 27 – Computer model of two R69-Cys monomers in complex with DNA.
- Figure 28 - Amino acid sequences of R69-Cys and R*69-Cys; the reaction scheme of labeling R69-Cys with pyrene maleimide; sequences of various synthetic oligonucleotides used in the R69 study
- Figure 29 – Determination of half-maximal binding for (R69Cys)₂.
- Figure 30 - Determination of half-maximal binding for (R*69Cys)₂.
- Figure 31 – Ion exchange chromatography analysis of the reaction mixture containing R69Cys and N-(1-pyrenyl-maleimide
- Figure 32 – The superposition of the absorbance spectrum of R69Cys-py and pyrene labelled N-acetyl-cysteine.
- Figure 33 - Fluorescence spectra of R69-Cys-py in the absence and in the presence of the specific and the non-specific DNAs.
- Figure 34 - Variation of relative fluorescence intensity (F/F₀) of R69-Cys-py at 495nm and at 383nm as a function of the specific oligonucleotide, the non-specific oligonucleotide, and other polyanions.

- Figure 35 - Effect of NaCl on the interaction of R69-Cys-py with the specific oligonucleotide, and the non-specific oligonucleotide.
- Figure 36 - Excitation spectra of R69-Cys-py in the presence and in the absence of aspecific oligonucleotide.
- Figure 37 - Fluorescence spectra of R69-Cys-py in the absence and in the presence of the hybrid oligonucleotide and P22O_{R1}.
- Figure 38 - Disulfide mediated dimerization of R69-Cys in the presence and absence of various specific and non-specific DNA oligonucleotide templates.
- Figure 39 - Extents of R69-Cys dimer formation in the presence of various specific and non-specific DNAs as a function of reaction time.
- Figure 40 - Hybrid DNA promotes disulfide-mediated hetero-dimer formation of R69-Cys and R*69-Cys.
- Figure 41 - ESI-MS analysis of the dimerized product of the reaction mixture R69-Cys, R*69Cys and 434O_{R1}-P22O_{R1} operator.
- Figure 42 - Summary of the mass spectrometric analysis of homo- and hetero- dimers of R69-Cys and R*69-Cys.
- Figure 43 - Comparison of the retention times of the purified peptides DMA-GG and RGG in RP-HPLC.
- Figure 44 - RPHPLC analysis of the crude and purified peptides DMA-GG, RGG, RGG2, KGG.
- Figure 45 - Time-course RPHPLC analysis of the tryptic digestion of peptides RGG and DMA-GG.
- Figure 46 - CD spectra of RGG and DMA-GG.
- Figure 47 - Cation exchange HPLC analysis of the crude synthesis product Nuc61Arg.
- Figure 48 - RPHPLC separation of the Nuc61Arg fraction collected in the IEHPLC).
- Figure 49 - ESI-MS analysis of the main peak from the RPHPLC of Nuc61Arg.
- Figure 50 - Cation exchange HPLC analysis of the crude synthesis product Nuc61Dma.
- Figure 51 - RPHPLC separation of the Nuc61Dma fraction collected in the IEHPLC.

Figure 52 - ESI-MS analysis of the main peak from the RPHPLC of Nuc61Dma.

Figure 53 – RPHPLC coinjection analysis of purified Nuc61Arg and Nuc61Dma polypeptides.

Figure 54 – Betascope images of the nitrocellulose and DEAE filters from a titration of ssDNA with KGG and DMA-GG peptides and from a titration with RNA of KGG and RGG peptides.

Figure 55 – Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiments obtained titrating a nucleic acid fragment (1nM) with increasing concentration of KGG, RGG, and DMA-GG peptides.

Figure 56 - Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiment obtained titrating the dsDNA fragment (1nM) with increasing concentration of KGG, RGG and DMA-GG peptide.

Figure 57 - Betascope image of the nitrocellulose and DEAE filters from a titration of the RNA fragment with Nuc38Arg, RGG and DMA-GG peptides.

Figure 58 - Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiment obtained titrating the RNA fragment (1nM) with increasing concentration of Nuc38Arg, RGG and DMA-GG peptides.

Figure 59 - Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiment obtained titrating the ssDNA fragment (1nM) with increasing concentration of Nuc61Arg and Nuc61Dma peptides and relative betascope image.

Figure 60 - Melting curves at 260 nm of the A₂₅*T₂₅ dsDNA oligonucleotide alone and in the presence of a 50-fold molar excess of the KGG, RGG and DMA-GG peptides.

Figure 61 - Circular dichroic spectra of MS2 phage RNA in the presence of increasing concentrations of RGG peptide.

Figure 62 - Interaction of RGG, DMA-GG, RGG2 and KGG peptides with MS2-phage RNA.

Figure 63 - Circular dichroic spectra of MS2 phage RNA in the presence of increasing concentrations of DMA-GG peptide.

Figure 64 – Interaction of RGG and DMA-GG peptides with tar-RNA.

Figure 65 - Interaction of RGG and DMA-GG peptides with yeast tRNA.

Figure 66 - Interaction of RGG and DMA-GG peptides with single stranded calf thymus DNA.

Figure 67 - Interaction of RGG, DMA-GG, RGG2 and KGG peptides with double stranded calf thymus DNA..

Figure 68 - Interaction of RGG peptide with MS2phage RNA at increasing NaCl concentrations.

Figure 69 - Interaction of RGG and DMA-GG peptides with the ssDNA oligonucleotide at increasing NaCl concentrations.

Table 1 - DNA base binding specificity of helix-turn-helix proteins according to the shape and size (small, medium, large, aromatic) of the amino acid involved.

Table 2 – A partial list of SWISS-PROT protein sequences containing the RGG-box.

Table 3 – Analytical data of the synthesized peptides based on the RGG motif of human nucleolin.

Table 4 – Summary table of the nucleic acid interaction studies performed with the various synthetic peptides.

Introduction

Protein-nucleic acid interactions occur at all levels of cellular life cycle ranging from DNA replication to gene expression. Nucleic acid binding proteins recognize their target nucleic acids at such a high specificity that allows vital processes being turned on and off without interfering with each other. In fact, proteins such as transcriptional activators, repressors, restriction endonucleases, snRNPs, tRNA synthetases, show extensively high specificity for their special target sites. The transcription factors for instance, are able to find a single site of 10-20 base pairs in a background of up to 10^6 - 10^9 or so base pairs. This recognition specificity is thought to be essential for the regulation of gene expression and hence for the control of cell growth and differentiation. On the other hand, many proteins that take part in equally well and precisely regulated processes, such as histones and polynucleotide polymerases, have very low or even no *in vitro* detectable sequence specificity. Our understanding of specific nucleic acid recognition is based on growing number of protein/nucleic acid complexes provided by X-ray crystallography and NMR spectroscopy. There are much less data on non-specific interactions.

The aim of my thesis is to study the non-specific interactions between proteins and nucleic acids, in order to find out how weak binding can contribute to highly specific interactions. I have used two model systems. The first one is the monomeric N-terminal domain of the 434 repressor, a peptide of 69 amino acids that exhibits a very weak binding to its cognate half site in solution. The second system is the RGG motif of human nucleolin; RGG motifs are known to bind RNA and DNA in various systems and undergo a post-translational dimethylation of the arginine residues to give N^ωN^ω-dimethylarginine. In this work, the protein substrates were produced by recombinant DNA and solid phase peptide synthesis methods, and the interactions were studied by various *in vitro* methods including circular dichroism and fluorescence spectroscopy.

The thesis consists of three sections. The introduction (this section) summarizes the structural motifs of nucleic acid recognition and the main mechanisms of specific and non-specific binding, as well as the relevant background on spectroscopic methods. At the end of the introduction there is a short summary on the aims of this thesis.

The Materials and Methods section summarizes the techniques and the experimental conditions used. The Results and Discussion section is composed of two subsections containing the experiments with the 434 repressor N-terminal domain and the RGG motif, respectively.

1. DNA and RNA recognition by proteins

The understanding of the molecular basis of protein-nucleic acid recognition requires a characterisation of the conformational properties of the protein, the nucleic acid target site, and the changes that ensue as a consequence of the interaction. A wealth of biochemical and structural information has been accumulated to illustrate the details of protein-nucleic acid interaction in numerous instances. Though from the structural studies some rules have been derived for individual structural families (Choo and Klug, 1997) or for a group of families that interact in similar ways with DNA (Suzuki, 1993; Suzuki, 1994; Suzuki et al., 1995a; Suzuki and Yagi, 1994), the complicated nature of recognition mechanism precludes a simple recognition code. The formation of a protein-nucleic acid complex presupposes that chemical nature and the three-dimensional arrangement of the functional groups of the protein must match those of the nucleic acid target site. The same physical rules that determine protein and nucleic acid structure govern their specific and non-specific interactions (Rhodes et al., 1996). General principles for recognition described below are mainly based on the physical and chemical properties of amino acids and nucleotides, including chemical and shape recognition. (for reviews see Pabo and Sauer, 1992; Rhodes et al., 1996; Sinden, 1994)

1.1 Intermolecular forces involved in the chemical recognition

Protein and DNA molecules will interact if there is a decrease in Gibbs free energy on the formation of a complex. The change in free energy (ΔG) during complex formation depends on the change in both entropy (ΔS) and enthalpy (ΔH , due to short range non-covalent interactions) such as that $\Delta G = \Delta H - (T\Delta S)$, where T represents temperature. Both the enthalpy (ΔH) and entropy (ΔS) terms depend on the shape of the surface between the protein and its target DNA (Rhodes et al., 1996)

The forces involved can be classified into four categories:

1.1.1 *Electrostatic forces: salt bridges*

Salt bridges are defined as electrostatic interactions between groups of opposite charge; they typically provide about 40 kJ mol^{-1} of stabilization per salt bridge. Generally they occur between the ionised phosphates of the nucleic acid and either the ϵ -ammonium group of lysine, the guanidinium group of arginine, or the protonated imidazole of histidine (this is the only amino acid which can exist in both acidic and basic form at physiological pHs). Salt bridges are influenced by the concentration of salts in the solution: an increase of salts weakens their strength, whereas, thanks to the high dielectric constant of water, their strength increases in the absence of water molecules between the ionised groups. Salt bridges can be considered as long-range interactions, the energy of the interaction is proportional to the inverse of the separation of the charges (Fersht, 1990) but is rather insensitive to their relative orientation. Salt bridges, therefore, confer little specificity to the protein-nucleic acid interaction. Especially in B-DNA, changes in sequence perturb the average structure subtly. Salt bridges alone are not able to discriminate different sequences of B-DNA but when arranged in patterns they are able to distinguish single-stranded from double-stranded nucleic acids (Blackburn and Gait, 1996).

1.1.2 *Dipolar forces: hydrogen bonds*

Hydrogen bonds are a result of dipole-dipole interactions:



Hydrogen bond energy diminishes with the inverse of the sixth power of the distance between the centers of the dipoles and also decreases greatly if the bond is bent (if X, H and Y are not co-linear) (Fersht, 1990). X and Y are almost always nitrogen or oxygen in biological macromolecules.

In protein-nucleic acid complexes the hydrogen bonds occur between amino acid side chains, the backbone amides and carbonyls of the protein, and the bases and sugar-phosphate oxygens of the nucleic acid, sometime with the mediation of water molecules. When biomolecules are not involved in a complex, all their exposed hydrogen bond donors (X) and acceptors (Y) form linear hydrogen bonds to water. When the complex forms, there is little change in the free energy due to hydrogen bond formation if the linear hydrogen bonds to water are replaced by similar ones between the macromolecules. Forming bent hydrogen bonds or not forming them at all carries a free energy penalty of around 20 kJ mol^{-1} (Fersht, 1990), thus hydrogen bonds are very important in making sequence-specific protein-nucleic acid interactions (Blackburn and Gait, 1996).

1.1.3 Entropic forces: hydrophobic interactions:

A molecule in water creates a curved interface, ordering a layer of water molecules around itself. Upon complex formation many of the ordered water molecules will be displaced and will become part of the disordered bulk water of the solvent (Garner and Rau, 1995). This phenomenon results in an entropic contribution which favours the protein-nucleic acid binding (Schwabe, 1997). The change in free energy is roughly proportional to the number of water molecules released and, therefore, to the surface area which gets buried (Blackburn and Gait, 1996).

1.1.4 Dispersion forces: stacking interactions

Nucleic acid base stacking in aqueous solution is caused by two factors: the hydrophobic interactions mentioned above and the dispersion forces. At any instant, the electronic charge distribution within atomic groups is asymmetric due to electron fluctuations (Saenger, 1984). Therefore, dipoles created in one group of atoms polarize the electronic system of neighboring atoms or molecules, thus inducing parallel dipoles which attract each other (London, 1930). The interaction energy decreases with the sixth

power of the distance (Fersht, 1990) and is therefore very sensitive to the thermal motion of the molecules involved .

Despite their extreme distance dependence, dispersion forces are clearly important in maintaining the structure of double stranded nucleic acids. Furthermore, they also help single stranded regions in nucleic acids bind to proteins because aromatic side chains can intercalate between the bases of a single stranded nucleic acid (Blackburn and Gait, 1996; Gabbay et al., 1972).

1.2 Chemical recognition at the protein-nucleic acid interface

The unique arrangement of the nucleotide functional groups within the major or minor groove provides the specificity used by proteins to discriminate regions of nucleic acid sequence. There does not appear to be a code for nucleic acid sequence recognition comparable to the triplet code for amino acids, but a number of specific interactions between certain amino acids and nucleic acid base pairs have been demonstrated. Chemical rules based merely on the intrinsic ability of a given residue

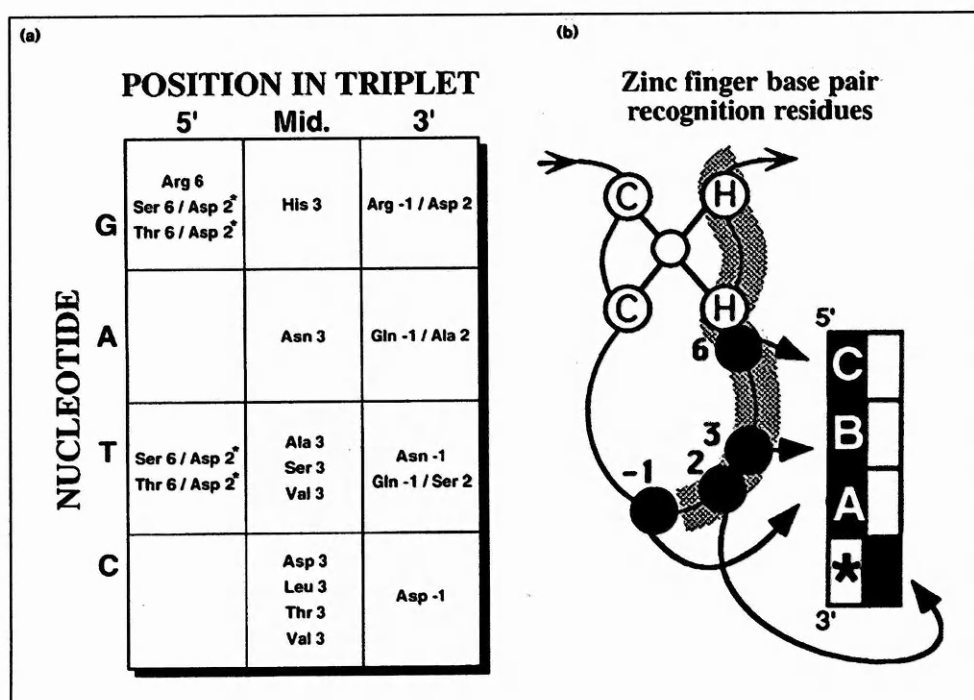


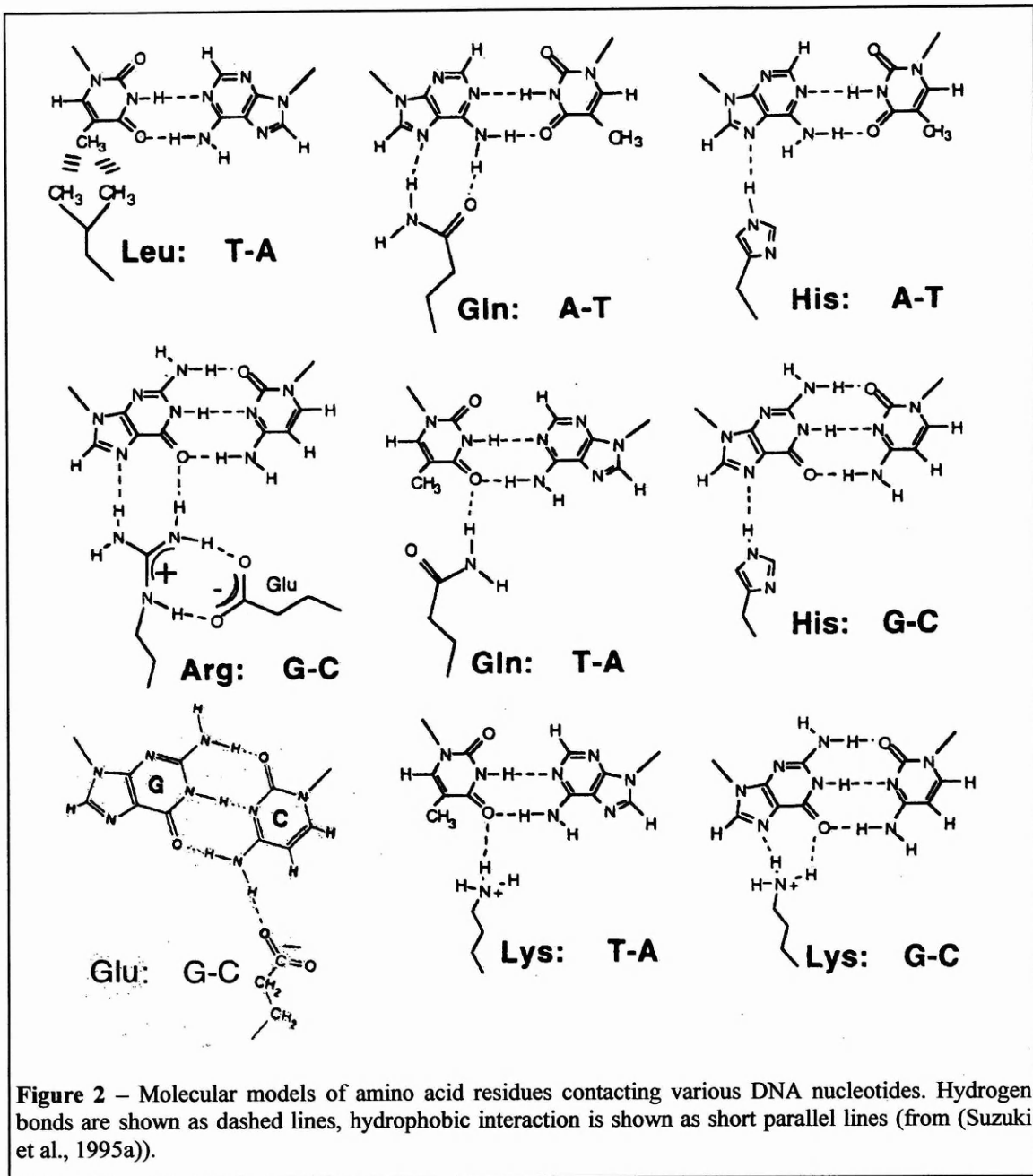
Figure 1- DNA-recognition by C2-H2 zinc-fingers: (a) Consensus zinc-finger recognition code derived from screening a zinc-finger library (Choo and Klug, 1994). The helical position follows the residue name. An asterisk indicates a base within the binding site of an adjacent zinc-finger. (b) Schematic representation of a zinc-finger showing the positions of the amino acid to base contacts (Rhodes et al., 1996).

and a base to produce a non-covalent interaction have limited applicability because there seems to be great variation in the way that each amino-acid is employed to interact with DNA base-pairs. For each of the different DNA-binding motifs, such as the helix-turn-helix, homeodomain, zinc-finger and hormone receptor, patterns of base-residue contacts, deduced from crystal and NMR structures, seem to be reasonably conserved for members of the same family (Rhodes et al., 1996). These patterns allow the construction of tables (**Figure 1** and **Table 1**) for single amino acid-single base contacts whose validity is apparently limited to members of the same type of DNA-binding domain.

	Small	Medium	Large	Aromatic
A	C, S, T	N, D, K	Q, E, R, K, M	Y, W
T	A, C, S, T	V, I, N, H	L, M, Q, R, K	Y, F, W
G	C, S, T	H, N	R, K, Q	Y
C	V, C, S, T	D, N, H, I	E, Q, L, M	Y, F, W

Table 1 - DNA base binding specificity of helix-turn-helix proteins according to the shape and size (small, medium, large, aromatic) of the amino acid involved (Suzuki, 1994; Suzuki et al., 1995a): specific partners are shown in bold

Some other recently published studies (Jones et al., 1999; Nadassy et al., 1999) analyzed a considerable number of protein-nucleic acid complexes in the PDB, from which some common themes could be deduced. Ninety percent of the protein-nucleic acid hydrogen bonds have the donor group on the protein and the acceptor group on the nucleic acid; in dsDNA the phosphate group is involved in 60% of hydrogen bonds, the deoxyribose sugar oxygens in 6% and the nucleobases provide the remaining 34%. The distribution is about the same in ssDNA complexes but in RNA ones, bonds to the sugar are much more abundant and the 2'-hydroxyl plays a major part (20%), second only to the phosphate. The bulky 5'-methyl group thymine is involved in van der Waals



contacts with the methyls of several amino acids, as well as creating a stereo hindrance for incorrect binding (Seeman et al., 1976; Suzuki and Yagi, 1994).

The neutral main chain NH group of the protein and the charged side chains of Arg and Lys are the major hydrogen bond donors and the two charged groups form salt bridges to the phosphates.

The most common contacts (**Figure 2**) directly to the nucleobases are the two hydrogen bonds formed between Gln (or Asn) amide groups and adenine in the major

groove, and similar interactions between Arg guanidinium group (or Lys amino group) and guanine.

The protein surface that contacts DNA is more polar than the average accessible surface, the positively charged arginine residue shows the highest propensity to stay at the interface, followed by the polar threonine and asparagine residues, and the lysine residue. The negatively charged aspartic and glutamic acids instead, are rarely found at interfaces (Jones et al., 1999) but sometime can contact the DNA backbone through divalent cations such as Mg^{2+} and Ca^{2+} (Pabo and Sauer, 1992).

1.3 Role of water molecules

Apart from the favourable entropic contribution following water displacement, water molecules play other roles in both DNA and RNA recognition. Ordered water molecules have been observed in most protein-DNA interfaces. Water can form base-specific hydrogen bonds only if it is forming at least two other H-bonds with donors or acceptors on the protein and is sequestered from the bulk solvent. In this way the water molecule can be specifically oriented by interactions with the protein turning it into a surrogate side-chain. In trp-repressor-DNA complex, there are three water molecules per half operator bound in the major groove between the protein and the DNA bases; at least two of them appear to be making H-bonds that specify base pairs from the dyad axis. (Otwinowski et al., 1988). Interestingly, the DNA in the absence of trp-repressor protein presents many of these water molecules in essentially identical positions (Shakked et al., 1994). Thus it appears that the protein is recognizing the DNA together with the associated water structure (Rhodes et al., 1996). In the glutaminyl-tRNA synthetase complex with tRNA, two buried water molecules are an integral part of the H-bond matrix presented in the shallow groove of the tRNA acceptor stem (Rould et al., 1989).

In non-specific complexes, water may also act as a kind of lubricant allowing the protein to scan along the DNA for the specific binding site (Schwabe, 1997). Further

examples of water molecules present at the protein-nucleic acid interfaces will be described in paragraph 3.2.1.

2. Structural complementarity of the interacting protein and nucleic acid

Structural studies have shown a remarkable shape complementarity as well as the local electrostatic and van der Waal complementarity at the protein nucleic acid interface. This complementarity is strictly dependent on the protein and nucleic acid structural parameters and on their higher or lower intrinsic propensity to adapt their tridimensional arrangement. Here follows a short description of the main structural characteristics which allow this complementarity.

2.1 Similarities and differences in the recognition by DNA and RNA helices

DNA-binding proteins interact with duplex DNA, which is generally in B-form. In B-DNA the major groove is wide enough (11.7 Å, defined as phosphate-phosphate distance diminished by 5.8 Å to account for van der Waals radii of phosphate groups) to accomodate either a α -helix or a β -ribbon, and the functional groups on the exposed

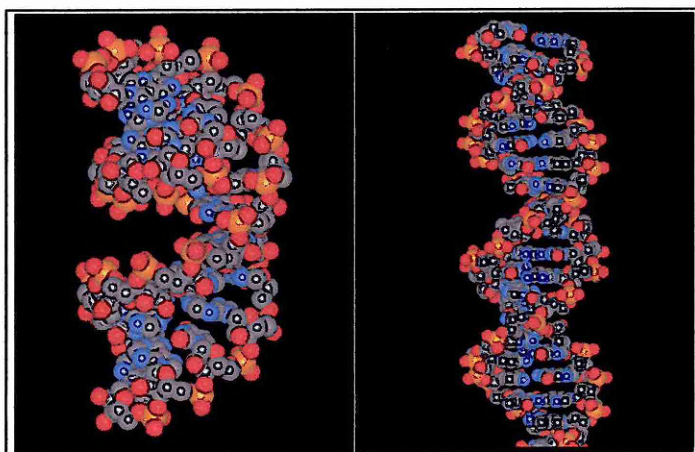


Figure 3 - Comparison of sizes of major and minor grooves in A-form and B-form double helices. Left panel: A-form RNA. Right panel: B-form DNA

edges of the base pairs can be directly contacted by side chains of the protein (**Figure 3**). The minor groove is deep and narrow (5.8 Å) and thus less accessible to secondary structures like α -helices, whose diameter is around 4.6Å (Schulz and Schirmer, 1979).

RNA is distinguished from DNA most readily by the presence of the ribose 2' hydroxyl and by the fact that the extended RNA double helix is predominantly A-form.

RNA in A-form shows an opposite geometry with respect to B-DNA: the minor groove is shallow and broad (10-11 Å), whereas the major groove is very deep and narrow (4 Å) (**Figure 3**). These general features allow non-specific dsRNA-binding proteins to recognize their correct nucleic acid substrate.

A second major consideration for the suitability of the major or minor grooves in both RNA and DNA for direct sequence recognition, is the degree of structural variation of the four base pairs as viewed from the two grooves. Seeman et al. (Seeman et al., 1976) pointed out that the distinction of AT from GC base pairs in the minor groove relies only on the N2 of guanine; on the other hand the pattern of hydrogen bond donors and acceptors in the major groove allows to distinguish all four base pairs. Furthermore RNA helices can present non-Watson-Crick base pairs like GU, or AG and UU, offering in both grooves hydrogen-bonding and shape differences not available in the four orientations of the two Watson-Crick base-pairs. RNA can be also targeted in the single-stranded regions within secondary structures or in sites of local distortions induced in double helical regions.

2.2 DNA recognition and protein folds

Many DNA-binding proteins achieve specific recognition through small, discrete, independently folded structural units. Most of the structures identified so far fall into a few types, each type having characteristic amino acid sequence and three dimensional structure. The main structure families are (**Figure 4**):

- 1) the helix-turn-helix (HTH) motif, including a large family of prokaryotic transcription factors and the eukaryotic homeodomain (Gehring et al., 1994; Wolberger, 1996);

- 2) the zinc-binding proteins, at least three structural classes have been identified including the class I zinc-finger (or Zn(Cys₂-His₂) domain, the Zn twist [Zn(Cys)₄]₂ domain, and the binuclear Zn cluster Zn(Cys)₆ domain;
- 3) the leucine zipper (bZIP) and basic helix-loop-helix (bHLH) motifs (Ellenberger, 1994).
- 4) the β -ribbon;
- 5) the TATA box binding proteins (Burley, 1996).

The helix-turn-helix motif, which is involved in the work of this thesis, is discussed below in detail. For other motifs, see also (Freemont et al., 1991; Harrison, 1991; Klug, 1993; Pabo and Sauer, 1992; Sinden, 1994) for reviews.

The majority of structures of protein-DNA complexes show interaction in the major groove due to the physical limitations of fitting secondary structures in the minor groove. Minor groove binding is only observed in complexes where the DNA structure is significantly distorted, resulting in the widening of the groove to allow the entry of β -sheets (e. g. TATA box binding protein, (Kim et al., 1993)) or alpha helices (e.g. $\delta\gamma$ -resolvase, (Yang and Steitz, 1995)). In some complexes simultaneous binding exist in the major and the minor groove. (e.g. Hin recombinase, (Feng et al., 1994) and RAP1, (Konig et al., 1996)) and the binding involves an helix-turn-helix motif in the major groove and a trailing loop in the minor groove; in these complexes the minor groove interactions are as important for base recognition as the HTH motif.

Most of the well-characterized families of DNA-binding proteins use α -helices to make base contact in the major groove. The overall shape and dimension of an α -helix allows it to fit into the major groove in a number of related but significantly different ways. Some lie in the middle of the major groove and have the axis of the α -helix approximately tangential to the local direction of the major groove. Others are tilted at different angles, and some are arranged so that only the N-terminal portion of the α -helix fits completely into the major groove. The surrounding region of the proteins helps to determine how these α -helices are positioned in the major groove, by

maintaining the global architecture of the protein or stabilizing the binding geometry through the non-specific phosphate backbone contacts (Pabo, 1984). Besides the intrinsic chemical ability, one must keep in mind that a favourable hydrogen bond can be made only if the interacting groups in both nucleic acid and amino acid side chains are in an appropriate position and orientation. These are affected by the binding geometry of the interacting molecules.

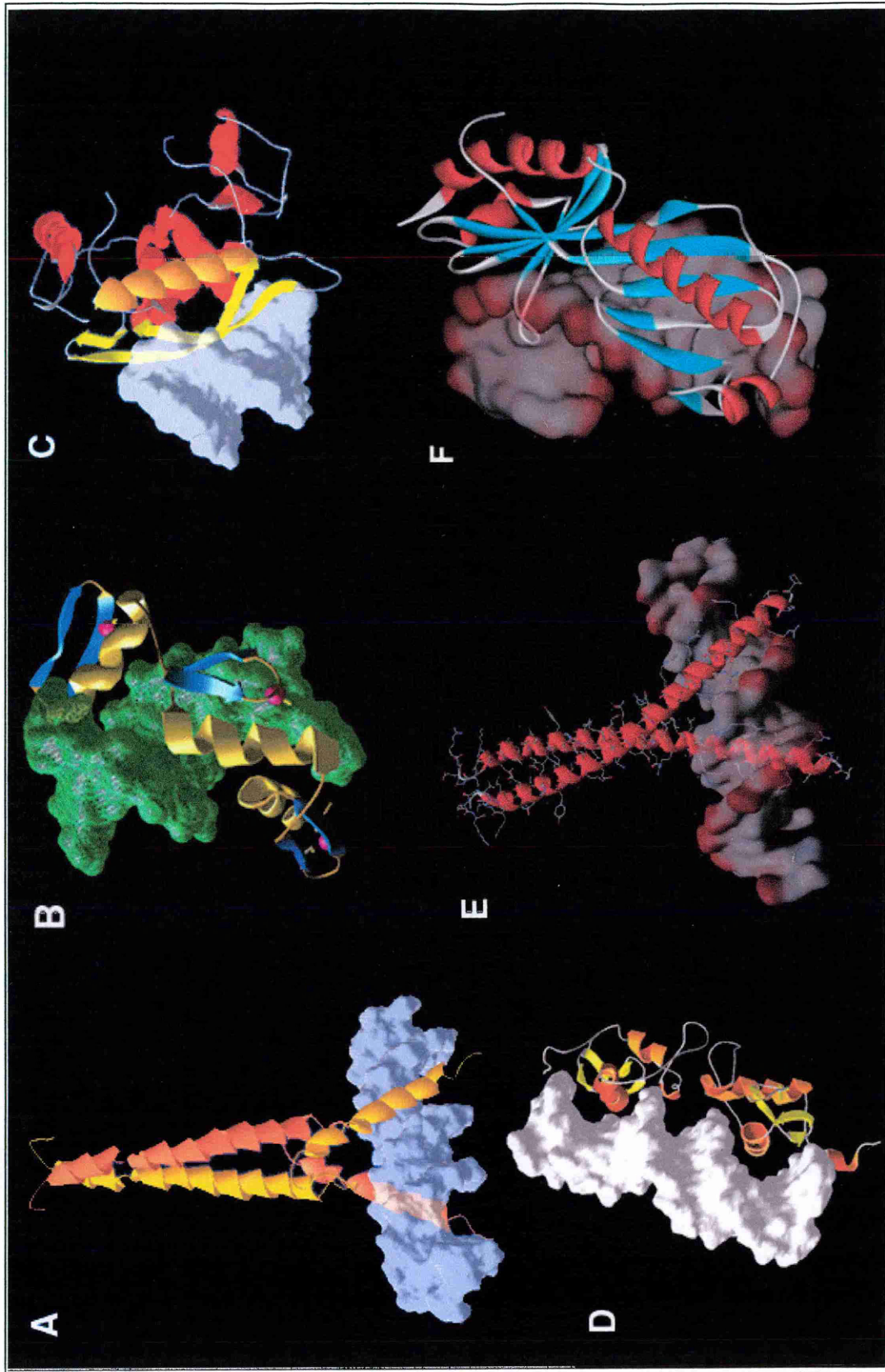


Figure 4 – Ribbon models of various protein DNA-binding motifs in complex with DNA, from resolved X-ray structures. (A) bHLH/z of Max homodimer (PDB entry 1AN2), (B) C2H2 zinc-finger of Zif268 (1ZAA), (C) β -ribbon of methionine repressor (1MJM), (D) C4 zinc-finger of glucocorticoid receptor (1GLU), (E) bZIP of yeast GCN4 (1DGC), (F) β -sheet of TATA box binding protein (1QN4). The models were created with the programs Swiss PdbViewer and WebLabViewer.

The DNA phosphate backbone, with relatively uniform shape and negative charge, plays an integral role in the site-specific recognition. In principle, any basic or neutral hydrogen-bonding side chain can be used to contact the phosphodiester oxygens, while it seems that short polar side chains and the peptide -NH may provide more stereo specificity than those mediated by the long flexible side chain of Arg or Lys. Backbone contacts may serve to hold the protein against the bases in a fixed arrangement and thereby enhance the specificity of side chain-base interaction, or help to establish the DNA conformation change according to the requirement of protein binding. In the specific recognition, the size of amino acid residues should be compatible with the requirement of the contacts. From a fixed position on the interaction surface, a long side chain can reach further and deeper in the DNA groove; whereas at a position very close to the DNA, a small residue can easily fit in but a bulky residue may not. Hence, contacts with the DNA backbone are important for the complementary recognition.

The specific interaction between proteins and DNA has most frequently been considered from the viewpoint of protein, due to the large structural diversity of proteins. However, the DNA conformation and configuration also have a profound influence in the process of recognition. Local or global DNA distortions (i.e. groove width, charge, bending or/and twisting) were observed in many protein-DNA complexes, so that the changed DNA structure can better accept a protein secondary structure.

The DNA double helix could adopt discrete conformations depending on the sequence and on the degree of hydration, so that the regular A and B forms do not adequately describe DNA in solution. Sequence dependence of DNA distortion is evident in both intrinsic curvature of DNA (Hagerman, 1990) and protein-induced DNA bending (Olson et al., 1998; Suzuki and Yagi, 1995). Tight bending of DNA requires that the stacking interactions of the base pairs can accommodate the deformations in the groove width. Generally the kinks observed in the complexes occur at pyrimidine-purine base steps. The most extreme example of DNA distortion is seen in the structure

of the TATA binding protein bound to its DNA target in which the DNA has two 90° bends (Kim et al., 1993); the structural properties of the TATAAAA sequence play an important role in this distortion (Rhodes et al., 1996). When proteins recognize a specific DNA sequence, they 'read' the base sequence either through direct interactions, or through recognizing features of the overall DNA structure that are dependent on the base sequence. The latter type of recognition has been termed *indirect readout* or *analogue recognition* in contradiction to the *direct readout* or *digital recognition* of individual bases. Primary sequence dictates local twist angle, the specific tilt and roll of bases, bends in DNA, and width of the major and minor grooves, as well as the ease or flexibility for structural changes. These structure parameters will precisely position the hydrogen bond donor and acceptor sites in the bases and the phosphate backbone in space. Most DNA binding proteins are designed to recognize a particular shape or flexibility of the double helix in addition to a direct readout of individual bases in the recognition site. In many cases, changes in the nucleotides within the binding site that are indirectly contacted by the protein strongly influence the binding affinity of a protein. (for review see Travers, 1989). Evidence for this is that sometimes bases are highly conserved in different DNA targets, yet are not in direct contact with the protein (Schwabe et al., 1993).

2.3 RNA recognition and protein folds

On the basis of the A-form structure the interactions with RNA sequences seem favoured via the minor groove. However, in the literature examples of protein-RNA interaction both in the minor and in the major groove can be found (Rould et al., 1989; Ruff et al., 1991). Although it is true that the edges of base-pairs are inaccessible in the major groove of A-form RNA in the central portion of a long duplex, most naturally occurring RNA molecules are single-stranded and tend to form irregular structures containing short duplex regions interspersed with hairpins, mismatches, bulges or loops, exposing the base pair edges in the major grooves at the ends of these RNA helices

(Cusack, 1999). RNAs, even more than DNAs, may be specifically recognized by *indirect readout*: the correct RNA can distort to fit the protein whereas the incorrect one can not and some irregularities (e.g. major groove widening) may be specifically induced or enhanced by the protein-RNA interaction (Cusack, 1999). In this respect, for instance, the sex lethal protein in complex with its RNA target presents two RNA-binding domains forming a V-shaped cavity; nine consecutive nucleotides are bound to the surface of the cavity forming a sharp turn stabilised also by hydrogen bonds between 2'-OH sugar hydroxyls of three nucleotides and the protein; as a result of this, the protein shows a 10^4 fold lower affinity for a similar DNA sequence (Handa et al., 1999; Kanaar et al., 1995).

Primary sequence analysis has led to the identification, in functionally diverse RNA-binding proteins, of a number of recurring RNA-binding modules (for reviews see (Antson, 2000; Cusack, 1999; Mattaj and Nagai, 1996)), the most well known being:

- 1) ribonucleoprotein fold (RNP, e.g. hnRNP A1, snRNP U1A, nucleolin, sex lethal and PABP proteins),
- 2) K homology domain (KH, e.g. hnRNP K and FMR1 proteins)
- 3) double-stranded RNA-binding domain (e.g. Xlrpba and protein kinase PKR proteins)
- 4) oligonucleotide/oligosaccharide binding (OB) fold (e.g. transcription factor Rho)
- 5) The RGG box (e.g., hnRNP U, SSB-1, hnRNP A1 and nucleolin proteins)

In several of the above cited protein-RNA complexes single-stranded RNA binds to a surface of the protein formed by one or multiple β sheets. One of the reasons for preference of β -sheet over α -helical surfaces could be the more stable nature of the β -sheet, favouring binding of an extended ligand molecule. In contrast, α -helices usually have a more flexible nature, which could prevent the binding of extended and flexible ligands.

In the complex of TRAP (Trp RNA-binding attenuation protein) bound to its target ssRNA, the protein interacts very specifically by direct and water-mediated hydrogen bonds with A2 and G3 nucleotides of a GAG triplet and the bases of these two nucleotides stack each other and with the aromatic ring of a phenylalanine (Antson et al., 1999). Indeed, in single stranded RNA portions, the predominantly hydrophobic bases are much more exposed and are not involved in base pairing, allowing them to make a significant number of interactions. Single stranded nucleic acid binding proteins will therefore have a much more hydrophobic interacting surface containing aromatic groups and a general electrostatic field to neutralize the charge of the phosphate backbone. The resolved protein structures reveal binding pockets determined by residues providing van der Waals and stacking interactions and base specificity generated by multiple hydrogen-bond donors/acceptors at the edges of these pockets (Antson, 2000).

Considerable interest has recently focused on the arginine rich peptides that are components of the viral and phage proteins involved in RNA recognition; these short peptides are primarily unstructured when free in solution, but adopt distinct folds on complexing with their RNA targets (see paragraph 3.3.1) (Patel, 1999). Initially identified as the RNA-binding domain of hnRNP U, the RGG box is a conserved motif which has been found in numerous RNA-binding proteins and, being the subject of our research, will be more extensively described in paragraph 6.

3. Formation of the complex and folding events

3.1 Multiple subsite recognition

Usually, a single structural unit, such as the α -helix can access only one side of DNA and contact with a short (3-5bp) DNA sequence so that is not sufficient to confer specific and high affinity binding. In nature, this problem is overcome by multiple subsites binding, which are carried out by two or more structural units. These units can

be part of the same polypeptide chain or be different subunits of multimeric proteins. The affinity of a protein for its DNA binding sites is a result of the number and strength of electrostatic and hydrophobic interactions between the protein and DNA. A larger binding site can provide a stronger interaction between the DNA and a protein than a smaller site.

There are several ways in which the size of a binding site can be increased so as to allow for strong and specific binding. The first is simply to add on arms or tails that recognize additional features of the DNA, particularly in the minor groove (eg. $\alpha 2$ and engrailed homeodomains, yeast RAP1 and Hin recombinase). The second is to form homo- or hetero-dimers (eg. Prokaryotic regulatory proteins of the HTH family such as phage λ and phage 434 repressor proteins and several eukaryotic regulatory proteins belonging to the bZIP and bHLH families). Protein oligomerization in fact increases specificity, by increasing the total buried area and the number of contacts between the protein monomers. The third is to employ multiple DNA binding domains either by using tandem repeats of the same type of DNA-binding motif or by linking together different types of motifs (e.g. the Zn-finger motif and Oct-1 POU domain) (Rhodes et al., 1996).

Multiple binding sites can also be used to modulate the binding affinity. Instead of a unique DNA sequence, most specific DNA-binding proteins recognize a set of related sequences with varied binding affinity. This is achieved in several ways:

- 1) Readjustment of the protein side chains or DNA structures so as to form a different interaction network between the functional groups in DNA and the protein.
- 2) Using the spacer between the subsites to direct the overall structural arrangement of the individual binding motifs.

3.2 The kinetics of protein-nucleic acid complex formation

The rate of formation of all protein-nucleic acid complexes is affected by two main factors: random thermal diffusion and long-range, directional electrostatic attraction. According to a simple model, a nucleic acid binding protein could find its cognate sequence by random association with the nucleic acid, followed by dissociation and reassociation elsewhere if the sequence is not correct. Such a model is a three-dimensional random walk through the contents of the entire cell. Inside prokaryotic and eukaryotic cells, there is an extremely high concentration of dissolved macromolecules (300-400 mg ml⁻¹). This means that the environment inside the cell is much more like that in a macromolecular crystal than in solution. This macromolecular crowding enhances the affinity of intermolecular interactions and reduces the macromolecular diffusion rates (Rhodes et al., 1996); in even the smallest bacterial genome, sequence-specific protein-DNA complexes form too rapidly for the model of random association to be correct (Blackburn and Gait, 1996). An alternative and generally proposed model which can account for the observed rates of complex formation is one-dimensional diffusion: the protein first binds non specifically to the DNA and then slides along the DNA until it finds the target sequence. In a study performed by Kim and coworkers (Kim et al., 1987) the equilibrium association and dissociation constants were measured for all six operators bound by Cro repressor both with operators of the same length and with operator-containing DNA fragments of increasing length. The collected data demonstrated that the association and dissociation rate constants increase as the length of the operator-containing fragment increases: on the other hand the affinity of Cro for the different operators remains essentially unaffected. These kinetic studies have shown Cro repressor binding consistent with the sliding mechanism (Kim et al., 1987).

3.2.1 *Non-specific and specific binding*

The model of one-dimensional scanning of the nucleic acid requires proteins to distinguish non-specific from specific binding but this mechanism is still unclear

because there are too few resolved structures of specific proteins bound to non-specific sites. Several DNA-binding proteins interact with DNA in two distinctly different fashions according to the specific or non-specific character of the bound nucleic acid and a number of common themes can be deduced from some of the resolved complexes.

In the glucocorticoid receptor complex structure (Luisi et al., 1991), the protein dimer is bound by its Cys₄ zinc domains to a cognate and to a non-cognate half-site; one of the two half sites in fact is 4bp apart instead of 3bp from the other one, forcing one subunit of the protein dimer out of register by 1bp; in the specific half-site, direct readout of the sequence occurs at four bases by direct and water-mediated hydrogen bonds and seven contacts are formed with the phosphate backbone. In the non-specific half-site, on the other hand, only one base is H-bonded and only five phosphates are contacted. Therefore the non-cognate half-site appears to be less tightly associated with the protein and the α -helical recognition element is less buried in the major groove (Luisi et al., 1991).

The EcoRV endonuclease structure was also solved in complex with cognate and non-cognate DNA; in the cognate complex the area of the nucleic acid-protein interface is 800 Å larger than in non-cognate complex; the complex shows a high rate of packing and three times as many inter-molecular hydrogen bonds as the non cognate complex which is poorly packed; furthermore, the specific complex requires a larger degree of DNA distortion (Jones et al., 1999).

In the structure of BamHI endonuclease dimer bound to non-cognate DNA (Viadiu and Aggarwal, 2000) the nucleic acid is accommodated loosely within a cleft at the bottom of the BamHI dimer and protrudes outside instead of being almost surrounded by the enzyme such as in the specific complex. The enzyme is tilted at a different angle resulting in markedly different DNA binding surfaces in the two complexes; the buried solvent-accessible surface area decreases dramatically from 4350 Å² to 1489 Å² going from the specific to non-specific structure; all of the base-specific interactions and DNA backbone contacts are lost, substituted by a remarkably smaller

number of water-mediated hydrogen bonds mostly with the DNA backbone. Much of the electrostatic stabilization of the non-specific complex appears to arise from the helix dipole moments of several alpha helices projecting towards the DNA backbone. The volume of the cavity at the protein-DNA interface is much larger in the non-specific complex and is filled with a higher number of water molecules. Often, a large number of water molecules are shown to be involved in non-specific DNA binding, probably cushioning the interface and allowing electrostatic interactions to dominate. This dominance is emphasized by a strong dependence of this mode of binding on salt concentration (Murphy IV and Churchill, 2000; Sidorova and Rau, 1996). Like many DNA binding proteins, BamHI finds its target site faster than the three-dimensional diffusion limit and shows an increase in the cleavage reaction as the length of non-specific DNA around the cognate site is increased (Nardone et al., 1986): the loose mode of binding and the overall lack of DNA backbone contacts may facilitate diffusion, reducing the lifetime of the non-specific complex and by lowering the activation energy for the breaking and reforming of DNA contacts as the enzyme moves (Viadiu and Aggarwal, 2000).

An engineered monomeric form of Cro repressor from phage λ has recently been crystallised bound to a not palindromic DNA duplex (Albright et al., 1998). The protein binds the nucleic acid in a sequence non-specific manner: the protein orientation relative to the DNA is markedly rotated with respect to the wild-type complex, no base-specific contacts are made by the helix-turn-helix motif which is responsible for all the specific interactions in the wild-type complex and only the turn of the HTH remains close to DNA; the nucleic acid lacks major conformational changes and the phosphate positions along the sequence are preserved. Indeed, the Cro monomer interacts with the DNA predominantly through the sugar-phosphate backbone and only a set of salt bridges to positively charged side chains in a channel on the surface of the protein are maintained, while the ones to non-charged side chains in another region of the complex are lost. The buried surface area of the non-specific complex is 25% less than the wild-type and there

is a substantial loss of hydrogen bonding and van der Waals interactions, the overall nature of the contacts is therefore more ionic in character (Albright et al., 1998). The increased ionic nature of the non-specific complex is reflected in its greater salt-sensitivity (Takeda et al., 1986)

3.3 Induced fit folding events

When two macromolecules interact, one or both partners often undergo structural rearrangements to establish a complementary binding interface. Such induced fit interactions (or adaptive binding) are quite diverse, ranging from fine adjustments of a few atoms to large-scale folding or unfolding reactions or to major domain rearrangements with consequently diverse energetic costs. Some conformational changes in nucleic acids have been already described in previous sections (DNA/RNA-binding structural motifs), so a short description of protein adaptive binding follows in this paragraph.

3.3.1 RNA-induced folding events

Induced fit appears to be a common theme in RNA-protein interactions and there are rather striking examples of how the RNA component of the complex can induce structure in a disordered or partially disordered protein (Frankel and Smith, 1998; Mogridge et al., 1998; Zheng and Gierasch, 1997). In the RNA complexes studied to date both partners have been observed to rearrange simultaneously or become stabilized; this rearrangement happens especially in proteins where the RNA-binding domain can be localized to a short, contiguous polypeptide region that recognizes its specific RNA site in the absence of an extensive protein scaffold such as peptide-RNA complexes in viral and phage systems (e.g. Rev peptide with RRE, BIV-tat and HIV peptides with TAR) (Frankel, 2000; Frankel and Smith, 1998; Williamson, 2000). Such complexes are characterized by the fact that the RNA tertiary structure generates scaffolds and binding pockets with the capability to envelope minimal elements of

protein structure (Patel, 1999). These peptides, mostly unstructured in solution, fold as isolated α -helices or β -hairpins within the RNA binding pocket; the binding energy is provided by interaction of the amino acid side chains with precisely positioned mismatches, triples, bulges and loops of the RNA architecture. Arginine plays a key role in these interactions with its long and flexible aliphatic sidechain carrying the charged guanidinium group with its pentadentate hydrogen donor potential: arginine rich peptides in this way can fold forming a selective combination of chemical (hydrogen bonds and salt bridges) and shape (van der Waals) recognition. Indeed, there seem to be a number of possible advantages of using unfolded or just partially folded proteins to recognize RNA: bacteriophage λ N protein, a transcriptional antiterminator, is entirely unfolded in vitro and is actively degraded by proteases in E.coli. Upon binding to its cognate boxB RNA the amino-terminal RNA-binding domain becomes structured and was suggested that RNA-induced folding may at least partially protect N protein from protease degradation (Van Gilst et al., 1997; Mogridge et al., 1998). In this way maintaining proteins in an unfolded state provides a system to monitor the bound status of a protein by means of proteases as a control for protein concentration. Furthermore induced folding or the ability to remodel surfaces allows interactions with multiple partners, facilitating the ordered addition of other components or signalling the binding events to other partners. This feature is expected to be very important for large ribonucleoprotein complexes where many components must be accurately assembled (Frankel and Smith, 1998).

3.3.2 *DNA-induced folding events*

The formation of a DNA-protein complex is usually associated with local folding events. Spolar and Record (Spolar and Record, 1994) reviewing the energetics of protein-DNA complexes formation by calorimetric data demonstrated a large increase in the heat capacity, similar to that one observed when protein molecules undergo a transition from a denatured to a folded state. Several calorimetric studies

have pointed to the lack of large heat capacity change on non-specific DNA binding (Ladbury et al., 1994; Takeda et al., 1992) in contrast to the large negative change observed on specific binding; this difference is generally taken to indicate a decrease in the solvent-accessible surface area removed from bulk water in non-specific versus specific DNA binding.

Conformational changes upon binding to DNA may be the rule, rather than the exception in DNA-binding proteins and DNA induced changes in secondary and tertiary structure have been described for a range of eukaryotic transcriptional regulators (Lefstin and R., 1998).

In some cases, conformational changes are induced in response to either specific or non-specific DNA as in Ets1 (Petersen, 1995) and MASH-1 (Mejerhans, 1995), whereas in others they are restricted to, or observed at a greater extent in specific DNA complexes as happens for the basic regions in bZIP and bHLH proteins that are usually disordered in solution but become α -helical upon binding to DNA (Ellenberger, 1994; Kohler et al., 1999; Patel et al., 1990; Weiss, 1990).

According to a number of studies (Berg and von Hippel, 1987; Spolar and Record, 1994; von Hippel and Berg, 1986) the HTH motif was generally believed to dock without major conformational changes, even to specific target sites. On the other hand, our group (Percipalle et al., 1995), reported an increase in the α -helix content in artificial single-chain repressors containing the HTH motif, upon interaction with non-cognate and cognate binding sites. More recently Ciubotaru and coworkers gave indication that both specific and non-specific DNA induce conformational changes in the full 434 repressor leading to alternative conformations of repressor dimers (Ciubotaru et al., 1999). The difference between the “rigid” HTH and “flexible” bZIP and bHLH motifs may thus be less dramatic.

Hence, the formation of a protein-DNA complex involves local folding events and these are coupled to the thermodynamics of binding.

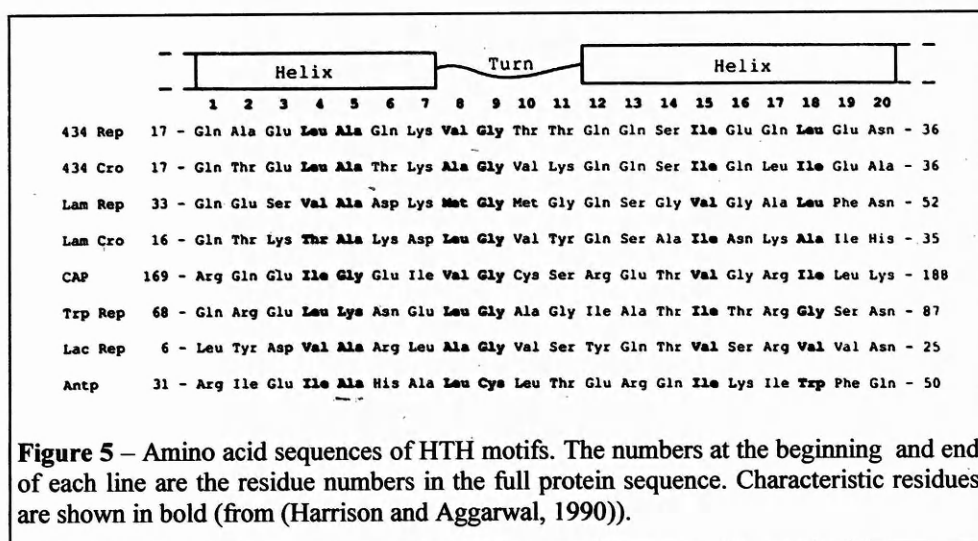
In addition to the thermodynamics upon protein-DNA interaction, the pathways which lead to protein binding and dimerization on DNA targets are also important for understanding how the transcription factors modulate their regulatory functions. Being one of the subject of our research on the DNA-binding motif of phage 434 repressor, this topic will be discussed in paragraph 5.6.

4. DNA Recognition by the Helix-Turn-Helix Motif

The helix-turn-helix (HTH) structure is the first discovered DNA-recognition motif, and is to date one of the most thoroughly studied. This motif has first been identified in many prokaryotic gene-regulatory proteins and more recently it has also been found in a wide range of other DNA-binding species, such as eukaryotic homeodomains and transcription factors serving several functions including the genetic control of development (see (Brennan, 1992; Brennan, 1991; Gehring et al., 1994; Harrison and Aggarwal, 1990; Sinden, 1994; Wintjens and Rooman, 1996; Wolberger, 1996) for review).

In its simplest form, the HTH motif consists of two nearly perpendicular α -helices connected by a short linker (turn) of three aminoacids (**Figure 10**). The second helix in the HTH motif is the so called “recognition” helix and enters into the major groove of the DNA target, allowing protein side-chains to make extensive base-specific hydrogen-bonds and contacts (Brennan, 1992; Brennan, 1991). A number of structural variations have been observed, in which the turn between the two helices is longer (ranging from 1 to 21 additional residues (Wintjens and Rooman, 1996)) and adopts a different conformation. The first identified HTH domains, in the prokaryotic gene-regulatory proteins Cro and repressor of bacteriophage λ and related phages (434 and p22), have a completely α -helical topology, but several α - β domains have been observed where β strands interrupt, precede or follow the helices involved in DNA-binding. These β strands form in general an anti-parallel β -sheet, packed against the helices of the motif and have been called “winged HTH domains”. The spatial arrangement of the recognition helix and the preceding helix in terms of root mean square deviation (r.m.s.d.) of the backbone is anyway strongly conserved among most of the HTH resolved structures; the relative orientation between the two helices thus seems to be indispensable for achieving DNA-binding, whereas the type of connection between them plays a minor role (Wintjens and Rooman, 1996).

The conventional designation of the motif based on the first structures solved (Harrison and Aggarwal, 1990) is a twenty-residue segment which consists of an 8 amino acid α -helix followed by a 3 amino acid right turn with an angle of about 120° followed by another α -helix of 9 amino acids. Some critical positions are highly conserved and believed to be responsible for the structure stability. Those conserved amino acids are Ala in 5th position of the first α -helix, the Gly in 9th position (the first amino acid at the turn), either Val or Ile at position 15 and a general hydrophobic character of amino acids at positions 4, 8, 10 and 18 (**Figure 5**). The amino acid composition of other positions shows a great deal of heterogeneity. Nevertheless, since an α -helix has a repeat length of 3,6 amino acid per turn, certain positions of the HTH face the body of the protein (a hydrophobic environment), while certain residues face the solvent or DNA (a hydrophilic environment) and a similarity in the type of amino acids at these positions has been observed (Suzuki et al., 1995b).



The HTH motif is not a stably folded structure on its own, and usually one or more extra helices from the rest of the protein should be engaged to stabilize this motif. The “recognition” helix of HTH lies in the major groove of DNA and carries the main amino acid residues responsible for specific binding, however, one can not conclude that the recognition involves only these local contacts (e.g. see paragraph 5.2). Several studies have revealed that other regions out of the HTH units can also have a significant

role in the recognition. For example, besides the contacts in the HTH motif, the λ repressor specifies the operator base pairs by using an extended peptide chain in the loop following the second helix and the N-terminal arm to wrap around the DNA (Jordan and Pabo, 1988).

The HTH motif is evolved in a dimension to fit into the major groove of DNA. The diameter of a typical α -helix is about 10Å, which exactly matches the 12Å wide and 6-8Å deep major groove of B-DNA. If an α -helix is parallel to the direction of the major groove, the straight α -helix can contact 4-6 bp before the bases move out of the plane of the amino acids in the α -helix. Thus the short interaction with no more than 4-6 bp of DNA specifies the binding of a particular sequence of amino acid to a unique DNA sequence.

According to a classification introduced by Suzuki and coworkers (Suzuki et al., 1995a) three types of residues are arranged into the recognition helix (numbering as residue 1 the first residue of the helix following immediately the three amino acid turn):

- 1) the DNA base contacting residues (usually residues 1, 2, 5 and 6 of the second helix), which are important for the specificity.
- 2) the phosphate backbone contacting residues (usually residues -1, 3, 9), which fix the binding geometry.
- 3) the other residues facing away from the DNA, which, even though they do not interact with DNA limit the rotation of the recognition helix by interacting with the rest of the protein (e.g. the strongly conserved hydrophobic residue 4 and often residues 7 and 8 which are on the same side of the helix) (Brennan, 1991; Suzuki et al., 1995b).

Crystal structures of prokaryotic HTH protein complexed with their specific DNA, (e.g. several phage repressor-operator complexes, Trp repressor, *E.Coli* CAP), have been resolved. These co-crystal structures show some common features of the HTH-protein-DNA interaction (Harrison and Aggarwal, 1990; Pabo and Sauer, 1984; Pabo and Sauer, 1992):

- 1) The repressors bind as a dimer. Each monomer recognizes one half of the binding site and the approximate symmetry of the DNA binding site is reflected in the approximate symmetry of the protein-DNA complex.
- 2) The conserved HTH unit contacts the DNA in each half of the operator site. There is no universal mode for docking the HTH motif against the major groove of DNA. This is due to the fact that not only does the precise "angle of attachment" vary from case to case but also the major groove itself has a variable geometry. For example, although 434 repressor (Aggarwal et al., 1988b; Mondragon and Harrison, 1991; Mondragon et al., 1989a; Mondragon et al., 1989b; Rodgers and Harrison, 1993; Shimon and Harrison, 1993) attach similarly to DNA backbone, they create different groove structures upon binding. However, despite these variations, there are some important regularities in the mode that HTH elements bind to DNA. The first helix of the HTH unit is somewhat "above" the major groove, but the N-terminus of this helix is in contact with the DNA backbone. The second helix of the HTH unit fits into the major groove and the N-terminal portion of this helix is closest to the edge of the base pairs.
- 3) Though there are some local distortions in different protein-DNA complexes, the operator sites are generally B-form DNA.
- 4) Side chains from the HTH unit make site-specific contacts with base pair groups in the major groove. Direct contacts, polar and non-polar, between amino acid side chains and the edges of base pairs in the major groove are principal sources of specificity. Various side chains (Gln, Asn, Ser, Arg, Lys, *etc.*) donate and accept hydrogen bonds. Many of the contacted base pairs interact with more than one amino acid side chain and many of these side chains interact with more than one base pair. Changes in directly contacted base pairs generally decrease affinity by at least one or two orders of magnitude.
- 5) Each complex has an extensive network of hydrogen bonds between the protein and the DNA backbone. Particularly noteworthy are the hydrogen bonds to non-esterified phosphate oxygens, especially from peptide -NH groups, neutral -NH₂ groups of Gln

and Asn side chains, and -OH groups of Ser and Thr. These interactions occur in the context of tight van der Waals complementarity that anchors the protein very precisely. Hydrogen bonds between positively charged amino acid side chains (Lys, Arg) and DNA phosphates appear only with modest frequency.

Some other factors also contribute to the recognition, which may vary from complex to complex, e.g. in 434 repressor, the free energetics of DNA conformation make an additional contribution to specificity and the Trp repressor appears to use several water molecules to provide critical contacts. In addition to the DNA binding domains, other domains in the HTH protein also play important role in regulating activities, such as the N-terminal domain of CAP allows dimer formation and also binds to cAMP, an allosteric effector of DNA binding. In the case of phage 434, P22 and λ repressors, the C-terminal domains are important for stable dimer formation and high affinity binding, as well as for the cooperative binding on the neighbouring operator sites.

5. Interaction Between Phage 434 Repressor and its Operator

5.1 Role in genetic switch

In temperate bacteriophages, such as lambda and 434, an efficient genetic switch regulates the choice between lysogeny and lytic growth (Ptashne, 1992). The switch requires differential affinity of two proteins, repressor and Cro, for six operator sites, designated as O_{R1} , O_{R2} , O_{R3} and O_{L1} , O_{L2} , O_{L3} . The O_R operator controls two distinct promoters, known as P_R and P_{RM} . Promoter P_R governs transcription of genes, including the one for Cro, which are important for initiating lytic growth. P_{RM} is the promoter for repressor transcription. In a lysogen, repressor dimers bind co-operatively to O_{R1} and O_{R2} blocking the RNA polymerase binding to P_R . Thus the promoter P_R is turned off, while the promoter P_{RM} is activated, which in turn starts the transcription of

its own message. High concentration of repressor leads to binding at O_{R3} and repression of P_{RM} . In the lytic phase, the synthesis of Cro protein turns off the repressor synthesis through binding on O_{R3} , which blocks the RNA polymerase binding to P_{RM} .

5.2 Structure of the repressor and its N-terminal domain (R1-69) operator complexes

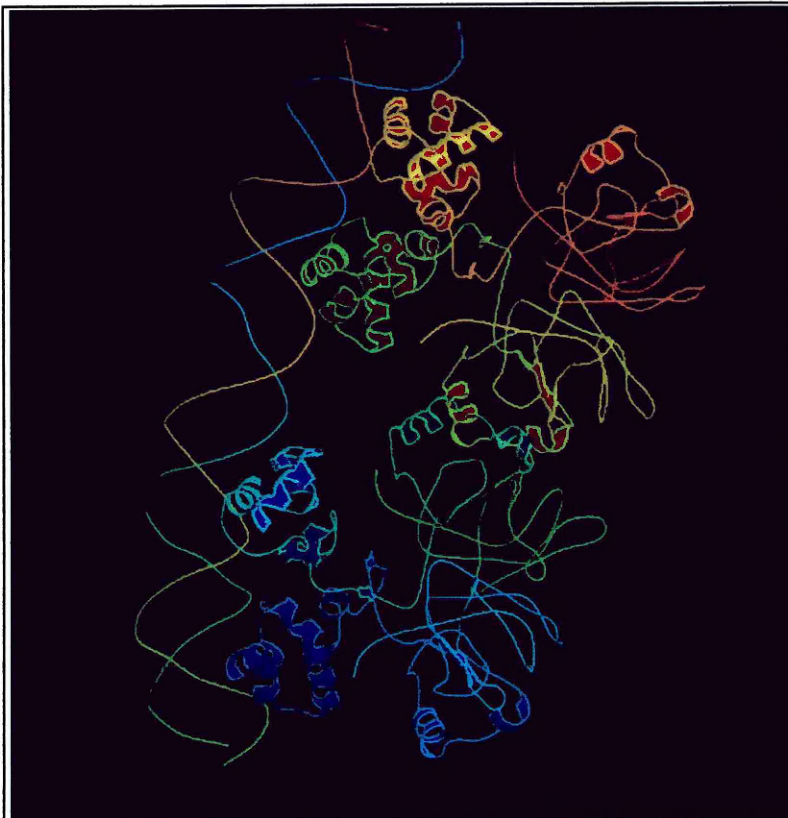


Figure 6 - Ribbon model of two 434 repressor dimers bound to a hybrid O_{R1}/O_{R2} site; the spatial arrangement of the N-terminal HTH is based on the crystal structure of the R1-69/ O_{R1} and O_{R2} complex. The C-terminal dimerization domain structure is a prediction model based on the assumption that it should possess a fold similar to the cI repressor of phage lambda (from the Brookhaven PDB entry 1RPD (R.Chattopadhyaya, K.Ghosh))

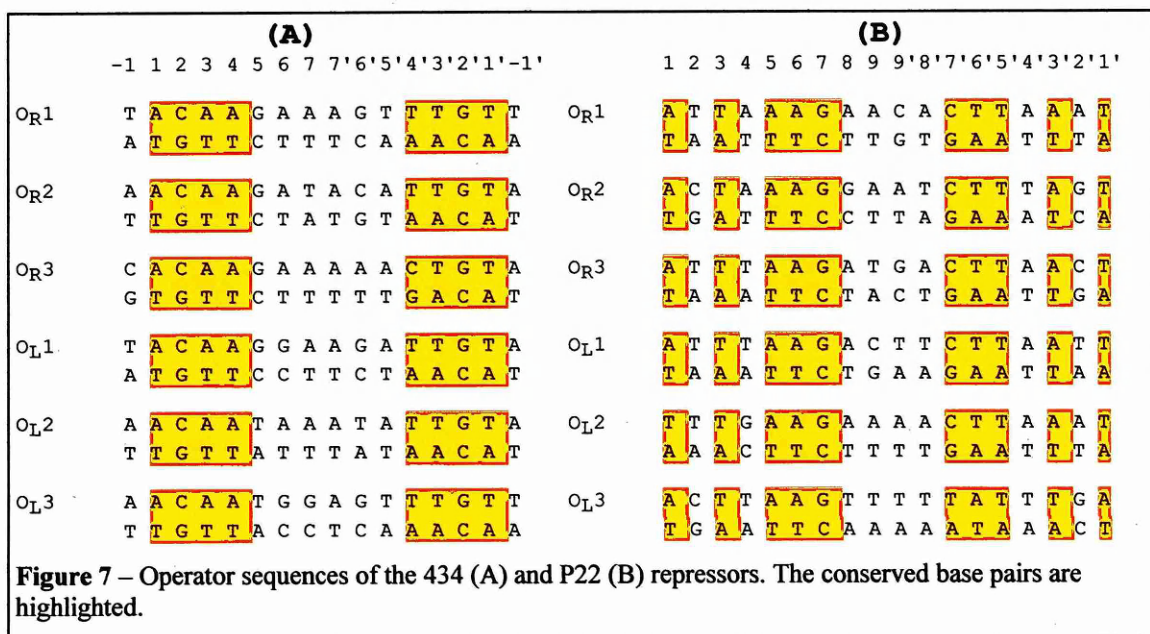
The 434 repressor binds as a dimer to the O_R operator sites with differential affinities: in the native context it shows the highest affinity for O_{R1} , next occupies the adjacent, weak O_{R2} site by cooperative interactions with an O_{R1} -bound repressor dimer and then repressor binds O_{R3} with lowest affinity (**Figure 6**). The lack of additional free energy provided by cooperative interactions

between adjacent bound dimers accounts for different behaviour of the repressor if the sites are present on separate DNA fragments: the highest affinity is displayed for O_{R1} , then O_{R3} , and finally O_{R2} . The carboxy-terminal of the repressor mediates the

dimerization (the native repressor dimerizes in solution with a dissociation constant of 10^{-6} M (Donner et al., 1998)) and is responsible for the formation of tetramers by cooperative binding to adjacent sites, while the amino-terminal (amino acids 1-69) is the DNA binding domain. R1-69 is a bundle of five α -helices linked by turns of varying length. The α -2 and α -3 helices form the HTH motif (**Figure 8**).

The six operator sites to which 434 repressor binds are 14 bp in length, with pseudo two-fold symmetry and are separated by five or six base-pairs. The outer four bases of each half sites are a conserved 5'ACAA/TTGT3' sequence, and the inner six base pairs of each operator are variable. Each monomer binds to a half site of the operator. **Figure 7** is a list of natural operator sequences of phage 434 and P22 repressors.

The DNA-binding affinity of R1-69 monomer is very low but this domain can be co-crystallized with cognate DNA. Crystal structures of the N-terminal domain (R1-69) of the phage 434 repressor complexed with its cognate operators O_{R1}, O_{R2}, O_{R3} have been solved at 2.5Å resolution..



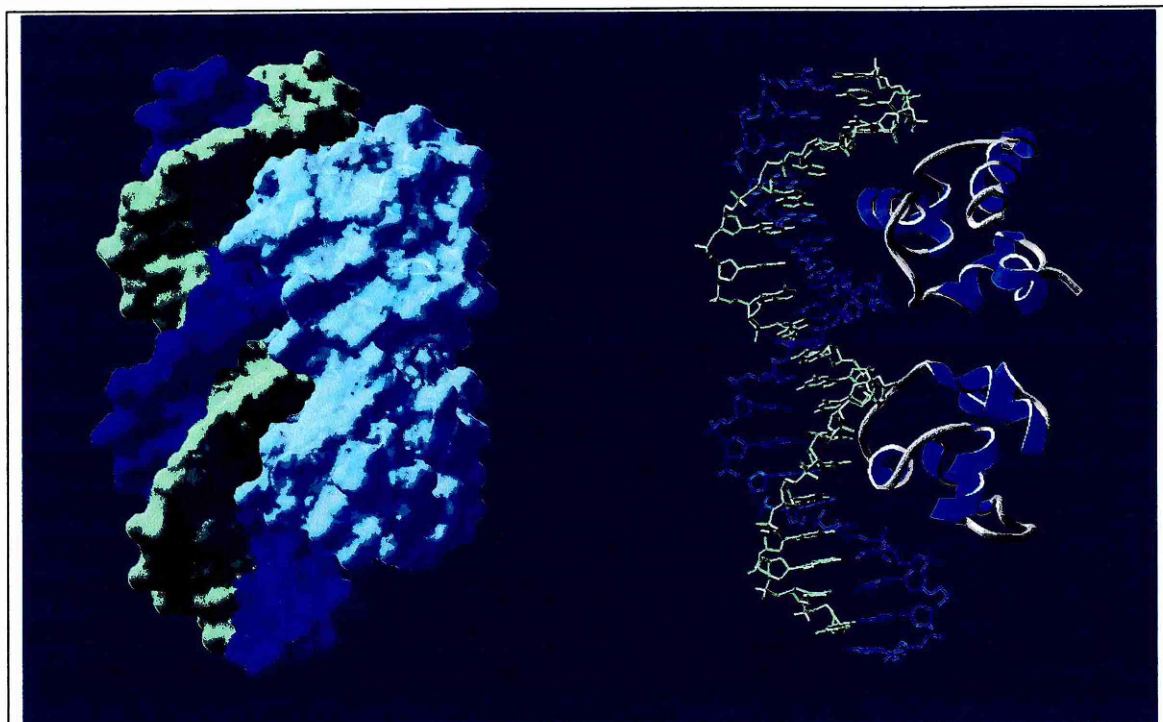
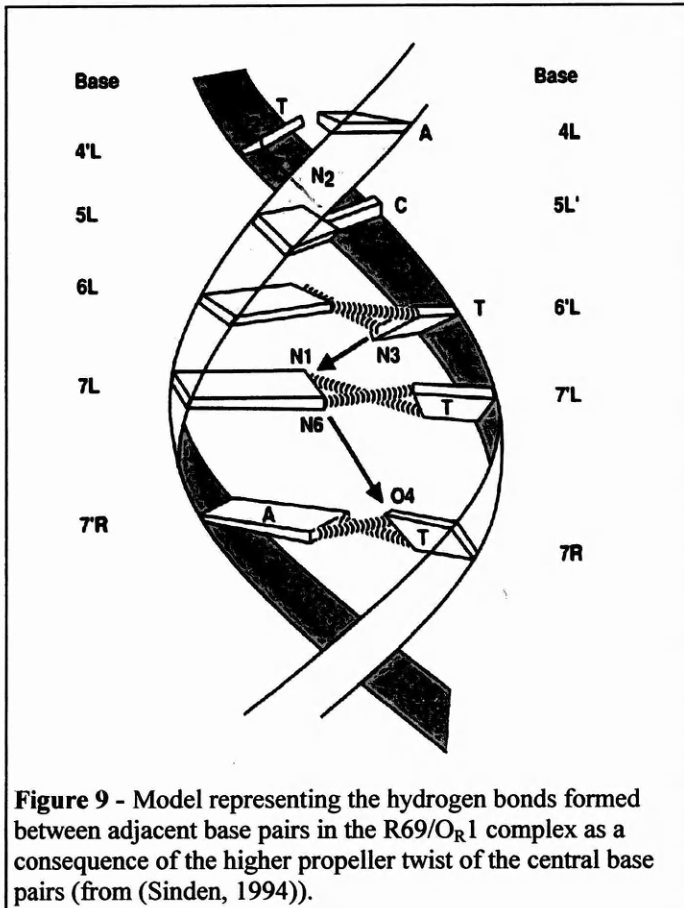


Figure 8 – Space-filling and ribbon representations of the R1-69 bound to OR1. The left model highlights the shape complementarity between repressor and DNA operator. In the right model is evident the α 3-helix of the HTH motif entering the DNA major groove and the DNA bending resulting in minor groove compression in the center of the operator. Both models were constructed with the program Swiss PdbViewer using the crystallographic coordinates (1R69) of the R1-69/OR1 complex.

In the R1-69/OR1 complex (Aggarwal et al., 1988), the B-type DNA is distorted by bending with variations in twist, and other helical parameters (**Figure 8**). The bent configuration permits contacts between the sugar-phosphate backbone and the NH₂-terminus of α -2 helix (along with Arg10), as well as the interactions of Gln28 and Gln29 with base pairs 1 and 2. A striking result of the bending is that the minor groove is compressed in the center and is gradually widening towards the ends. The width of the minor groove (defined by phosphate-phosphate distance) is compressed to 8.8Å in the center while opened to 14Å at the ends of the 14bp operator (the normal width of minor groove in B-DNA is 11.5Å). Thus the DNA is overwound in the center and underwound at the ends. Near the center of the operator between nucleotides 7L and 7R,

Arg 43 extends from the loop between α -3 helix and α -4 helix into the minor groove. Arg 43 from each monomer forms asymmetric contacts with the bases, sugars and phosphates. The presence of the positively charged side chains in the minor groove is thought to stabilize the minor groove compression, which bridge negatively charged phosphates close together (**Figure 8**).



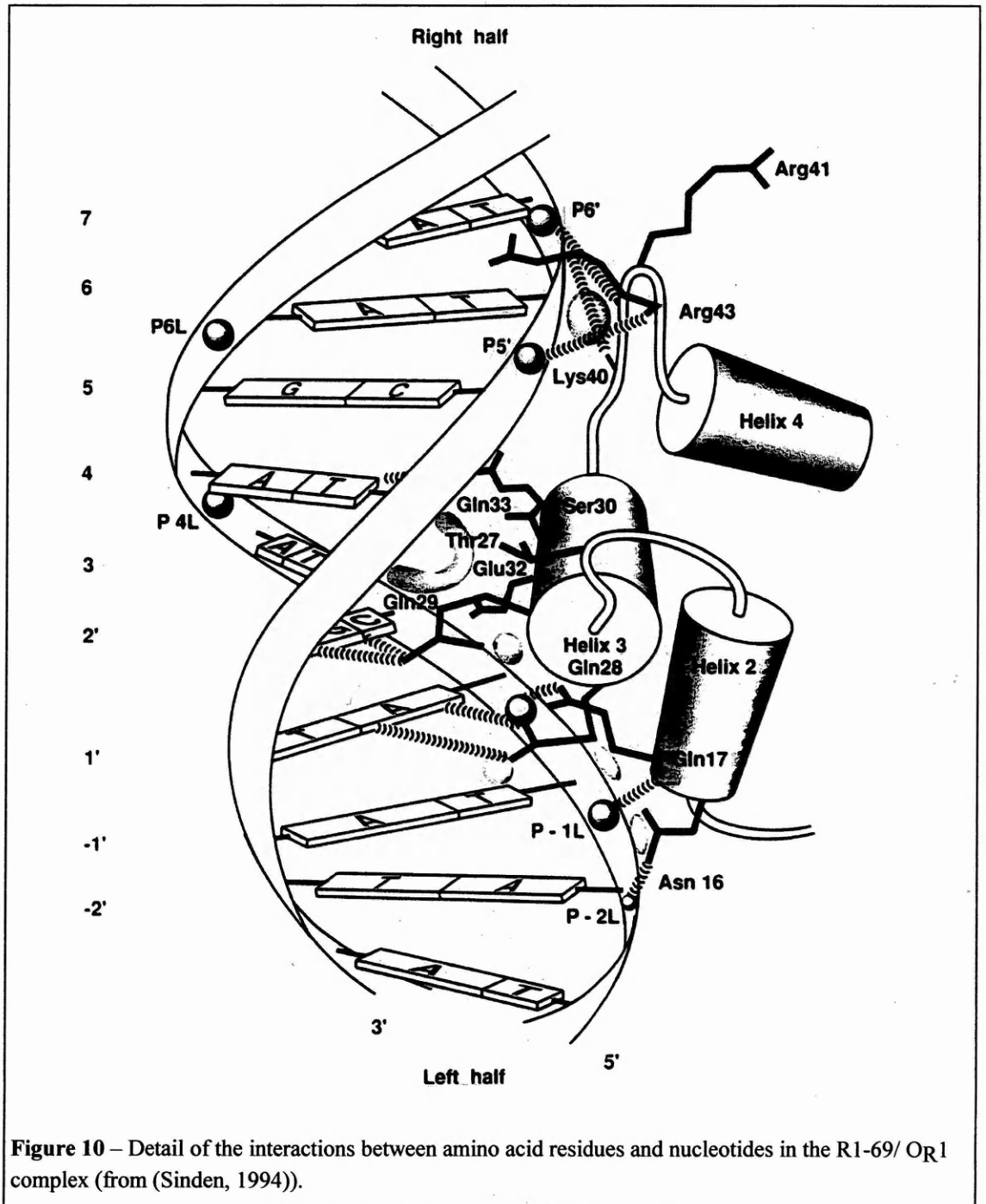
The central base pairs show a higher propeller twist as compared to the classical B-DNA. The propeller twist positions functional groups involved in Watson-Crick hydrogen bonding close to groups on an adjacent base pair. As a result, three bifurcated hydrogen bonds are formed between three adjacent base pairs (O4 of T7R and N6 of A7L, N2 of G5L and O2 of T4L, N3 of T6L and N1 of A7L) at the

bend center (**Figure 9**). The capacity to form bifurcated hydrogen bonds (also observed in R1-69/O_R3 complex) is thought to be able to compensate for distortion on operator by the repressor binding and it is dependent on the DNA sequence.

The α -3 helix lies in a parallel orientation in the major groove. The NH₂-termini of helices 2 and 4 is close to the sugar phosphate backbone, and the loop from helix 3 to helix 4 follows the sugar phosphate backbone at the center of the operator. There is no large-scale conformational change of the protein upon binding, but it does reveal significant local adjustment of side chain configuration and a small shift in the turn

from α -2 helix to α -3 helix. Side chains that rearranged on binding include Gln29, Gln32, and Arg43, all of them interact with DNA.

The protein-DNA interaction involves several hydrogen bonds and hydrophobic interactions, see **Figure 10** for the illustration and **Figure 11** for the base-amino acid contacts.



DNA position	Amino acid	Contact ^a
Major groove hydrogen bond contacts (shown in Figure 8.11)		
#1 A·T	Gln 28	Bidentate hydrogen bonds are formed between the NH ₂ of Gln 28 and N7 of A, and between the C=O of Gln 28 and the N6 of A
#2 C·G	Gln 29	The terminal NH ₂ of Gln 29 makes bidentate contacts with the O6 carbonyl group and the N7 position of guanine
#4 A·T	Gln 33	The O4 carbonyl of T forms a hydrogen bond with the NH ₂ of Gln 33
Major groove van der Waals contacts		
#-1 T·A	Gln 28	The methyl group of thymine at position -1 (outside the 14-bp operator) contacts the methyl groups on the side chain of Gln 28
#3 A·T	Thr 27, Gln 29	Side chain methyl groups of Thr 27 and Gln 29 form a van der Waals pocket to bind the methyl group of thymine
#4 A·T	Gln 29, Ser 30	The methyl group on thymine contacts the side chain methyl groups of Gln 29 and Ser 30
P1, Sugar 1	Asn 16, Gln 17	van der Waals contacts made between the PO ₄ and sugar and the CH ₂ side chains of the amino acids
Hydrogen bonds to the phosphate backbone		
P-2	Asn 16	The P2 phosphate (on the 5' side of A 2, between T 3 and A 2) forms a hydrogen bond with the terminal NH ₂ of Asn 16
P-1	Gln 17, Arg 10	The P1 phosphate (between T 1 and A 2) hydrogen bonds to the NH ₂ of Arg 10 (not shown) and the main chain NH of Gln 17, at the end of helix 2
P1	Gln 17, Asn 36	The terminal NH ₂ of Gln 17 forms a hydrogen bond to P1; Asn 36 also forms a hydrogen bond with P1 (not shown)
P5'	Arg 43	The main chain NH group of Arg 43 forms a hydrogen bond with P5'
P6'	Lys 40, Arg 41	The main chain NH groups of Lys 40 and Arg 41 form hydrogen bonds with P6'

^aData from Aggarwal *et al.* (1988).

Figure 11 – Contacts between the 434 repressor and operator (from(Sinden, 1994)).

5.2.2 R1-69/O_{R2} complex

In the R1-69/O_{R2} complex (Shimon and Harrison, 1993), both the protein and the DNA backbone conformations are very similar as in the R1-69/O_{R1} complex, and same extensive network of hydrogen bonds anchors the repressor to the DNA backbone. The major groove contacts between the outer four conserved base pairs and critical amino acid side chains are essentially identical in the O_{R1} and O_{R2} complexes. However, the R1-69/O_{R2}, has relatively coplanar base pairs in the center of the operator, and has no 'bifurcated' non-Watson-Crick hydrogen bonds as observed in the O_{R1} and O_{R3} complexes. This conformational variation is due to the sequence difference in the two operator center. O_{R1} and O_{R3} have runs of poly (dA)(dT) ("A-

tract") sequence, while O_{R2} has a alternating poly (dAT)(dAT) like sequence, which may have a cross-strand stereo clash if propeller twist is introduced.

5.2.3 R1-69/O_{R3} complex

The O_{R3} operator (Rodgers and Harrison, 1993) contains consensus sequence (5'ACAA) in one half-site of the operator, whereas there is one base pair mutation in the other half-site (5'ACAG). The structure of the R1-69/O_{R3} for the consensus half site is essentially identical to that seen in the O_{R1} and O_{R2} complexes. However, there is an unexpected extensive structure change in the non-consensus half site. The most marked change is at the DNA backbone from position 3' to 4'. The backbone bows out towards the protein, bringing the phosphates closer to residues in the α -2- α -3 helix turn, as well as the amino terminus of α -3 helix. In addition, the protein monomer rotates relatively to the other monomer that interacts with the consensus half site. This monomer rotation further decreases the gap between the DNA backbone and the α -2- α -3 helix turn. The substitution of G-C for A-T at position 4 also causes a rearrangement of the Gln33 side chain, the interaction between Gln33 and the base at position 4 is weakened comparing with the consensus half site, and no direct hydrogen bond is present. It was demonstrated that recognition of the base at position 4 by repressor is not an independent event and, more over, this process is influenced by the sequence of bases not contacted by repressor (Bell and Koudelka, 1993; Bell and Koudelka, 1995).

5.3 The recognition properties of 434 repressor

From the structure of the R1-69 /operator complexes, together with biochemical studies (Bell and Koudelka, 1993; Bell and Koudelka, 1995; Koudelka, 1991; Koudelka and Carlson, 1992; Koudelka et al., 1988; Koudelka et al., 1987; Koudelka and Lam, 1993), several conclusions could be drawn on the recognition properties of the 434 repressor.

In the native 434 operators, all twelve half sites contain ACA at position 1 to 3, and eleven have A at position 4. Biochemical studies demonstrated that any substitution for ACAA at base pair 1 to 4 caused the reduction of binding affinity by at least 100 fold (Anderson et al., 1987; Koudelka and Lam, 1993). Both the structural (Aggarwal et al., 1988) and biochemical studies show that these conserved positions are specified by the direct contacts between a series amino acid side chains and the base pairs. No base pair substitution can conserve the complementarity. Some indirectly contacted amino acids also play important roles in the specificity determination. Koudelka and Lam showed that at least three mutations are needed to eliminate repressor's position 4 base specificity. These amino acids are the direct contacted Gln33, and indirectly contacted Thr27 and Gln32 (Koudelka and Lam, 1993).

There are no direct amino acid side chain-base pair contacts at position 5. However there is a solvent mediated network of hydrogen bonds between phosphate 5 and Gln33. It seems that the base pair 5 influences the configuration of Gln 33 and its interaction with base pair 4.

The 434 repressor binds to the operators with A-T or T-A at the central base pairs more tightly than those with G-C or C-G. This effect is stronger at position 7 than at position 6. Water mediated Arg43 contacts in the minor groove of the operator center are the only interactions between the repressor and DNA in this region. However, these interactions do not appear to account for the central base pair specificity, since similar hydrogen bonds could also be made with other sequences. Koudelka and collaborators have reported that a Arg43 to Ala mutant has the similar preference for the central sequence as the wild type repressor. (Koudelka et al., 1987)

The conformation of the uncomplexed 434 operators vary with central sequence, while the conformation of the DNA-phosphate backbone in the protein-DNA complexes is the same and independent of central base sequence (Koudelka et al., 1995; Koudelka and Carlson, 1992). The central base pair preference is most probably due to the ease for the operators to readjust the local conformation distortion upon repressor

binding. The base pairs at the 6th and 7th positions are configured to bring the half-site of the operator into proper alignment with the protein, thus allowing each monomer of the bound dimer to make optimal contacts within each operator half-site. The imposed sugar-phosphate backbone conformations do not appear to vary with nucleotide sequence, but the adjustment of the base pairs does. Therefore, the repressor binding should be favoured by those sequences for which the conformation constrained by the overall structure requirement is energetically least costly. AT-rich sequences seem to accommodate the distortions more readily than GC-rich sequence in the 434 repressor-operator complexes. X-ray studies of a number of structures have shown that at runs of G and C the minor groove of the helix was exceptionally wide or could easily become wide, whereas at certain runs of A and T the minor groove was narrow or could easily become narrow (Drew and McCall, 1990). A sequence dependent likelihood of flexure was established based on the DNase I digestion experiment on nucleosome DNA. There is a strong correlation between the observed affinity of the operator for the repressor and the predicted likelihood of flexure (Travers and Klug, 1990).

5.4 The helix swap experiment

Wharton and collaborators have substituted the putative recognition helix of 434 repressor with the putative recognition helix of 434 Cro protein and created a hybrid protein named repressor* (Wharton et al., 1984).

In a second “helix swap” experiment of Wharton and Ptashne (Wharton and Ptashne, 1985), the solvent exposed residues of the 434 repressor recognition helix ($\alpha 3$ helix) were replaced with the corresponding residues from the recognition helix of the Salmonella phage P22. The resulting protein 434R($\alpha 3$ (P22R)) bound specifically and with high affinity to P22 operators. The subsequent experiments demonstrated that combining the 434 and 434R($\alpha 3$ (P22R)) repressor monomers, an heterodimer can be formed which specifically recognises a chimeric P22/434 operator that lacks two-fold rotational symmetry (Hollis et al., 1988). The 434 and 434R($\alpha 3$ (P22R)) repressors are

also able to form stable heterodimer complex *in vivo*, and efficiently bring about repression through binding to the P22/434 hybrid operator in *E.coli* (Webster et al., 1992). The repression level is comparable to the 434 and 434R(α 3(P22R)) homodimers binding to their cognate operators. Later studies on single-chain variants (Chen et al., 1997; Simoncsits et al., 1997) showed that the mutants are highly selective even though the operator sites of the 434 and the P22 repressor share a number of features.

5.5 Contacts between R69 monomers

Dimerization in the natural 434 repressor is mediated by the C-terminal domains, in its absence, a weak dimer is formed in the presence of DNA on crystallization. A set of contacts between monomers has been observed in the crystal structure of R69 bound to a 20bp DNA fragment containing the O_{R1} binding site (Aggarwal et al., 1988). A patch of hydrophobic residues, Leu⁴⁵, Pro⁴⁶, Val⁵⁶, and Leu⁶⁰ from both monomers is involved plus a salt bridge from Arg⁴¹ to Glu⁴⁷ and a hydrogen bond between Arg⁴¹L and the carbonyl of Arg⁴³R. The point of closest approach for the two subunits is the Pro⁴⁶ with a C α to C α distance of 9.9Å.

The inter-domain contacts are strong enough to define the structure of the bound dimer but insufficient to maintain its integrity when not associated with DNA, as a consequence of this R69 is a monomer in solution; a microcalorimetry study could not indeed detect any cooperativity between two R69 domains covalently linked but could not exclude the formation of important interdomain contacts upon DNA interaction; a few specific contacts might contribute significantly to the DNA-binding affinity and specificity. (Ruiz-Sanz et al., 1999)

5.6 Assembly pathways of dimeric transcription factors on DNA

Transcriptional regulation relies on the formation of large multicomponent protein-DNA complexes (Ptashne and Gann, 1998; Tijan and Maniatis, 1994). Many transcription factors bind to DNA to form dimeric protein-DNA complexes. Examples include basic region leucine zipper proteins (bZIP), helix-loop-helix zipper (bHLHZip) and some representatives of prokaryotic and eukaryotic transcription factors binding DNA by HTH motifs. The protein complexes recognize short stretches of DNA with remarkable specificity, in spite of the large number of competing non-specific protein-DNA and protein-protein interactions. The so-called recruitment model (Ptashne and Gann, 1998) predicts that DNA acts as a template for the formation of the protein complex, and the components sequentially bind to each other in forming the complex. Many representatives of the above cited protein families have been observed to form dimers in solution in absence of DNA but, in principle, for these dimeric proteins, there exist two limiting pathways that may describe the route of complex assembly: the protein can dimerize first, then associate with DNA (dimer pathway) or can follow a pathway in which two monomers bind DNA sequentially and assemble their dimerization interface while bound to DNA (monomer pathway) (Berger et al., 1998; Bray and Lay, 1997; Holmbeck et al., 1998; Kim and Little, 1992; Kohler et al., 1999; Metallo and Schepartz, 1997; Rastinejad et al., 1995; Rentzeperis et al., 1999; Wendt et al., 1998; Wu et al., 1998). In vitro studies suggested that activating transcription factor 2 (ATF-2, a bZIP protein) and Max (a bHLHZip protein) (Kohler et al., 1999) as well as the Arc repressor of the P22 phage (Rentzeperis et al., 1999) bind as monomer to their target sites, and that the sequential binding of monomers is faster and more specific than the alternative pathway wherein the monomers dimerize before binding to DNA (Kohler et al., 1999). This finding was based on the observation that protein and DNA competitors failed to decrease the rate of the monomeric pathway as opposed to that of

the dimeric pathway (Kohler et al., 1999) and that the dimerization rate does not affect the rate of DNA recognition (Berger et al., 1998).

Dimerization interfaces in proteins are often kept together by strongly distance-dependent and precisely oriented van der Waals hydrophobic interactions, whether electrostatic potential which is a main component of protein-DNA interactions, decreases with only the inverse of the distance between the interacting partners and is not very directional. From this derives the hypothesis that the basic protein monomers first bind DNA via electrostatic interactions and then proceed to form the specific dimeric complex. This model could account for measured association rates above the theoretical three-dimensional diffusion limits (Metallo and Schepartz, 1997).

6. The RGG motif

Kiledjian and Dreyfuss (Kiledjian and Dreyfuss, 1992) identified by deletion analysis the RNA binding activity of the hnRNP U, a protein whose sequence is missing a canonical consensus sequence RNA binding domain (RBD). The RNA binding activity was found in the C-terminal part and precisely in a 26 amino-acid long fragment containing a cluster of Arg-Gly-Gly repeats: this region, necessary and sufficient for RNA-binding of the U protein, was termed the "RGG box".

The RGG box has been subsequently found in several other RNA-binding proteins involved in crucial physiological functions (**Table 2**) including many nucleolar proteins such as SSB-1, nucleolin and fibrillarin, other hnRNP proteins (hnRNP A1 and hnRNP G), and a number of RNA helicases. Two characteristics are almost invariably conserved in the primary sequence of the RGG boxes:

i) the strong positive charge of the RGG motif always relies on arginine residues and no lysines are present; this implies a conserved and unique function for this residue. By analogy to a series of arginine rich peptides (e.g. HIV-1 TAT RNA-binding protein, HIV1 Rev, P22box B etc.) the arginine makes unique contacts that cannot be made by other aminoacid side chains. Furthermore, in metazoa, the Arg residue can be methylated by arginine N-methyltransferases to NG, NG-dimethylarginine (ADMA), suggesting that methylation of Arg might modulate the interaction of these proteins with ligands.

ii) the presence of aromatic residues, Phe or Tyr, interspersed with the RGG repeats; these amino acids may be involved in the unstacking of RNA and DNA base pairs since it has been shown that aromatic residues insert partially between the bases of a nucleic acid helix (Gabbay et al., 1972).

The importance of this domain in protein-nucleic acid and in protein-protein interactions is well documented. On the other hand, little information is available on the energetics of the interaction of these motifs with RNA or DNA and on the effect of arginine methylation. Here follows a short resume of the known functions of RGG motifs.

Among the protein families containing this domain, fibrillarins have been isolated in a number of species from eukaryotes to archeobacteria (Narcisi et al., 1998). Eukaryotic fibrillarins are longer than the homologues from archeobacteria and are characterized by the extra RGG domain that seems to be involved in the nuclear localization of this protein (Amiri et al., 1994, Bult et al., 1996.). Fibrillarin is the most abundant protein in the fibrillar regions of the eukaryotic cell nucleolus where the early stages of pre-rRNA processing take place (Eichler and Craig, 1994). It is also a component in many small nucleolar ribonucleoprotein particles involved in the processing of pre-rRNA, but its biochemical function is not entirely understood (Maxwell and Fournier, 1995).

The GAR1 proteins are also involved in the processing of rRNA in yeast and have RGG motifs located in the N-terminal portion and in the central part of the molecule (Girard et al., 1992). The heterogeneous ribonucleoprotein family in turn modulates packaging of pre-mRNA into particles and the transport of poly A⁺ mRNA from the nucleus to the cytoplasm. This family comprises over twenty distinct hnRNP proteins, with a variety of specificities and roles within RNA metabolism (Dreyfuss et al., 1993). hnRNP A1 is one of the few of this family that shuttles continuously between the nucleus and cytoplasm carrying poly A⁺ mRNA (Michael et al., 1995; Buvoli et al., 1990). All the protein members of this family contain, at the C-terminus, a region rich with RGG repeats involved in RNA binding (Kumar et al., 1990), protein-protein interactions (Cartegni et al., 1996), nuclear localization (Michael et al., 1995) and RNA strand annealing activity (Pontius and Berg, 1992). In hnRNPA1 a potential mechanism of regulation of the RGG-box RNA-binding activity was attributed to phosphorylation, the binding of the hnRNPA1 C-terminal domain was in fact abrogated by phosphorylation of a serine just N-terminal to the RGG-box (Municio et al., 1995).

Other proteins belonging to smaller families and containing the RGG motifs are involved in RNA binding, such as the cold-inducible RNA binding protein (Sheikh et al., 1997) or the EWS protein involved in tumorigenic processes (Plougastel et al., 1993) and its similar protein FUS/TLS (Ichikawa et al., 1994). The RGG domain is also part of the proteins involved in the binding of telomeric DNA, like the GBP2 or P67 from yeast (Konkel et al., 1995; Kondo et al., 1992), suggesting that this domain is also able to bind DNA fragments.

Several proteins containing the RGG motif are RNA helicases or putative RNA helicases. P68 is a nuclear protein of 68 kDa that plays an important role in the regulation of cell growth and division and has been shown to possess RNA helicase

activity (Hirling et al., 1989). Nucleolin and G3BP are also helicases with ATP-dependent RNA and DNA unwinding activity and contain RGG repeats (Tuteja et al., 1991; Tuteja et al., 1995, Costa et al., 1999). The RGG motif may be an important structural requirement for the mechanism of RNA recognition and unwinding.

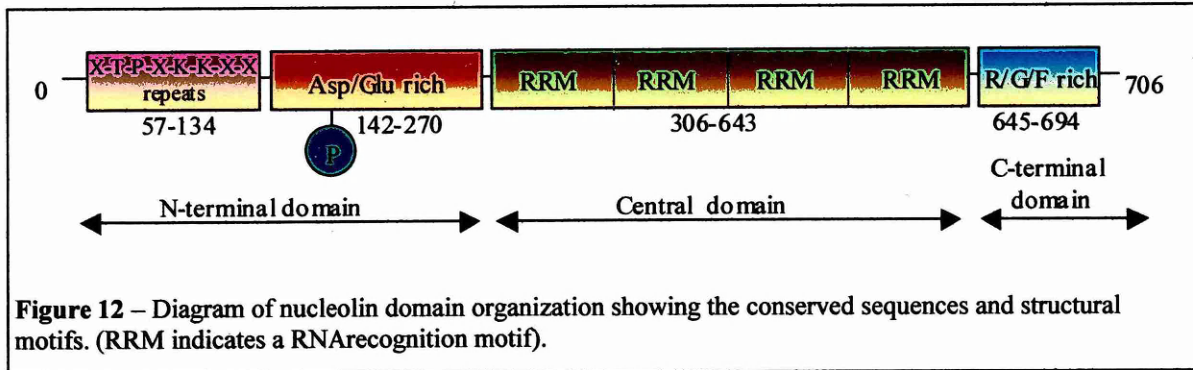
6.1 Possible involvement of the RGG box in various disorders

Several RGG-box containing proteins are involved in various human abnormalities and malignancies. For example, a protein implicated in fragile X syndrome (FMR-1) contains an RGG-box as well as two KH domains, the Epstein-Barr virus nuclear antigen -1 (EBNA-1) which is necessary for the pathogenicity of Epstein-Barr virus also contains an RGG box, suggesting that this protein is involved in RNA metabolism in addition to transcription and DNA replication. The EWS gene is involved in Ewing's sarcomas and malignant melanomas of soft parts and a closely related gene, TLS, is involved in myxoid liposarcomas. Both of these genes encode proteins with three RGG boxes as well as an RNP.

7. Nucleolin

One of the most abundant proteins in the nucleoli of vertebrate cells is the highly conserved protein nucleolin. Nucleolin seems to be involved in a wealth of metabolic processes ranging from DNA/RNA unwinding to transcription repression, cell proliferation and growth, embryogenesis and differentiation and is an autoantigen in systemic autoimmune disorders; beside all these diverse functions, nucleolin probably plays a key role in ribosome biogenesis, a process that is still largely unknown. In eukaryotes ribosome biogenesis takes place in the nucleolus, a plurifunctional subnuclear organelle and a major site for several nuclear functions including rRNA transcription and processing, ribosome assembly, import-export of RNA and proteins.

Nucleolus contains proteins of pre-ribosomes and other specific nucleolar proteins such as RNA polymerase I, protein kinases, phosphatases, methylases and nucleases. Among these nucleolin represents up to 5% of the nucleolar proteins in exponentially growing cells. Mammalian nucleolin is 709 amino acids in length and consists of an unusual grouping of sequence and structural motifs (**Figure 12**).



7.1 The N-terminal domain

The N-terminal region houses several long stretches of acidic residues (**Figure 12**) which cause nucleolin to run in SDS-PAGE with an apparent molecular mass of 105 kDa, although the actual calculated mass is 77 kDa from the cDNA sequence.

The amino-terminal domain, variable and less conserved in different species, is structured in alpha helical domains comprising four lengthy acidic stretches, similar to those of certain high-mobility group proteins and similarly inducing nucleolar chromatin decondensation through ionic interactions and displacement of histone H1 (Erard et al., 1988, 1990). The N-terminal domain has been involved in many protein-protein interactions, such as with the U3 snoRNP (Ginisty et al, 1998) and some ribosomal proteins (Bouvet et al., 1998). This region contains also some basic and repeated octapeptide motifs (XTPXKKXX, X=non-polar residue, similar to an analogous sequence of histone H1) which could be responsible for modulation of DNA condensation (**Figure 12**). The N-terminal region contains in vivo phosphorylation sites for casein kinase II, cdc2 and protein kinase interspersed with basic lysine residues,

susceptible to proteolysis. The transcription of rDNA genes starts only when the serine residues are phosphorylated by CKII and the proteolytic site cleavage is concordantly enhanced, suggesting that nucleolin stability is dependent on phosphorylation (Warrener et al, 1991). A bipartite nuclear localization signal is situated between the amino-terminal domain and the central RNA binding domain, and is necessary for nucleolin to enter the nucleus (Xue et al., 1993).

7.2 The central globular domain

The central domain of nucleolin, involved in pre-rRNA recognition, condensing and packaging (Creancier et al., 1993; Ghisolfi et al., 1992b; Sapp et al., 1989) exhibits alternating hydrophobic and hydrophilic stretches, and also contains four consensus RNA-binding domains (CS-RBD) also called RNA recognition motif (RRM) (**Figure 12**).

A typical CS-RBD contains 80-90 amino acid residues with two highly conserved sequences, the RNP1 octapeptide (R/K)G(F/Y)(G/A)(F/Y)VX(F/Y) and the RNP-2 (L/I)(F/Y)(V/I)(G/K)(G/N)L hexapeptide motifs (Query et al., 1989). CS-RBDs can be distinguished according to their target preferences in two classes. In the first class they interact specifically with single-stranded RNA sequences; this class includes hnRNP A1, hnRNP C1, poly(A) binding protein, splicing factor ASF/SF2 and sex-lethal protein. The minimal single stranded RNA required ranges from 5 nucleotides (PABP) to 10 nucleotides (ASF). In the second class CS-RBDs interact with stem-loop structures and include U1-snRNP A protein, U2-snRNP B" protein, U1-snRNP (70 kDa) and nucleolin. The crystal structure of the CS-RBD domain of U1-snRNP A, revealed four antiparallel beta-strands packed against two perpendicularly orientated alpha-helices. The RNP-1 and RNP-2 motifs are located in two adjacent beta strands and conserved aminoacids within these two motifs have been shown to be implicated in direct nucleotide contacts (Jessen et al., 1991). Nucleolin interacts specifically (Kd from 100 to 5 nM) with an RNA stem-loop structure (NRE) and with in vitro selected RNAs

(containing UCCCGA) through its two first CS-RBDs. This property may account for nucleolin's association with pre-rRNA in the nucleolus. Taken individually, none of the four CS-RBDs interacts significantly with the RNA target, but a peptide that contains the first two adjacent CS-RBDs is sufficient to account for nucleolin RNA-binding specificity. The full integrity of these two domains is required and mutation of conserved amino acids within the RNP-1 sequence of CS-RBD 1 or 2 drastically decrease the specific RNA interaction, whereas mutations in CS-RBD 3 and 4 have no apparent effect in this context (Serin et al., 1997).

Central and carboxy-terminal domains together bind strongly to G-rich DNA and ATP (Miranda et al., 1995). These two domains together, and in particular the portion containing the RBD3, RBD4 and the RGG, are also able to mediate nucleolin self-association and to accelerate annealing of complementary oligonucleotides including sequences carrying mismatched bases (Hanakahi et al., 2000).

7.3 The C-terminal domain

Spaced Arg-Gly-Gly (RGG) repeats interspersed with aromatic amino acids, mainly phenylalanine residues, constitute the extended carboxy-terminal domain (Figure 12).

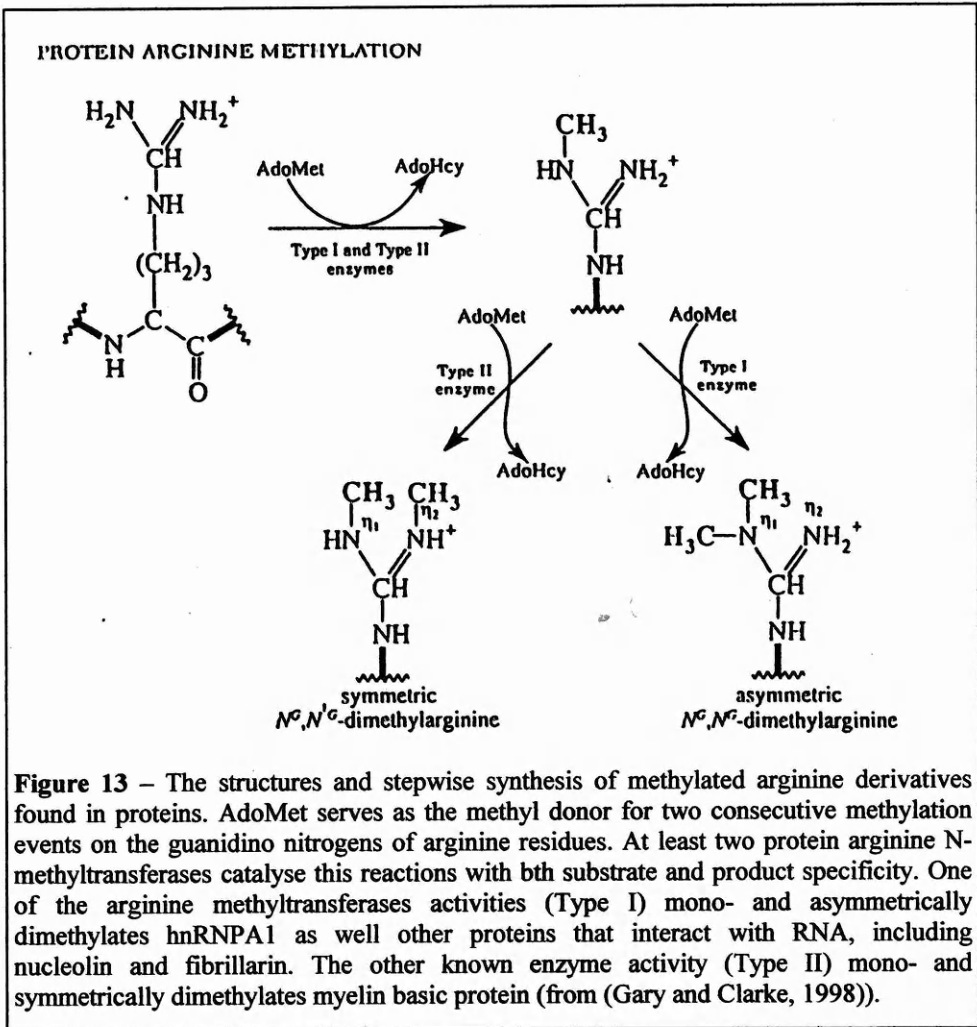
The length of this RGG domain is variable among nucleolins, with its sequence and arrangements of the repeats not well conserved. Studies with the C-terminal fragment of hamster nucleolin have shown that this 52 AA peptide containing nine RGG motifs is capable of binding RNA, ssDNA and dsDNA in a non-specific manner and the non-specific interaction with RNA leads to unstacking and unfolding of this RNA (Ghisolfi et al., 1992). Circular dichroism and infrared spectroscopy provide evidence that a peptide rich in RGG motifs can adopt a repeated β -turn structure, by addition of consecutive β -turns, with the phenylalanine and arginine side chains projecting away from the spiral axis. (Ghisolfi et al., 1992).

The presence of the RGG domain does not influence the binding specificity for the NRE sequence, but its removal results in a considerable drop in the association constant which, starting from $2 \times 10^6 \text{ M}^{-1}$ and $0.1 \times 10^6 \text{ M}^{-1}$ for specific RNA and non-specific RNA respectively, decreases by one order of magnitude (Ghisolfi et al., 1992b). An intrinsic protease activity, which leads to nucleolin autodegradation, is attributed to the carboxy-terminal two-thirds of the protein (Fang and Yeh, 1993). Nucleolin is accumulated within the nucleolus by virtue of its binding to other nucleolar components (probably rRNA) via the two central RBDs or one RBD together with the GAR domain in addition to the nuclear localization sequence (Creancier et al., 1993). A recombinant 41-amino-acid long C-terminal fragment of nucleolin both alone and in conjunction with the central RBDs, was reported to bind G-G paired rDNA (G2 and G4 quartets) with very high affinity (Hanakahi et al., 1999). The RGG domain seems involved also in protein-protein interactions, in particular in hamster nucleolin the C-terminal was shown to bind a subset of ribosomal proteins (Bouvet et al., 1998) and *in vitro* assays showed that the RGG domain was sufficient for the interaction with one of them; protein-protein interactions had been already demonstrated with the RGG rich fragment of hnRNPA1, which was shown to interact with itself and with other hnRNPs (Cartegni et al., 1996).

Sequence analysis of a 53-amino acid tryptic peptide derived from the carboxyterminal region of nucleolin, identified 10 asymmetric dimethylarginine residues, but no mono- or symmetric dimethylarginines, nor any unmodified arginine residues (Lischwe et al., 1985).

7.3.1 Arginine methylation

N^ω-methylation of arginine residues is a widespread but poorly understood post-translational protein modification in eukaryotes (Aletta et al., 1998). Over 50 different proteins are found to be enzymatically methylated in extracts of human P12 cells, and about 90% of methylation occurs at the guanidine side-chain of arginine residues



(Najbauer et al., 1993). The predominant (>90%) product of arginine methylation is N^ω, N^ω-dimethylarginine (asymmetric dimethylarginine or ADMA); the methyl group, donated by S-adenosylmethionine is transferred to arginine by the enzyme type I N-methyltransferase (PRMT1) through an N^ω-monomethylarginine (MMA) intermediate (Gary and Clarke, 1998) (**Figure 13**).

A number of reports have recently shown the association of PRMT1 with a number of proteins involved in signal transduction, highlighting the potential roles of this post-translational modification (Aletta et al., 1998). *In vitro* methylation studies and sequence comparisons suggest that in sites undergoing asymmetric methylation by PRMT1 at an Arg residue (position 0), +1 Gly and +2 Gly are conserved to 100% and 80%, respectively (Liu and Dreyfuss, 1995). The term “RGG” is used to describe the actual site of asymmetric methylation, and the last Gly, even if not 100% conserved, is

included to distinguish this site from the myelin basic protein symmetric methylation site Arg-Gly-Leu (Gary and Clarke, 1998).

Several nucleolar proteins that interact with RNA, such as hnRNP A1, fibrillarin and nucleolin have been proved to be substrates for PRMT type I, suggesting a functional role of methylation in the RNA-protein interactions (Gary and Clarke, 1998). Nucleolin has been described in Paragraph 7; fibrillarin is a 34 kDa polypeptide component of a nucleolar small nuclear riboprotein involved in the first processing step of preribosomal RNAs and contains around 13 moles of ADMA per mole of protein, with no mono or symmetric DMA (Lischwe et al., 1985); hnRNP A1 is a 34 kDa polypeptide and a major component of the 40S ribonucleoprotein particle, it has a role in the differential selection of proximal or distal 5' splice sites and *in vitro* can promote alternative exon skipping in some constructs; this protein contains probably 4 residues of ADMA (Kim et al., 1997).

Methylation occurs in conserved glycine and arginine rich (GAR) sequences (Table 2) of these proteins that usually contain several RGG- motifs (Gary et al., 1998; Klein et al., 2000) and both the dimethylated and the unmethylated forms of GAR domains are known to occur *in vivo*. Unusual sites of asymmetric arginine methylation have been reported in RXR-motifs of Poly(A)-binding protein II (Smith et al., 1999). Sequence analysis of a tryptic fragment of basic fibroblast growth factor (bFGF), a mammalian protein that can stimulate growth in a wide range of cell types, suggested the presence of three dimethylarginine residues in the context of a GAR domain. Potential interactions between arginine methylation sites and serine phosphorylation sites within SRGG domains raise additional possibilities for combining and tuning diverse signalling mechanisms (Siebel and Guthrie, 1996).

Arginine methylation is known to facilitate nuclear export of hnRNP proteins (Shen et al., 1998) but at present there is no clear understanding of the exact role that arginine methylation may play in nucleic acid-protein interactions: the specific RNA binding of Hrp1 protein, a member of hnRNPs, is not affected by methylation (Valentini

et al., 1999), on the contrary, the binding property of recombinant hnRNP A1 protein to single-stranded nucleic acid is reduced upon enzymatic methylation (Rajpurohit et al., 1994). GAR domains themselves seem to non-specifically bind to nucleic acids; however, the presence of a C-terminal GAR domain is essential for sequence specific RNA binding to occur in such diverse proteins as nucleolin (Valentini et al., 1999; Yang et al., 1994), hnRNP A1 (Kiledjian and Dreyfuss, 1992) and hnRNP U (Cobianchi et al., 1988). The structure of the GAR domain does not seem to be crucial for the methylation, because the RGG consensus seems sufficient to direct asymmetric arginine methylation, as seen in the *in vitro* methylation of short peptides (Najbauer et al., 1993). It is therefore important to investigate the nature of aspecific interactions between RGG motifs and nucleic acids and to clarify the effect of methylation on such interactions. Apart from steric differences, seemingly adverse physicochemical effects may play a role here: i) Dimethylation is known to render the arginine molecule more basic, as the pI values of DMA was determined to be 10.77 as opposed to 10.02 of arginine (Paik, 1983). On the other hand, N^ω-alkylation was shown to decrease the pK_a of the arginine side/chain (Kennedy et al., 2000) ii) Methylation may decrease the hydrogen bonding ability through the replacement of the available H atoms by methyl groups. The increase in side-chain bulkiness and the loss of H-bonding ability is also expected to change the solvation environment of the guanidinium group (Kennedy et al., 2000).

The analysis of the physicochemical effects of methylation is somehow impaired from the fact that *in vitro* methylation catalysed by PRMT1 on various substrates is not entirely quantitative and often produces a mixture of products MMA and DMA (Liu et al., 1995; Tang et al., 1998).

8. Spectroscopic methods used in this work to analyze protein and nucleic acid structure

Proteins and protein-nucleic acid complexes can be studied with different spectroscopic techniques which can give various and significantly different information, but nevertheless complementary to each other.

8.1 Circular dichroism spectroscopy

Circular dichroism (CD) and optical rotatory dispersion (ORD) are two optical phenomena that are based on the asymmetric nature of biological macromolecules which are able to rotate the plane of light polarization due to the presence of centers of chirality (carbon atoms). ORD is the property of a molecule to rotate the plane of linearly polarized light as a function of the wavelength. CD is a measure of the differential absorbance between left- and right- circularly polarized light; circularly polarized light and the linearly or plane-polarized light are readily interconvertible. A plane polarized light beam consists of right- and left- circularly polarized beams of equal intensity and, conversely, a circularly polarized beam consists of two orthogonal plane-polarized beams 90° out of phase. A linearly polarized light beam becomes elliptically polarized after passing through an optically active medium (**Figure 14**) (Fasman, 1996). Since the conformation of a macromolecule affects its optical activity, these properties can be utilized to obtain information about the structural organization of such molecules in solution. Theoretical calculations and several experimental measurements have demonstrated high sensitivity of CD spectra toward protein and nucleic acid secondary structure.

8.1.1 Expression of CD

CD can be expressed in two ways; one way is to measure the difference in absorbance of the right- and left-circularly polarized components. The other is to measure the ellipticity. The ratio of the minor axis to the major axis of the ellipse defines the tangent of ellipticity. Because the difference in absorbances of the two components is only a small fraction of the average absorbance, the ellipse is extremely elongated.

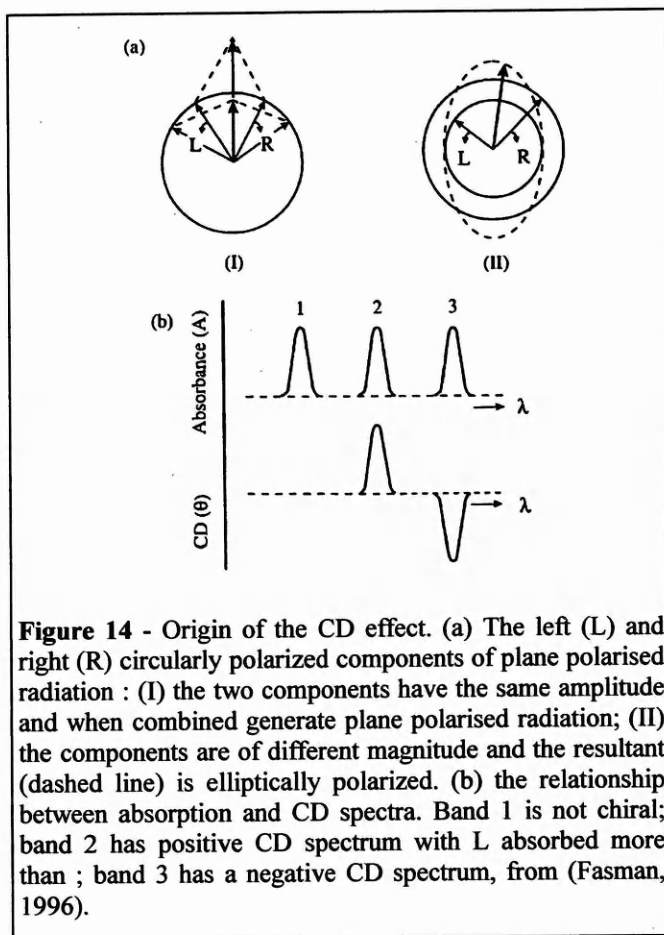


Figure 14 - Origin of the CD effect. (a) The left (L) and right (R) circularly polarized components of plane polarised radiation : (I) the two components have the same amplitude and when combined generate plane polarised radiation; (II) the components are of different magnitude and the resultant (dashed line) is elliptically polarized. (b) the relationship between absorption and CD spectra. Band 1 is not chiral; band 2 has positive CD spectrum with L absorbed more than ; band 3 has a negative CD spectrum, from (Fasman, 1996).

Thus, tangent (ellipticity) can be approximated as ellipticity. Traditionally CD can be calculated from:

$$\epsilon_L - \epsilon_R = (A_L - A_R) / (l * C)$$

and

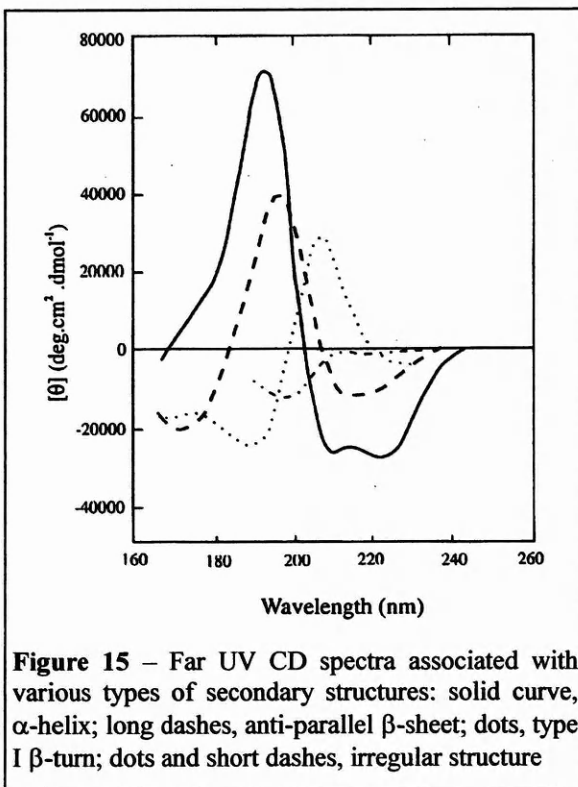
$$[\theta] = \phi * 100 / C * l = 3298 \Delta \epsilon$$

where A_L and A_R are absorbances, l is the light path in cm, C is the molar concentration, ϕ is the ellipticity in degrees., It is important to specify the scale on which the molar concentration C is calculated regardless of which one is used. It is common practice for macromolecules to use the mean residue concentration especially when the far-UV CD of polypeptides and proteins is reported.

8.1.2 CD determination of protein secondary structure

CD bands of proteins occur in two spectral regions. The far-UV or amide region (170-250 nm) is dominated by contributions of the peptide bonds, whereas CD bands in near-UV region (250-300 nm) reflect the contributions of the aromatic side chains disulfide bonds and induced CD bands of prosthetic groups (Chakrabarty et al., 1993; Schmid, 1989; Yang et al., 1986). The two spectral regions give different types of information about protein structure.

The signals observed in the amide region provide information about the peptide bond and the secondary structure of a protein. They are often used to monitor structural transitions due to changes in secondary structures and subdomain folding (Lumb et al., 1994) and also to investigate structural stability of proteins (Thompson et al., 1993).



Optical activity in the region of the spectrum between 190 and 230 nm is essentially dominated by the protein backbone. A number of experiments have shown that the nature of particularly aliphatic side chains does not markedly affect the CD spectrum in this region, therefore, to first approximation one can consider a protein as a linear combination of backbone regions with α -helical, β -sheets or random coil structures (**Figure 15**).

All- α proteins show a strong double minimum at 222 and 208-210 nm and a stronger maximum at 191-193 nm. All- β proteins usually have a single, negative (between 210 and 225 nm) and a single, positive CD band (between 190 and 200 nm), whose intensities are much lower than those of α -helix. Other types of all- β proteins such as α -chymotrypsin, elastase and soybean

trypsin inhibitor were reported to have a strong negative band near 200 nm similar to that found for unordered proteins (**Figure 15**).

For proteins therefore the major objective is to deduce average secondary structure by CD data. There are at least three established methods to determine this:

A semi-empirical approach was derived by Wu et al. (Wu et al., 1981) and makes use of known structures of synthetic polypeptides as reference spectra. An alternative approach is to compute reference spectra from CD spectra of proteins of known secondary structure (Provencer, 1984). The third method concerns the direct analysis of a protein as a linear combination of the CD spectra of proteins of known secondary structure, thus avoiding the problem of defining reference spectra of individual conformations (Schmid, 1989).

The semiempirical method (Wu et al., 1981), in particular is based on the theoretical values of mean residue ellipticity $[\theta]_{MRW}$ (based on the concentration of the sum of the amino acids in the protein solution), calculated for an infinite α -helix and an infinite random-coil. This method gives a rough idea about the α -helical content of a given protein, especially if the investigated protein is made essentially of α -helices.

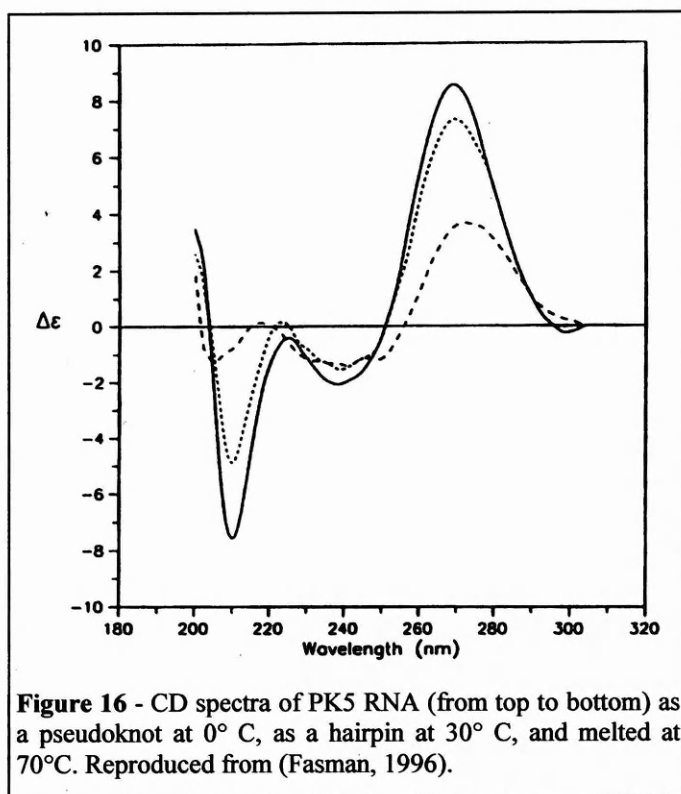
8.1.3 *CD of protein-nucleic acid complexes*

The CD spectra of interacting proteins and nucleic acids complexes are usually analyzed in the ultraviolet wavelength region below 320 nm, where nucleic acids and proteins have optical activity as a result of their secondary structure.

Several protein-DNA complexes have been studied by circular dichroism since the technique is very sensitive (a measured ellipticity of 10 mdeg is equivalent to a ΔA of only 0.0003 absorbance unit) and furthermore it allows difference spectroscopy to be performed. The underlying principle is that frequently the protein has a small but not negligible CD when compared with the complex or DNA alone. In this case it is therefore more convenient to compute the difference spectrum in order to determine the average secondary structure of the protein in the presence of DNA. A comparison of the

difference CD spectrum with that of the protein alone indicates whether structural transitions have occurred in the presence of the inducing agent.

Most of the studies reported in literature investigate eukaryotic transcription factors such as bZIP and bHLH proteins (Ellenberger, 1994) providing indications on the sequence requirements to form coiled coils (Hu et al., 1990) and the structural transitions in the presence of cognate DNA for GCN4 (Talanian et al., 1990; Weiss, 1990), Jun/Fos heterodimer (Patel et al., 1990).

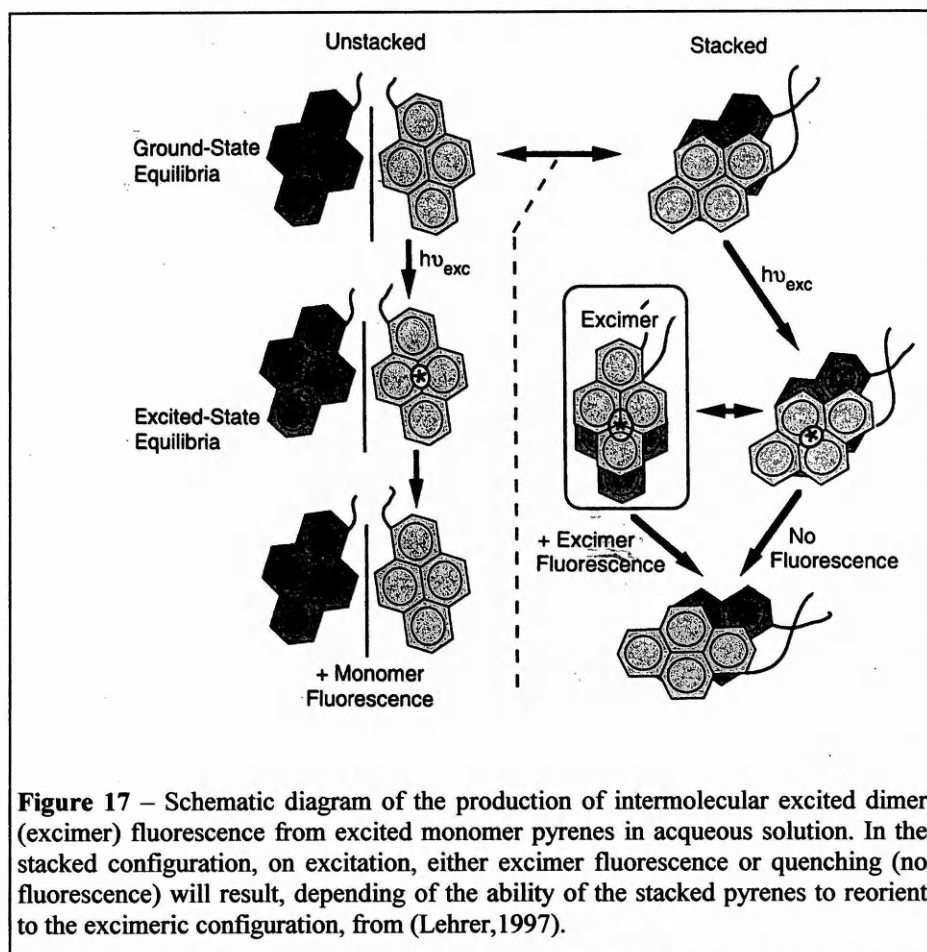


Since protein secondary structures generally dominate CD spectra at wavelengths below 250 nm, the region from about 250 nm to 320 nm provides a valuable spectral “window” for detecting the secondary structure of nucleic acids that are complexed with proteins. Of course aromatic amino acid side chains contribute to the CD in the wavelength region 250-320 nm,

but this contribution is usually small relative to the CD of polymeric nucleic acids. Moreover the CD difference spectra (of complexes minus components) in this wavelengths region do not generally exhibit the complex features that can be attributed to the transitions of the aromatic amino acids. Therefore the CD effects observed above 250 nm on forming protein-nucleic acid complexes are usually considered to be reflective of changes in the nucleic acid secondary structure. In **Figure 16** is reproduced the CD spectrum of a nucleic acid as a function of temperature, showing the effect of temperature melting over the dichroic signal.

8.2 Pyrene excimer fluorescence: a probe for proximity

In addition to emitting fluorescence from the excited monomer state, some fluorophores can form an excited-state dimer or excimer, by a specific interaction between the excited monomer and a ground-state monomer. When attached to proteins at specific sites, fluorophores that can form excimers can be used to provide information about changes in proximity between attachment sites. The most useful fluorophores for



this purpose are pyrene and its derivatives. Amino acid residues that selectively react with pyrene derivatives can be substituted in regions of the protein sequence that could get in close proximity upon interaction with other molecules and/or structural rearrangements. The monomeric state of the fluorophore pyrene exhibits a fluorescence spectrum with characteristic fine structures between 350 and 450 nm

The excimer, with a precise symmetrical configuration, is formed if a pyrene monomer, on excitation into its absorption band ($\lambda_{\max} = 340\text{-}343\text{ nm}$), interacts during its fluorescence lifetime in a specific manner forming parallel pairs about 3.5 \AA apart (Forster, 1969) with a neighbouring ground-state unexcited pyrene (**Figure 17**).

When the two molecules of pyrene are in proximal position, they exhibit a broad fluorescence spectrum between 450 and 600 nm (**Figure 18**); in aqueous solutions only monomer fluorescence will be produced if the pyrene molecules are in the unstacked configuration. For the stacked configuration, on excitation, either excimer fluorescence or quenching (no fluorescence) will result, depending on the ability or disability of the stacked pyrenes to reorient to the excimeric configuration according to protein conformational restraint (Lehrer, 1997). Pyrene excimer fluorescence occurs on nanosecond time scale and is determined by the microenvironment of the fluorophore which makes it ideal to probe even weak and dynamic interactions (Lakowicz, 1994).

The fluorescence property of pyrene has been successfully used to study protein/nucleic acid interactions involving RNA polymerase (Johnson et al., 1979; Preuss et al., 1997) and RecA filaments (Wittung et al., 1994)

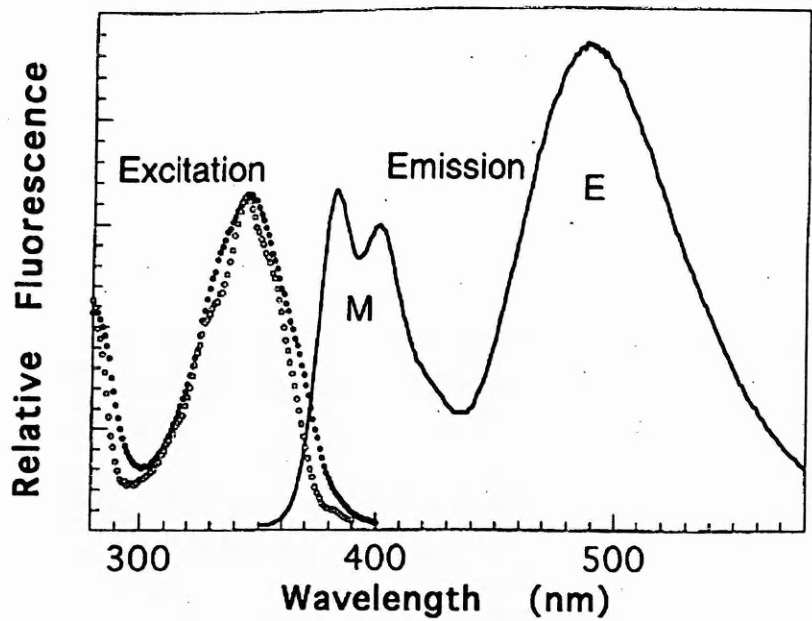


Figure 18 – Fluorescence spectra of pyrene-labeled tropomyosin in aqueous solution. M, monomer fluorescence band; E, excimer fluorescence band. The broader excitation spectrum (●) of the excimer respect to the monomer (○) is due to ground-state hydrophobic interactions between pyrenes, (from (Lehrer, 1997)).

Aim of this work

The aim of this work is to provide data on the weak interactions between proteins and nucleic acids, and to elucidate their potential contributions to specific processes, using two model systems.

It is generally supposed that proteins would first bind DNA via non-specific electrostatic interactions and then move to the specific binding site via a quasi one-dimensional diffusion. Ptashne's recruitment hypothesis supposes that components of protein complexes bind to the nucleic acids as monomers and then the specific complex is assembled on the surface of the nucleic acid (Ptashne and Gann, 1998). We were particularly interested in studying the elementary steps of this assembly using the N-terminal DNA binding domain (R69) of the cI repressor of bacteriophage 434 as a model system. In this work we use two experimental approaches to study the interaction of R69 molecules, modified via C-terminally attached additional cysteine residues (R69-Cys), in the presence and absence of cognate and non-cognate oligonucleotides. The first of these, pyrene excimer fluorescence, is an ideal system to probe even weak and dynamic interactions of pyrene-labelled R69-Cys molecules, with themselves and with DNA, thanks to the sensitivity of the fluorophore to the microenvironment and to the excimer fluorescence phenomenon. The other technique used in this study, disulfide crosslinking, relies on the assumption that formation of inter-subunit disulfide bridges is favoured if two R69-Cys molecules are bound appropriately to their specific binding sites, i.e. DNA should act as a template for disulfide crosslinking. Our experimental design is schematically shown in **Figure 19**.

The fluorescence and disulfide/crosslinking experiments described in **Section I** show that cognate DNA acts as an efficient template for the dimerization of the monomeric 434 repressor. On the other hand, a related (P22-like) DNA site does not lead to appreciable dimerization. Pyrene-fluorescence studies were found to allow the

separate monitoring of ionic (non-specific) interactions - these could be detected both with cognate and with non-cognate DNA.

The second aim of this work is the study of the interactions between the RGG motif and nucleic acids and to understand how these interactions are affected by asymmetric methylation of the arginine guanidinium group, a natural post-translational modification of some RGG sequences. As current methods for enzymatic methylation *in vitro* do not allow the production of homogeneously methylated peptides in sufficient quantities, we decided to develop a synthetic method for this purpose and use it for the synthesis of model peptides derived from the sequence of human nucleolin including the 61 aa long C-terminal domain.

The peptide representing residues 676-692 of human nucleolin (Ac-GRGGFGGRGGFRGGGRGG-NH₂,) and a number of rationally designed analogues (R=DMA or K, F=G) were constructed, and their interactions with nucleic acids studied using double filter binding, melting-curve analysis, circular dichroism difference spectroscopy. The experiments described in **Section II** show that peptides containing RGG motifs are able to modify nucleic acid conformation, and this effect is very much decreased if the RGG motif is methylated. On the other hand, dimethylation does not seem to relevantly influence the strength of binding.

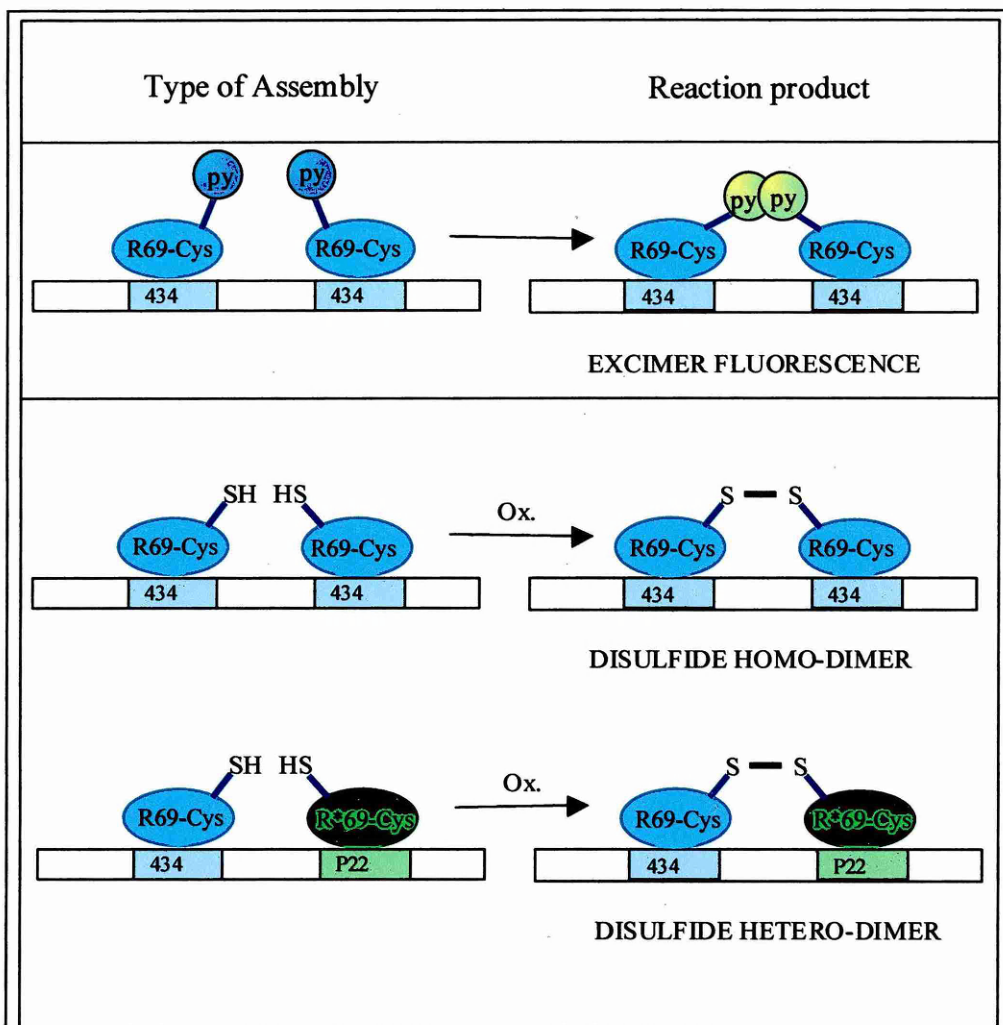


Figure 19 - Experimental design: specific DNA should act as a template for recruiting R69 molecules, leading to the appearance of excimer fluorescence (for R69-Cys-pyrene) and to the formation of disulfide dimers (for R69-Cys).

MATERIALS AND METHODS

9. MATERIALS

9.1 Solutions

General Buffers

- 1) TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0
- 2) 10x TBE buffer: 900 mM Tris base, 900 mM boric acid, 20 mM Na₂EDTA, pH 8.3
- 3) 10x SDS running buffer: 250 mM Tris base, 1.9 M glycine, 1% SDS, pH 8.3
- 4) SDS gel staining buffer: 150 ml methanol + 150 ml water + 21 ml acetic acid + 9 ml 1% Coomassie Brilliant Blue R250 (in water : methanol (1:1 v/v))
- 5) SDS gel destaining buffer: 1 litre buffer contains 200 ml methanol and 70 ml acetic acid
- 6) CD measurement buffer: 20 mM Na/K phosphate, pH 7.2

Gel-loading buffers

- 1) SDS gel-loading buffer: 125 mM Tris-HCl, pH 6.8, 30% (v/v) glycerol, 2% SDS, 6M urea, 1 M β -mercaptoethanol
- 2) EMSA gel-loading buffer: 30% (v/v) glycerol, 20-50 mM Tris-HCl, pH 7.2, low concentration bromophenol blue

Antibiotic stocks:

- 1) Ampicillin: 50 mg/ml in sterile water, store at -20 °C.
- 2) Chloramphenicol: 30 mg/ml in EtOH, store at -20 °C.

9.2 Enzymes, chemicals and purification kits

Restriction endonucleases and DNA modifying enzymes were ordered from New England Biolabs Inc., Boehringer Mannheim, Pharmacia Biotech, GIBCO BRL (Life Technologies, Inc.), Promega and Perkin Elmer. Radiochemicals were from Amersham. Protein and DNA molecular weight markers were from Pharmacia Biotech. All other chemicals were from Merck, Sigma, Aldrich, Fluka or Boehringer Mannheim. The QIAquick nucleotide removal kit, QIAquick gel extraction kits were from QIAGEN.

9.3 Oligonucleotides

Oligonucleotides were synthesized by Primm s.r.l. (Milan, Italy) or MWG Biotech (Germany). Some of the oligonucleotides have been purified by RPHPLC on a 5RP18 300Å column using a gradient of acetonitrile (0-60% in 60') in 50mM triethylammonium acetate buffer pH 5.5.

Oligonucleotides and other nucleic acids used in this work:

434O _R 1:	5' -CATACAAGAAAGTTTGTAT GTATGTTCTTTCAAACAATA-5'
434O _R 2:	5' -CAAACAAGATACATTGTAAT GTTTGTCTATGTAACATTA-5'
434O _R 3:	5' -CACACAAGAAAACTGTAAT GTGTGTTCTTTTTGACATTA-5'
GRE:	5' -GAAAAAAATGAGTCATCCG CTTTTTTTECTCAGTAGGC-5'
P22O _R 1:	5' -TATTTAAGTGTTCTTAAATG ATAAATTCACAAGAATTTAC-5'
434O _R 1-P22O _R 1:	5' -CATACAAGAAAACTTAAATAT GTATGTTCTTTTGAATTTATA-5'

Biolabs), was incubated at 37 °C for 1 hr. The RNA was precipitated by 0.3 volumes of sodium acetate buffer pH 5.2 in 3 volumes of ethanol 70% and then re-suspended in water.

10.3 RPHPLC purification:

Ultrapure water was produced by a Waters MilliQ apparatus, acetonitrile was from BDH or Riedel De Haen, trifluoroacetic uvasol from Fluka. HPLC analytical columns were from Merck (Lichrosphere 4.6x150mm 5RP18 300Å), Waters (Shodex SP-825 7x70mm) and Hewlett Packard (Zorbax Stablebond 4.6x150mm 5RP18 300Å). Preparative columns were from Waters (Radial Compression Module 25x100mm and 40x100mm, 15RP18 300Å or 15RP4 300Å). Reverse phase HPLC was generally performed using a gradient from water to acetonitrile both containing 0.1% TFA. The HPLC equipment used consist in two Waters (501 and 510 pumps) gradient analytical systems and a Gilson (306 pumps) preparative system. Cation exchange chromatography was performed on a MonoS HR10/10 column 10x100mm, on a SP Sepharose column 26x100 mm (both from Amersham Pharmacia Biotech) or on a Shodex SP-825 7x70 mm (Waters). The instrument used was an AKTA basic (Amersham Pharmacia Biotech).

10.4 Electrospray Ionization Mass Spectrometry (ESI-MS)

Mass spectrometry analysis was performed using a Perkin Elmer API 150EX electron spray, single quadrupole mass spectrometer. The samples have been analysed in positive mode, by direct syringe infusion.

The samples were either HPLC fractions dissolved in 0.1% TFA in water/MeCN or solid phase extracted from solutions containing other buffer components using ZipTipC18 (Millipore) and eluted in 0.1% TFA in water/MeCN, or RPHPLC purified, lyophilized and redissolved in a mixture of 0.1% HCOOH in MeCN/water. The MWs of

the peptides were calculated by deconvolution of the ion spectra using the software BioMultiview (Perkin Elmer).

10.5 Expression and purification of R69Cys and R*69Cys

pSETR69Cys and pSETR*69Cys plasmids encoding respectively the R69Cys and R*69Cys sequences (**Figure 28**) were constructed as described (Ruiz-Sanz et al., 1999) and were kindly provided by Dr. A. Simoncsits (Protein Structure Group, ICGEB). The plasmids were freshly transformed to BL21(DE3) pLysS strain (Novagen) to obtain about 1000 to 2000 small colonies on LB plates containing 75 mg/l ampicillin and 25 mg/l chloramphenicol after 12 to 14 hr incubation at 37°C. The colonies were suspended and grown in 3.6 l LB medium containing antibiotics as described above. When the cell culture reached A₆₀₀ of 0.6-1, IPTG was added to 0.4 mM in order to start induction. After 2 hr of induction, cells were harvested by centrifugation and resuspended in 120 ml TE buffer. The suspension was frozen at -80°C and thawed, then sonicated briefly to reduce the viscosity and centrifuged. A first step of cation exchange chromatography was performed on the supernatant using a Pharmacia Biotech SP-Sepharose column by using a linear gradient of KCl (from 50 mM to 500 mM) in TE and checking the eluted fractions by non-reducing SDS-PAGE (**Figure 20** and **Figure 21**). The collected fractions were diluted three fold with TE and purified (in two portions) on a Mono S HR 10/10 column (Pharmacia Biotech) by using a linear gradient of KCl (from 50 mM to 500 mM) in TE. The proteins eluted again in two main peaks (monomer and disulfide crosslinked dimer).

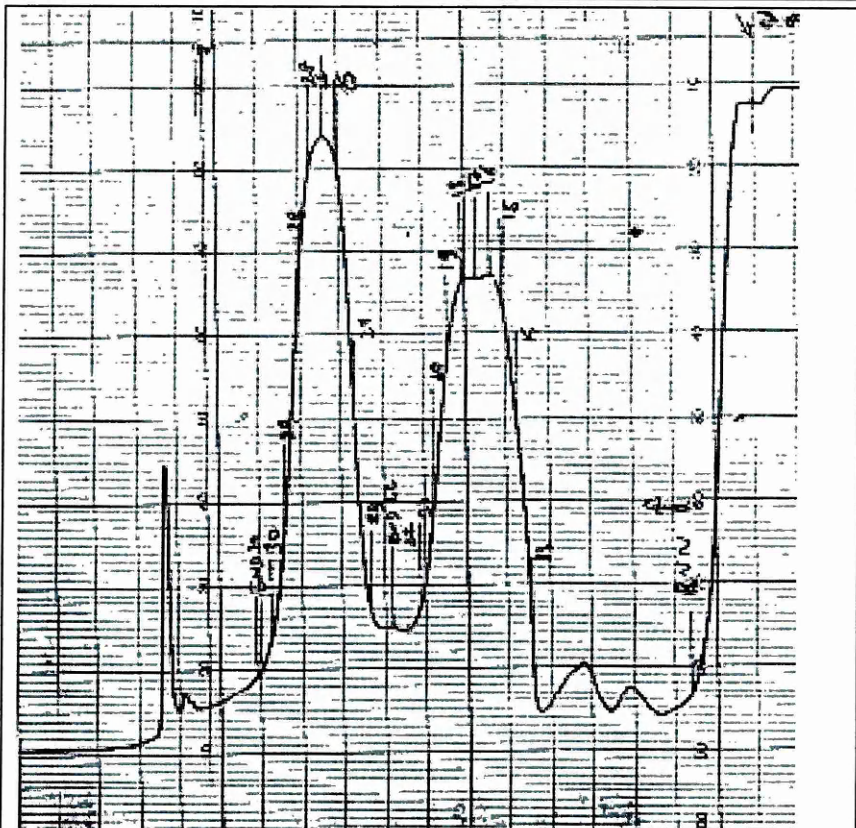
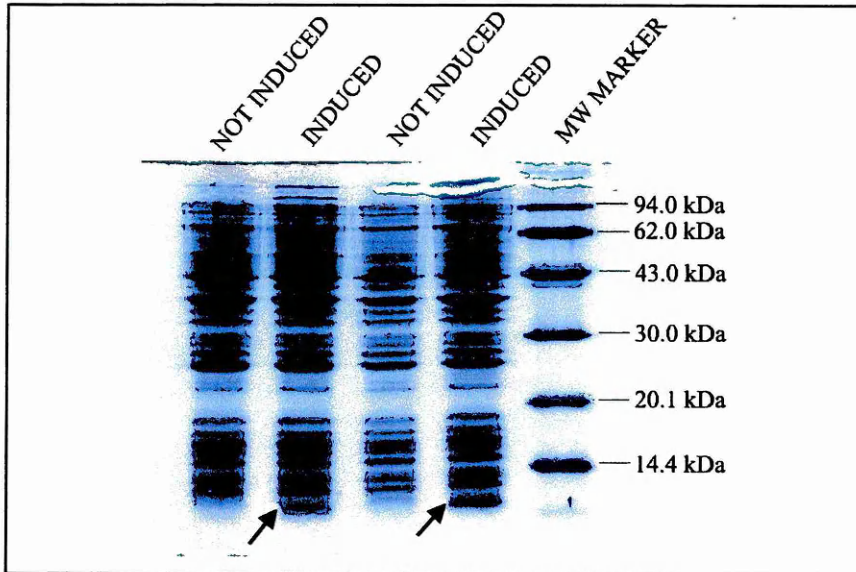
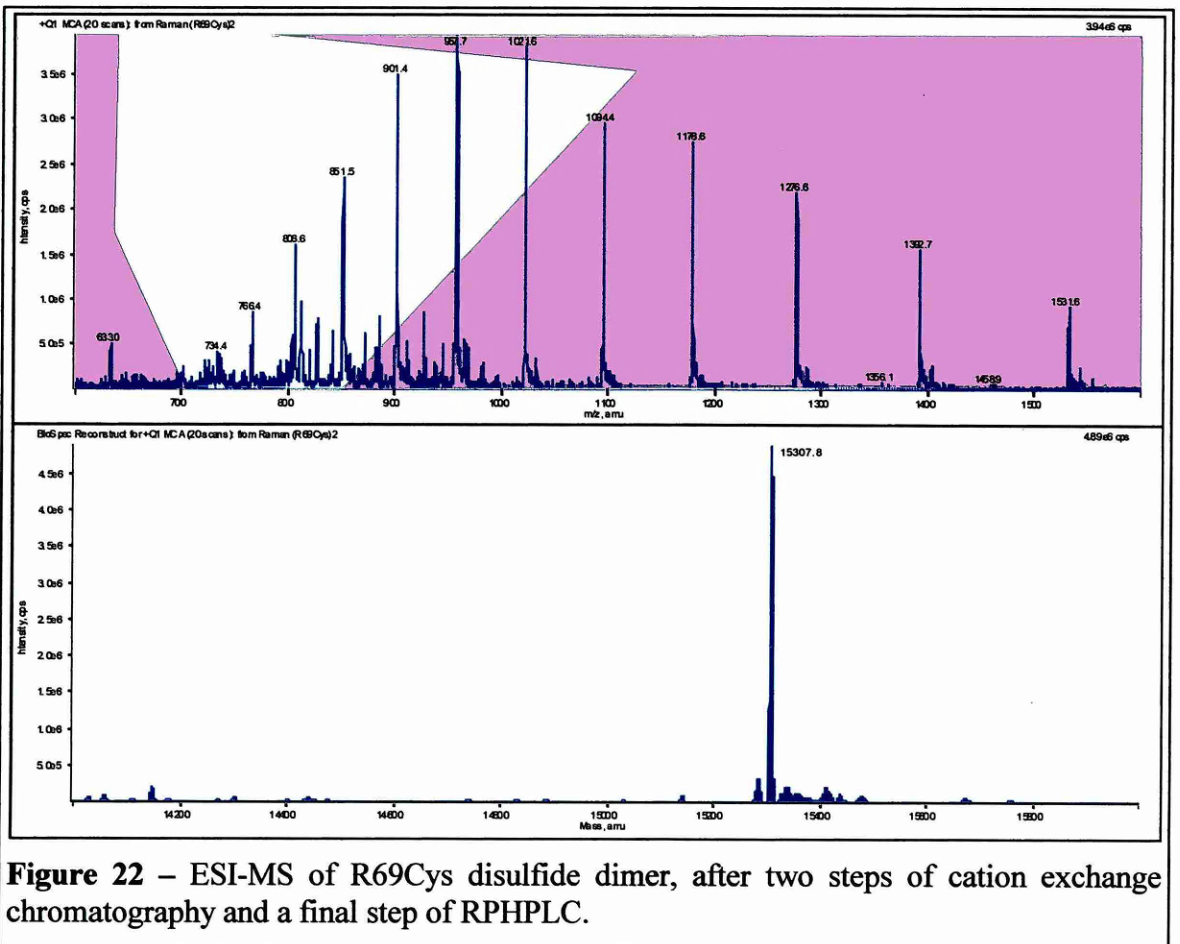
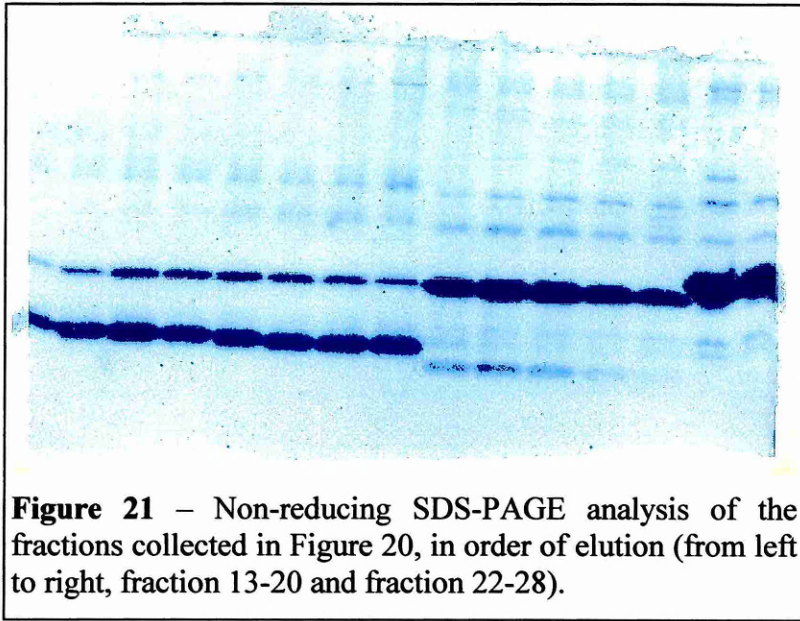


Figure 20 – *Top panel*: SDS-PAGE gel of the lysate of the bacterial expression of R69Cys before and after induction with IPTG. The arrow indicates the R69Cys band. *Bottom panel*: Cation exchange purification (SP Sepharose) of R69Cys from bacterial extract. In order of elution (from right to left), the two main peaks are R69Cys monomer and R69Cys dimer.



Both proteins were then analyzed by RPHPLC (purity > 95%, **Figure 24**) and their molecular weight confirmed by mass spectrometry (**Figure 22** and **Figure 23**).

Dimers	Average calc. mass (Da)	Observed mass (Da)
Homo-dimer of R69-Cys	15307.4	15307.8
Homo-dimer of R*69-Cys	15081.2	15081.4

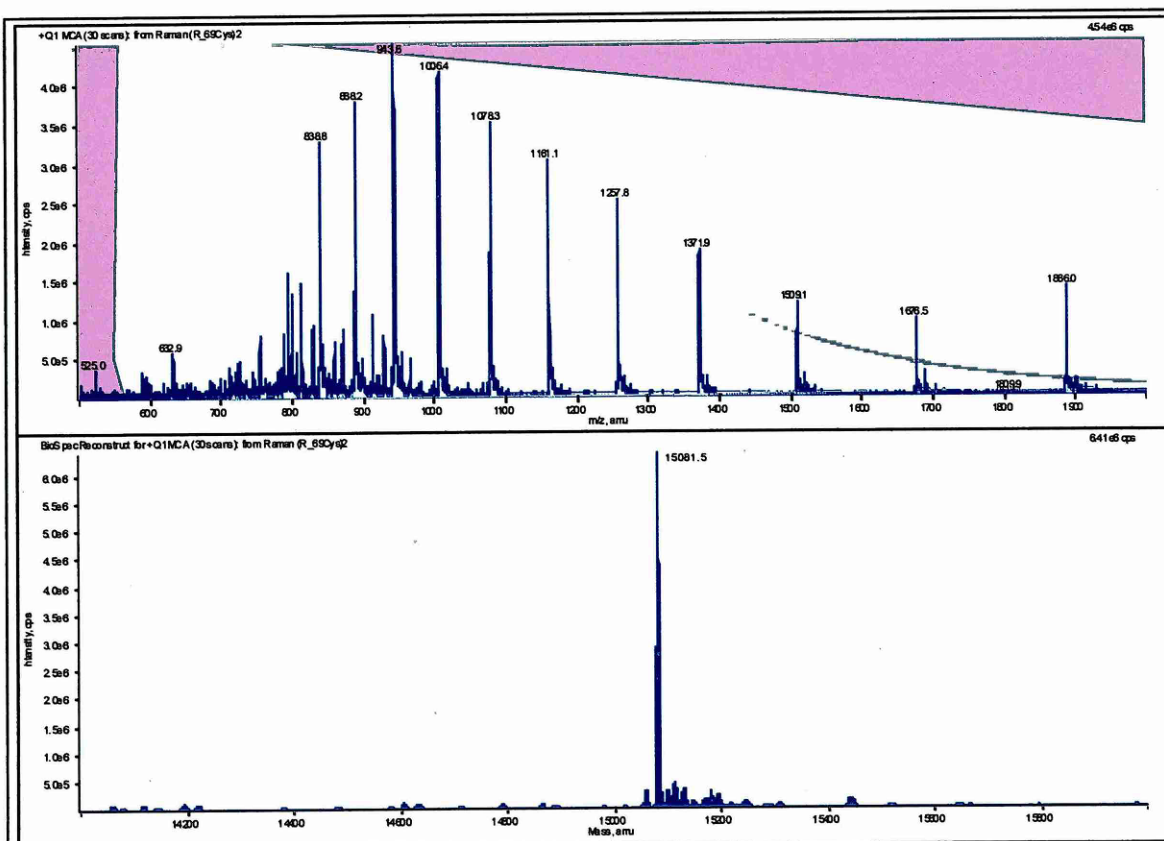


Figure 23 – ESI-MS of R*69Cys disulfide dimer after two steps of cation exchange chromatography and a final step of RPHPLC.

10.6 Electrophoretic mobility shift assay (EMSA)

Binding buffers

Buffer: 50 mM NaCl, 5 mM MgCl₂, 0.2 mM EDTA, 20 mM HEPES (N-[2-hydroxyethyl] piperazine-N-[2-ethanesulfonic acid]) pH 7.9 and 5% glycerol.

Binding reaction and electrophoresis

Binding reaction stock was prepared in 1x binding buffer containing 2.5 µg/ml sonicated salmon sperm DNA, 100 µg/ml bovine serum albumin, 6% (v/v) glycerol, <2 pM radioactively labelled DNA probe and bromophenol blue at the minimal visible concentration. Repressors were diluted with 1x binding buffer B in a concentration of ten times higher than required. 5 µl ten times concentrated repressor was added to 45 µl binding stock, and the mixture was incubated at room temperature for 1 hr. Then identical volume of reaction samples were loaded on 8% polyacrylamide gel (29:1), (prerun at 4°C in 0.5 x TBE), electrophoresis with 25 V/cm for about 2 hr. The gel was fixed in 10% (v/v) acetic acid, dried and autoradiographed. Apparent K_d values were obtained by determining the protein concentrations at half-maximal binding in protein titration experiments as described ((Simoncsits et al., 1997)).

10.7 Pyrene Excimer Fluorescence Studies

Materials: N- (1-pyrenyl) maleimide, DL-cystine hydrochloride and dithiothreitol (DTT) were purchased from Sigma Chemical Company, USA. The ds oligonucleotides were dissolved in 20 mM sodium phosphate buffer pH 7.4, and used as stock solutions for the experiments.

*Preparation of R69-Cys and R*69-Cys*

R69-Cys was prepared and expressed as described above. The samples were found homogenous by SDS-PAGE. The reduced subunits were prepared via incubation with 5mM DTT for at least 12 hrs and desalted by gel filtration on an FPLC desalting column (HiPrep 26/10 from Pharmacia) and reanalyzed by RPHPLC (**Figure 24**).

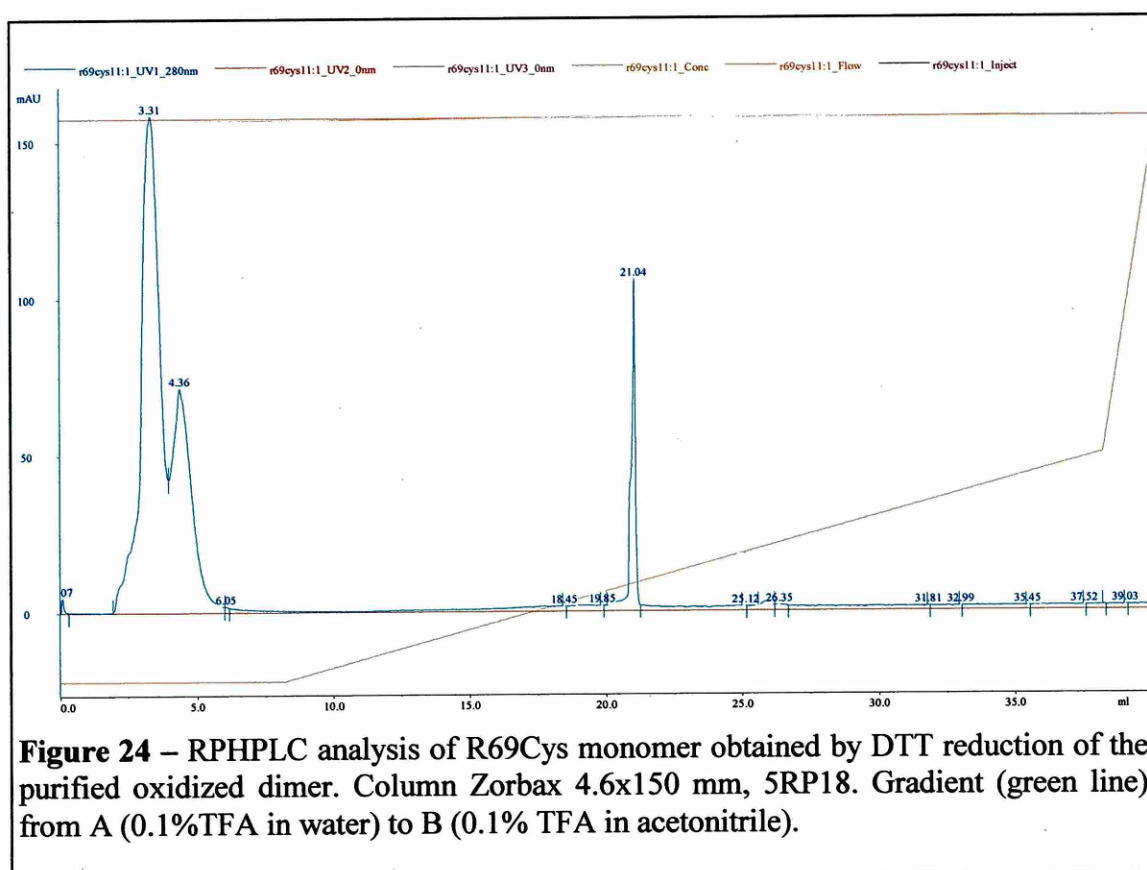


Figure 24 – RPHPLC analysis of R69Cys monomer obtained by DTT reduction of the purified oxidized dimer. Column Zorbax 4.6x150 mm, 5RP18. Gradient (green line) from A (0.1%TFA in water) to B (0.1% TFA in acetonitrile).

Labeling of R69-Cys with pyrene:

The R69-Cys molecule was covalently labeled with the fluorescent probe pyrene at its cysteine residue (R69-Cys-py). 10 mg of N- (1-pyrenyl) maleimide (Sigma), dissolved in 0.5 ml of dimethylsulfoxide (Merck) were added to 1 ml of R69-Cys solution (2 mg/ml, in 25 mM sodium potassium phosphate buffer pH 7.4 containing 5

mM DTT). The mixture was incubated at room temperature (25°C) for 24 hrs. Insoluble material was removed by centrifugation. The labeled protein was separated from the unreacted reagent by gel filtration on an FPLC desalting column (HiPrep 26/10 from Pharmacia). The labeled protein fractions were pooled and concentrated using Centricon-3 filters (Amicon). The concentration of the protein stock solution was determined by amino acid analysis. The labeling was found quantitative as determined by measuring the optical density of the probe at 344 nm, using an extinction coefficient of $44 \text{ mM}^{-1} \text{ cm}^{-1}$.

Fluorescence studies

Fluorescence spectra were recorded using a Hitachi F-4000 fluorescence spectrophotometer. The excitation and emission bandpasses were set at 5 and 3 nm respectively. Fluorescence spectra were recorded at room temperature in corrected spectrum mode with the excitation wavelength set at 335 nm. R69-Cys-py (1 μM , 0.5 ml) in 20 mM sodium phosphate buffer, pH 7.2, was titrated with increasing concentrations of the oligonucleotides or polyanion (5 μl of a 50 μM stock solution of the oligonucleotides per addition). To study the effect of NaCl, a mixture of 1 μM of R69-Cys-py and 5.5 μM of the oligonucleotide in 20 mM sodium phosphate buffer, pH 7.2, was adjusted to different salt concentrations.

10.8 Intermolecular Disulfide Crosslinking

R69-Cys or R*69-Cys (100 μM) in 20mM sodium potassium phosphate buffer (pH 8.3) was incubated with 5 mM DTT for at least 12 hrs to ensure complete reduction of the subunits (**Figure 24**). On prolonged storage, R69-Cys and R*69-Cys in solution spontaneously oxidized to disulfide dimers. We found that in the presence of Tween-20 and under controlled oxidizing conditions with thiol-exchange catalysts, they remain unoxidized monomers. The reduced subunits (13.5 μM) were incubated at room

temperature in the absence or in the presence of 44 μM oligonucleotides, in 20 mM sodium potassium phosphate buffer (pH 7.4) containing 0.04% Tween-20. The reaction was initiated by adding DL-cystine hydrochloride to a final concentration of 1.2 mM. Small aliquots (20 μl) of the samples were withdrawn at different time intervals and the reaction was quenched by mixing with 2 μl of 0.44 M acetic acid and 7 μl of SDS-sample buffer without beta-mercaptoethanol). The disulfide crosslinked dimer was separated from the unreacted monomer by electrophoresis on a 15% SDS-polyacrylamide gel. The protein bands were visualized by Coomassie brilliant blue staining, and densitometric quantification was performed on scanned images using the PACKARD Optiquant software. The reaction mixture of the assembly experiment carried on with equal amounts of R69Cys and R*69Cys on hybrid operator 434O_{R1}-P22O_{R1} was subjected, after 60 min of incubation, to a fast desalting step by RPHPLC (from 0.1%TFA in water to 0.1%TFA in MeCN in 5 min) to separate the protein fraction from the small molecular weight compounds and the nucleic acid. The full fraction containing the proteins was then analyzed by ESI-MS.

10.9 Peptide synthesis

9-Fluorenylmethyloxycarbonyl (Fmoc) protected aminoacid building blocks were ordered from Novabiochem, Inbios s.r.l., ChemImpex Inc., Bachem, Fluka, and Perseptive Biosystems. Fmoc-aDma(Mts)-OH and Fmoc-Gly(Dmb)-Gly-OH (Dmb is a 2,4-dimethoxybenzyl group) were synthesized in our laboratory as described later in the methods section. Resins for SPPS were from Novabiochem and Fluka. The TBTU (O-(benzotriazol-1-yl)N,N,N',N'-tetramethyluronium tetrafluoroborate) and PyBop (benzotriazol-1-yl-oxy-tris-(pyrrolidino)-phosphonium-hexafluorophosphate) activators were from Novabiochem, HATU (O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate) activator from Perseptive Biosystems, the

HOBt (1-hydroxybenzotriazole) from Fluka. Anhydrous N, N'-dimethylformamide (DMF), N-methyl-pyrrolidone (NMP) and piperidine (PIP) were from Biosolve; diisopropylethylamine (DIPEA), trifluoroacetic acid (TFA), ethanedithiol (EDT), thioanisole (TA), triisopropylsilane (TIPS) and trimethylsilylbromide (TMSBr) from Fluka. Peptides were synthesized by solid phase method using Fmoc chemistry. Side chain protecting groups were: Trityl (Trt for Asn, Cys, Gln), t-Butyl (But for Ser, Thr), t-Butyloxycarbonyl (Boc for Lys, Tyr, Trp), tert-butylester (OBut for Asp, Glu), 2,2,4,6,7-petamethyldihydro-benzofuran-5-sulfonyl (Pbf for Arg), mesitylene sulfonyl (Mts for aDma). The resins used, generally with substitution around 0.1-0.2 mmoles/g, were TentaGel Sieber amide (Novabiochem) for peptides with C-terminal amide, TentaGelA (Novabiochem) or PolyethyleneGlycolePolystyrene (PEG-PS, Perseptive Biosystems) for C-terminal acid, Tentagel Chlorotrytyl (TRT, Novabiochem or Fluka) for side chain protected C-terminal acid peptides or peptides with C-terminal sequence prone to diketopiperazine formation. The synthesis scales used were 0.05 mmoles and 0.1 mmoles and aminoacid excess was 6x. According to the difficulty presented by the peptide sequence, automatic or manual synthesis was adopted. A Milligen 9050 Pepsynthesizer or a Protein Technologies PS3 synthesizer was used for automatic mode.

Manual synthesis was performed in 6 ml or 20 ml polypropylene syringes (IST) with polyethylene frits, mounted on a Solid Phase Extraction (Macherey Nagel) vacuum chamber for the solvent washes and subjected to continuous nitrogen bubbling. A typical protocol for the manual or automatic elongation cycle was:

Step	Reagent	Time	Repeats	Amount
Resin wash	DMF	1'	3	10ml
Fmoc deprotection	20%PIP in DMF	7'	2	15ml
Resin wash	DMF	1'	3	10ml
Aminoacid acylation	AA/TBTU/DIPEA(6x/6x/12x) in DMF	60'	1	AA conc. 0.3M
Resin wash	DMF	1'	3	10ml

After each coupling cycle, before Fmoc deprotection, a manual ninhydrin test for detection of free amines was performed to check the completeness of the coupling; if positive the coupling was repeated changing the activator (HATU or PyBop) and increasing the coupling time until completeness of the acylation.

Resin cleavage and side chain deprotection:

The resin was carefully washed in dimethylformamide, dichloromethane, methanol and dried under vacuum. The peptides were then cleaved from the resin and side chain deprotected using different mixtures (20ml per gram of resin):**A)** TFA/TIPS/water (95/2.5/2.5) for 2hrs, RT: peptides with no Cys, Trp, Met, Arg or Dma
B) TFA/TIPS/EDT/water (90/2.5/5/2.5) for 3 hrs, RT: peptides with Cys, Trp, Met or Arg
C) TFA/TIPS/EDT (92.5/2.5/5) for 1 hour, RT and then TMSBr/TA/EDT (50/25/25) for 15' to 60' in ice: peptides with multiple Arg, Dma and Gly(Dmb)Gly residues.

The cleavage mixture was separated from the resin by filtration on a fritted syringe and the resin further washed with TFA (3 x 2ml). The excess TFA was then removed by evaporation on a rotavapor (Heto) to leave 2 or 3 ml of solution and the peptide was then precipitated by cold diethyl ether (15-20 volumes); the precipitation mixture was left on ice for 30' and then centrifuged for 10' at 3000 rpm. The pellet was then washed with diethylether and centrifuged for four times or re-dissolved in water and extracted four times with equal volumes of diethyl ether. Finally the aqueous phase was collected and lyophilised.

10.10 Synthesis of Fmoc-Dma(Mts)-OH

Fmoc-Dma(Mts)-OH was prepared in two steps starting from Boc-Orn-OH (or Z-Orn-OH). The first thiomethyl group of a S,S-dimethylarensulfonyliminodithiocarbonimidates ($\text{Mts-N}=\text{C}(\text{SMe})_2$) reacts with the

sodium salt of the N α -protected ornithine in near to quantitative yield. The second thiomethyl group in the so formed S-methylisothiourea reacts with dimethylamine in the presence of heavy metal salts (Ag⁺ - or Hg²⁺) generating N α -protected -Dma(Mts)-OH. The N α -protecting group is then removed to introduce the Fmoc. The total yield of the synthesis is 94%. This method, developed in our group (Szekely et al., 1999), is generally effective for the synthesis of other N $^{\omega}$ -substituted arginines.

10.11 Synthesis of Fmoc-Gly(Dmb)Gly-OH

Fmoc-Gly-Cl (1) was prepared from Fmoc-Gly-OH and SOCl₂ in the presence of catalytic amount of DMF according Carpino et al. (Carpino et al., 1986).

20mmol of (1) were dissolved in 40 ml dry THF to give **solution A**.

Tms-(Dmb)Gly-OTms. (2)

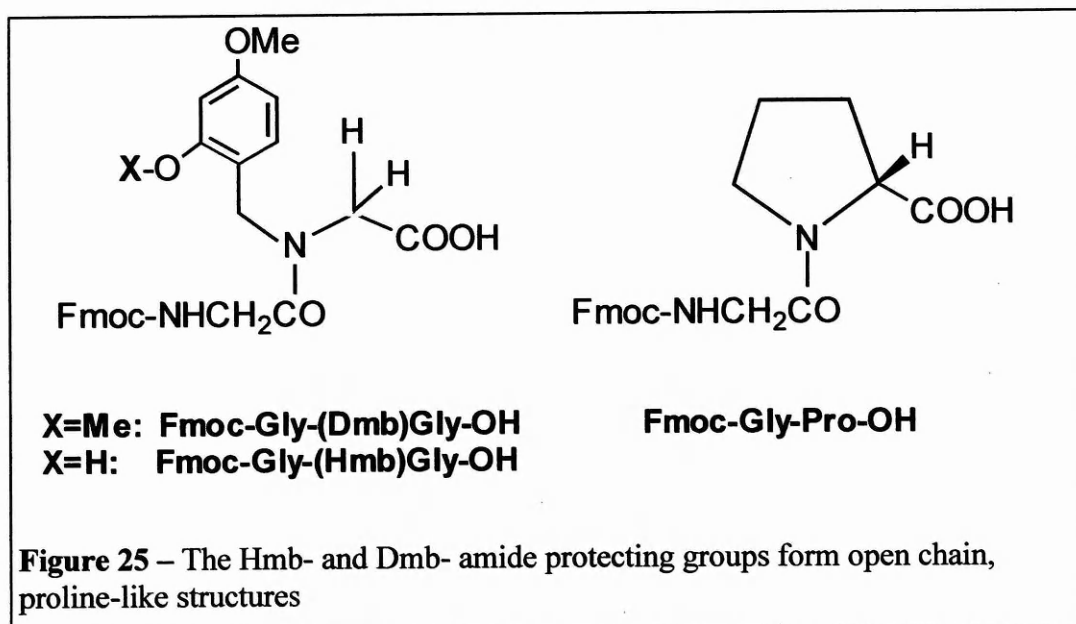
30 mmol (1.5 eq excess) of H-(Dmb)Gly-OH (Dmb is a 2,4-dimethoxybenzyl group) was prepared according to Johnson et al. (Johnson et al., 1995) with some modification and crystallised from water with yield 80-85%. It was then silylated with 15 mmol Bis-trimethylsilylacetamide in 40 ml DCM, containing 0.5 ml Tms-Cl until dissolved (15 min) and then 30 mmol DIEA were added. (**solution B**)

Fmoc-Gly(Dmb)Gly-OH (Figure 25):

Both solutions **A** and **B** were cooled to -10 -20 ° C and then the solution A was added to well stirred solution B at temperature <-10 ° C. The reaction mixture was stirred 1 h at room temperature and the organic solvent was evaporated. 100 ml Et₂O and 10 ml of a 10 % solution of citric acid were added and stirred for 30 min. Then the pH was adjusted to 8.5-9 with saturated solution of K₂CO₃ and extracted 2 times with 200 ml Et₂O. The water phases were acidified with sulfuric acid to pH 2.5 - 3 and the product was extracted with EtOAc (3x100 ml). Collected organic phases were washed with water, brine, dried with Na₂SO₄, filtered and evaporated in vacuo to dryness. Yield 9.92 g (98.1%, oil) - RPHPLC purity (87%-94%) . The product was dissolved in CHCl₃ (10

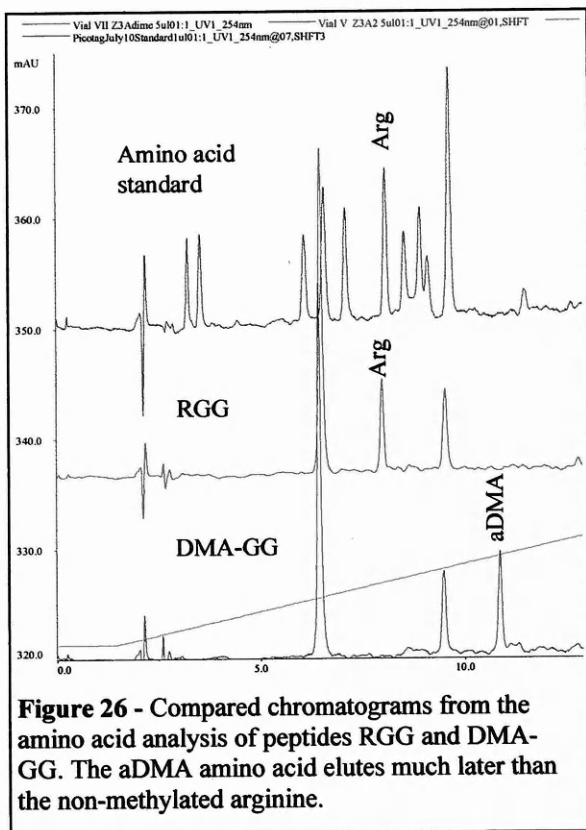
ml) and applied on a column (15X5 cm I.D.) filled with Kieselgel 60 Merck type 9385 (230-400 mesh) equilibrated in CHCl₃. The column was eluted with CHCl₃ and then with gradient from CHCl₃/MeOH/AcOH=96/3.5/0.5 to CHCl₃/MeOH/AcOH=93/6.5/0.5 (1.2 l each one).

The pure fractions, according TLC (CHCl₃/MeOH/AcOH=9/1/0.1), R_f=0.42) or RPHPLC analysis (R_t=26.4 min, 1ml/min, 3% slope), were combined and evaporated in vacuo, dissolved in 200 ml EtOAc and washed with 0.01M H₂SO₄ (3X100 ml), with water, brine, dried over MgSO₄, filtered,



evaporated to ~ 30 ml and precipitated with petroleum ether (10 volumes). The product was separated from the solvents and dried at 40 °C in vacuo. Yield 7.05 g (69.7%) with purity > 98.5% and 2.5 g with purity ~ 70%. The latter fraction was applied together with the product from the other batch, for re-purification.

Fmoc-Gly(Dmb)Gly-OH (C₂₈H₂₈N₂O₇, MW 504.19) analysed by ESMS gave several peaks corresponding to: 1047.4 [non covalent dimer +K]; 1031.2[non covalent dimer +Na]; 1009.5[non covalent dimer]; 543.3[MK+]; 527.2[MNa+]; 505.2[MH+]



10.12 Aminoacid analysis:

Aminoacid standards, HCl 6N constant boiling and PITC were from Pierce. The HPLC column used in the aminoacid identification step was a Waters Picotag Column 4.6x300mm.

5-20ul (containing few hundred picomoles) of RPHPLC purified peptide fractions were subjected, after freeze-drying, to overnight hydrolysis with HCl at 110°C in glass vials sealed under vacuum in the presence of a small crystal of phenol.

After a step of resuspension in EtOH/TEA/water (2/1/1) to reach a basic pH, the hydrolyzate was derivatized with PITC/EtOH/TEA/water (1/7/1/1), evaporated to dryness and analyzed in RP-HPLC (Äkta, Pharmacia Biotech) at 254nm with a gradient of acetonitrile in Na-acetate buffer pH 6.4 in the column thermostated at 46°C. The amounts of the aminoacids were calculated against a calibration curve constructed with a standard containing 19 aminoacids (see example in **Figure 26**).

10.13 Trypsin digestion of RGG and DMA-GG peptides

75 nmoles of RGG or DMA-GG peptide were dissolved in 1500 µl of 20mM ammonium acetate, pH 8, containing 1 mM calcium chloride. Trypsin sequencing grade (Promega) was dissolved in the same buffer and added at an enzyme substrate ratio of

1:50 (mol/mol). The solution was incubated at 37° C and aliquots of 100 µl were taken and immediately injected in RPHPLC at regular intervals.

10.14 Double filter binding assays

Double filter binding experiments were carried out as described by Wong and Lohman (Wong and Lohman, 1993). Nitrocellulose, DEAE membranes and the 96-wells dot-blot apparatus were from Schleicher & Schuell. The DEAE membrane was subjected to a washing step in 0.1M Na₄EDTA (10'), then for three times in 1M NaOH (1') followed by 1M NaCl (10') and finally equilibrated in the binding buffer (20mM Na-phosphate pH 7.2). The nitrocellulose membrane was washed once in 0.1M KOH (10') and then equilibrated in the binding buffer.

The binding mixtures were prepared starting from stock solutions of peptides (200uM) with concentration determined by amino acid analysis. ³²P labelled DNA, and RNA synthetic oligonucleotides (1 nM) were titrated with increasing amounts of non-methylated and methylated peptide in a 20 mM sodium phosphate buffer (pH 7.2 at 25 °C) in a final volume of 60 ul. After one hour incubation to reach the equilibrium, 25ul of the binding mixture was applied per each well of the dot-blot apparatus after a pre-wash of the membrane with 100ul of the binding buffer. After vacuum application a new wash of 100 ul of binding buffer was applied to rinse the well. At the end of the assay, both membranes were exposed by phosphoimaging on a Canberra Packard Cyclone instrument.

The amounts of radioactive oligonucleotides bound to the membranes have been quantified by the Canberra Packard Optiquant software and the data analyzed and

plotted with the software Kaleidograph. The estimated error of the assay is in terms of $\text{DNA}_{\text{bound}}/\text{DNA}_{\text{total}}$ ratio ± 0.03 .

10.15 DNA melting curves

Melting temperatures were measured on a Pharmacia Biotech Ultraspec 3000 spectrophotometer equipped with a Peltier heated cell holder and a temperature control unit under computer control. Quartz micro cuvettes of 1 cm path length and 200 μl working volume were used. Absorbance was monitored at 260 nm between 30 and 60 $^{\circ}\text{C}$, with heating and cooling rates of 0.5 $^{\circ}\text{C}/\text{min}$. T_m values were calculated using the SWIFT software (Pharmacia Biotech) with an estimated error of ± 0.5 $^{\circ}\text{C}$. The double stranded $A_{25}T_{25}$ DNA was prepared from equimolar amounts of the corresponding synthetic oligonucleotides and annealing at 65 $^{\circ}\text{C}$ for 5 min. The concentration of dsDNA in samples was calculated from the absorbance at 260 nm at 60 $^{\circ}\text{C}$, assuming a molar extinction coefficient of 590000 calculated from the base composition. Peptide concentration was calculated from amino acid analysis. All samples were in sodium phosphate buffer 20 mM, pH 7. DNA concentration was 1.7 μM and measurements were repeated with a peptide/DNA molar ratio of 150, 50, and 10.

10.16 Circular dichroism spectroscopy

The peptides were dissolved in 20 mM sodium potassium phosphate buffer (pH 7.2). The concentrations of these stock solutions of peptides were estimated by amino acid analysis. A Jasco J-650 spectropolarimeter was used. 1 ml solutions of the particular nucleic acids (34 $\mu\text{g}\cdot\text{ml}^{-1}$ MS2-phage RNA, tar-RNA and tRNA, 50 $\mu\text{g}\cdot\text{ml}^{-1}$ ss or ds DNA) in 20 mM sodium potassium phosphate buffer (pH 7.2) was taken and was

titrated with increasing concentrations of the peptides. The dilution of the samples upon titration was about 5%. CD spectra were recorded at each addition of the peptides.

The spectra were corrected for solvent contributions and for dilution. The interaction of the peptides with nucleic acids was monitored by the change in the circular dichroism spectra of nucleic acids, subtracting the spectrum of the peptide in buffer solution from the spectrum of the complex in solution.

Results:

Section I: Dimerization of the DNA-binding domain of bacteriophage 434 repressor on DNA

11. Computer modelling

The planned experimental design relies on a fundamental structural requirement: the C-terminal ends of two R69-Cys molecules (sequences in **Figure 28A**), upon

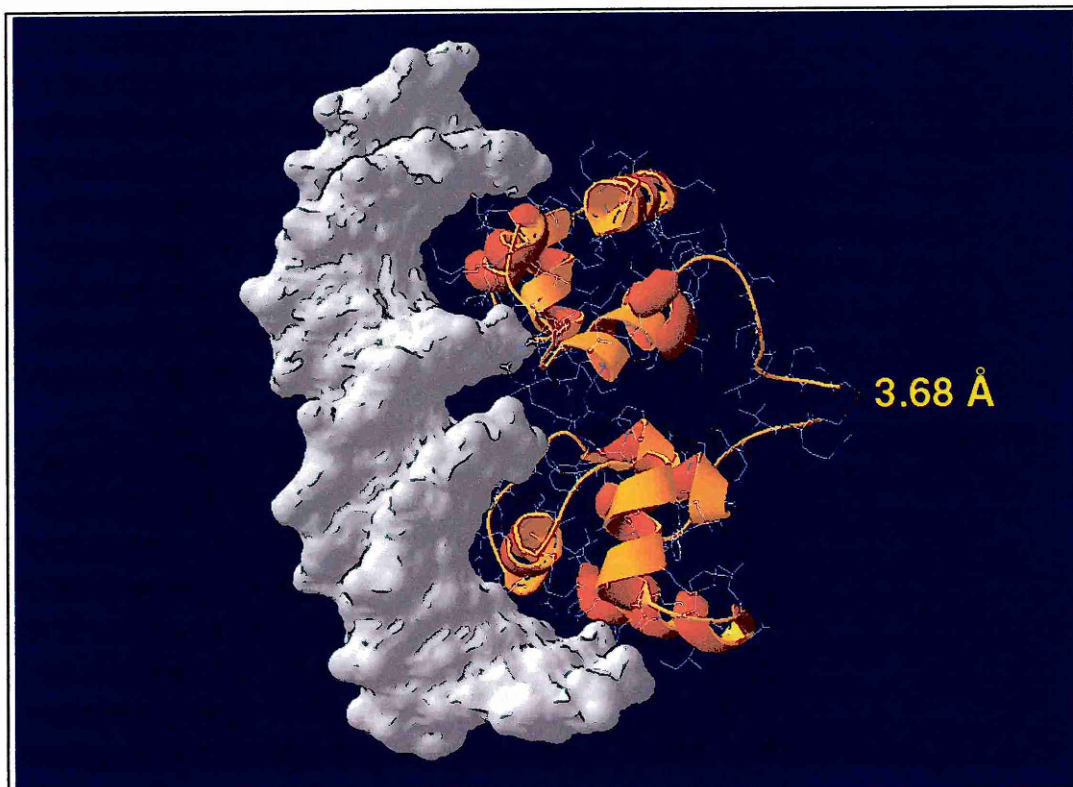


Figure 27 – Computer model of two R69-Cys monomers in complex with DNA. The coordinates of residues 1-69 were taken from the NMR structure of R69 in solution (Brookhaven PDB 1PRA) and superimposed (1-63 residues) on the peptide backbone of the resolved X-ray R69/ O_R2 complex (1RPE); 3.68 Å refers to the distance of the sulphur atoms of the monomers in one of the possible conformers after addition of the two cysteine residues (Cys70). The model was constructed using the program Swiss-PdbViewer v3.6b3.

eventual DNA binding, should not be impaired by structural constraints in order to allow the correct positioning of the pyrene moieties and of the cysteine residues at a distance compatible with excimer fluorescence and disulfide formation.

In order to verify this hypothesis we constructed a computer model of the R69-Cys molecule interacting with the cognate O_R2 DNA on the base of two structures deposited

in the Brookhaven Protein Data Bank, 1PRA (Neri et al., 1992) and 1RPE (Shimon and Harrison, 1993). The NMR structure of R69 in solution (1PRA) has been necessary in the calculation due to the fact that in the X-ray R69/ O_R2 complex structure (1RPE), no coordinates are included for the last six residues of the protein since they are disordered in the crystal and cannot be seen in the electron density map. Indeed these six residues show a very high degree of flexibility also in the NMR structure, which includes twenty possible conformers. The peptide backbones of the common residues (1-63) of the two structures have been superimposed and after adding the Cys residue in position 70 we verified that among the possible NMR resolved conformers of the C-terminal tail there was clear indication of a narrow neighbouring of the two cysteines (**Figure 27**) in some of the computed structures.

12. Expression and purification of the R69 molecules

The R*69Cys molecule was designed modifying the R69Cys sequence according to the “helix swap” experiment carried on by Wharton & Ptashne (Wharton and Ptashne, 1985) and described in the introduction. Amino acids at positions -1, 1, 2, and 5 of the α 3-helix of the 434 repressor were replaced with the corresponding residues of the P22 phage c2 repressor. This modification was previously shown to confer P22-like specificity to the native 434 repressor (Wharton and Ptashne, 1985) and to homodimeric single-chain constructs (Simoncsits et al., 1997).

R69-Cys and R*69-Cys were obtained by E.coli expression of a pET vector containing the cloned sequences as described in Materials and Methods. The proteins (sequences shown in **Figure 28A**) were purified as disulfide crosslinked dimers by FPLC on a Pharmacia SP-Sepharose ion-exchange column followed by a Mono-S HR10/10 column. The samples were found homogenous by SDS-PAGE and RPHPLC to over 95% and their molecular weights were confirmed by mass spectrometry.

When necessary, both the reduced monomeric subunits were prepared via incubation with 5mM DTT for at least 12 hrs and desalted by gel filtration on an FPLC

desalting column. The protein concentrations were estimated by using the calculated molar extinction coefficient at 280 nm ($\epsilon = 5690$).

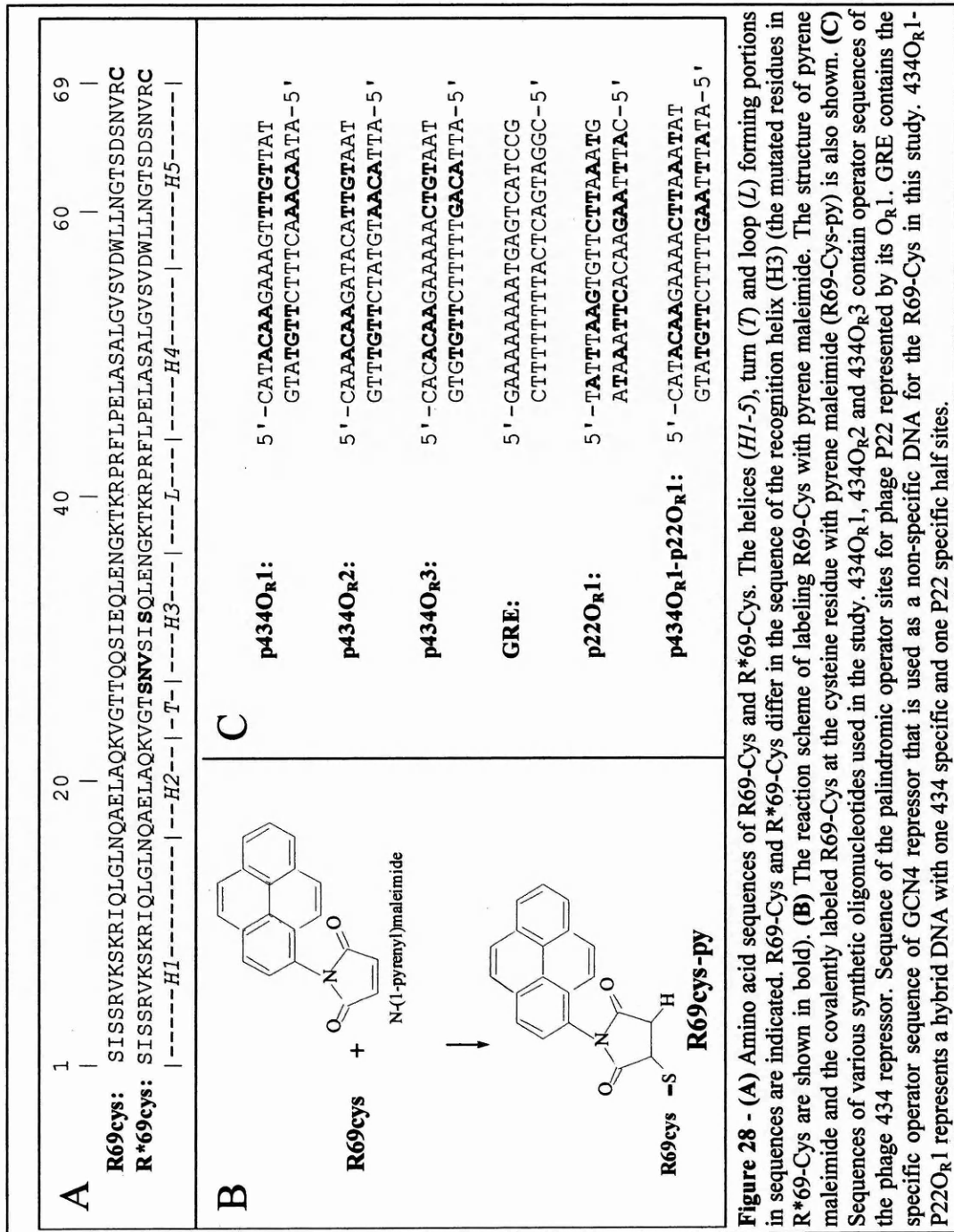
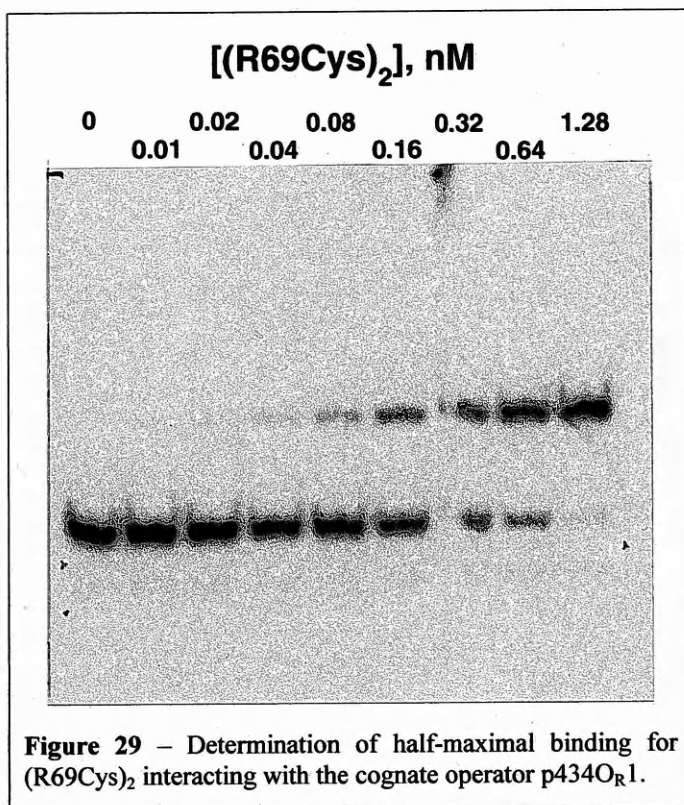


Figure 28 - (A) Amino acid sequences of R69-Cys and R*69-Cys. The helices (H1-5), turn (T) and loop (L) forming portions in sequences are indicated. R69-Cys and R*69-Cys differ in the sequence of the recognition helix (H3) (the mutated residues in R*69-Cys are shown in bold). **(B)** The reaction scheme of labeling R69-Cys with pyrene maleimide. The structure of pyrene maleimide and the covalently labeled R69-Cys at the cysteine residue with pyrene maleimide (R69-Cys-py) is also shown. **(C)** Sequences of various synthetic oligonucleotides used in the study. 434O_{R1}, 434O_{R2} and 434O_{R3} contain operator sequences of the phage 434 repressor. Sequence of the palindromic operator sites for phage P22 represented by its O_{R1}. GRE contains the specific operator sequence of GCN4 repressor that is used as a non-specific DNA for the R69-Cys in this study. 434O_{R1}-P22O_{R1} represents a hybrid DNA with one 434 specific and one P22 specific half sites.

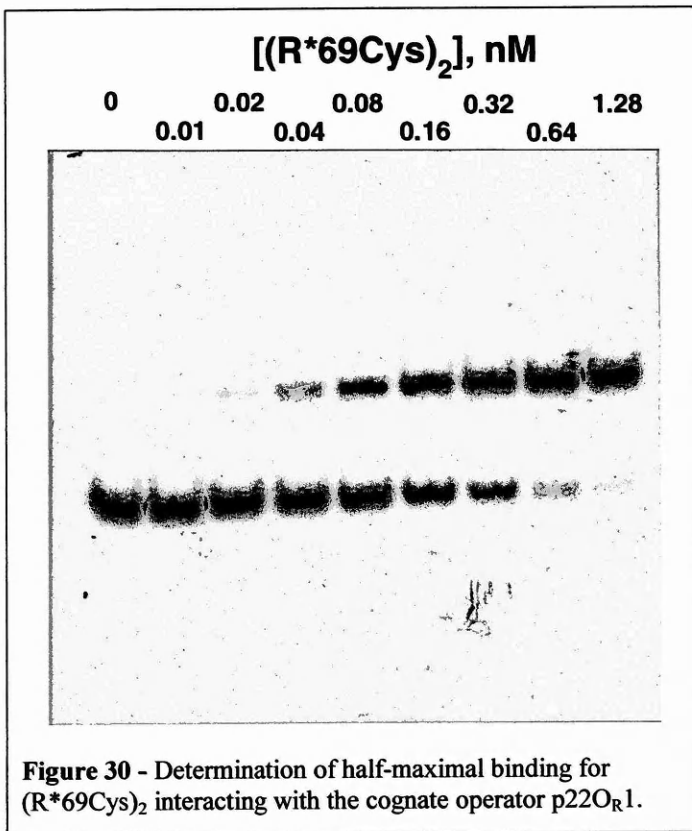
13. Disulfide dimers of R69Cys and R*69Cys bind *in vitro* to their cognate operators



In our design the final product of the correct assembly of two R69-Cys molecules on DNA would be the disulfide dimer. We decided to test the stability and specificity of this complex using preformed disulfide dimers in a DNA-binding experiment *in vitro*. The combination of two 434 DNA-binding domains to yield artificial DNA-binding dimers is not new in itself. A previous work already showed that covalent dimers connecting two 434 N-

terminal domains (residues 1-63) either through the C-terminus (ChD) or as direct sequence repeats in a recombinant fragment (ReD) can bind their specific targets with high affinity (K_d were 2-3 x 10⁻⁸ M and 5-6 x 10⁻⁹ M respectively) (Percipalle et al., 1995). The arrangement of our disulfide dimers resembles the ChD framework, but with a major difference due to the length of the domains and of the linker used. In the previous study the domains connection was indeed represented by a synthetic flexible linker constituted by two residues of aminohexanoic acid linked at each amino group of a lysine for a total length of around 36 Å, in the present study the connection is limited to the Cys 70 disulfide bond.

Electrophoresis mobility shift assay (EMSA) was used to demonstrate the specific association of the disulfide dimers of R69Cys and R*69Cys with the respective



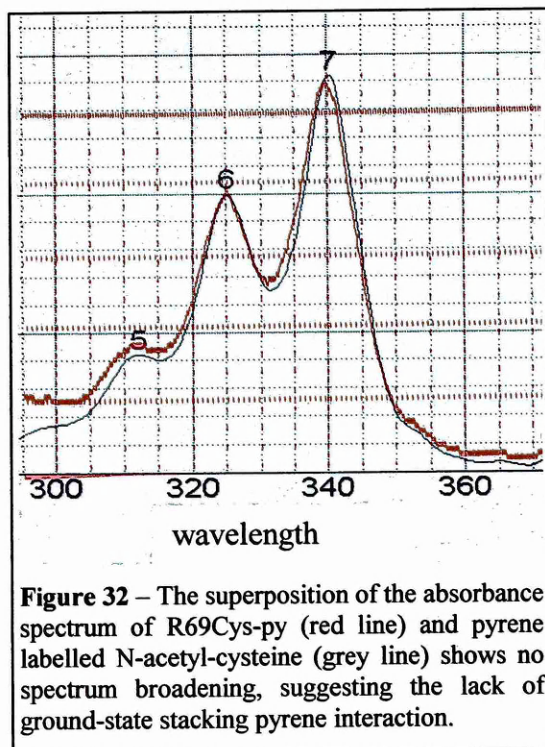
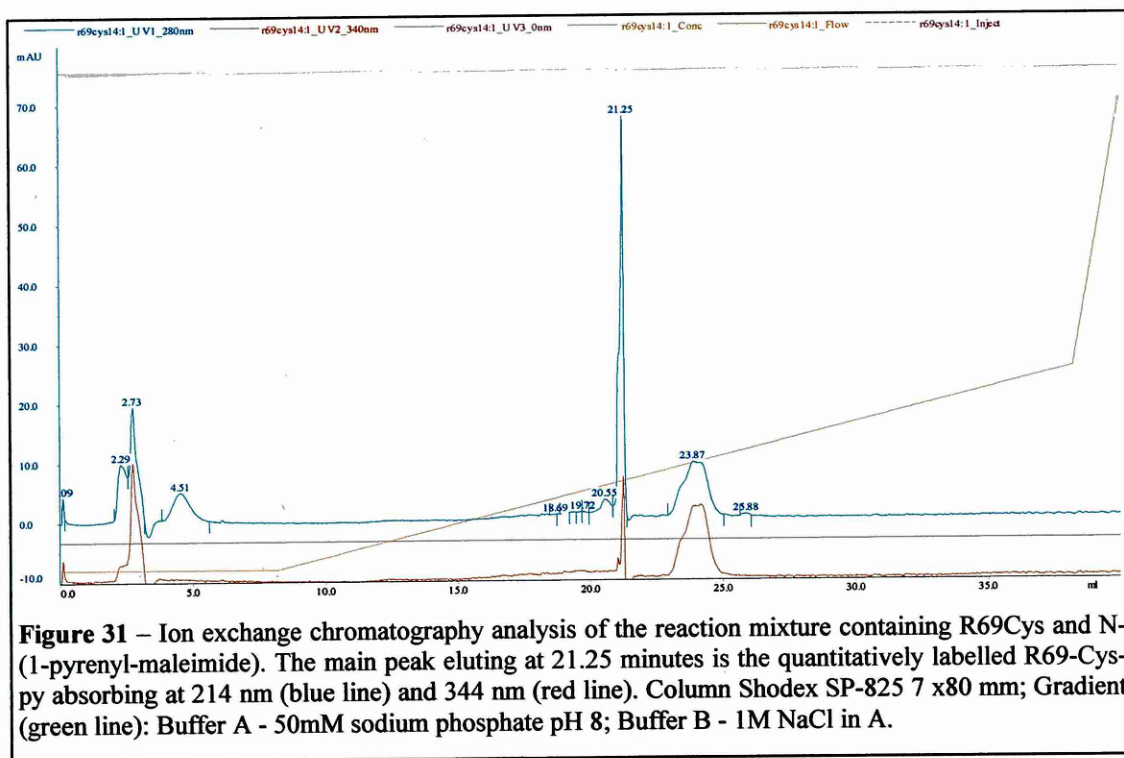
natural operator sites, $p434O_{R1}$ and $p22O_{R1}$. The sequences of the operators are shown in **Figure 28C**. The experiments were carried out increasing the protein concentration in the presence of 1 pM operator DNA and with a 1000 fold excess of competitor salmon sperm DNA. Both proteins showed specific binding activities towards their cognate binding sites. The protein concentration necessary to bind 50% of the labelled DNA

represents the apparent equilibrium dissociation constant (K_d) if the DNA probe concentration is negligible compared to the protein concentration and the binding equilibrium is not perturbed under the assay conditions (Carey, 1991). In this assay the R69Cys dimer showed an approximate K_d value of 3.2×10^{-10} M and the R^*69Cys dimer of 1.6×10^{-10} M (**Figure 29, Figure 30**).

These figures are in close agreement with previously published data on engineered protein frameworks based on direct repeat of two R69 or R^*69 molecules (1×10^{-9} M and 5×10^{-10} M respectively (Simoncsits et al., 1997)) and with those reported for naturally dimerized 434 and P22 repressors (Bell and Koudelka, 1993; Hollis et al., 1988; Koudelka and Lam, 1993; Wharton and Ptashne, 1985).

14. Fluorescence Studies

14.1 Labeling of R69-Cys with pyrene:



The R69-Cys molecule was covalently labeled with the fluorescent probe pyrene at its cysteine residue (R69-Cys-py) according to the scheme in **Figure 28B** using N-(1-pyrenyl) maleimide. The labeling was found quantitative as determined by RPHPLC (**Figure 31**) and by measuring the optical density of the probe at 344 nm, using an extinction coefficient of $44 \text{ mM}^{-1}\text{cm}^{-1}$ (Hangland, 1992).

14.2 Cognate operators induce monomer fluorescence quenching and excimer fluorescence of R69-Cys-py

Pyrene fluorescence was studied in the presence of operator and non-operator (non-specific) DNA (see **Figure 28C** for the sequences). Fluorescence spectra were

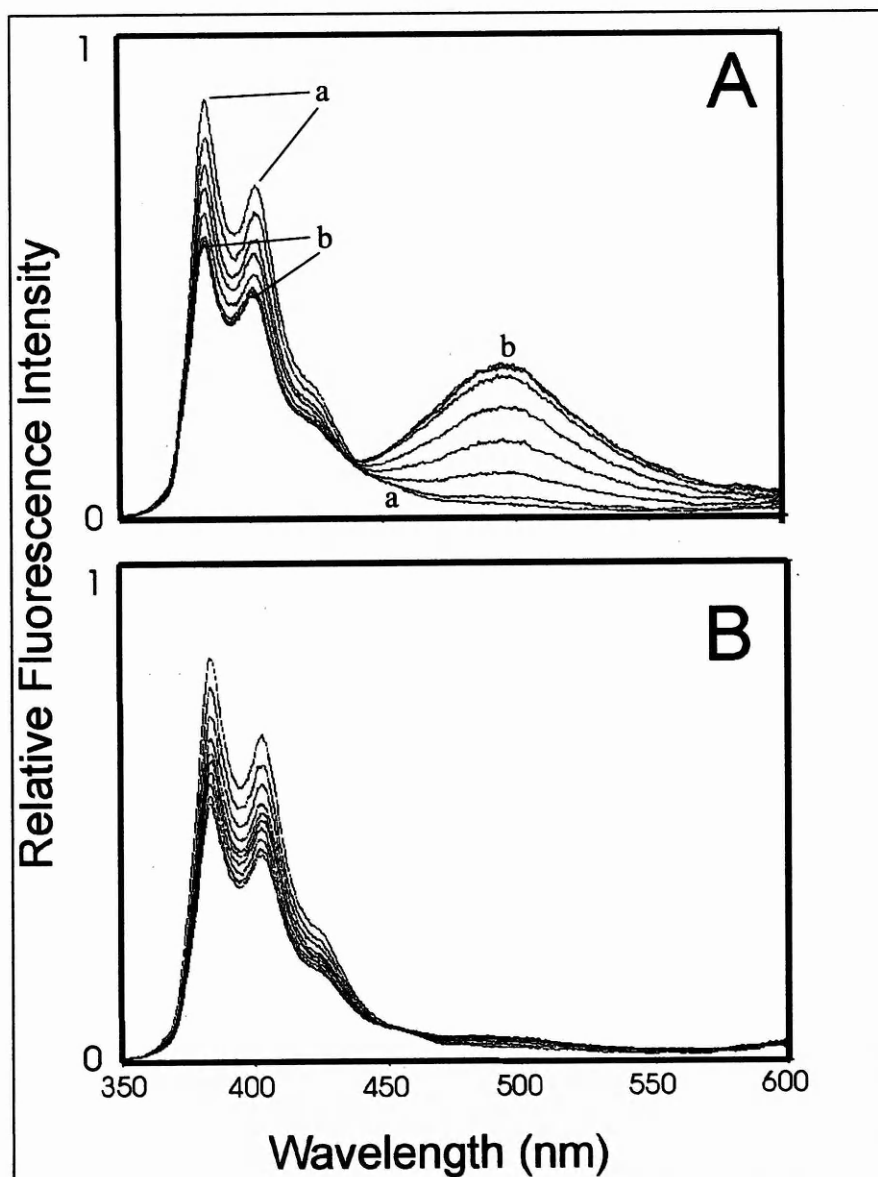


Figure 33 - Fluorescence spectra of R69-Cys-py (1 μM) in the absence and in the presence of (A) the specific (434OR1) and (B) the non-specific (GRE) DNAs. Spectra represented correspond to increasing DNA concentrations: 0.00, 0.48, 0.94, 1.40, 1.85, 2.29, 2.73 and 3.15 μM DNAs respectively. Spectrum a is in the absence of DNA, spectrum b is obtained at the highest (3.15 μM) DNA concentration. Spectra overlap above 2.29 μM of the specific DNA. Spectra corresponding to 4.90 and 6.66 μM of the non-specific DNA are also shown in panel B. -

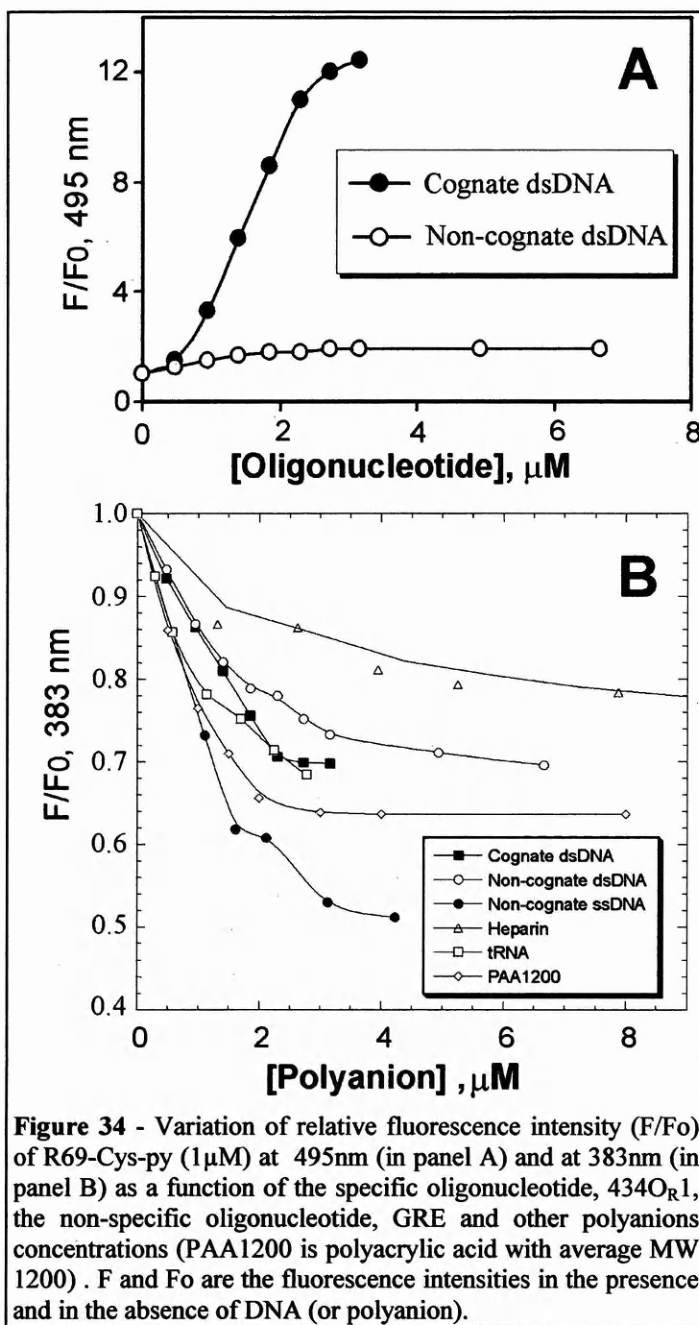
recorded at room temperature in corrected spectrum mode with the excitation wavelength set at 335 nm. R69-Cys-py (1 μ M, 0.5 ml) in 20 mM sodium phosphate buffer, pH 7.2, was titrated with increasing concentrations of the oligonucleotides (5 μ l of a 50 μ M stock solution of the oligonucleotides per addition).

Figure 33A shows the fluorescence spectra of R69-Cys-py in the absence and in the presence of increasing concentrations of the operator DNA, 434O_{R1}.

In the absence of DNA, R69-Cys-py shows an absorbance (300-370nm, **Figure 32**) and a fluorescence spectrum (**Figure 33**) similar to pyrene labelled N-acetylcysteine; the fluorescence spectrum shows two prominent sharp fluorescence peaks centred at 383 and 405 nm, respectively, and exhibits almost no fluorescence between 450 and 600 nm which is the known range of excimer fluorescence resulting from stacked fluorescent labels. These data indicate that R69-Cys-py does not undergo detectable self-association and exists as monomer in the absence of operator DNA.

Upon increasing the concentration of the specific DNA, 434O_{R1}, the fluorescence intensity below 450 nm decreases with a concomitant appearance of excimer fluorescence between 450 and 600 nm (**Figure 33A**).

The relative fluorescence (F/F_0) titration curves for the monomeric (383 nm) and for the dimeric species (495 nm) can be plotted separately (**Figure 34**).

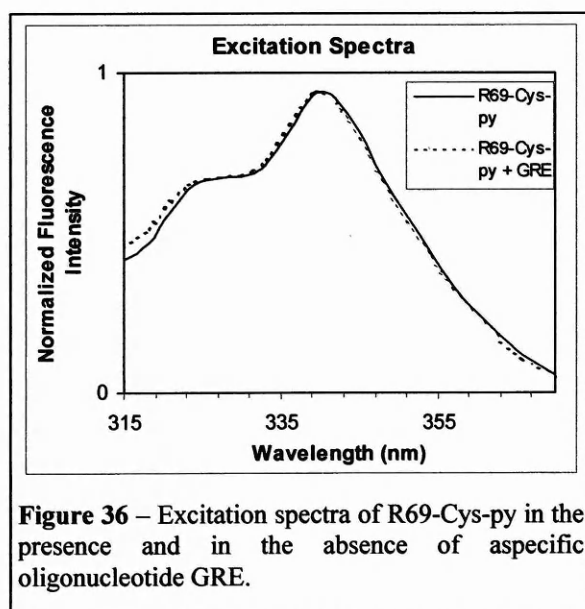
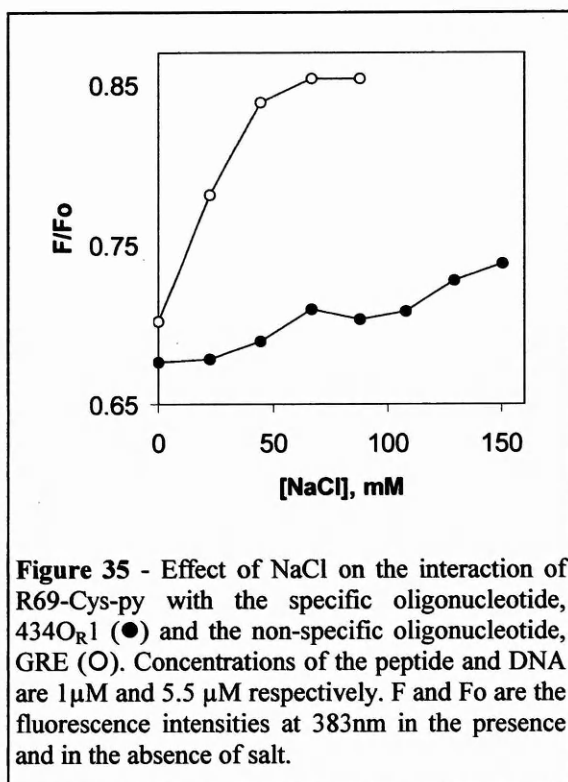


The quantity of the dimeric species increases in a saturating manner with the concomitant decrease of the initial fluorescence at 383 nm (**Figure 34**). Thus, the result shows that the pyrene moieties at the C-termini of the two R69 molecules are in a proximity sufficient for stacking interaction. Such a proximity of the C-termini is possible if the individual subunits occupy the half sites of the operator sequence in a configuration similar to what is observed in the crystal structure of their complex

(Aggarwal et al., 1988). We term this kind of dimerization as "productive assembly". This result also suggests that R69 is capable of recognizing the operator half sites in the absence of the C-terminal domain of the 434 repressor, and that dimerization occurs in complex with operator.

14.3 Monomer fluorescence is quenched upon addition of non-cognate DNA or other polyanions.

Similar titrations were carried out with the non-specific oligonucleotide GRE (Figure 33B). The fluorescence of pyrene below 450 nm was found to decrease in the same manner as observed in the case of specific DNA (Figure 34B). However, no significant excimer fluorescence was observed in the region of 450-600 nm (Figure 34A), suggesting that the fluorescent labels are not in each other's vicinity. The F/F_0 at 383 nm decreases with increasing concentrations of the non-specific DNA and reaches saturation only at slightly higher concentrations than in the case of specific DNA



(**Figure 34B**). The absence of excimer fluorescence above 450 nm shows that non-operator DNA does not induce productive dimer assembly as observed in the case of specific operator sequences. On the other hand, the decrease in the monomeric pyrene fluorescence (below 450nm, **Figure 34B**) shows that interaction with DNA does occur.

As the quantum yield of pyrene monomers is known to decrease in polar environments such as DNA (Preuss et al., 1997; Wittung et al., 1994), we tested the effect of a number of polyanions on the spectrum of R69-Cys-py (**Figure 34B**). It appears that tRNA, single stranded DNA, polyacrylic acid, and to a lesser extent, heparin all produce a concentration dependent decrease in the monomeric pyrene spectrum quite similar to the one produced by non-cognate ds DNA, so we conclude that this decrease of monomeric pyrene fluorescence is at least partly due to electrostatic interactions, i.e. the presence of negative charges in the vicinity of the fluorescent group might be responsible for quenching the pyrene fluorescence.

Figure 35 shows that addition of NaCl exhibits an effect on the fluorescence at 383 nm that is more pronounced in the case of the R69-Cys-py/non-operator DNA complex than in the case of the R69-Cys-py/O_R1 DNA complex. Interactions of proteins with the DNA phosphate backbone are sensitive to salt concentration and thus they are thought to be primarily electrostatic in nature (Boschelli, 1982; Matthew and Ohlendorf, 1985; Ohlendorf and Matthew, 1985; Takeda et al., 1986). The salt sensitivity of the decrease in fluorescence at 383 nm may be thus due to processes mediated by ionic interactions between DNA and the protein. In this respect, in **Figure 36** is reported the excitation spectrum of R69-Cys-py in the absence and in the presence of the non-operator DNA. The absorption spectrum, after GRE addition, does not seem broadened, relative to the spectrum of R69-Cys-py alone. This is a further indication that the lack of excimer fluorescence upon non-specific DNA interaction is probably due to a real lack of pyrene proximity and not to the impossibility of the pyrenes to reorient on excitation

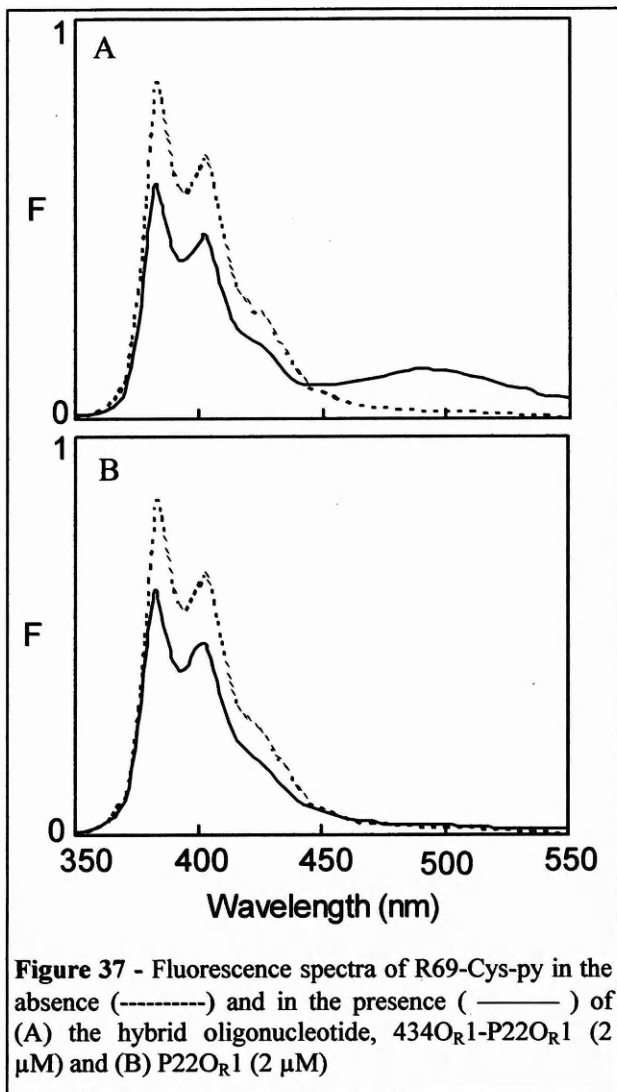


Figure 37 - Fluorescence spectra of R69-Cys-py in the absence (-----) and in the presence (——) of (A) the hybrid oligonucleotide, 434OR1-P22OR1 (2 μM) and (B) P22OR1 (2 μM)

14.4 Interaction with a hybrid operator induces significant excimer fluorescence of R69-Cys-Py.

We have also investigated the assembly of R69-Cys-py on a hybrid operator DNA (434OR1-P22OR1) composed of one 434 half site and one consensus P22 half site. It is noted that the P22 subsite TTAA shares an AA sequence with the 434 subsite ACAA (Figure 28C) so it can be expected that R69Cys-py will, to some extent, bind to the P22 subsite. Figure 37A in fact shows that R69-Cys-py exhibits significant excimer fluorescence in the presence of the hybrid operator,

434OR1-P22OR1. The F/F₀ value (at 495 nm) increases to a maximum value of 5.2 (at 2 μM oligonucleotide), which is significantly smaller than the value of >12 (Figure 34A) obtained in the presence of the 434OR1 oligonucleotide. This result suggests that the 434 operator half-site promotes the formation of the productive dimer since its characteristic excimer fluorescence is readily detectable when a properly spaced, partly similar half-site is present.

No detectable excimer fluorescence is seen in the presence of the oligonucleotide P22OR1 which is composed of two P22-operator subsites (Figure 37B); in this case the situation is similar to that observed with the non-operator DNA, GRE (Figure 34A). In

other words, the assembly of 434 DNA-binding domain R69 on the P22 operator site is not detectable under the experimental conditions used here.

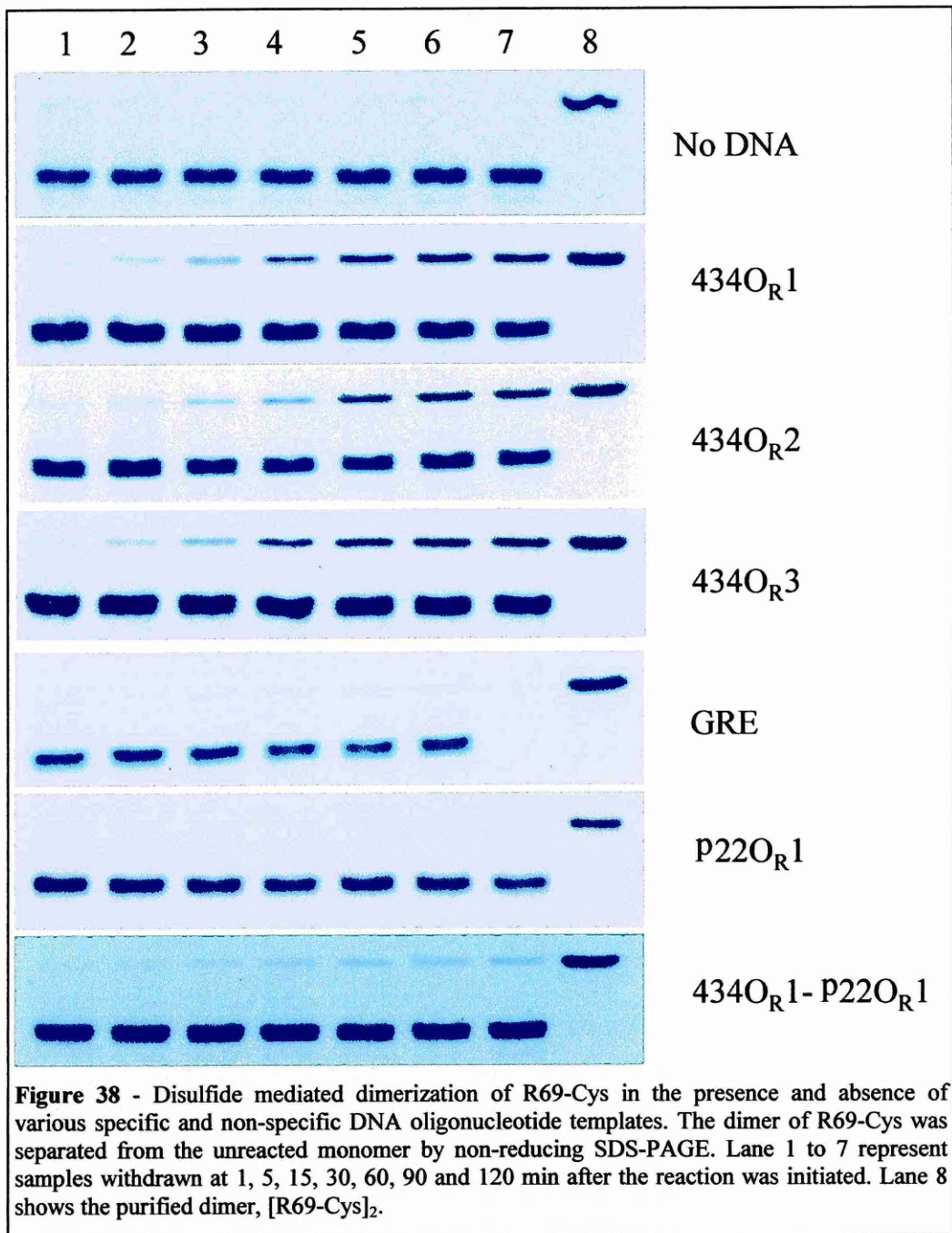
15. Intermolecular Disulfide Crosslinking

15.1 Cognate and hybrid DNA promote R69cys oxidative dimerization

The C-terminal cysteine residue of the R69-Cys molecule offers an alternative route to probe subunit assembly. We hypothesized, that if two R69-Cys molecules bind to operator DNA in the same way as seen in the crystallographic structure of the R69/DNA complexes and confirmed by our fluorescence experiments, the two C-terminal thiol groups get near each other so that oxidation may be greatly facilitated. In other terms, cognate DNA will serve as a template for dimerization.

The oxidative dimerization of R69-Cys molecules was carried out incubating in 20 mM sodium phosphate buffer (pH 7.2) the reduced R69Cys monomers in the absence and in the presence of various specific and non-specific oligonucleotides. The oxidative process was initiated by adding DL-cystine hydrochloride. Aliquots of the reaction mixtures were withdrawn at different intervals, analysed by non-reducing SDS-PAGE (**Figure 38**) and the dimer bands measured by densitometric analysis (**Figure 39**). Under our experimental conditions, evident $[R69-Cys]_2$ formation could be detected only in the presence of specific DNA, composed of two cognate half sites. It is observed, that oligonucleotides corresponding to various natural operator sites (434O_{R1}, 434O_{R2} and 434O_{R3}), promoted subunit assembly to a similar extent, while non-specific oligonucleotides GRE and P22O_{R1} did not show any relevant effect as compared to the DNA-free reaction.

It is significant to note that the oligonucleotide P22O_{R1}, which corresponds to the P22 repressor cognate sequence, did not promote the dimerization of R69-Cys, while dimer formation could be detected in the presence of the hybrid operator 434O_{R1}-P22 O_{R1} (**Figure 38 and Figure 39**). This result confirms the conclusion of the

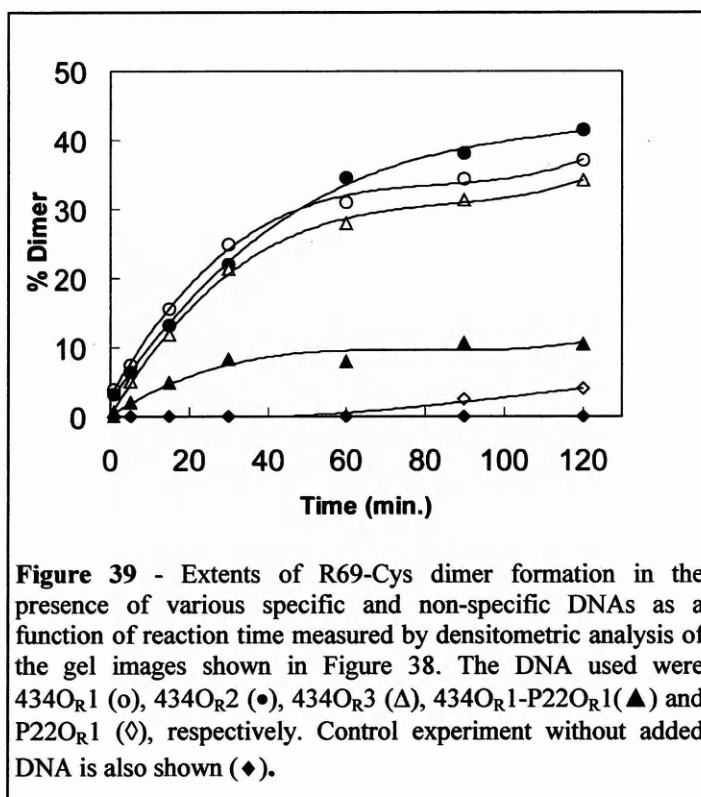


fluorescence studies that one 434 operator subsite might be sufficient to promote a measurable subunit assembly when an adjacent, partly similar P22 subsite is present.

For modelling the assembly of different molecules recruited by adjacent subsites, we used the engineered subunit, R*69-Cys (sequence in **Figure 28A**), in which the DNA-contacting residues of the recognition helix were replaced by the corresponding residues of the P22 repressor.

In solution and in the absence of DNA, both R69-Cys and its engineered mutant R*69-Cys can be oxidized to the respective homodimers ($[R69-Cys]_2$ and $[R^*69-Cys]_2$).

When a 1:1 mixture of R69-Cys and R*69-Cys are oxidized in the absence of DNA, the result is a mixture of homo and heterodimers that can be distinguished by SDS/PAGE (**Figure 40B**) and especially mass spectrometry. If specific DNA acts as a template for disulfide mediated crosslinking, then the presence of specific DNA in the reaction mixture will influence the distribution of the products.



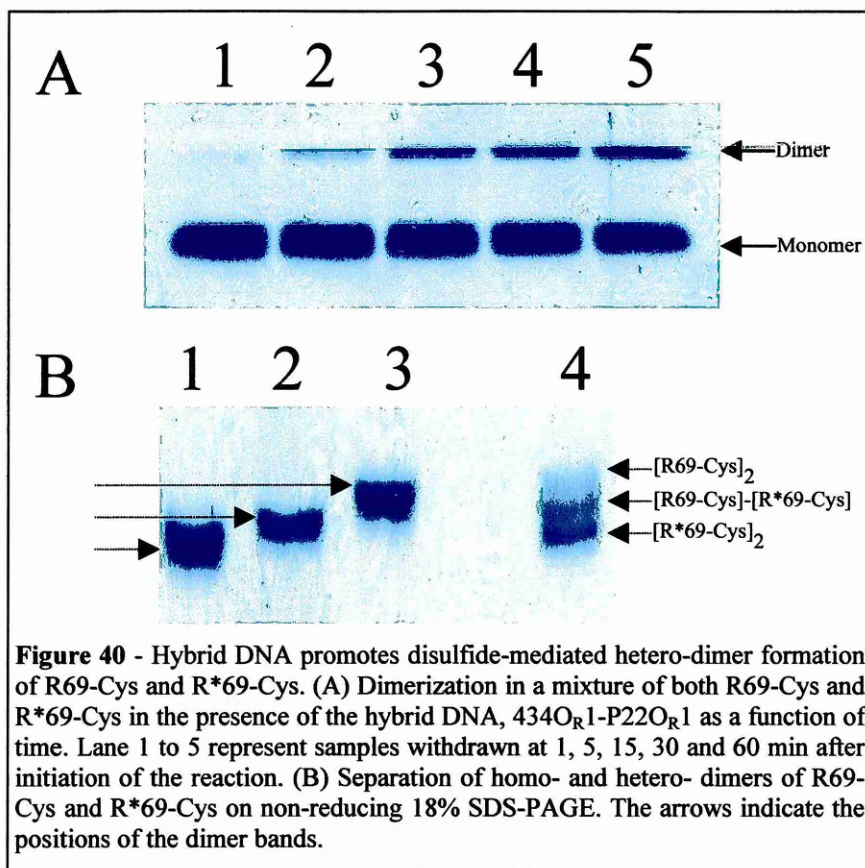


Figure 40 - Hybrid DNA promotes disulfide-mediated hetero-dimer formation of R69-Cys and R*69-Cys. (A) Dimerization in a mixture of both R69-Cys and R*69-Cys in the presence of the hybrid DNA, 434O_R1-P22O_R1 as a function of time. Lane 1 to 5 represent samples withdrawn at 1, 5, 15, 30 and 60 min after initiation of the reaction. (B) Separation of homo- and hetero- dimers of R69-Cys and R*69-Cys on non-reducing 18% SDS-PAGE. The arrows indicate the positions of the dimer bands.

When the 1:1 mixture of R69-Cys and R*69-Cys was mixed with the hybrid 434O_R1-P22O_R1 operator, a single new product of intermediate electrophoretic mobility appeared (**Figure 40A**).

Mass spectrometric analysis (**Figure 41** and **Figure 42**) confirmed that the new product corresponded to the heterodimeric $[R69-Cys]-[R^*69-Cys]$. Based on the sensitivity of this analytical technique we can estimate that the quantity of homodimers, if formed, must have been far below 1% of the heterodimer. These results, thus, show that sequences as short as the operator half sites are sufficient to specifically recruit protein molecules.

We mention that according to the results obtained by excimer fluorescence (**Figure 37**) R69-Cys itself is capable of binding to the P22 subsite of the hybrid operator to a certain extent as a result of such recruitment. The fact that no R69

homodimers could be observed either by SDS-PAGE or by mass spectroscopy shows that R*69 successfully competes for the binding when applied in a 1:1 ratio.

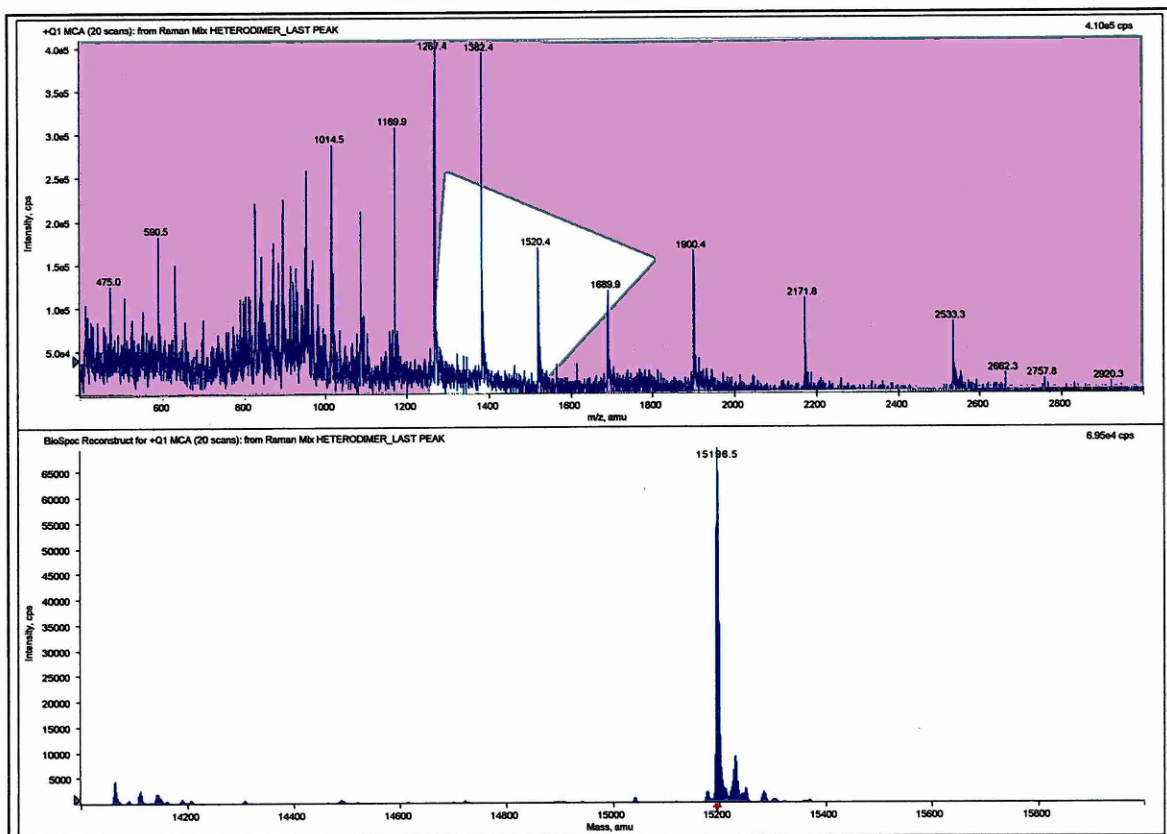


Figure 41 – ESI-MS analysis of the dimerized product of the reaction mixture R69-Cys, R*69Cys and 434O_R1-P22O_R1 operator. The measured molecular weight (15196.5) is in good agreement with the calculated average molecular weight of (R69-Cys)-(R*69Cys) disulfide heterodimer (15194.0).

Dimers	Expected mass ¹ (Da)	Observed mass ² (Da)
Homo-dimer of R69-Cys	15307.4	15307.8
Homo-dimer of R*69-Cys	15081.2	15081.4
Hetero-dimer*	15194.0	15196.2

Figure 42 – Summary of the mass spectrometric analysis of homo- and hetero- dimers of R69-Cys and R*69-Cys.

¹ Mass expected from the amino acid sequence

² Mass obtained by electron spray mass spectrometry.

* Dimer formed in the presence of the hybrid (434O_R1-P22O_R1)

16. Discussion

In this work, assembly of the R69-Cys molecule on DNA was investigated by two different approaches: i) by attaching a fluorescent label, pyrene to the C-terminus of R69-Cys and studying the changes of the fluorescence spectrum upon DNA addition, and ii) by oxidative crosslinking via the C-terminal cysteine residue.

Computer modelling done on the base of the 3-D structure of the 434 repressor DNA-binding domain (R69)/operator DNA complex and on the base of the NMR structure of R69 in solution (Neri et al., 1992) (Shimon and Harrison, 1993) (Aggarwal et al., 1988) shows that the C-termini of the two protein monomers are in a close proximity upon specific DNA interaction (Figure 27). As a further proof we demonstrated by EMSA that, apart exhibiting wild type specificity, the disulfide dimers show binding constants in perfect agreement with the published data on natural and single chain repressors, molecules in general not subjected to structural restraints (Figure 29 and Figure 30).

The proximity of the C-termini is in principle sufficient a) for C-terminally attached pyrene residues to produce a stacking interaction and excimer fluorescence, and, b) for the C-terminally attached cysteines to form an intermolecular disulfide bridge that leads to covalent dimerization. We hypothesized, that such a conformation - that we term a productive assembly - would be promoted by operator DNA. According to the methodology used in this work, R69-Cys in solution does not show appreciable levels of dimerization in the absence of DNA, a fact supported both by the lack of excimer fluorescence and by the disulfide crosslinking experiments (Figure 33, Figure 38 and Figure 39). On the other hand, operator DNA produces a productive assembly, while non-operator DNA does not do so. Hybrid operators, in turn promote the assembly of heterodimers (Figure 40).

Non-operator DNA did not influence the oxidative crosslinking results, however it gave rise to a decrease of the pyrene fluorescence at 383 nm, which we term, for brevity, as 383 nm signal. Interestingly, also operator DNA produced this signal, in

parallel with the characteristic excimer fluorescence at 495 nm (Figure 33 and Figure 34). Consequently, the signal at 383nm must be due to an interaction that occurs both with operator and with non-operator DNA. In contrast to the excimer fluorescence signal of operator DNA complexes, the 383 nm signal is salt sensitive suggesting that the 383 nm signal may be the result of electrostatic interactions (Figure 35). The salt sensitivity of the 383 nm signal was more pronounced in the case of non-operator DNA. This leads us to the assumption that the phenomenon underlying the 383 nm signal occurs in operator as well as in non-operator complexes. These observations can be interpreted in several ways. First, DNA may promote the formation of non-productive dimers of R69 by attracting the molecules to its surface via electrostatic forces. Alternative dimer conformations induced by interaction with specific versus non-specific DNA were suggested for the intact 434 repressor (Ciubotaru et al., 1999). Second, electrostatic interactions with the sugar-phosphate backbone may produce a decrease in pyrene fluorescence, which should occur both with operator and with non-operator DNA. Third, a conformational change in the R69 protein may be the reason of the 383 nm signal. Both operator and non-operator DNA is known to produce such conformational changes for the covalent dimers of R69 (Percipalle et al., 1995) as well as for the intact 434 repressor (Ciubotaru et al., 1999). The fact that the excitation spectrum of R69-Cys-py in the presence of the non-operator DNA does not seem broadened with respect to the spectrum of R69-Cys-py alone, suggests that the R69-Cys-py molecule could interact with DNA either as a monomer or in a dimeric conformation lacking of pyrene proximity (Figure 36).

The results presented here indicate that the sequential binding of monomers, previously suggested for ATF-2 and Max (Kohler et al., 1999) as well as for the P22 Arc repressor (Rentzeperis et al., 1999), is a plausible pathway also for the N-terminal domain of the 434 repressor, R69. This peptide shows no appreciable dimerization in the absence of DNA, and its binding to individual operator half sites is not detectable.

However, two adjacent half sites can guide the assembly of specific dimers as was demonstrated by the formation of a 434-P22 heterodimer. Interactions with non-operator DNA, which are believed to be mostly ionic in nature, do not lead to such properly oriented, dimerization-competent assembly of the subunits. On the other hand, such non-specific interactions may be expected to guide the assembly via one-dimensional diffusion along the DNA-chain (Berger et al., 1998). One-dimensional diffusion, recognition coupled specific assembly and subsequent stabilization might thus form hierarchical steps in this process.(Percipalle et al., 1995) The assertion that conformational changes of either the monomeric or dimeric DNA binding domains take place during these steps is consistent with the studies performed with the intact 434 repressor (Ciubotaru et al., 1999).

Section II: N^ω-methylation of arginine residues modulates the non-specific interaction of the RGG box with nucleic acids

17. Synthesis and characterization of the peptides DMA-GG, RGG, KGG, RGG2

For our experimental studies we have synthesized a peptide Ac-GRGGFGGRGGFRGGGRGG-NH₂, corresponding to residues 676-692 of human nucleolin (Table 3), both in dimethylated (4 aDMA residues, DMA-GG peptide) and in unmethylated (4 Arg residues; RGG peptide) forms. This sequence contains four RGG repeats, a number comparable to the native RGG motifs of several proteins (see Table 2). In order to better understand the role of the different aminoacids in the motif context, we synthesized other two analogues presenting some aminoacid substitutions: KGG (all Arg substituted by Lys), RGG2 (all Phe substituted by Gly).

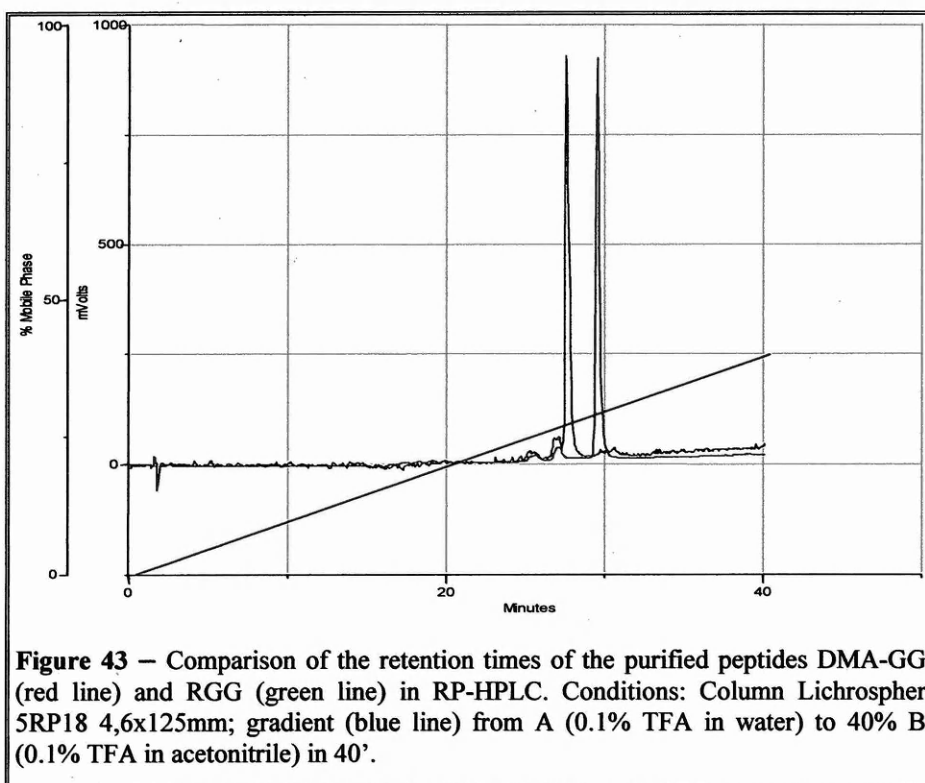
The KGG analogue was constructed to investigate whether the role that arginine plays in interaction with nucleic acids is essentially due to its positive charge or whether other components, such as its unique hydrogen bonding capability are involved.

The RGG2 analogue was created to check the role of phenylalanines in the interaction due to their potential ability to stack with the nucleobases.

The peptides were synthesized on solid phase using Fmoc-chemistry by manual synthesis (DMA-GG) or in an automated continuous flow system. The use of Tentagel-Sieber amide resin, a hyper acid labile solid support suited for the synthesis of peptide amides, characterized by a high degree of solvation and an average level of substitution (0.15 mmoles/g), allowed us to increase the coupling efficiency and to overcome the strong tendency of these repetitive sequences to undergo a conformational collapse on the resin and to be poorly solvated. We experienced that the length of these peptides represents more or less the limit for classical Fmoc stepwise synthesis, because this tendency increases strongly with the growing of the peptide chain reducing dramatically the acylation rate.

Currently there is no commercially available side-chain protected derivative of aDMA and the synthesis of long peptide sequences, especially in the presence of small, activated species such as Gly, requires the protection of the arginine guanidinium side-chain. For this purpose our group has recently developed an efficient method for the synthesis of N^ω-substituted arginines (Szekely et al., 1999) suitable for Fmoc peptide chemistry, which allowed the preparation of the building block Fmoc-aDMA(Mts)-OH used in the DMA-GG synthesis (see Materials and Methods).

The four cleaved and deprotected peptides were all very soluble and hydrophilic and were purified by RP-HPLC, eluting the column with a linear gradient from A (0.1% TFA in water) to 40% B (0.1% TFA in acetonitrile); see **Figure 44** for the RPHPLC analyses. The expected more hydrophobic character of the dimethylarginine residue is reflected in the increased retention time in RPHPLC of DMA-GG as compared to RGG (see **Figure 43**)



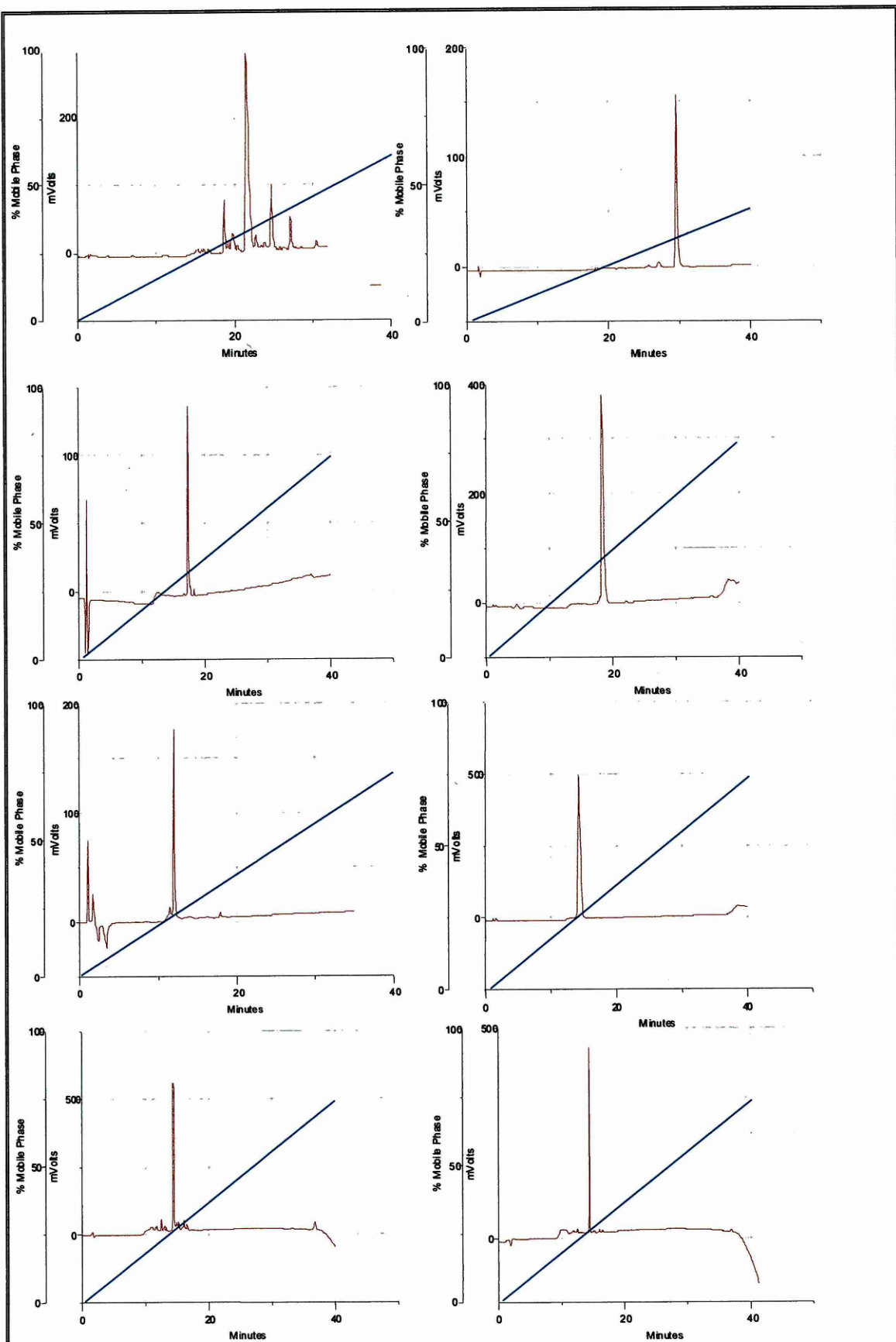


Figure 44 - RPHPLC analysis of the crude (left column) and purified (right column) peptides DMA-GG, RGG, RGG2, KGG (in this order). Conditions: Column Lichrospher 5RP18 4,6x125mm; gradients (blue lines) from A (0.1% TFA in water) to B (0.1% TFA in acetonitrile) in 40'.

After purification, stock solutions of all the peptides were characterized by amino acid analysis on a custom made system developed on the base of the Waters Pico-Tag System (see Materials and Methods). This step, apart from confirming the amino-acid composition, was essential for the determination of the peptide concentration used in the nucleic-acid interaction studies. Further analysis performed by ESI-MS confirmed the expected mass theoretical values and the absence of RPHPLC co-eluting deletion peptides.

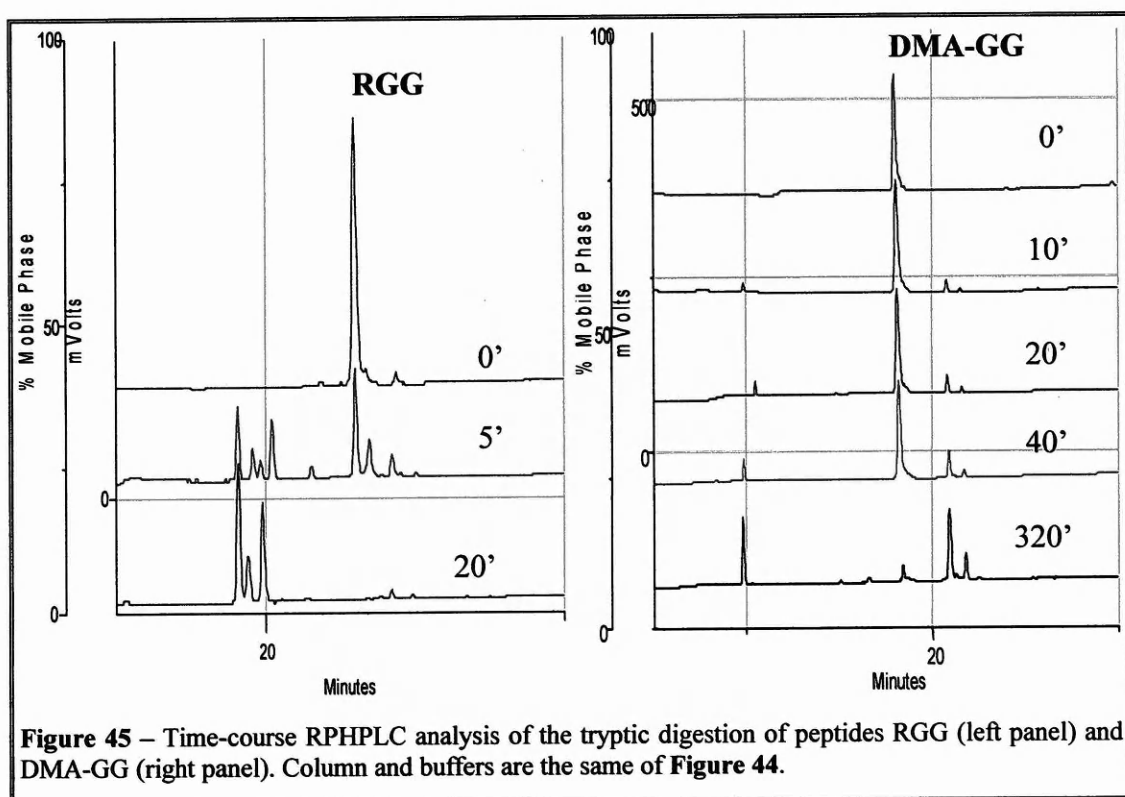
Analytical data concerning all peptides are shown in **Table 3**

17.1 Proteolytic cleavage resistance of RGG and DMA-GG peptides.

RPHPLC analysis of a tryptic digestion mixture of the RGG and DMA-GG peptides suggests that the two peptides have different resistance to proteolytic cleavage.

The left panel of **Figure 45** shows the time-course of a tryptic digestion (1.5 nmoles of trypsin) of 75 nmoles of RGG in 1500ul of 20mM ammonium pH 8 buffer solution at 37°C.

The analysis shows that already after 20' from the beginning of the reaction the

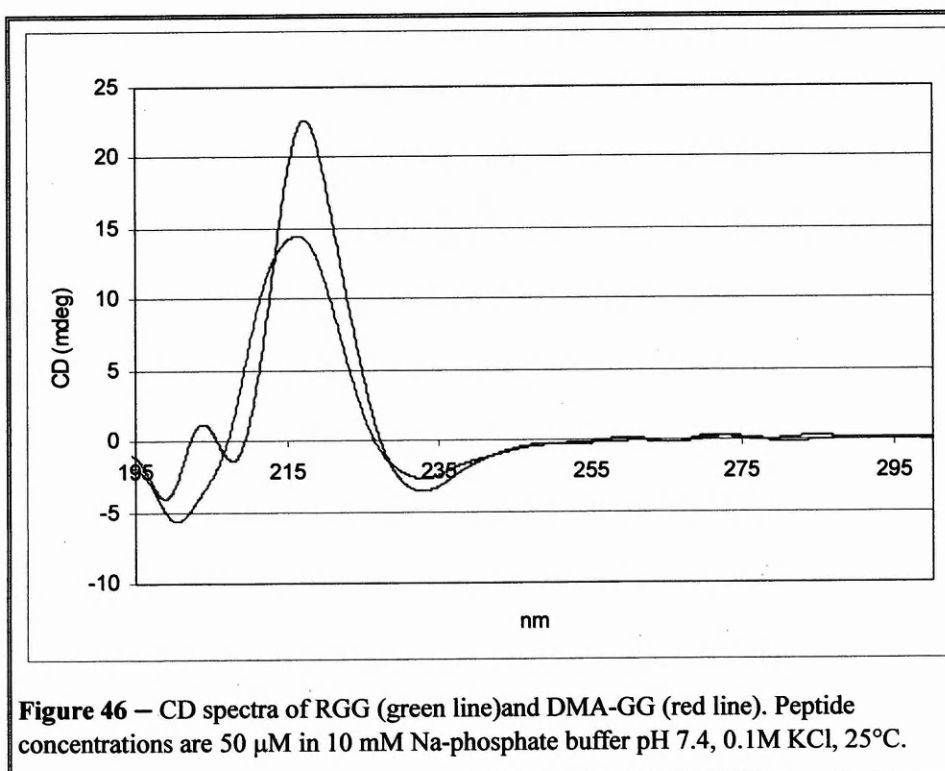


peak corresponding to the RGG peptide has practically disappeared; on the contrary the analysis of the DMA-GG tryptic cleavage (right panel, same conditions and concentrations of peptide and trypsin) shows small amounts of the starting product even after 320' from the beginning of the reaction.

This result is consistent with a work published by Olson and co-workers (Olson et al., 1990), in which, during the proteolysis of a natural preparation of rat nucleolin, a 48-Kda fragment containing the central and carboxy-terminal domains was preferentially cleaved near the boundaries of the central CS-RBDs, while the C-terminal part was found resistant to the cleavage by trypsin even at the highest enzyme concentrations used.

17.2 Circular dichroic spectra of RGG and DMA-GG peptides are indicative of β -turn structure.

The circular dichroic spectra of the two peptides RGG and DMA-GG (50 μ M in 10 mM Na-phosphate buffer pH 7.4) show a maximum around 215 nm, a minimum around 200nm and a second weaker negative band around 230nm (**Figure 46**). A



similar type of CD spectrum resembles the spectra characteristic of some β turn structures observed in linear peptides and described in literature as type D, a red-shifted variant of the so-called B-spectrum characteristic of type II β turns (Fasman, 1996) (see also Figure in introduction); indeed the Gly-Gly sequence has already been described to form β turns in the published crystal structure of the peptide YGGFL (Smith and Griffin, 1978) and a 100 aa long recombinant fragment comprising the 50 aa long Gly-Arg rich region of hamster nucleolin has also shown CD and FTIR spectra indicative of this conformation (Ghisolfi et al., 1992a). The repetition of sequential β turns in long polypeptides could also lead to formation of supersecondary structures such as β spirals (Ghisolfi et al., 1992a).

18. Synthesis and purification of the polypeptides Nuc61Arg and Nuc61Dma

Nuc61Arg and Nuc61DMA correspond to the 61 aa long fragment (residues 646-706) of the C-terminal domain of human nucleolin containing unmethylated and asymmetrically dimethylated arginine residues respectively (Table 3). The sequence contains ten arginines participating in nine RGG repeats and six phenylalanine residues, four of which follow immediately after an RGG sequence.

Thanks to the positive experience with the RGG and DMA-GG peptides and to the already mentioned difficulties both in terms of yield and homogeneity of the methylation reactions *in vitro*, we decided to synthesize chemically the polypeptide Nuc61Dma. The same chemical method was decided for Nuc61Arg after that, in the first stages of our work, only a poor level of expression of the nucleolin C-terminal fragment could be obtained by using an *E. coli* expression system. It must be noted that the Nuc61Arg synthesis, apart representing an ideal test for the development of all the strategies that lead to the Nuc61Dma synthesis, allows a direct comparison of the behaviour of the two polypeptides thanks to the common pathways of synthesis and purification, as happens for the shorter peptides.

The establishment of a reliable synthesis protocol for both peptides took around 2 years of different attempts and the development of novel synthesis strategies. In the note¹ follows a short resume of the various stages, which were fundamental for the final result.

Note ¹

1st trial. Automated stepwise solid phase peptide synthesis of Nuc61Arg.

The synthesis was performed with a protocol similar to that described for the RGG peptide, but the result of the synthesis was negative with crude peptide yield below 2% and an elevated number of incorrect peptides due to side-reactions, which rendered the purification impractical.

The possible problems were the solubility of peptide resin (which lead to conformational collapse), acylation of the side chain of Arg and/or auto-capping (termination of the peptide chain) due to the coupling reagent, a known side-reaction in peptides with elevated number of Gly residues.

To resolve the last problem in the subsequent synthesis we decided to use 10-20% excess of Gly respect to the coupling reagent and to use less powerful activating reagents for coupling (TBTU or PyBOP instead of HATU).

2nd trial. Convergent manual solid phase peptide synthesis of Nuc61Dma

18.1 Stepwise synthesis of Nuc61Arg and Nuc61Dma

The large amount of data accumulated with the failed syntheses (see Note 1) pointed to the fact that a backbone amide protection (provided by the 2-hydroxy-4-methoxy benzyl (Hmb) group) was necessary to overcome the peptide chain solubility problems but other drawbacks of this synthesis system (i. e. the acylation of the Hmb hydroxy group and the consequent slow O to N acyl transfer) had to be avoided. To overcome this problem we used the 2,4-dimethoxybenzyl (Dmb) group as backbone protection instead of the Hmb group. Another positive remark was the use of a dipeptide as a building block to increase the coupling rate; on the base of these indications was constructed the building block Fmoc-Gly-(Dmb)-Gly-OH (see Materials and Methods) used successfully in the synthesis of both peptides. Nuc61Arg and Nuc61Dma syntheses (0.1 mM scale) were performed each on 660 mg of Fmoc-Gly-TG Sieber Amide resin with substitution 0.15 mmoles/g. The peptide sequences and the positions of the Fmoc-Gly-(Dmb)-Gly-OH are shown in **Table 3**; the solvent employed was N-methylpyrrolidone, the amino acid excess 4X and a systematic double coupling was used for all the residues. Nuc61Arg was synthesized first and in four different points the

The convergent strategy implies the division of the full sequence in portions that are synthesized as N-terminal and side-chain protected peptide fragments; then the segments are sequentially activated and coupled on the first segment synthesized and still attached on the solid phase. For the synthesis of the fragments we used an hyper acid labile resin (Fmoc-Gly-TGT) that in virtue of its cleavable trityl linker allows the release of the peptide fragments from the solid phase under very mild acid conditions, leaving the side-chain protecting groups unaffected. The peptide segments (see underlined fragments from N1 to N6 in Table 3) were constructed using the amino acid Fmoc-Dma(Tos)-OH and the C-terminal amino acid was always Gly in order to avoid racemization problems and to limit the sterical hindrance at the condensation step. The protected blocks were all synthesized in good yield and RPHPLC purified (>95% purity) and condensation carried on using 1.5 to 2 excess of the activated (by TBTU or PyBOP) fragment with respect to the solid phase component in DMF, NMP, DMSO or mixtures of these solvents under increased temperature and sonication. The result of the condensation was anyway negative due to general problems of solubility of the fragments and the synthesis was stopped.

3rd - 4th trial. Convergent manual solid phase peptide synthesis of Nuc61Dma by backbone amide protection strategy

The strategy was very similar to the previous synthesis, but the Fmoc-Gly-OH had been substituted by Fmoc-(Hmb)-Gly-OH. The Hmb group (2-hydroxy-4-methoxy benzyl), employed every 4 or 5 residues of the sequence, is described as a useful backbone amide protecting agent to prevent secondary structure formation of the growing peptide chain and to improve the solubility in organic solvents (Johnson et al., 1995). The peptide fragments (non-underlined sequences N1 to N6 in Table 3) span the same residues as in the previous synthesis apart the dimethylarginine side chain protecting group (Fmoc-Dma(Mts)-OH). The use of the Hmb group indeed resolved our problems of peptide chain solubility but its usefulness was limited by two major drawbacks. As described also in literature (Sampson et al., 1999), the coupling time of the amino acid subsequent to the Hmb group has to be enormously prolonged: we overcame this problem employing directly the dipeptide Fmoc-Gly-(Hmb)Gly-OH as building block, a strategy which has also the advantage of increasing the peptide chain of two aa per coupling; the second major drawback is the fact that the activated fragment or amino acid first enters in the hydroxy group of Hmb and is then very slowly transferred to the amino terminal group and this mechanism originates a very complicate picture of side-reactions especially when the acylating group is a large protected peptide fragment carrying other Hmb groups by its own.

synthesis was stopped so that the intermediate products could be cleaved and analysed to check the progress in the yield and purity. At the synthesis step 38, around 100 mg of resin were subtracted and cleaved to obtain peptide Nuc38Arg (see Table 1) used later in some nucleic acid binding studies.

Nuc61Arg and the intermediate products were cleaved by a combined procedure (TFA/TMSBr) to thoroughly remove the Mts and Dmb protecting groups (see Materials and Methods), and analysed by HPLC. Nuc61Arg was subjected to a first stage of purification by cation exchange chromatography on a sulfopropyl column (**Figure 47**), the last major peak contains the correct peptide and the approximate yield is around 15% of the total peak area.

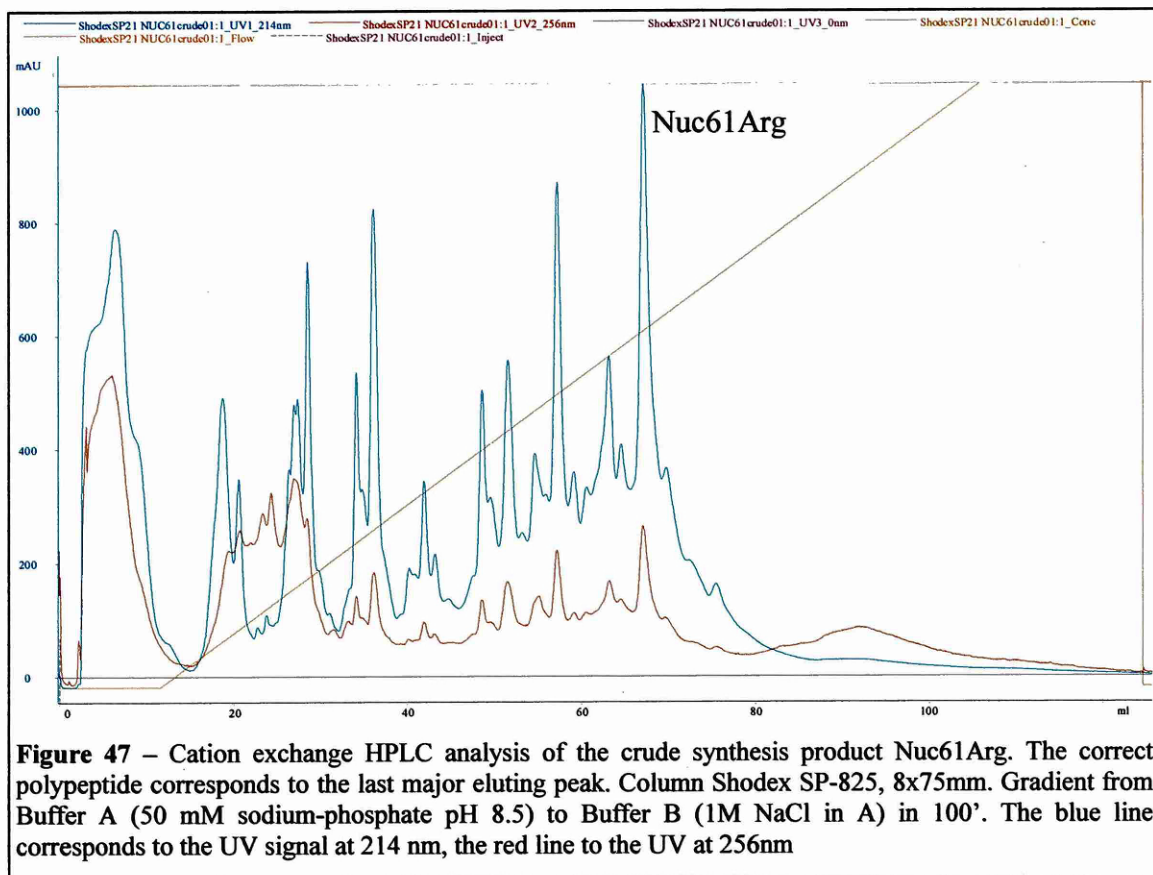


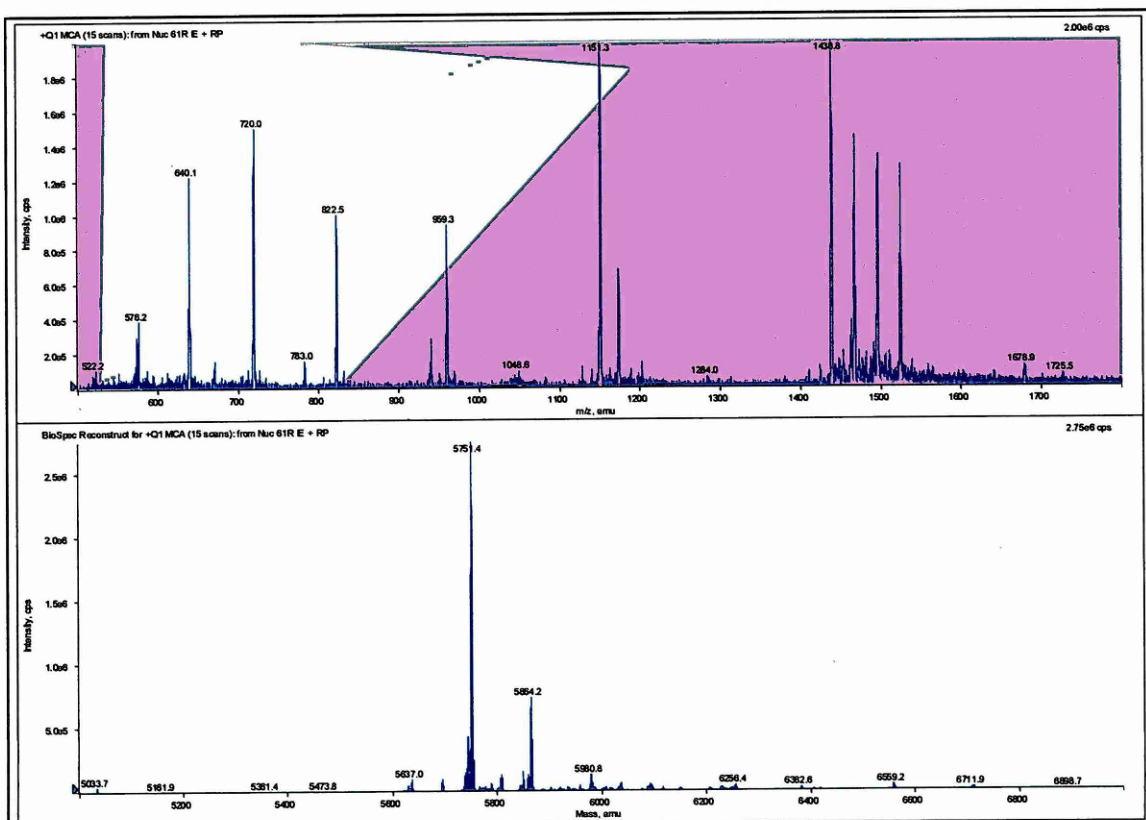
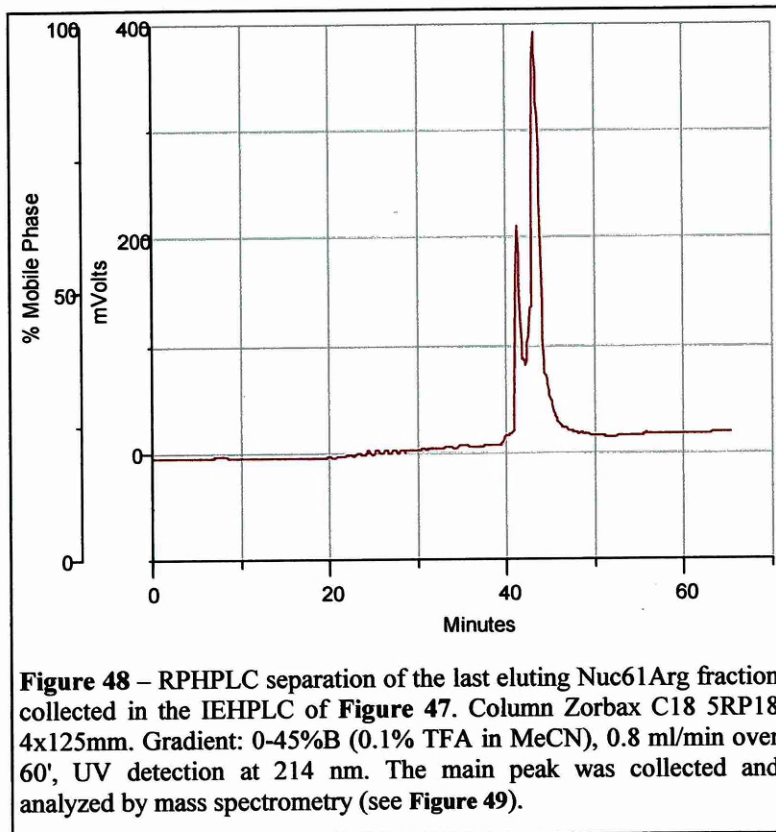
Figure 47 – Cation exchange HPLC analysis of the crude synthesis product Nuc61Arg. The correct polypeptide corresponds to the last major eluting peak. Column Shodex SP-825, 8x75mm. Gradient from Buffer A (50 mM sodium-phosphate pH 8.5) to Buffer B (1M NaCl in A) in 100'. The blue line corresponds to the UV signal at 214 nm, the red line to the UV at 256nm

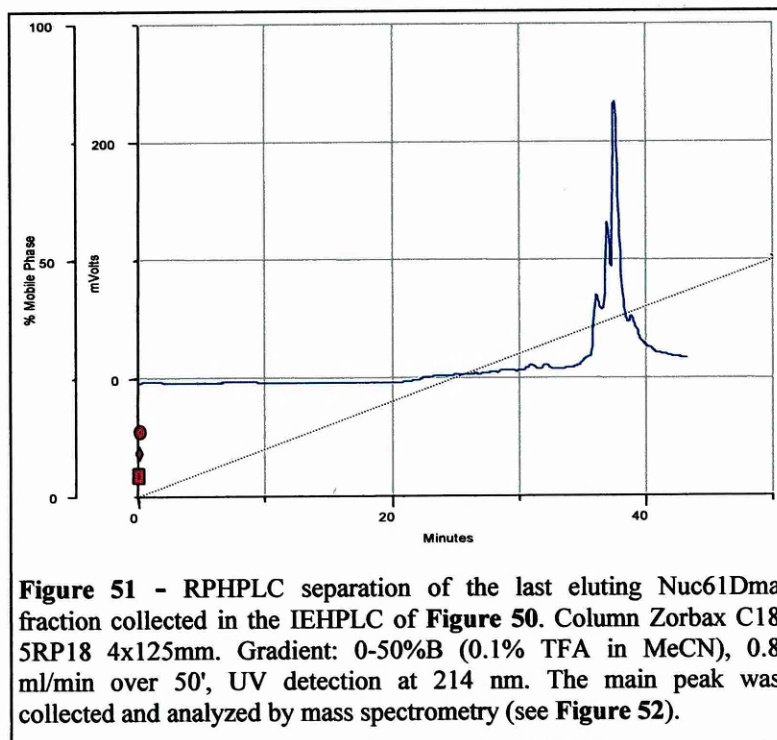
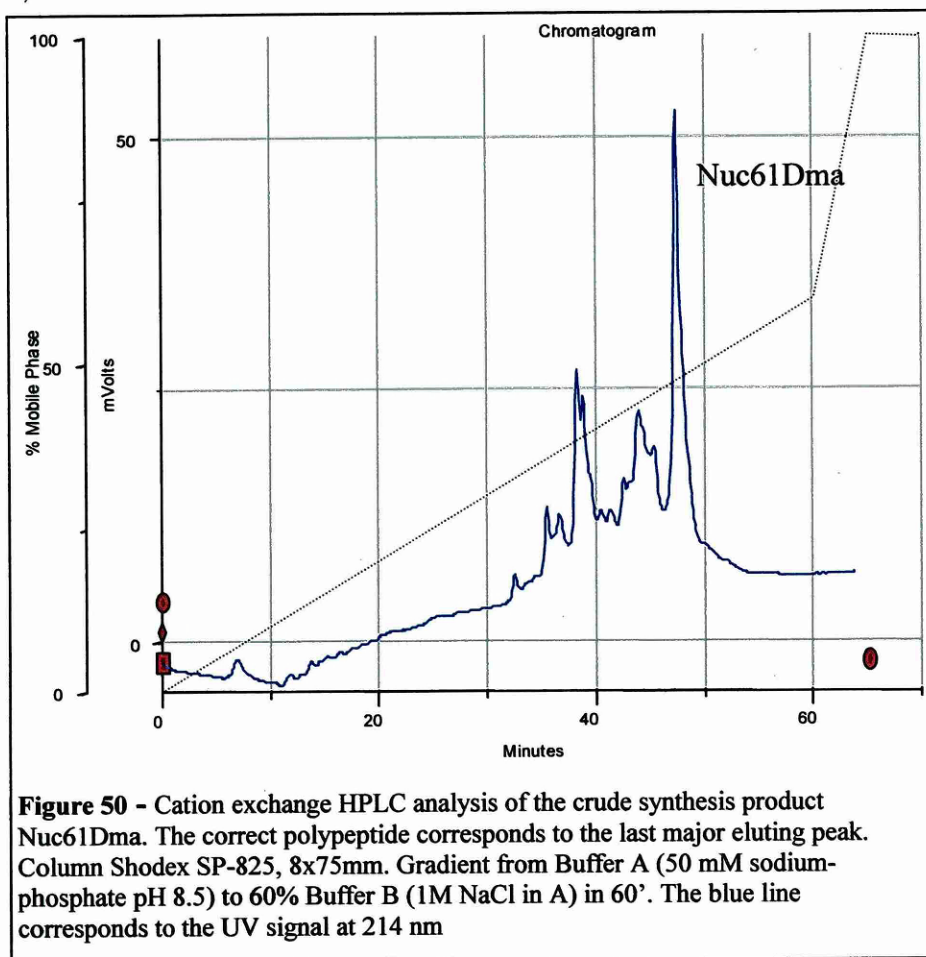
A subsequent step of RPHPLC (**Figure 48**) allowed the separation of a main product peak from closely related impurities not resolved in the ion exchange step; the

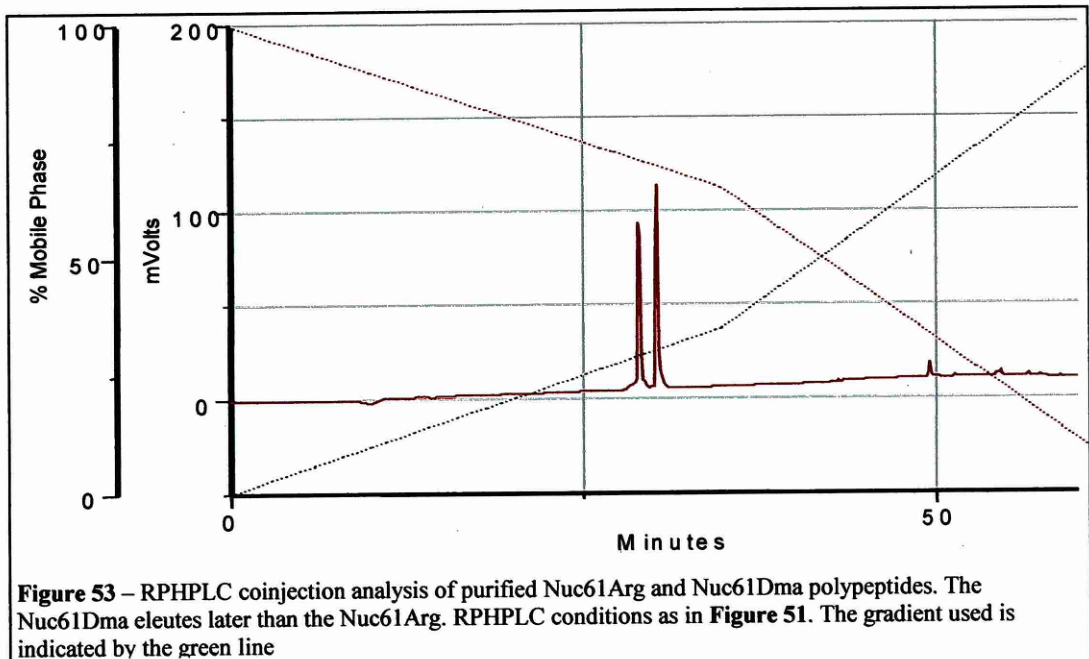
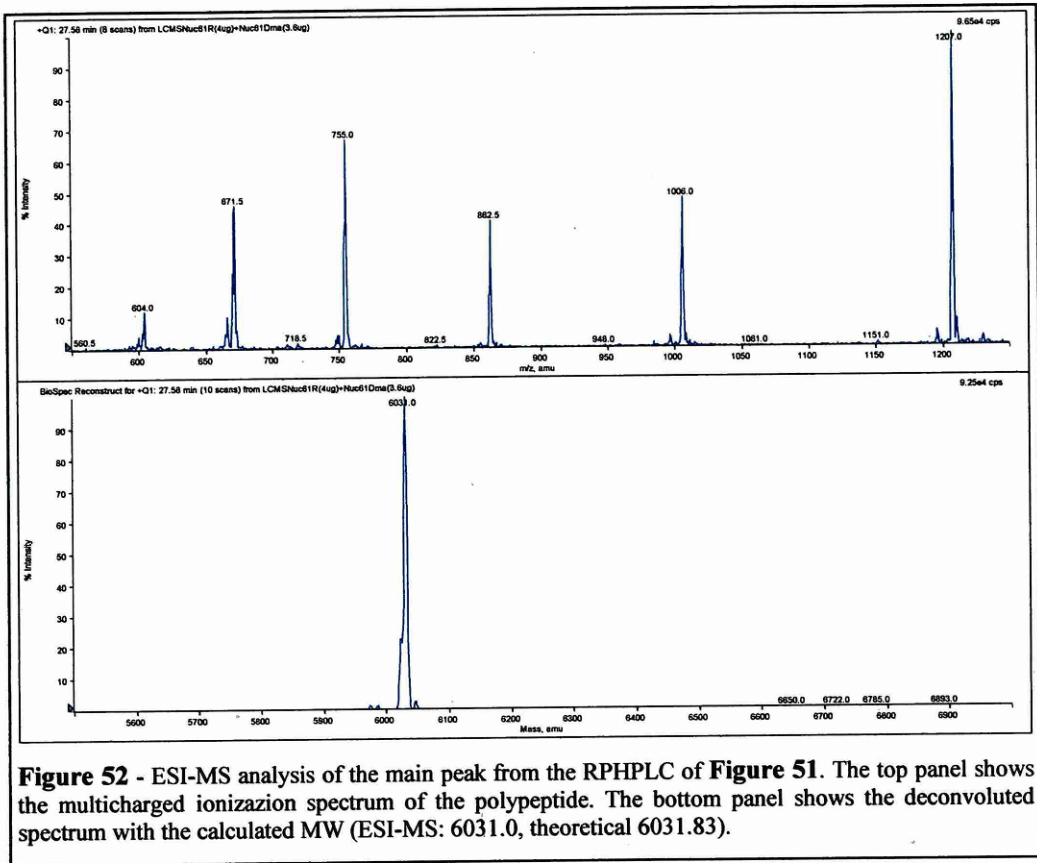
ESI-MS analysis of this main peak gave the expected theoretical mass of Nuc61Arg (**Figure 49**).

The synthesis of Nuc61DMA was performed under the same conditions as the previous synthesis, but with further improvements thanks to the identification of the most troublesome stages of the Nuc61Arg synthesis originating a number of deletion peptides (the major peaks eluting before Nuc61Arg in **Figure 47**), which could be easily identified by ESI-MS. The first step of cation exchange chromatography immediately showed a cleaner crude product characterized by a main peak representing around 35% of the total area (**Figure 50**). The ESI-MS analysis of this fraction after a second step of RPHPLC separation (**Figure 51**) confirmed the expected MW theoretical value for Nuc61DMA (**Figure 52**).

The Nuc61Arg and Nuc61DMA peptides have been involved in few nucleic acid interaction studies, due to the fact that their purification and characterization after the positive results in the respective syntheses were achieved a few weeks before the submission of this thesis.







19. Nucleic acid interaction studies

19.1 Double filter binding assays

We carried out double filter binding assays to determine the affinity of the RGG peptide and of its analogues towards a number of nucleic acid substrates.

The double filter binding method is a variation of the classical nitrocellulose filter binding: in this experimental design, a cationic (DEAE) membrane placed directly below the first nitrocellulose sheet in a 96 well dot-blot apparatus, collects the labelled nucleic acid probe not retained in complex with the peptide on the first sheet. Both membranes are then exposed in a phosphoimager and the densities of the spots are measured. The additional data provided by the DEAE membrane enable a precise normalization of the amount of labelled probe adsorbed on the nitrocellulose with respect to the total amount filtered in each well, improving the data accuracy.

The oligonucleotides used in these assays were prepared by chemical synthesis and their sequences are:

ssDNA	GAT CTC GCA TCA CGT GAC GAA GAT C
RNA	UGU GUG UGU GUG UGU GUG UGU GUG
dsDNA	5'-GAT CTC GCA TCA CGT GAC GAA GAT C-3' 3'-CTA GTG CGC AGT GCA CTG CTT CTA G-5'

The binding reactions were performed with two-fold serial dilutions of the peptides in 25 mM Na-phosphate buffer pH 7,2 containing 4mM MgCl₂ (60 µl reaction volume in the presence of 1nM labelled probe).

19.1.1 Arginine dimethylation does not affect the nucleic acid binding properties of the peptide RGG

In a first set of binding experiments 1nM ssDNA and RNA were titrated with RGG, DMA-GG and KGG peptides.

Figure 54 shows the β -emission image of both the nitrocellulose and DEAE membranes of two of these binding assays. The data measured by betascope imaging for all the binding assays were plotted on graphs to produce the binding curves shown in **Figure 55** and **Figure 56**.

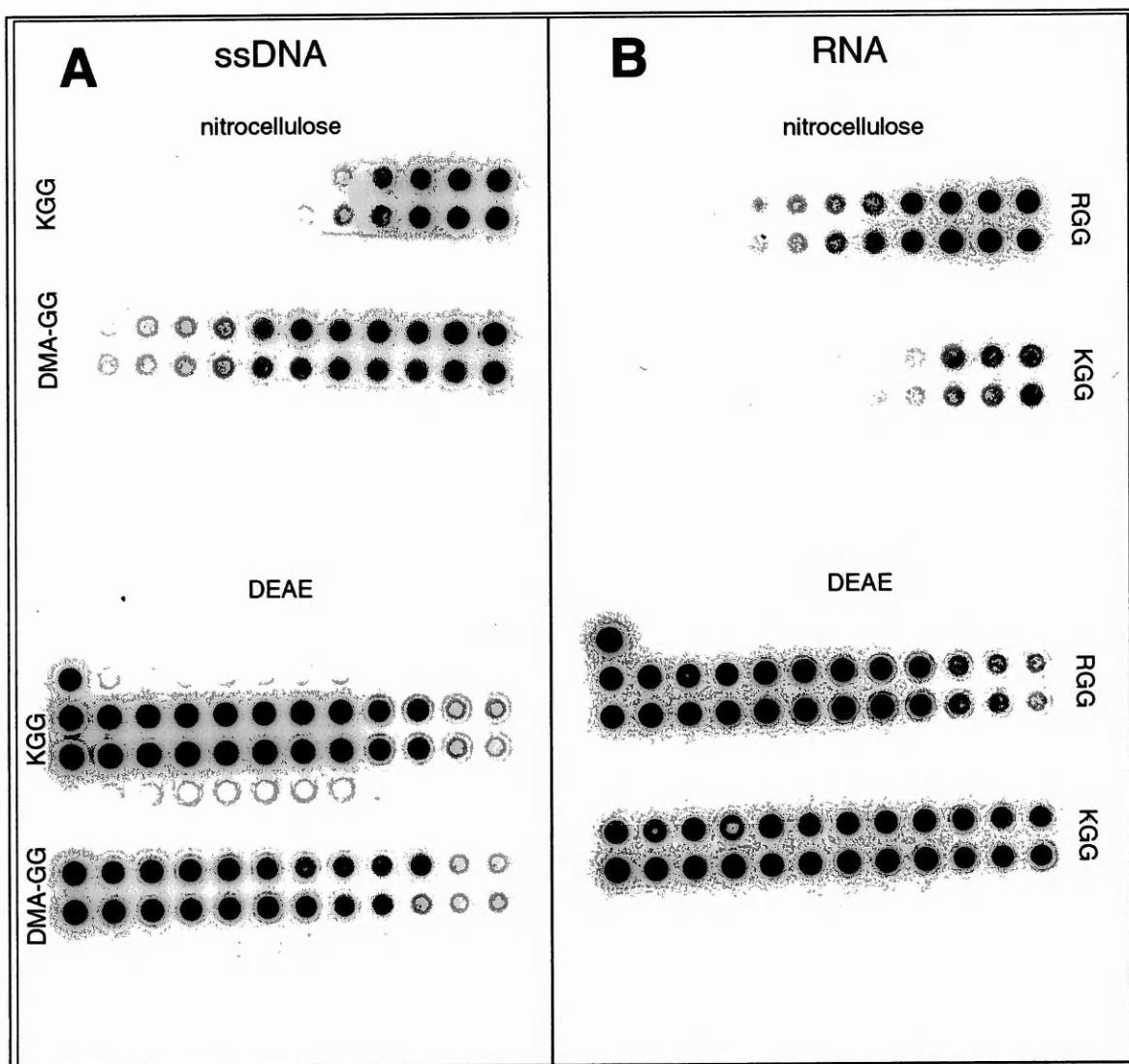


Figure 54 – Betascope images of the nitrocellulose and DEAE filters from a titration of ssDNA with KGG and DMA-GG peptides (panel A) and from a titration with RNA of KGG and RGG peptides (panel B). Wells from left to right contained progressively higher concentrations of peptide (0, 0.16, 0.31, 0.62, 1.09, 1.55, 3.10, 4.65, 7.75, 15.50, 38.75, 77.50, 155 μ M for both RGG and DMA-GG peptides; 0, 7, 14, 28, 49, 70, 140, 209, 349, 698, 1744, 3488 μ M for the KGG peptide). Each group of two rows contains duplicate sets of data.

Under our experimental conditions, the K_d of the interaction (which can be approximated to the point of half maximal binding) of the RGG peptide can be estimated around 10 μM for RNA binding and 3 μM for ssDNA binding (**Figure 55**). These values are not substantially altered by arginine methylation as can be seen from DMA-GG interaction with both substrates (**Figure 55**). The RGG peptide seems to bind slightly stronger in each case, however the differences are within the estimated experimental error.

On the other hand, KGG, the peptide analogue that contains lysine residues in place of all arginine residues, shows an approximate 20 to 50 fold weaker binding affinity for the RNA and ssDNA molecules respectively. The RGG2 peptide behaves similarly to the RGG and DMA-GG peptides with both substrates (data not shown).

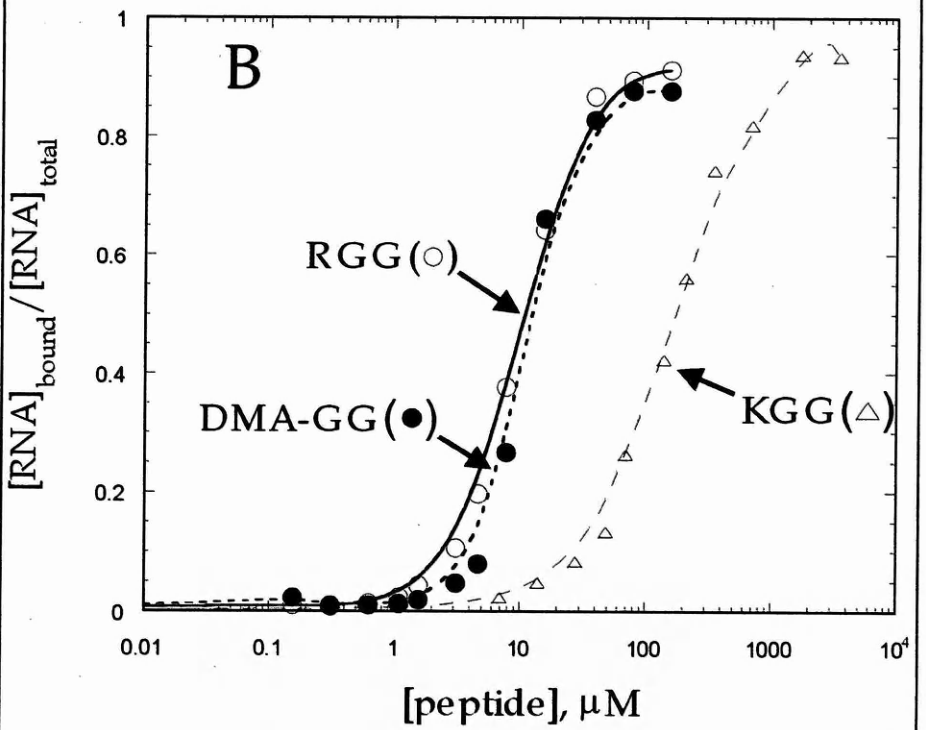
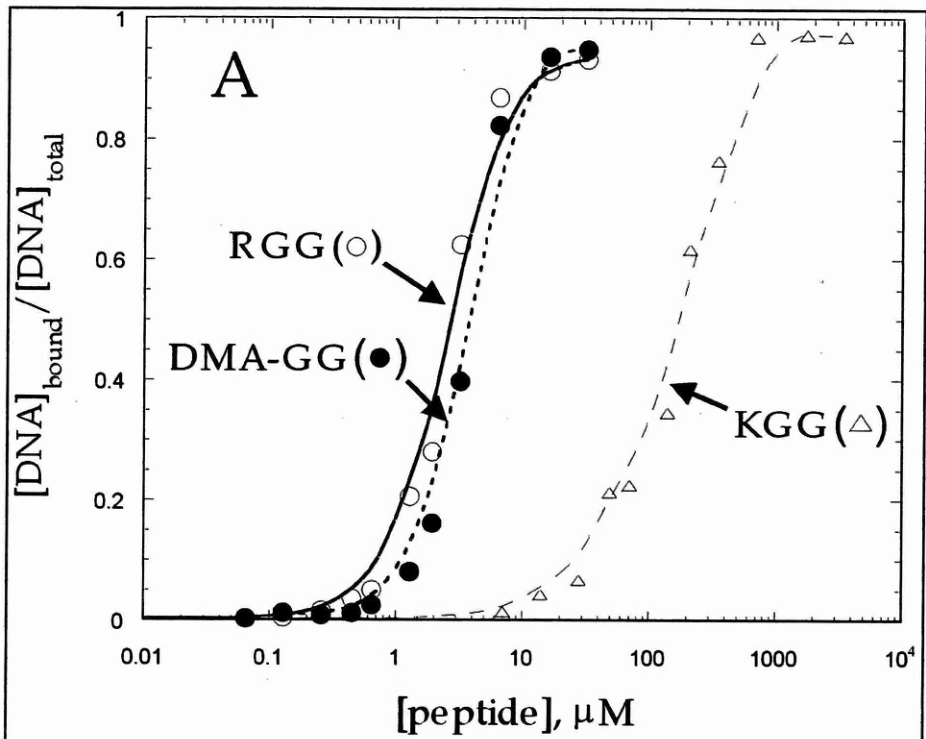
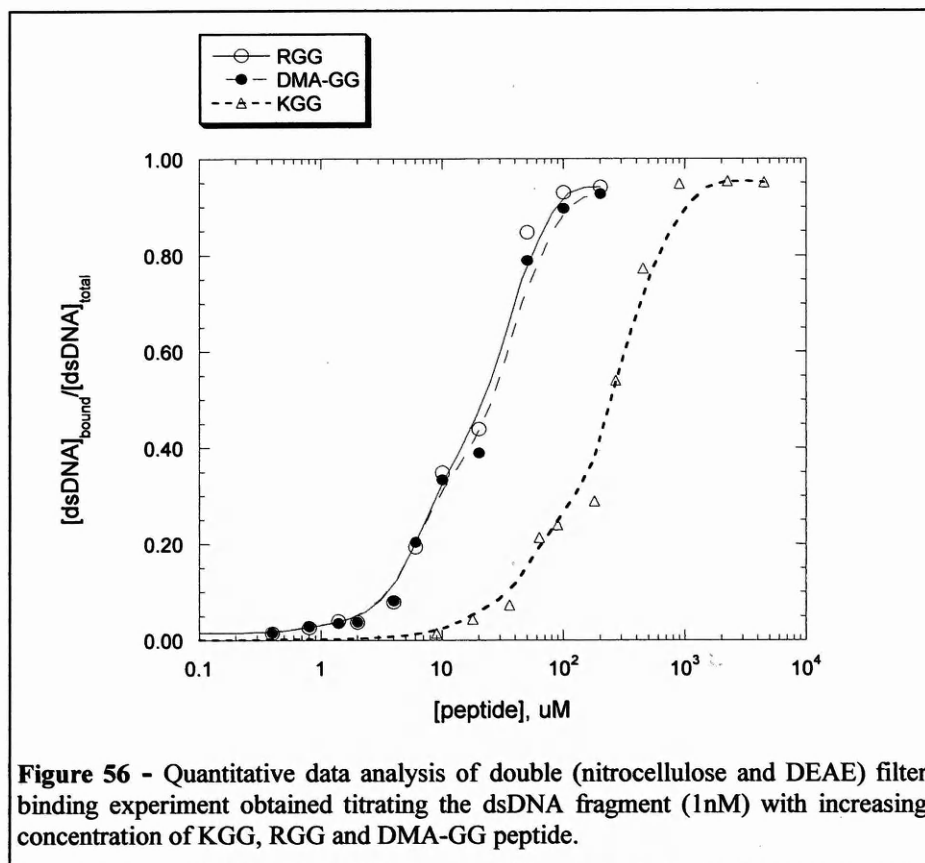


Figure 55 – Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiments obtained titrating a nucleic acid fragment (1nM) with increasing concentration of KGG (Δ), RGG (\circ), and DMA-GG (\bullet) peptide. Panel A shows the experiment in the presence of ssDNA, panel B in the presence of RNA. The data are average counts from duplicate sets of data in three different experiments.

19.1.2 RGG and DMA-GG bind dsDNA with lower affinity

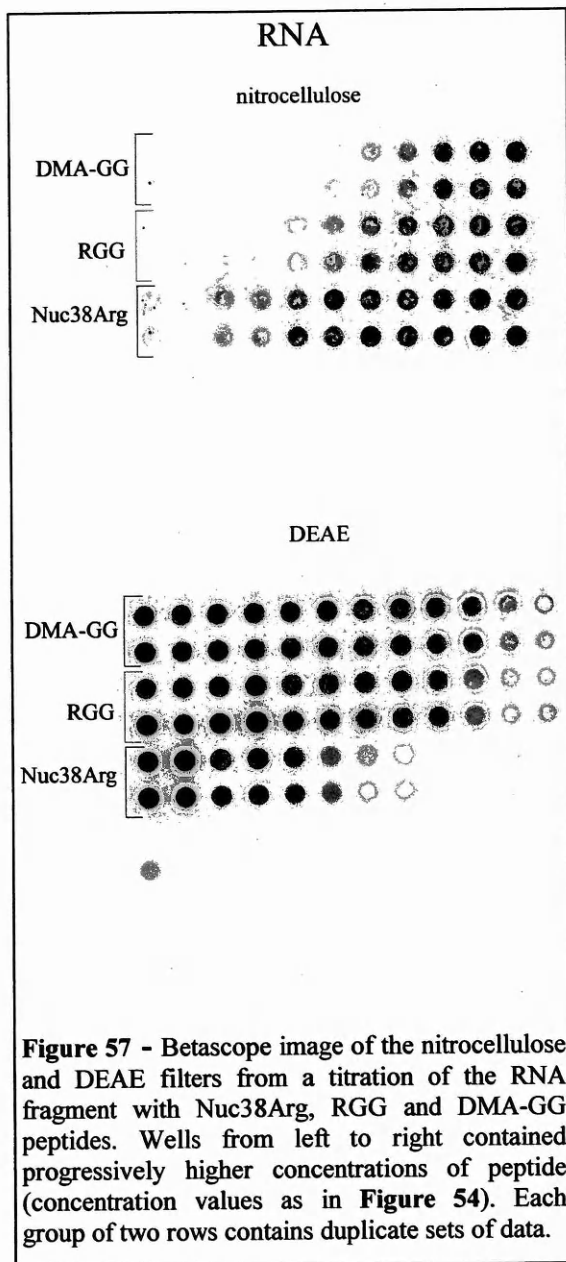
The situation with dsDNA is very similar from the point of view of the relative affinities of the various peptides, but both RGG and DMA-GG show an approximate ten-fold lower propensity to bind dsDNA (**Figure 56**) respect to ssDNA suggesting the



fact that the RGG and DMA-GG peptides are somehow sensitive to the geometry of the nucleic acid.

The binding assays taken in complex indicate that the RGG peptide can efficiently bind RNA and ssDNA and to a weaker extent dsDNA. The dimethylation on the other hand does not significantly affect its binding properties, but the substitution of Arg with Lys dramatically decreases its affinity for the nucleic acids.

19.1.3 *Nuc38Arg, Nuc61 Arg and Nuc61Dma binding to RNA and ssDNA*



Nuc38Arg corresponds to residues 669-706 of human nucleolin and is an intermediate peptide produced during the stepwise synthesis of Nuc61Arg. The sequence contains six arginines involved in five RGG sequences (the RGG peptide has four Arg residues in four RGG) and a C-terminal tail composed of twelve aminoacids of heterogeneous composition (Table 1). We examined the binding properties of this peptide (purified as described for Nuc61Arg) in a double filter binding assay in the presence of the RNA substrate and in direct comparison with RGG and DMA-GG (**Figure 57**).

The collected data, plotted in **Figure 58**, indicate an approximate five-fold higher affinity of Nuc38Arg for RNA with respect to both the shorter peptides (RGG, DMA-GG).

At the present time Nuc 61Arg and Nuc61Dma have been assayed in a single filter binding experiment in the presence of ssDNA, the binding curves in **Figure 59** show an approximate K_d of 1.5×10^{-7} M for both polypeptides confirming the hypothesis that longer RGG boxes show increased affinity for the nucleic acids. Again arginine dimethylation did not reveal relevant influence over binding affinity.

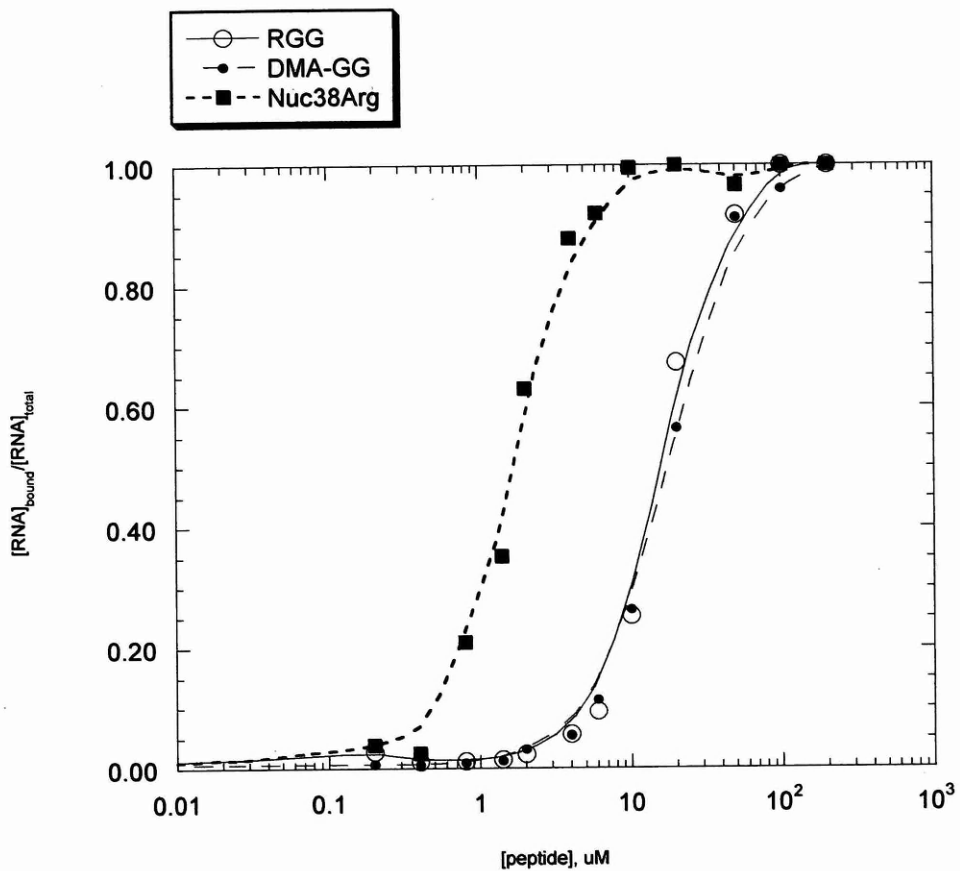


Figure 58 - Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiment obtained titrating the RNA fragment (1nM) with increasing concentration of Nuc38Arg, RGG and DMA-GG peptides.

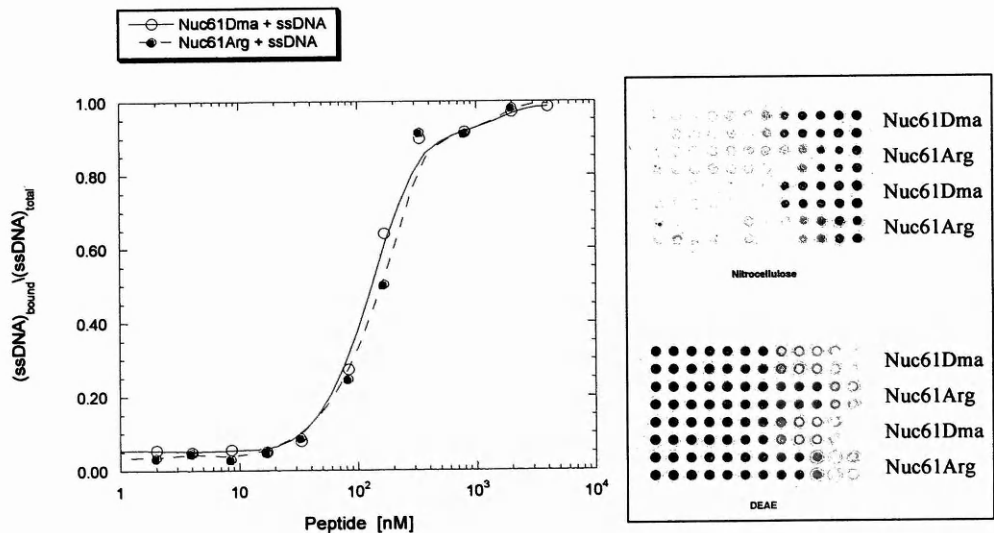
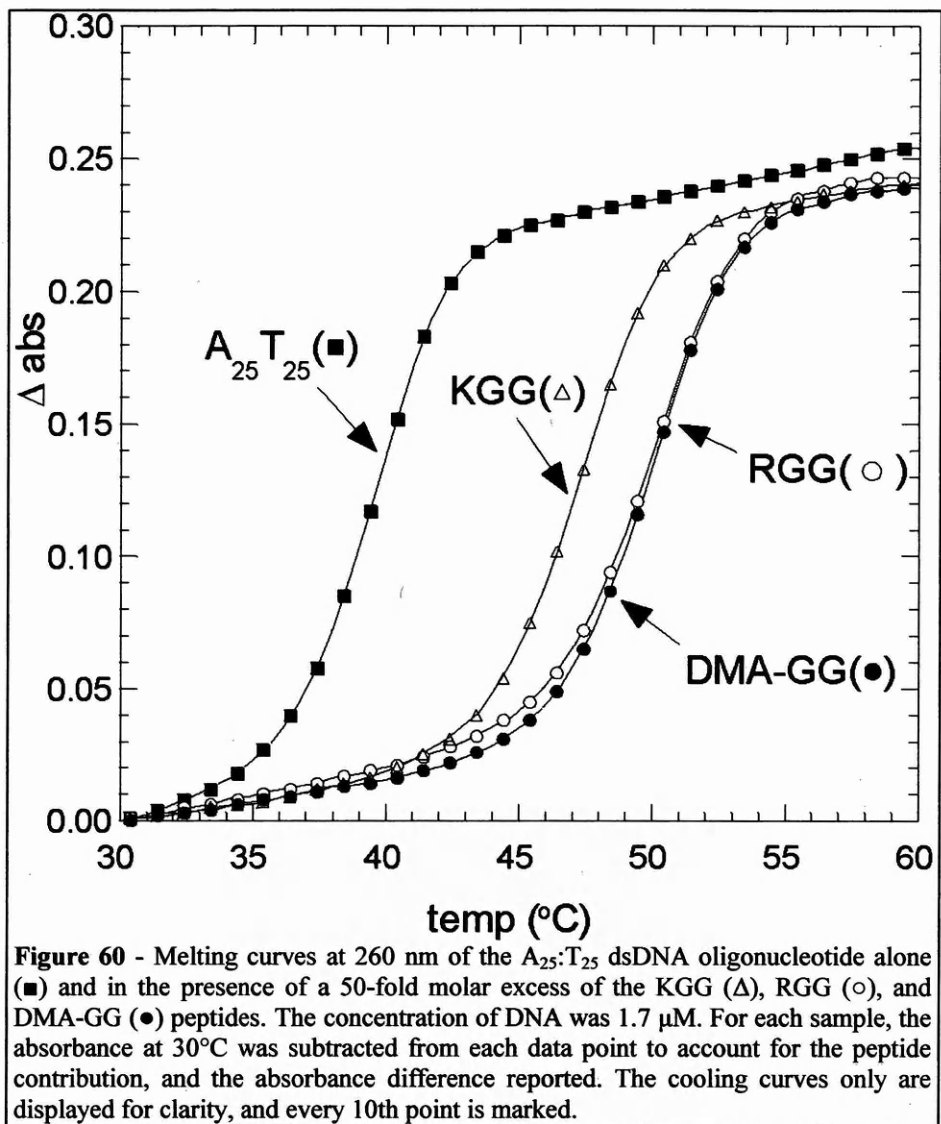


Figure 59 Left panel: Quantitative data analysis of double (nitrocellulose and DEAE) filter binding experiment obtained titrating the ssDNA fragment (1nM) with increasing concentration of Nuc61Arg and Nuc61Dma peptides. Right panel: Betascope image of the nitrocellulose and DEAE filters from the titration of the ssDNA fragment with Nuc61Arg and Nuc61Dma. Peptide concentrations are increasing twofold from left to right starting from 1nM to 2000nM (Nuc61Dma) and from 0.5nM to 1000nM (Nuc61Arg).

19.2 DNA melting curves

As the filter-binding assays suggested that RGG and DMA-GG peptides behave similarly upon nucleic acid binding while the binding of KGG-peptides is weaker, we conducted an additional test: the measurement of DNA melting temperature in the presence of the various peptides. These experiments were carried out with a ds-DNA oligomer, A₂₅T₂₅, which was chosen for his low T_m. In the buffer used, A₂₅T₂₅ has a reversible melting curve with a unique midpoint at 39.5 °C (**Figure 60**). The presence of the peptides increased the T_m, but at different levels. For a 50-fold peptide excess of the RGG and DMA-GG peptides, the T_m value raised to 50.0 °C whereas the KGG peptide increased the T_m value to 47.5 °C. The T_m showed some dependence on the peptide concentration. In the case of a 150-fold excess of RGG, some turbidity appeared and the T_m could not be measured reliably. Neither DMA-GG nor KGG produced a turbidity at similar concentrations. Otherwise, all the transitions appeared to be reversible and unique. The actual magnitude of ΔT_m varied with the peptide concentration, but the relative differences between the various peptides remained the same (not shown). Similar, stabilizing effects are described in literature in the interaction of protamines with DNA (Saenger, 1984). Protamines are small, highly positively charged peptides containing rows of arginines, and thus bear some similarity with the Gly/Phe/Arg rich peptides from nucleolin. It has been shown that protamines are capable of raising the T_m of dsDNA and that arginine-containing peptides are more effective than their lysine equivalents, probably because of the stronger hydrogen bonding capabilities of arginine.

It has also been observed that at proper concentrations, protamines are able to precipitate DNA (Saenger, 1984). Interestingly, we also observed some turbidity formation in the presence of 150-fold excess of the RGG peptide, but not in the presence of the DMA-GG or the KGG peptide.



19.3 Circular dichroism studies

The non-specific interaction of a protein domain with a nucleic acid can be experimentally demonstrated by circular dichroism (CD) spectroscopy (see introduction chapter). Previous published works showed that a recombinant polypeptide comprising of the glycine-arginine rich domain of nucleolin modifies the structure of RNA

molecules, a phenomenon that can be followed by the negative changes in the CD spectra of the nucleic acids (Ghisolfi et al., 1992a; Ghisolfi et al., 1992b).

In order to better characterize these structural changes we performed a number of circular dichroism studies with the synthetic peptides RGG, DMA-GG, KGG and RGG2 in the presence of the following various nucleic acid substrates:

RNA:

MS2phage RNA (1200 bases)

TAR-RNA: GGC CAG AUC UGA GCC UGG GAG CUC UCU GGC C

tRNA from yeast

DNA:

Calf thymus ssDNA

Calf thymus dsDNA

ssDNA: GGG ATC GAA ACG TAG CGC CTT CGA TCC C (hairpin like)

19.3.1 The interaction with the RGG peptide affects the nucleic acid dichroic spectrum.

In **Figure 61** is shown the dichroic spectrum of MS2 phage RNA in the absence and in the presence of increasing concentrations of the RGG peptide. Above 250 nm only RNA and not the peptides exhibits a CD signal, a positive peak centered at 268 nm. With the increase of the peptide concentration, the peak shifts towards higher wavelength with a reduction in the ellipticity.

Similar modifications of the spectrum have been observed previously with the recombinant fragments of nucleolin (Ghisolfi et al., 1992a; Ghisolfi et al., 1992b). The decrease in the ellipticity, indicates unstacking of the bases and unfolding of the

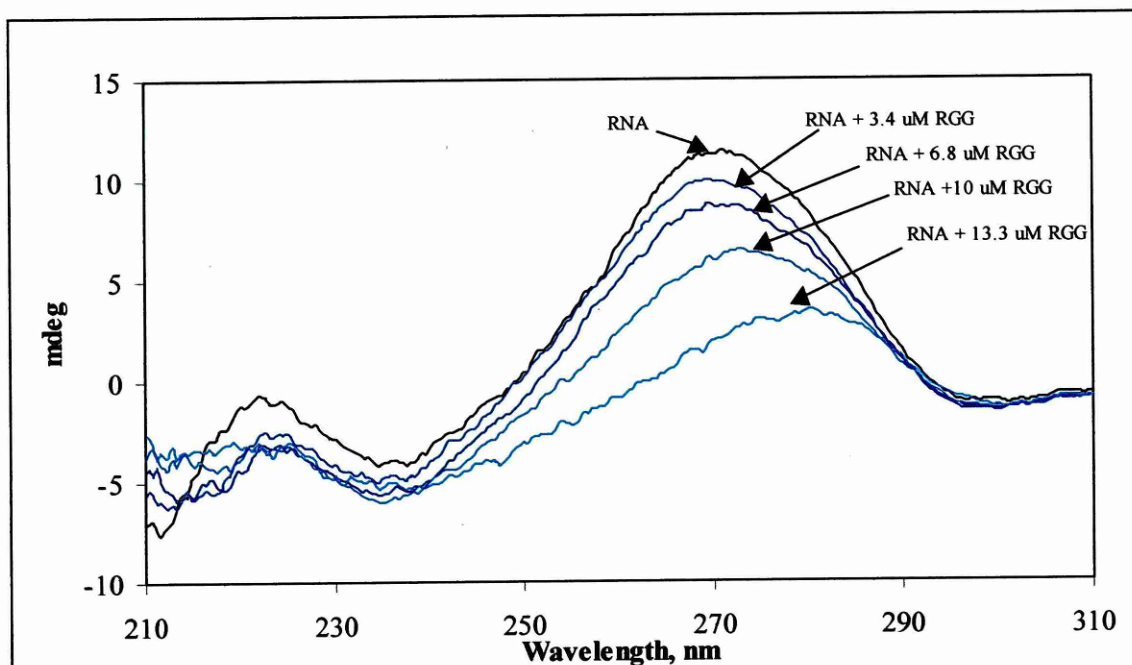


Figure 61 - Circular dichroic spectra of MS2 phage RNA in the presence of increasing concentrations (curves from top to bottom) of RGG peptide. A $34 \mu\text{g}\cdot\text{ml}^{-1}$ solution of RNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used.

secondary structure and has been ascribed to an increased base-base distance and a substantial tilt of the bases in the complex (Ghisolfi et al., 1992a; Ghisolfi et al., 1992b). We used this decrease in RNA mean residue mass ellipticity ($\Delta[\theta]_{\text{MRM}}$) at 268 nm to characterize the interactions between the peptides and various nucleic acid samples.

19.3.2 DMA-GG and KGG alter the nucleic acid conformation to a much lesser extent than RGG and RGG2

Figure 62 compares the interactions of RGG, DMA-GG, KGG and RGG2 peptides with MS2-phage RNA as a function of the peptide concentration. The profile shows a sharp increase in $\Delta[\theta]_{\text{MRM}}$ around 10 μM of RGG peptide. The solution was found to become turbid above the indicated concentrations of the peptide suggesting that the RNA-peptide complex is insoluble or is prone to aggregation.

Accordingly to this turbidity phenomenon, the second interesting feature of the spectrum of **Figure 61**, the large red shift of the maximum from around 268 to 275 nm, was previously also described in literature as a characteristic of the spectrum of one of the condensed forms of RNA, due to the inherent ability of RNA to pack in various tertiary structures (Steely et al., 1986). On the other hand, DMA-GG produced a much lower level of $\Delta[\theta]_{268}$, (**Figure 62** and **Figure 63**) and neither a sharp increase in $\Delta[\theta]_{268}$ nor turbidity was observed, even if the peptide was added in high concentrations. KGG behaviour was practically identical to the DMA-GG one, while RGG2 showed an effect

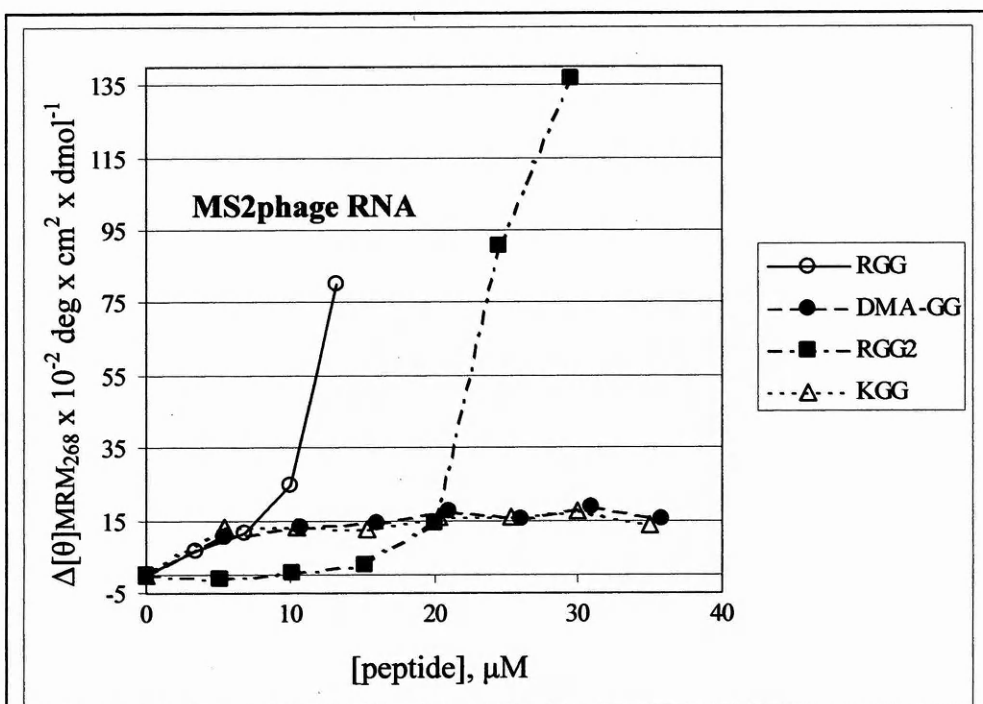


Figure 62 - Interaction of RGG, DMA-GG, RGG2 and KGG peptides with MS2-phage RNA. $\Delta[\theta]_{MRM,268nm}$ is the difference in mean residue mass ellipticity at 268 nm of RNA upon the interaction with the peptides. A $34 \mu\text{g}\cdot\text{ml}^{-1}$ solution of RNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used.

on the RNA dichroic signal similar to RGG but characterized by a minor influence on $\Delta[\theta]_{268}$ up to 20 μM concentration and by a very sharp decrease of the 268 nm signal after this point, reaching higher peptide/RNA ratios than in the RGG assay with no evident turbidity in the solution.

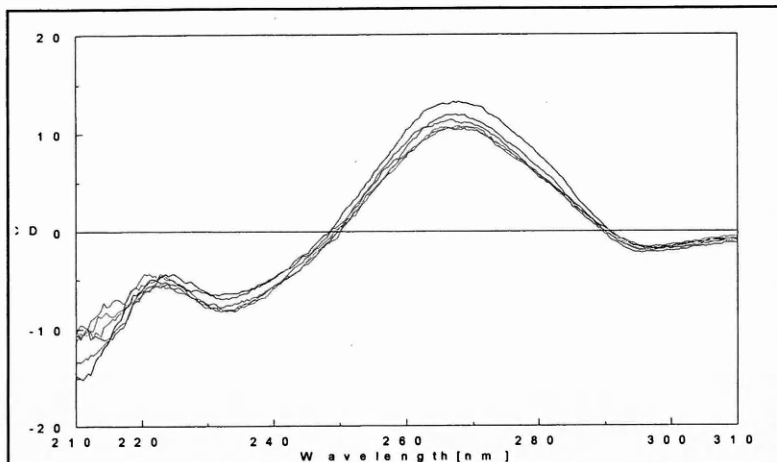


Figure 63 Circular dichroic spectra of MS2 phage RNA in the presence of increasing concentrations (curves from top to bottom are 0, 5.4, 16, 26, 36 μM) of DMA-GG peptide. A 34 $\mu\text{g}\cdot\text{ml}^{-1}$ solution of RNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used.

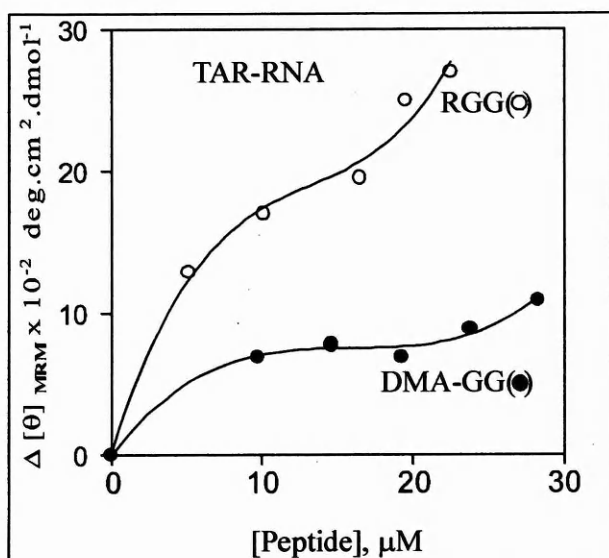


Figure 64 – Interaction of RGG and DMA-GG peptides with tar-RNA. A solution of 34 $\mu\text{g}\cdot\text{ml}^{-1}$ of tar-RNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used. $\Delta[\theta]_{\text{MRM}}$ is the difference in mean residue mass ellipticity at 265 nm of the nucleotides upon their interaction with the peptides.

The difference in the behaviour of RGG in comparison with its dimethylated counterpart DMA-GG was qualitatively reproduced upon interaction with the short, synthetic tar-RNA molecule (Figure 64), and with yeast tRNA (Figure 65) with the exception that no turbidity was observed in these cases at high peptide/RNA ratios.

Circular dichroism experiments performed with single stranded and double stranded calf thymus DNA (Figure 66 and Figure 67 respectively) confirmed in all of these cases that the RGG peptide also perturbs the DNA secondary structure while DMA-GG and KGG produce a characteristically weaker effect than the RGG; the RGG2 peptide again behaves similarly to RGG but at higher peptide/nucleic acid ratios.

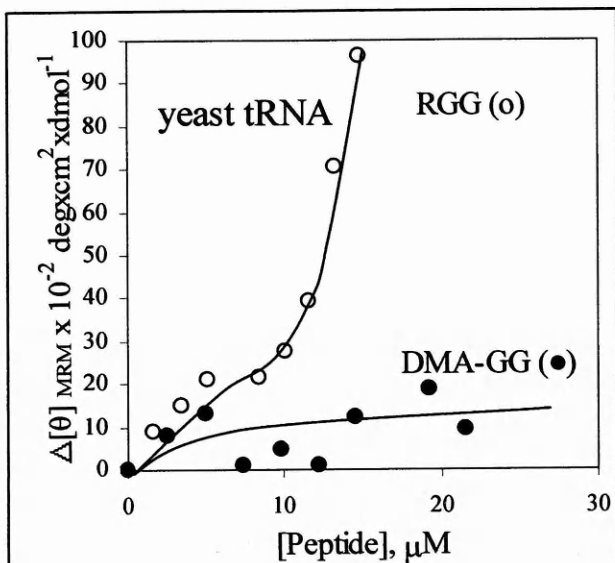


Figure 65 - Interaction of RGG and DMA-GG peptides with yeast tRNA. A solution of 34 $\mu\text{g}\cdot\text{ml}^{-1}$ of RNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used. $\Delta[\theta]_{\text{MRM}}$ is the difference in mean residue mass ellipticity at 265 nm of the nucleotides upon their interaction with the peptides.

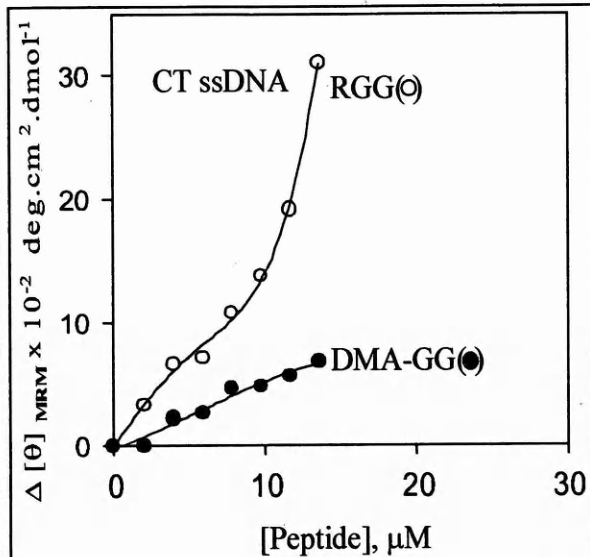


Figure 66 - Interaction of RGG and DMA-GG peptides with single stranded calf thymus DNA. A solution of 50 $\mu\text{g}\cdot\text{ml}^{-1}$ of ss DNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used. $\Delta[\theta]_{\text{MRM}}$ is the difference in mean residue mass ellipticity at 282 nm of the nucleotides upon their interaction with the peptides.

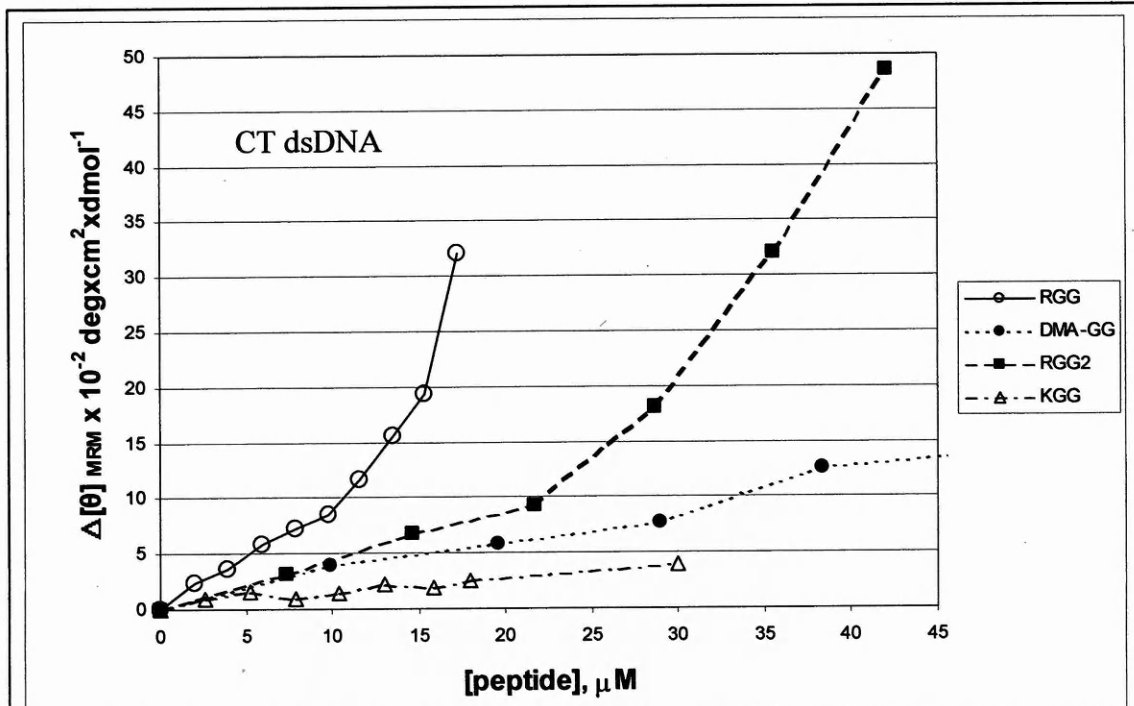


Figure 67 - Interaction of RGG, DMA-GG, RGG2 and KGG peptides with double stranded calf thymus DNA. A solution of 50 $\mu\text{g}\cdot\text{ml}^{-1}$ of ds DNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used. $\Delta[\theta]_{\text{MRM}}$ is the difference in mean residue mass ellipticity at 282 nm of the nucleotides upon their interaction with the peptides.

Peptide	Apparent K _d (M) ^a				Molar ratio (peptide/100 nucleotides) causing 50% decrease in the dichroic signal [% of maximum decrease observed, correspondent molar ratio] ^b						Δ T _m °C
	RNA	ssDNA	dsDNA	MS2 phage RNA	Yeast tRNA	TAR-RNA	ssDNA oligo	Calf thymus ssDNA	Calf thymus dsDNA	A25*T25	
RGG	1 x 10 ⁻⁵	3 x 10 ⁻⁶	2 x 10 ⁻⁵	12 [66%, 13.5 then solution turbidity]	12 [80%, 14]	50% not reached [20%, 23]	50% not reached [44%, 5.8]	7.8 [83%, 9.5]	10.6 [67%, 11.5]	10.5 °C	
Nuc38Arg	2 x 10 ⁻⁶										
DMA-GG	1 x 10 ⁻⁵	3 x 10 ⁻⁶	2 x 10 ⁻⁵	50% not reached [15%, 30]	50% not reached [16.5%, 18]	50% not reached [7%, 28]	50% not reached [14%, 13]	50% not reached [18%, 8.8]	50% not reached [24%, 26]	10.5 °C	
KGG	1 x 10 ⁻⁴	1 x 10 ⁻⁴	2 x 10 ⁻⁴	50% not reached [13.5%, 30]	ND	ND	ND	ND	50% not reached [8%, 20]	8.0 °C	
RGG2	1 x 10 ⁻⁵	3 x 10 ⁻⁶	2 x 10 ⁻⁵	22 [100%, 28]	ND	ND	ND	ND	16.4 [100%, 28.5]	ND	

Table 4 – Summary table of the nucleic acid interaction studies performed with the various synthetic peptides. ^a The apparent K_d is measured on the base of the double filter binding data. ^b The % decrease is referred to the initial [θ]_{MRM} (representing 100%) of the nucleotides at 268 nm (RNA) or 282nm (DNA) measured in absence of peptides. ^c The Δ T_m is referred to a peptide/DNA ratio of 50, respect to a T_m of 39.5 °C in the absence of peptide. ND = data not determined.

In **Table 4** the data of the nucleic acid interaction studies performed with the various peptides and the various nucleic acid substrates are summarised. It is interesting to note that, in our conditions, a molar ratio of 8-12 molecules of the RGG peptide for a 100 nucleotides fragment is sufficient to cause a 50% decrease of the dichroic signal in four of the six RNA and DNA substrates assayed reaching maximum observed values over the 80%. The ssDNA oligo would probably have entered in this group (it reached a 44% decrease with a 5.8 molar peptide excess) extending the peptide concentration range. The TAR-DNA on the contrary revealed a certain resistance to the structural modification with a 20% decrease in the presence of 23 molar excess of the peptide. The DMA-GG and the KGG peptides could never cause a decrease higher than 24% or respectively 14% of the initial dichroic signal of the assayed substrates.

19.3.3 Influence of ionic strength on the RGG- and DMA-GG-nucleic acid interaction

As a last point of our investigation we wanted to understand to which extent the interaction of the RGG and DMA-GG peptides with the nucleic acid substrates is affected by the ionic strength of the solution. We characterized this influence in two CD assays performed in the presence of MS2 phage RNA and of the ssDNA oligonucleotide at increasing NaCl concentrations (**Figure 68** and **Figure 69**).

The results indicate that an increase in ionic strength diminishes the absolute effect of both peptides on $\Delta[\theta]_{\text{MRM}}$. At 50 mM NaCl a fourfold concentration of RGG is necessary to cause the same decrease in the 268nm dichroic signal of MS2 phage RNA as in the buffer alone and the interaction with ssDNA shows a similar situation. On the other hand, the relatively higher capacity of RGG in comparison to DMA-GG to perturb the nucleic acid spectrum remains unaffected.

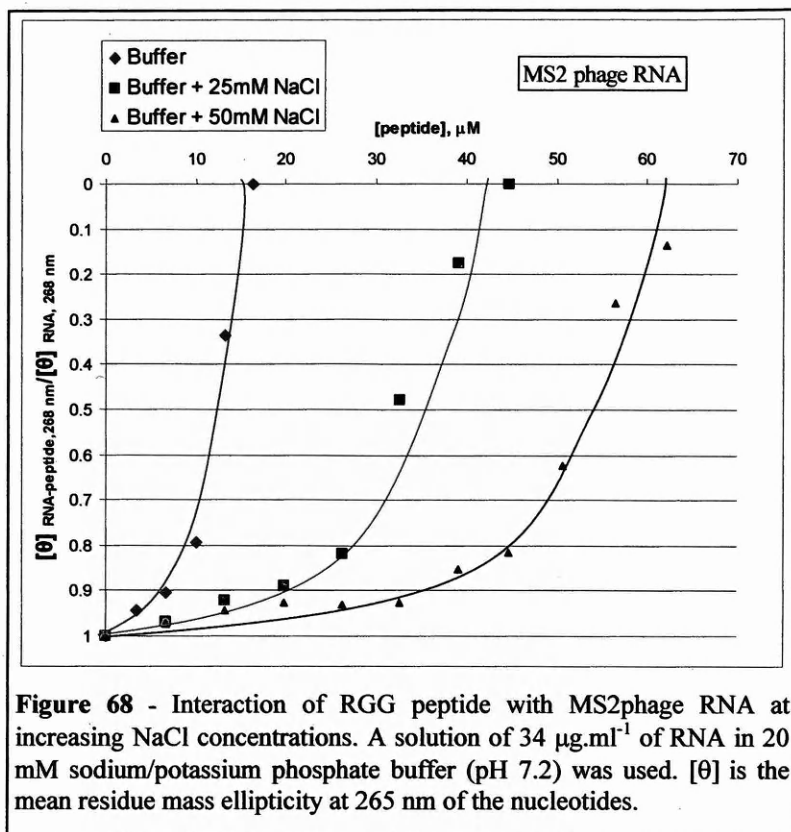


Figure 68 - Interaction of RGG peptide with MS2phage RNA at increasing NaCl concentrations. A solution of $34 \mu\text{g}\cdot\text{ml}^{-1}$ of RNA in 20 mM sodium/potassium phosphate buffer (pH 7.2) was used. $[\theta]$ is the mean residue mass ellipticity at 265 nm of the nucleotides.

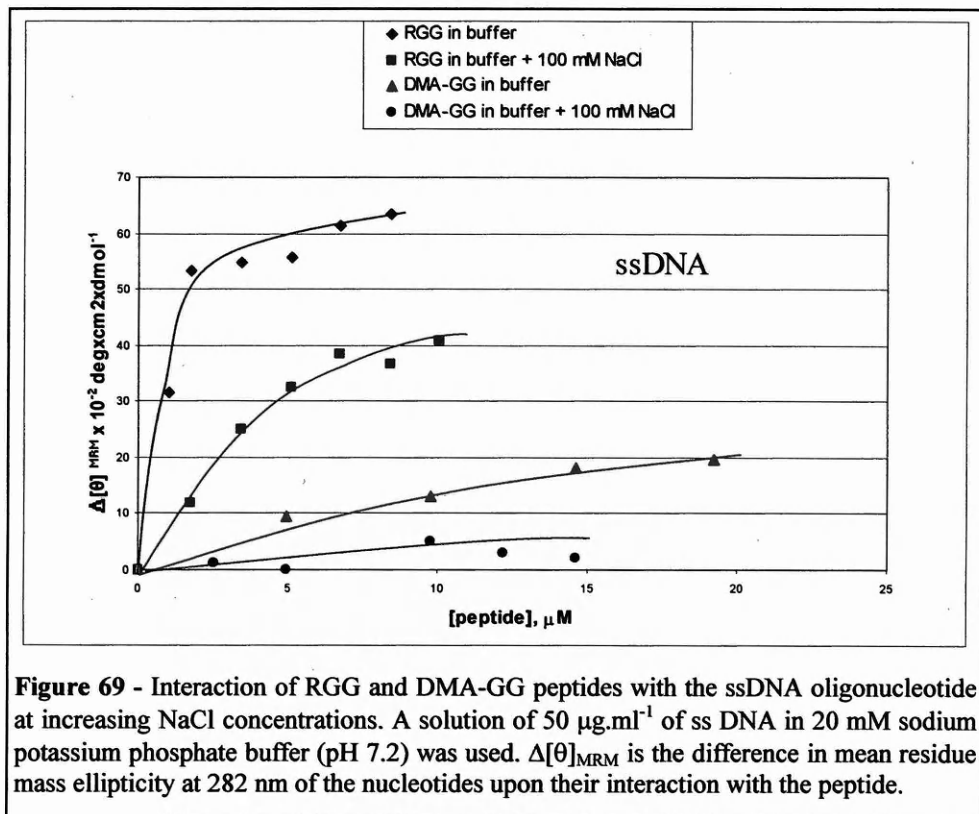


Figure 69 - Interaction of RGG and DMA-GG peptides with the ssDNA oligonucleotide at increasing NaCl concentrations. A solution of $50 \mu\text{g}\cdot\text{ml}^{-1}$ of ss DNA in 20 mM sodium potassium phosphate buffer (pH 7.2) was used. $\Delta[\theta]_{\text{MRM}}$ is the difference in mean residue mass ellipticity at 282 nm of the nucleotides upon their interaction with the peptide.

20. Discussion

The nucleolin RGG motif, a domain essential for efficient but non-specific binding of RNA, has been shown to destabilize the conformation of ribonucleic acid. This domain is also a natural substrate for the arginine N-methylation performed by Type I arginine methyltransferase, which results in the formation of N^ω, N^ω-dimethylarginine (aDMA).

Arginine residues have five possible proton donor groups and have a strong tendency to form hydrogen bonds, especially double hydrogen bonded structures. In order to form two hydrogen bonds to an arginine group, an acceptor molecule will have to contain either one hydrogen bond acceptor (resulting in a bifurcated hydrogen bond) or two acceptors (resulting in a bidentate hydrogen bond, practically two hydrogen bonds) (Shimoni and Glusker, 1995). The interaction of arginine with the nucleic acid phosphates can fulfill both possibilities combined in various patterns thanks to the N^δ and the two N^ω atoms.

Dimethylation on the other hand drastically reduces the number of hydrogen bonds that arginine can form, and some of the patterns with double H-bonds are excluded. Furthermore, dimethylation increases the hydrophobicity, the accessible surface, and the molecular volume of the guanidine moiety, but does not alter the net charge of the residue (Kennedy et al., 2000). We were thus tempted to experimentally investigate the role of arginine dimethylation on the interaction between RGG rich sequences and nucleic acids, using filter binding assays, circular dichroism spectroscopy, and melting experiments.

In order to perform our studies we had to construct by solid phase peptide synthesis a number of peptides spanning segments of different length of the nucleolin C-terminal domain. One of the major achievements of this work is indeed the

establishment of a reliable set of techniques and novel peptide synthesis strategies that lead to the total chemical synthesis of the 61 aa long nucleolin C-terminal domain both in non-methylated (Nuc61Arg) and methylated versions (Nuc61Dma).

The peptide **RGG** (representing residues 676-692 of human nucleolin (Ac-GRGGFGGRGGFRGGRGG-NH₂)) and a number of rationally designed analogues, **DMA-GG** (R=DMA), **KGG** (R=K), **RGG2** (F=G) constructed to investigate the role of the different amino acids in the motif context, were used in the nucleic acid interaction studies.

According to filter binding assays the peptides RGG and DMA-GG appear to be sensitive to the geometry of the nucleic acid as seen that they bind dsDNA ($K_d = 2 \times 10^{-5}$ M) with an approximate ten-fold lower affinity than ssDNA ($K_d = 2 \times 10^{-6}$ M) (Figure 47 and Figure 48). This result is different from the previous data published by Ghisolfi and collaborators (Ghisolfi et al., 1992a) in which the 52aa long Gly-Arg rich region of hamster nucleolin did not show any preference for the substrate geometry ($K_d = 2 \times 10^{-6}$ M for both substrates). This difference in the binding affinity of the peptide RGG for ssDNA and dsDNA is instead consistent with recently published results: recombinant nucleolin fragments containing the central RBDs and the RGG domain stimulated the annealing of ssDNA to give dsDNA and this resulted in the release of DNA from the nucleolin fragment. The presence of the RGG domain was essential to perform this annealing and dissociation actions (Hanakahi et al, 2000).

On the other hand filter-binding assays reveal no appreciable differences between the nucleic acid binding of the Arg-dimethylated and the unmethylated peptides (Figure 55, Figure 56, Figure 58 and Figure 59). This confirms that the non-specific binding of RGG peptides to the nucleic acid is primarily ionic in nature. These findings also confirm the results of Serin et al (Serin et al., 1997) and Valentini et al.(Valentini et al., 1999) who did not find appreciable differences between the binding of dimethylated and non/methylated proteins to nucleic acids. On the other hand, both filter binding assays and DNA melting experiments showed that the KGG peptide

behaves differently from both the RGG and the DMA-GG peptides. Similar results, showing a minor affinity of lysine for DNA in comparison to arginine have been reported in previous works using model peptides and equilibrium dialysis (Standke and Brunnert, 1975).

However, circular dichroism measurements reveal a clear difference in the behaviour of the methylated and the non-methylated peptides. Dimethylation substantially decreases the ability of the RGG peptide to modify the conformation of the nucleic acid substrate, and this difference is especially conspicuous in the case of the interaction with the MS2 phage RNA (Figure 61, Figure 62 and Figure 63). The decrease in the intensity of the CD band at 268 nm, and its shift to higher wavelengths (Figure 63), caused by the RGG peptide, is usually interpreted as base unstacking and destabilization of the nucleic acid structure (Fasman, 1996; Ghisolfi et al., 1990; Ghisolfi et al., 1992b). Furthermore, the red shift of the spectrum and the observed turbidity of the solution at higher peptide/nucleic acid ratios, according to some authors (Ghisolfi et al., 1992b; Steely et al., 1986) are characteristic of a nucleic acid condensed phase and reflect the inherent ability of RNA to pack in various tertiary structures described as long fibers. On the other hand, DNA melting studies on a simple dsDNA model show that the presence of the RGG or DMA-GG peptides increases, rather than decreases the stability of the DNA helix, and suggest that the changes in the CD spectra are due to conformational changes connected to local perturbations in base stacking but not to unwinding and destabilization of the double helical structure (Figure 60).

The difference in the capacity of the RGG and DMA-GG peptides in producing the condensed phase could be explained with the different hydrogen bonding capabilities of Arg and Dma. In the case of protamine-DNA complexes it was suggested that one guanidinium group could cross-link two DNA helices in DNA-protamine fibers (Saenger, 1984). A similar cross-linking ability could be compromised in the case of

aDMA containing peptides, at least from the point of view of the number of possible hydrogen bonding patterns.

The RGG2 peptide has a similar behaviour to the RGG peptide from the point of view of the nucleic acid binding but produces similar effects on nucleic acid CD spectra at higher peptide/nucleic acid ratios (Figure 62 and Figure 67). It is possible that the lack of the phenylalanine residue, substituted by a glycine, somehow affects the capacity of the RGG peptide to alter the base stacking. Some data are available in literature for model di- and tri-peptides composed by a phenylalanine flanked by a charged amino acid (Arg or Lys); these peptides were thought to bind to single-stranded polynucleotides mainly by the charged groups and to stack the Phe residue with bases. With double-stranded DNA, the insertion was thought only partial, the Phe acting as a wedge and causing bending of the double helix (Gabbay et al., 1972).

Phenylalanine residues are usually buried in the hydrophobic core of proteins, and rarely occupy positions that are exposed to the solvent. In protein/DNA complexes, aromatic residues are frequently found in intercalating positions between two base-pairs.

One could plausibly speculate that the non-specific binding of RGG peptides to the nucleic acid follows a two step scenario similar to the formation of specific protein/DNA complexes: a peptide first establishes ionic interactions with the nucleic acid backbone, and this is followed by the formation of specific H-bonds and a conformational change of the nucleic acid partner. According to this simple model, the non-methylated peptide would be able to carry out both steps. The methylated peptide, on the other hand, would bind to the nucleic acid with the same affinity, mainly controlled by the overall charge, but would not be able to proceed to the second step as it can not establish as many H bonds as the non-methylated peptide, and perhaps more importantly, several H-bond patterns would be excluded.

We therefore think that binding of the RGG peptide brings about a local deformation of the nucleic acid structure, and it is important to notice that neither DMA, nor lysine can replace arginine in this respect, as shown by the CD results presented here. The sequence of the RGG motifs in different proteins is quite variable, however there are constant features that suggest that these motifs may form precise contacts with the nucleic acid backbone. For example, the nearly-conserved spacing of arginine groups and a phenylalanine residue, that in protein/DNA complexes is frequently found to perturb the helical structure, is invariably present.

The conserved aromatic residues of the RGG motifs might be essential for the intercalation into the RNA structure, but the H-bonds necessary to precisely position them could not form if the peptide is methylated. Even though the arginine-glycine rich domains are not involved in specific nucleic acid recognition mechanisms, their presence in all these protein families, involved in a variety of different processes, is indispensable to increase the affinity for RNA or DNA. Since the binding of the methylated and unmethylated peptides appears to be similar in strength, one might hypothesize that the subtle conformational changes brought about by local structural rearrangements of the nucleic acid bases can play a role in regulating these processes.

References

- Aggarwal, A.K., Rodgers, D.W., Drottar, M., Ptashne, M. and Harrison, S.C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*, **242**, 899-907.
- Albright, R.A., Mossing, M.C. and Matthews, B.W. (1998) Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: Possible implications for nonspecific binding by the wild-type protein. *Protein Science*, **7**, 1485-94.
- Aletta, J.M., Cimato, T.R. and Ettinger, M.J. (1998) Protein methylation: a signal event in post-translational modification. *Trends Biochem Sci*, **23**, 89-91.
- Amiri, K. A. (1994). Fibrillar-like proteins occur in the domain Archaea. *J Bacteriol* **176**(7), 2124-7.
- Anderson, J.E., Ptashne, M. and Harrison, S.C. (1987) Structure of the repressor-operator complex of bacteriophage 434. *Nature*, **326**, 846-52.
- Antson, A.A. (2000) Single-stranded-RNA binding proteins [see comments]. *Curr Opin Struct Biol*, **10**, 87-94.
- Antson, A.A., Dodson, E.J., Dodson, G., Greaves, R.B., Chen, X. and Gollnick, P. (1999) Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature*, **401**, 235-42.
- Bell, A.C. and Koudelka, G.B. (1993) Operator sequence context influences amino acid-base-pair interactions in 434 repressor-operator complexes. *J Mol Biol*, **234**, 542-53.
- Bell, A.C. and Koudelka, G.B. (1995) How 434 repressor discriminates between OR1 and OR3. The influence of contacted and noncontacted base pairs. *J Biol Chem*, **270**, 1205-12.
- Berg, O.G. and von Hippel, P.H. (1987) Selection of DNA binding sites by regulatory proteins. Statistical mechanical theory and application to operators and promoters. *J. Mol. Biol.*, **193**, 723-50.

- Berger, C., Piubelli, L., Haditsch, U. and Bosshard, H.R. (1998) Diffusion-controlled DNA recognition by an unfolded, monomeric bZIP transcription factor [published erratum appears in FEBS Lett 1998 Jun 12;429(2):221]. *FEBS Lett*, **425**, 14-8.
- Blackburn, G.M. and Gait, M.J. (1996) *Nucleic Acids in Chemistry and Biology*.
- Boschelli, F. (1982) Lambda phage cro repressor. Non-specific DNA binding. *J Mol Biol*, **162**, 267-82.
- Bouvet, P., Diaz, J.J., Kindbeiter, K., Madjar, J.J. and Amalric, F. (1998) Nucleolin interacts with several ribosomal proteins through its RGG domain. *J Biol Chem*, **273**, 19025-9.
- Bray, D. and Lay, S. (1997) Computer-based analysis of the binding steps in protein complex formation. *Proc Natl Acad Sci U S A*, **94**, 13493-8.
- Brennan. (1992) DNA recognition by the helix-turn-helix motif. *Current Opinion in Structural Biology*, **2**, 100-108.
- Brennan, R.G. (1991) Interactions of the helix-turn-helix binding domain. *Curr Opin Struct Biol*, **1**, 80-88.
- Bouvet, P., Diaz, J. J., Kindbeiter, K., Madjar, J. J. & Amalric, F. (1998). Nucleolin interacts with several ribosomal proteins through its RGG domain. *J Biol Chem* **273**(30), 19025-9.
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. & Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**(5278), 1058-73.
- Burley, S.K. (1996) The TATA box binding protein. *Curr. Opin. Struct. Biol.*, **6**, 69-75.

- Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G. & Riva, S. (1990). Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *Embo J* **9**(4), 1229-35.
- Carey, J. (1991) Gel retardation. *Methods Enzymol*, **208**, 103-17.
- Carpino, L.A., Cohen, B.J., K.E., S., Sadat-Aalee, S.Y., Tien, J.-H. and D.C., L. (1986) (9-Fluorenylmethyl)oxy)carbonyl (Fmoc) Amino Acid Chlorides. Synthesis, Characterization, and Application to the Rapid Synthesis of Short Peptide Segments. *J. Org. Chem.*, **51**, 3734-45.
- Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S. and Biamonti, G. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol*, **259**, 337-48.
- Chakrabarty, A., Kortemme, T., Padmanabhan, S. and Baldwin, R.L. (1993) *Biochemistry*, **32**, 5560-5.
- Chen, J., Pongor, S. and Simoncsits, A. (1997) Recognition of DNA by single-chain derivatives of the phage 434 repressor: high affinity binding depends on both the contacted and non- contacted base pairs. *Nucleic Acids Res*, **25**, 2047-54.
- Choo, Y. and Klug, A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage [published erratum appears in Proc Natl Acad Sci U S A 1995 Jan 17;92(2):646]. *Proc Natl Acad Sci U S A*, **91**, 11163-7.
- Choo, Y. and Klug, A. (1997) Physical basis of a protein-DNA recognition code. *Curr Opin Struct Biol*, **7**, 117-25.
- Ciubotaru, M., Bright, F.V., Ingersoll, C.M. and Koudelka, G.B. (1999) DNA-induced conformational changes in bacteriophage 434 repressor. *J Mol Biol*, **294**, 859-73.
- Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1988) Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. Large-scale overproduction in *Escherichia coli* and cooperative binding to single-stranded nucleic acids. *J Biol Chem*, **263**, 1063-71.

- Costa, M., Ochem, A., Staub, A. & Falaschi, A. (1999). Human DNA helicase VIII: a DNA and RNA helicase corresponding to the G3BP protein, an element of the ras transduction pathway. *Nucleic Acids Res* **27**(3), 817-21.
- Creancier, L., Prats, H., Zanibellato, C., Amalric, F. and Bugler, B. (1993) Determination of the functional domains involved in nucleolar targeting of nucleolin. *Mol Biol Cell*, **4**, 1239-50.
- Cusack, S. (1999) RNA-protein complexes. *Curr Opin Struct Biol*, **9**, 66-73.
- Donner, A.L., Paa, K. and Koudelka, G.B. (1998) Carboxyl-terminal domain dimer interface mutant 434 repressors have altered dimerization and DNA binding specificities. *J Mol Biol*, **283**, 931-46.
- Drew, H.R. and McCall, M.J. (1990) *New approaches to DNA in the crystal and in solution*. Cold Spring Harbor Laboratory Press.
- Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321.
- Eichler, D. C. & Craig, N. (1994). Processing of eukaryotic ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* **49**, 197-239.
- Ellenberger, T. (1994) Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains. *Curr. Opin. Struct. Biol.*, **4**, 12-21.
- Erard, M. S., Belenguer, P., Caizergues-Ferrer, M., Pantaloni, A. & Amalric, F. (1988). A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. *Eur J Biochem* **175**(3), 525-30.
- Erard, M., Lakhdar-Ghazal, F. & Amalric, F. (1990). Repeat peptide motifs which contain beta-turns and modulate DNA condensation in chromatin. *Eur J Biochem* **191**(1), 19-26.
- Fang, S.H. and Yeh, N.H. (1993) The self-cleaving activity of nucleolin determines its molecular dynamics in relation to cell proliferation. *Exp Cell Res*, **208**, 48-53.
- Fasman, G.D. (1996) *Circular Dichroism and the Conformational analysis of Biomolecules*. Plenum Press, New York.

- Feng, J.A., Johnson, R.C. and Dickerson, R.E. (1994) Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science*, **263**, 348-55.
- Fersht, A. (1990) *Enzyme Structure and Mechanism*, W. H. Freeman Company.
- Forster, T. (1969) *Angew. Chem. Int. Ed. Engl.*, **8**, 333-343.
- Frankel, A.D. (2000) Fitting peptides into the RNA world. *Curr Opin Struct Biol*, **10**, 332-40.
- Frankel, A.D. and Smith, C.A. (1998) Induced folding in RNA-protein recognition: more than a simple molecular handshake. *Cell*, **92**, 149-51.
- Freemont, P.S., Lane, A.N. and Sanderson, M.R. (1991) Structural aspects of protein-DNA recognition. *Biochem J*, **278**, 1-23.
- Gabbay, E.J., Sanford, K. and Baxter, C.S. (1972) Specific interaction of peptides with nucleic acids. *Biochemistry*, **11**, 3429-35.
- Garner, M.M. and Rau, D.C. (1995) Water release associated with specific binding of gal repressor. *Embo J*, **14**, 1257-63.
- Gary, J.D. and Clarke, S. (1998) RNA and protein interactions modulated by protein arginine methylation. *Prog Nucleic Acid Res Mol Biol*, **61**, 65-131.
- Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resnedez-Perez, D., Affolter, M., Otting, G. and Wüthrich, K. (1994) Homeodomain-DNA recognition. *Cell*, **78**, 211-223.
- Ghisolfi, L., Joseph, G., Amalric, F. and Erard, M. (1992a) The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J Biol Chem*, **267**, 2955-9.
- Ghisolfi, L., Joseph, G., Erard, M., Escoubas, J.M., Mathieu, C. and Amalric, F. (1990) Nucleolin--pre-rRNA interactions and preribosome assembly. *Mol Biol Rep*, **14**, 113-4.
- Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F. and Erard, M. (1992b) Concerted activities of the RNA recognition and the glycine-rich C- terminal domains of

- nucleolin are required for efficient complex formation with pre-ribosomal RNA. *Eur J Biochem*, **209**, 541-8.
- Ginisty, H., Amalric, F. & Bouvet, P. (1998). Nucleolin functions in the first step of ribosomal RNA processing. *Embo J* **17**(5), 1476-86.
- Girard, J. P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D. & Lapeyre, B. (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. *Embo J* **11**(2), 673-82.
- Hagerman, P.J. (1990) Sequence-directed curvature of DNA. *Annu Rev Biochem*, **59**, 755-81.
- Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High Affinity Interactions of Nucleolin with G-G paired rDNA. *J Biol Chem*, **274**, 15908-12.
- Hanakahi, L. A., Bu, Z. & Maizels, N. (2000). The C-terminal domain of nucleolin accelerates nucleic acid annealing. *Biochemistry* **39**(50), 15493-9.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579-85.
- Hangland, R.P. (1992) *Handbook of Fluorescent Probes and Research Chemicals*. Molecular Probes, Inc., Eugene, OR.
- Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715-9.
- Harrison, S.C. and Aggarwal, A.K. (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem*, **59**, 933-69.
- Hirling, H., Scheffner, M., Restle, T. & Stahl, H. (1989). RNA helicase activity associated with the human p68 protein. *Nature* **339**(6225), 562-4.
- Hollis, M., Valenzuela, D., Pioli, D., Wharton, R. and Ptashne, M. (1988) A repressor heterodimer binds to a chimeric operator. *Proc Natl Acad Sci U S A*, **85**, 5834-8.
- Holmbeck, S.M.A., Dyson, H.J. and Wright, P.E. (1998) DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J Mol Biol*, **284**, 533-9.

- Hu, O'Shea and Kim. (1990) *Science*, **250**, 1400-3.
- Ichikawa, H., Shimizu, K., Hayashi, Y. & Ohki, M. (1994). An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation. *Cancer Res* **54**(11), 2865-8.
- Jessen, T.H., Oubridge, C., Teo, C.H., Pritchard, C. and Nagai, K. (1991) Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *Embo J*, **10**, 3447-56.
- Johnson, A.D., Meyer, B.J. and Ptashne, M. (1979) Interaction between DNA-bound repressors govern regulation by the λ phage repressor. *Proc Natl Acad Sci USA*, **76**, 5061-65.
- Johnson, T., Quibell, M. and Sheppard, R.C. (1995) N,O-bisFmoc derivatives of N-(2-hydroxy-4-methoxybenzyl)-amino acids: useful intermediates in peptide synthesis. *J Pept Sci*, **1**, 11-25.
- Jones, S., van Heyningen, P., Berman, H.M. and Thornton, J.M. (1999) Protein-DNA interactions: A structural analysis. *J Mol Biol*, **287**, 877-96.
- Jordan, S.R. and Pabo, C.O. (1988) Structure of the Lambda Complex at 2.5A Resolution: Details of the Repressor-Operator Interactions. *Science*, **242**, 893-8.
- Kanaar, R., Lee, A.L., Rudner, D.Z., Wemmer, D.E. and Rio, D.C. (1995) Interaction of the sex-lethal RNA binding domains with RNA. *Embo J*, **14**, 4530-9.
- Kennedy, K.J., Lundquist, J.T.t., Simandan, T.L., Kokko, K.P., Beeson, C.C. and Dix, T.A. (2000) Design rationale, synthesis, and characterization of non-natural analogs of the cationic amino acids arginine and lysine. *J Pept Res*, **55**, 348-58.
- Kiledjian, M. and Dreyfuss, G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *Embo J*, **11**, 2655-64.
- Kim, B. and Little, J.W. (1992) Dimerization of a specific DNA-binding protein on the DNA. *Science*, **255**, 203-6.
- Kim, J.G., Takeda, Y., Matthews, B.W. and Anderson, W.F. (1987) Kinetic studies on Cro repressor-operator DNA interaction. *J Mol Biol*, **196**, 149-58.

- Kim, S., Merrill, B.M., Rajpurohit, R., Kumar, A., Stone, K.L., Papov, V.V., Schneiders, J.M., Szer, W., Wilson, S.H., Paik, W.K. and Williams, K.R. (1997) Identification of N(G)-methylarginine residues in human heterogeneous RNP protein A1: Phe/Gly-Gly-Gly-Arg-Gly-Gly-Gly/Phe is a preferred recognition motif. *Biochemistry*, **36**, 5185-92.
- Kim, Y., Geiger, J.H., Hahn, S. and Sigler, P.B. (1993) Crystal structure of a yeast TBP/TATA-box complex [see comments]. *Nature*, **365**, 512-20.
- Klein, S., Carroll, J.A., Chen, Y., Henry, M.F., Henry, P.A., Ortonowski, I.E., Pintucci, G., Beavis, R.C., Burgess, W.H. and Rifkin, D.B. (2000) Biochemical analysis of the arginine methylation of high molecular weight fibroblast growth factor-2. *J Biol Chem*, **275**, 3150-7.
- Klug, A. (1993) Co-chairman's remarks: protein designs for the specific recognition of DNA. *Gene*, **135**, 83-92.
- Kohler, J.J., Metallo, S.J., Schneider, T.L. and Schepartz, A. (1999) DNA specificity enhanced by sequential binding of protein monomers. *Proc Natl Acad Sci U S A*, **96**, 11735-9.
- Kondo, K. & Inouye, M. (1992). Yeast NSR1 protein that has structural similarity to mammalian nucleolin is involved in pre-rRNA processing. *J Biol Chem* **267**(23), 16252-8.
- Konig, P., Giraldo, R., Chapman, L. and Rhodes, D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, **85**, 125-36.
- Konkel, L. M., Enomoto, S., Chamberlain, E. M., McCune-Zierath, P., Iyadurai, S. J. & Berman, J. (1995). A class of single-stranded telomeric DNA-binding proteins required for Rap1p localization in yeast nuclei. *Proc Natl Acad Sci U S A* **92**(12), 5558-62.
- Koudelka, G.B. (1991) Bending of synthetic bacteriophage 434 operators by bacteriophage 434 proteins. *Nucleic Acids Res*, **19**, 4115-9.

- Koudelka, G.B., Bell, A.C. and Hilchey, S.P. (1995) *Indirect effects of DNA sequence on protein-DNA interactions*. Adenine Press, Inc., New York.
- Koudelka, G.B. and Carlson, P. (1992) DNA twisting and the effects of non-contacted bases on the affinity of 434 operator for 434 repressor. *Nature*, **355**, 89-91.
- Koudelka, G.B., Harbury, P., Harrison, S.C. and Ptashne, M. (1988) DNA twisting and the affinity of bacteriophage 434 operator for bacteriophage 434 repressor. *Proc Natl Acad Sci U S A*, **85**, 4633-7.
- Koudelka, G.B., Harrison, S.C. and Ptashne, M. (1987) Effect of the non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, **326**, 886-888.
- Koudelka, G.B. and Lam, C.-Y. (1993) Differential recognition of OR1 and OR3 by bacteriophage 434 repressor and Cro. *J. Biol. Chem.*, **268**, 23812-23817.
- Kumar, A., Casas-Finet, J. R., Luneau, C. J., Karpel, R. L., Merrill, B. M., Williams, K. R. & Wilson, S. H. (1990). Mammalian heterogeneous nuclear ribonucleoprotein A1. Nucleic acid binding properties of the COOH-terminal domain. *J Biol Chem* **265**(28), 17094-100.
- Ladbury, J.E., Wright, J.G., Sturtevant, J.M. and Sigler, P.B. (1994) A thermodynamic study of the trp repressor-operator interaction. *J Mol Biol*, **238**, 669-81.
- Lakowicz, J.R. *Principles of fluorescence spectroscopy*. Plenum Press, New York, USA.
- Lefstin, J.A. and R., Y.K. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature*, **392**, 885-8.
- Lehrer, S. S. (1997). Intramolecular pyrene excimer fluorescence: a probe of proximity and protein conformational change. *Methods Enzymol* **278**, 286-95.
- Lischwe, M.A., Cook, R.G., Ahn, Y.S., Yeoman, L.C. and Busch, H. (1985) Clustering of glycine and NG,NG-dimethylarginine in nucleolar protein C23. *Biochemistry*, **24**, 6025-8.
- Liu, Q. and Dreyfuss, G. (1995) In vivo and in vitro arginine methylation of RNA-binding proteins. *Mol Cell Biol*, **15**, 2800-8.

- London, F. (1930) Über die Eigenschaften und Anwendungen der Molekularkräfte. *Z. Phys. Chem Abt.*, **11**, 222-51.
- Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R. and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA [see comments]. *Nature*, **352**, 497-505.
- Lumb, K.J., Carr, C.M. and Kim, P.S. (1994) *Biochemistry*, **33**, 7361-7.
- Mattaj, I.W. and Nagai, K. (1996) *RNA-protein interactions*. Oxford University Press, New York.
- Matthew, J.B. and Ohlendorf, D.H. (1985) Electrostatic deformation of DNA by a DNA-binding protein. *J Biol Chem*, **260**, 5860-2.
- Maxwell, E. S. & Fournier, M. J. (1995). The small nucleolar RNAs. *Annu Rev Biochem* **64**, 897-934.
- Mejerhans, D.e.a. (1995) DNA binding specificity of the basic-helix-loop-helix protein MASH-1. *Biochemistry*, **34**, 11026-36.
- Metallo, S.J. and Schepartz, A. (1997) Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA [letter]. *Nat Struct Biol*, **4**, 115-7.
- Michael, W. M., Choi, M. & Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature- dependent nuclear protein export pathway. *Cell* **83**(3), 415-22.
- Miranda, G.A., Chokler, I. and Aguilera, R.J. (1995) The murine nucleolin protein is an inducible DNA and ATP binding protein which is readily detected in nuclear extracts of lipopolysaccharide-treated splenocytes. *Exp Cell Res*, **217**, 294-308.
- Mogridge, J., Legault, P., Li, J., Van Oene, M.D., Kay, L.E. and Greenblatt, J. (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol Cell*, **1**, 265-75.
- Mondragon, A. and Harrison, S.C. (1991) The phage 434 Cro/OR1 complex at 2.5 Å resolution. *J Mol Biol*, **219**, 321-34.

- Mondragon, A., Subbiah, S., Almo, S.C., Drottar, M. and Harrison, S.C. (1989a) Structure of the amino-terminal domain of phage 434 repressor at 2.0 Å resolution. *J Mol Biol*, **205**, 189-200.
- Mondragon, A., Wolberger, C. and Harrison, S.C. (1989b) Structure of phage 434 Cro protein at 2.35 Å resolution. *J Mol Biol*, **205**, 179-88.
- Municio, M. M., Lozano, J., Sanchez, P., Moscat, J. & Diaz-Meco, M. T. (1995). Identification of heterogeneous ribonucleoprotein A1 as a novel substrate for protein kinase C zeta. *J Biol Chem* **270**(26), 15884-91.
- Murphy IV, F.V. and Churchill, M.E. (2000) Non-sequence specific DNA-recognition: a structural perspective. *Structure*, **8**, R83-9.
- Nadassy, K., Wodak, S.J. and Janin, J. (1999) Structural features of protein-nucleic acid recognition sites. *Biochemistry*, **38**, 1999-2017.
- Najbauer, J., Johnson, B.A., Young, A.L. and Aswad, D.W. (1993) Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins. *J Biol Chem*, **268**, 10501-9.
- Narcisi, E. M., Glover, C. V. & Fechheimer, M. (1998). Fibrillarin, a conserved pre-ribosomal RNA processing protein of Giardia. *J Eukaryot Microbiol* **45**(1), 105-11.
- Nardone, G., George, J. and Chirikjian, J.G. (1986) Differences in the kinetic properties of BamHI endonuclease and methylase with linear DNA substrates. *J Biol Chem*, **261**, 12128-33.
- Neri, D., Billeter, M. and Wuthrich, K. (1992) Determination of the nuclear magnetic resonance solution structure of the DNA-binding domain (residues 1 to 69) of the 434 repressor and comparison with the X-ray crystal structure. *J Mol Biol*, **223**, 743-67.
- Ohlendorf, D.H. and Matthew, J.B. (1985) Electrostatics and flexibility in protein-DNA interactions. *Adv Biophys*, **20**, 137-51.

- Olson, M.O., Kirstein, M.N. and Wallace, M.O. (1990) Limited proteolysis as a probe of the conformation and nucleic acid binding regions of nucleolin. *Biochemistry*, **29**, 5682-6.
- Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A*, **95**, 11163-8.
- Otwinowski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F. and Sigler, P.B. (1988) Crystal structure of trp repressor/operator complex at atomic resolution [published erratum appears in Nature 1988 Oct 27;335(6193):837]. *Nature*, **335**, 321-9.
- Pabo, C.O. and Sauer, R.T. (1984) Protein-DNA recognition. *Annu. Rev. Biochem.*, **53**, 293-321.
- Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, **61**, 1053-95.
- Paik, W.K., Farooqui, J.Z., Roy, T and Kim, S. (1983) Determination of pI values of variously methylated amino acids by isoelectric focusing. *J. Chromatogr.*, **256**, 331-334.
- Patel, D.J. (1999) Adaptive recognition in RNA complexes with peptides and protein modules. *Curr Opin Struct Biol*, **9**, 74-87.
- Patel, L., Abate, C. and Curran, T. (1990) Altered protein conformation on DNA binding by Fos and Jun. *Nature*, **347**, 572-5.
- Percipalle, P., Simoncsits, A., Zakhariyev, S., Guarnaccia, C., Sanchez, R. and Pongor, S. (1995) Rationally designed helix-turn-helix proteins and their conformational changes upon DNA binding. *Embo J*, **14**, 3200-5.
- Petersen, J.M.e.a. (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an alpha helix. *Science*, **269**, 1866-9.
- Plougastel, B., Zucman, J., Peter, M., Thomas, G. & Delattre, O. (1993). Genomic structure of the EWS gene and its relationship to EWSR1, a site of tumor-associated chromosome translocation. *Genomics* **18**(3), 609-15.

- Pontius, B. W. & Berg, P. (1992). Rapid assembly and disassembly of complementary DNA strands through an equilibrium intermediate state mediated by A1 hnRNP protein. *J Biol Chem* **267**(20), 13815-8.
- Preuss, R., Dapprich, J. and Walter, N.G. (1997) Probing RNA-protein interactions using pyrene-labeled oligodeoxynucleotides: Qbeta replicase efficiently binds small RNAs by recognizing pyrimidine residues. *J Mol Biol*, **273**, 600-13.
- Provencer, S.W. (1984) *EMBL technical report, DA07*.
- Ptashne, M. (1992) *A Genetic Switch*. Cell Press and Blackwell Scientific Publications, Cambridge, MA.
- Ptashne, M. and Gann, A. (1998) Imposing specificity by localization: mechanism and evolvability. *Curr Biol*, **8**, R812-R822.
- Query, C.C., Bentley, R.C. and Keene, J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell*, **57**, 89-101.
- Rajpurohit, R., Paik, W.K. and Kim, S. (1994) Effect of enzymic methylation of heterogeneous ribonucleoprotein particle A1 on its nucleic-acid binding and controlled proteolysis. *Biochem J*, **304**, 903-9.
- Rastinejad, F., Perlmann, T., Evans, R.M. and Sigler, P.B. (1995) Structural determinants of nuclear receptor assembly on DNA direct repeats [see comments]. *Nature*, **375**, 203-11.
- Rentzeperis, D., Jonsson, T. and Sauer, R.T. (1999) Acceleration of the refolding of Arc repressor by nucleic acids and other polyanions. *Nat Struct Biol*, **6**, 569-73.
- Rhodes, D., Schwabe, J.W., Chapman, L. and Fairall, L. (1996) Towards an understanding of protein-DNA recognition. *Philos Trans R Soc Lond B Biol Sci*, **351**, 501-9.
- Rodgers, D.W. and Harrison, S.C. (1993) The complex between phage 434 repressor DNA-binding domain and operator site OR3: structural differences between consensus and non-consensus half-sites. *Structure*, **1**, 227-240.

- Rould, M.A., Perona, J.J., Soll, D. and Steitz, T.A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution [see comments]. *Science*, **246**, 1135-42.
- Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**(5013), 1682-9.
- Ruiz-Sanz, J., Simoncsits, A., Toro, I., Pongor, S., Mateo, P.L. and Filimonov, V.V. (1999) A thermodynamic study of the 434-repressor N-terminal domain and of its covalently linked dimers. *Eur J Biochem*, **263**, 246-53.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo.
- Sampson, W.R., Patsiouras, H. and Ede, N.J. (1999) The synthesis of 'difficult' peptides using 2-hydroxy-4-methoxybenzyl or pseudoproline amino acid building blocks: a comparative study. *J Pept Sci*, **5**, 403-9.
- Sapp, M., Richter, A., Weisshart, K., Caizergues-Ferrer, M., Amalric, F., Wallace, M.O., Kirstein, M.N. and Olson, M.O. (1989) Characterization of a 48-kDa nucleic-acid-binding fragment of nucleolin. *Eur J Biochem*, **179**, 541-8.
- Schmid, F.X. (1989) In Creighton, T.E. (ed.) *Protein Structure*. IRL Press, Oxford, New York, Tokyo, pp. 251-284.
- Schulz, G.E. and Schirmer, R.H. (1979) *Principles of Protein Structure*.
- Schwabe, J.W., Chapman, L., Finch, J.T. and Rhodes, D. (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell*, **75**, 567-78.
- Schwabe, W.R. (1997) The role of water in protein-DNA interactions. *Curr Opin Struct Biol*, **7**, 126-34.
- Seeman, N.C., Rosenberg, M.J. and Rich, A. (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA*, **73**, 804-808.

- Serin, G., Joseph, G., Ghisolfi, L., Bauzan, M., Erard, M., Amalric, F. and Bouvet, P. (1997) Two RNA-binding domains determine the RNA-binding specificity of nucleolin. *J Biol Chem*, **272**, 13109-16.
- Shakkeed, Z., Guzikevich-Guerstein, G., Frolow, F., Rabinovich, D., Joachimiak, A. and Sigler, P.B. (1994) Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature*, **368**, 469-73.
- Sheikh, M. S., Carrier, F., Papathanasiou, M. A., Hollander, M. C., Zhan, Q., Yu, K. & Fornace, A. J., Jr. (1997). Identification of several human homologs of hamster DNA damage- inducible transcripts. Cloning and characterization of a novel UV- inducible cDNA that codes for a putative RNA-binding protein. *J Biol Chem* **272**(42), 26720-6.
- Shen, E.C., Henry, M.F., Weiss, V.H., Valentini, S.R., Silver, P.A. and Lee, M.S. (1998) Arginine methylation facilitates the nuclear export of hnRNP proteins. *Genes Dev*, **12**, 679-91.
- Shimon, L.J. and Harrison, S.C. (1993) The phage 434 OR2/R1-69 complex at 2.5 Å resolution. *J.Mol.Biol.*, **232**, 826-838.
- Shimoni, L. and Glusker, J.P. (1995) Hydrogen bonding motifs of protein side chains: descriptions of binding of arginine and amide groups. *Protein Sci*, **4**, 65-74.
- Sidorova, N.Y. and Rau, D.C. (1996) Differences in water release for the binding of Eco RI to specific and nonspecific DNA sequences. *Proc. Natl Acad. Sci.*, **93**, 12272-7.
- Siebel, C. W. & Guthrie, C. (1996). The essential yeast RNA binding protein Np13p is methylated. *Proc Natl Acad Sci U S A* **93**(24), 13641-6.
- Simoncsits, A., Chen, J., Percipalle, P., Wang, S., Toro, I. and Pongor, S. (1997) Single-chain repressors containing engineered DNA-binding domains of the phage 434 repressor recognize symmetric or asymmetric DNA operators. *J Mol Biol*, **267**, 118-31.
- Sinden, R.R. (1994) *DNA Structure and Function*. Academic Press Inc., San Diego, CA, USA.

- Smith, D. and Griffin, J.F. (1978) Conformation of [Leu⁵]enkephalin from X-ray diffraction: features important for recognition at opiate receptor. *Science*, **199**, 1214-6.
- Smith, J.J., Rucknagel, K.P., Schierhorn, A., Tang, J., Nemeth, A., Linder, M., Herschman, H.R. and Wahle, E. (1999) Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3. *J Biol Chem*, **274**, 13229-34.
- Spolar, R.S. and Record, M.T. (1994) Coupling of local folding to site specific binding of proteins to DNA. *Science*, **263**, 777-84.
- Standke, K.C. and Brunnert, H. (1975) The estimation of affinity constants for the binding of model peptides to DNA by equilibrium dialysis. *Nucleic Acids Res*, **2**, 1839-49.
- Steely, H.T., Jr., Gray, D.M., Lang, D. and Maestre, M.F. (1986) Circular dichroism of double-stranded RNA in the presence of salt and ethanol. *Biopolymers*, **25**, 91-117.
- Suzuki, M. (1993) Common features in DNA recognition helices of eukaryotic transcription factors [published erratum appears in EMBO J 1993 Oct;12(10):4042]. *Embo J*, **12**, 3221-6.
- Suzuki, M. (1994) A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure*, **2**, 317-26.
- Suzuki, M., Brenner, S.E., Gerstein, M. and Yagi, N. (1995a) DNA recognition code of transcription factors. *Protein Eng*, **8**, 319-28.
- Suzuki, M. and Yagi, N. (1994) DNA recognition code of transcription factors in the helix-turn-helix, probe helix, hormone receptor, and zinc finger families. *Proc Natl Acad Sci U S A*, **91**, 12357-61.
- Suzuki, M. and Yagi, N. (1995) Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res*, **23**, 2083-91.

- Suzuki, M., Yagi, N. and Gerstein, M. (1995b) DNA recognition and superstructure formation by helix-turn-helix proteins. *Protein Eng*, **8**, 329-38.
- Szekely, Z., Zakhariiev, S., Guarnaccia, C., Antcheva, N., Pongor, S. (1999) A highly effective method for synthesis of N^ω-substituted arginines as building blocks for Boc/Fmoc peptide chemistry. *Tetrahedron Lett.*, **40**, 4439-4442.
- Takeda, Y., Kim, J.G., Caday, C.G., Steers, E., Jr., Ohlendorf, D.H., Anderson, W.F. and Matthews, B.W. (1986) Different interactions used by Cro repressor in specific and nonspecific DNA binding. *J Biol Chem*, **261**, 8608-16.
- Takeda, Y., Ross, P.D. and Mudd, C.P. (1992) Thermodynamics of Cro protein-DNA interactions. *Proc Natl Acad Sci U S A*, **89**, 8180-4.
- Talanian, R.V., McKnight, C.J. and Kim, P.S. (1990) Sequence-specific DNA binding by a short peptide dimer. *Science*, **249**, 769-71.
- Tang, J., Gary, J.D., Clarke, S. and Herschman, H.R. (1998) PRMT 3, a type I protein arginine N-methyltransferase that differs from PRMT1 in its oligomerization, subcellular localization, substrate specificity, and regulation. *J Biol Chem*, **273**, 16935-45
- Thompson, K.S., Vinson, C.R. and Freire, E. (1993) *Biochemistry*, **32**, 5491-6.
- Tijan, R. and Maniatis, T. (1994) Transcriptional Activation: A Complex Puzzle with Few Easy Pieces. *Cell*, **77**, 5-8.
- Travers, A.A. (1989) DNA conformation and protein binding. *Annu Rev Biochem*, **58**, 427-52.
- Travers, A.A. and Klug, A. (1990) *Bending of DNA in nucleoprotein complexes*. Cold Spring Harbor Laboratory Press.
- Tuteja, N., Rahman, K., Tuteja, R. & Falaschi, A. (1991). DNA helicase IV from HeLa cells. *Nucleic Acids Res* **19**(13), 3613-8.
- Tuteja, N., Huang, N. W., Skopac, D., Tuteja, R., Hrvatic, S., Zhang, J., Pongor, S., Joseph, G., Faucher, C., Amalric, F. & et al. (1995). Human DNA helicase IV is nucleolin, an RNA helicase modulated by phosphorylation. *Gene* **160**(2), 143-8.

- Valentini, S.R., Weiss, V.H. and Silver, P.A. (1999) Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3'-end formation. *Rna*, **5**, 272-80.
- Van Gilst, M. R., Rees, W. A., Das, A. & von Hippel, P. H. (1997). Complexes of N antitermination protein of phage lambda with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry* **36**(6), 1514-24.
- Viadiu, H. and Aggarwal, A.K. (2000) Structure of BamHI bound to nonspecific DNA: a model for DNA sliding. *Mol Cell*, **5**, 889-95.
- von Hippel, P.H. and Berg, O.G. (1986) On the specificity of DNA-protein interactions. *Proc Natl Acad Sci U S A*, **83**, 1608-12.
- Warrener, P. & Petryshyn, R. (1991). Phosphorylation and proteolytic degradation of nucleolin from 3T3-F442A cells. *Biochem Biophys Res Commun* **180**(2), 716-23.
- Webster, C., Merryweather, A. and Brammar, W. (1992) Efficient repression by a heterodimeric repressor in *Escherichia coli*. *Mol. Microbiol.*, **6**, 371-377.
- Weiss, M.A.e.a. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature*, **347**, 575-8.
- Wendt, H., Thomas, R.M. and Ellenberger, T. (1998) DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem*, **273**, 5735-43.
- Wharton, R.P., Brown, E.L. and Ptashne, M. (1984) Substituting an alpha-helix switches the sequence-specific DNA interactions of a repressor. *Cell*, **38**, 361-9.
- Wharton, R.P. and Ptashne, M. (1985) Changing the binding specificity of a repressor by redesigning an alpha-helix. *Nature*, **316**, 601-5.
- Williamson, J.R. (2000) Induced-fit in RNA-protein recognition. *Nat Struct Biol*, **7**, 834-7.
- Wintjens, R. and Rooman, M. (1996) Structural Classification of HTH DNA-binding Domains and Protein-DNA Interaction Modes. *J Mol Biol*, **262**, 294-313.
- Wittung, P., Norden, B., Kim, S.K. and Takahashi, M. (1994) Interactions between DNA molecules bound to RecA filament. Effects of base complementarity. *J Biol Chem*, **269**, 5799-803.

- Wolberger, C. (1996) Homeodomain interaction. *Curr. Opin. Struct. Biol.*, **6**, 62-68.
- Wong, I. and Lohman, T.M. (1993) A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci U S A*, **90**, 5428-32.
- Wu, C.S., Ikeda, K. and Jang, J.T. (1981) *Biochemistry*, **20**, 566-70.
- Wu, X., Spiro, C., Owen, W.G. and McMurray, C.T. (1998) cAMP response element-binding protein monomers cooperatively assemble to form dimers on DNA. *J Biol Chem*, **273**, 20820-7.
- Xue, Z., Shan, X., Lapeyre, B. and Melese, T. (1993) The amino terminus of mammalian nucleolin specifically recognizes SV40 T-antigen type nuclear localization sequences. *Eur J Cell Biol*, **62**, 13-21.
- Yang, J. T., Wu, C. S. & Martinez, H. M. (1986). Calculation of protein conformation from circular dichroism. *Methods Enzymol* **130**, 208-69.
- Yang, T.H., Tsai, W.H., Lee, Y.M., Lei, H.Y., Lai, M.Y., Chen, D.S., Yeh, N.H. and Lee, S.C. (1994) Purification and characterization of nucleolin and its identification as a transcription repressor. *Mol Cell Biol*, **14**, 6068-74.
- Yang, W. and Steitz, T.A. (1995) Crystal structure of the site-specific recombinase gamma delta resolvase complexed with a 34 bp cleavage site. *Cell*, **82**, 193-207.
- Zheng, N. and Gierasch, L.M. (1997) Domain interactions in E. coli SRP: stabilization of M domain by RNA is required for effective signal sequence modulation of NG domain. *Mol Cell*, **1**, 79-87.

Published papers related to the thesis project:

Raman B, Guarnaccia C, Nadassy K, Zakhariev S, Pintar A, Zanuttin F, Frigyes D, Acatrinei C, Vindigni A, Pongor G, Pongor S. (2001) N(omega)-arginine dimethylation modulates the interaction between a Gly/Arg-rich peptide from human nucleolin and nucleic acids. *Nucleic Acids Res*, **29**, 3377-84.

Cimato TR, Tang J, Xu Y, Guarnaccia C, Herschman H, Pongor S, Aletta JM (2001) NGF-Mediated Increases in Protein Methylation Occur Predominantly at Type I Arginine Methylation Sites and Involve Protein Arginine Methyltransferase 1 (PRMT1) *J Neurosci Res*, in press

Szekely Z., Zakhariev S., Guarnaccia C., Antcheva N., Pongor S. (1999) A highly effective method for synthesis of N(omega)-substituted arginines as building blocks for Boc/Fmoc peptide chemistry. *Tetrahedron Letters*, **40**, 4439-4442

Percipalle P., Pongor S., Zakhariev S., Guarnaccia C., Saletti R., Foti S., Fisichella S. (1997) Synthesis and mass spectrometric characterization of the N-terminal (1-63) DNA-binding domain of the bacteriophage 434 repressor cI. *European Mass Spectrometry*, **3**, 151-159

Zahariev S., Szekely Z., Guarnaccia C., Antcheva N., Pongor S. (1999) A highly effective method for synthesis of N(omega)-substituted arginine derivatives In Peptides for the New Millenium, in *Proceedings of the 16th American peptide Symposium, Minneapolis, Minnesota, Kluwer Academic Publishers*, 74-75

Guarnaccia C., Zahariev S., Toro I., Simoncsits A., Pongor S. (1998)

Synthesis by chemical Ligation of Homo- and Heterodimeric polypeptides Based on the bZIP and HTH DNA-binding Motifs, in *Peptides 1996, Proceedings of the 24th European peptide Symposium*, 441-442

Percipalle P., Simoncsits A., Zakhariiev S., Guarnaccia C., Sanchez R., Pongor S. (1995)

Rationally designed helix-turn-helix proteins and their conformational changes upon DNA binding. *EMBO J.* ,14, 3200-5.

Manuscripts submitted or in preparation:

Guarnaccia C., Raman B. (joint first autorship), Zakhariiev S., Simoncsits A., Pongor S.

Dimerization of the DNA-binding domain of bacteriophage 434 repressor on DNA.

Manuscript in preparation