

# **Structural studies of regulators of protein expression**

Pablo Gutiérrez

Department of Biochemistry

McGill University, Montreal

January 2005

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Ph.D.

© Pablo Gutiérrez, 2005



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-21649-1*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-21649-1*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

*A vos que me salvaste tantas veces*

## Abstract

In most species, highly sophisticated global regulatory networks regulate the expression of genes in response to environmental and physiological demands. The mechanisms devised control virtually every event involved in transcription and translation, as well as influencing mRNA degradation, protein stability, protein localization, protein-protein interactions, and protein function. In order to understand the general mechanism used by several organisms to control gene expression, three regulatory proteins were investigated using nuclear magnetic resonance (NMR) techniques. The systems studied were proteins involved in the control of gene expression at the transcriptional (YaeO, Chapter 2), post-transcriptional (CsrA, Chapter 3) and translational levels (aIF2 $\beta$  Chapter 4).

YaeO is a protein involved in the regulation of Rho-dependent transcription termination in bacteria. The solution structure of YaeO was solved, the first for a Rho inhibitor, and it was demonstrated that that YaeO binds to the primary RNA binding domain of Rho. This study suggests that YaeO inhibits transcription termination by blocking the primary RNA binding site in Rho.

The carbon storage regulator A from *E. coli* is a protein involved in the post-transcriptional control of numerous genes involved in carbohydrate metabolism in bacteria. It has been shown that CsrA prevents translation of its target mRNA by blocking access to the ribosome binding site. The binding studies, together with the structure of CsrA, reveal the potential RNA binding region. A model for CsrA recognition of its target mRNAs is also proposed.

The last section focuses on the archaeal translation initiation factor IF2 $\beta$ . This protein is a key regulator of overall protein synthesis. The structure of aIF2 $\beta$  was solved and bound zinc was proved to be necessary for the structural integrity of this translation factor. The function of aIF2 $\beta$  in translation initiation was rationalized in terms of its structure and available structural data for translation factors aIF2 $\alpha$  and aIF2 $\gamma$ .

## Résumé

L'expression de gènes est régulé par de mécanismes très sophistiqués en réponse aux demandes environnementales et physiologiques. Ces mécanismes contrôlent pratiquement chaque évènement impliqué dans la transcription et la traduction génétique. Afin de comprendre le mécanisme général utilisé par plusieurs systèmes lors du contrôle de l'expression génétique, trois protéines furent analysées en utilisant des techniques de spectroscopie RMN. Les systèmes étudiés étaient des protéines impliquées dans le contrôle de l'expression génétique aux niveaux transcriptionnel (YaeO, 2ième chapitre), post-transcriptionnel (CsrA, 3ième chapitre) et traductionnel (aIF2 $\beta$ , 4ième chapitre).

YaeO est impliquée dans la régulation de la transcription Rho-dépendante. La structure du YaeO fut déterminée, et il fut démontré que YaeO s'attache au domaine attachant l'RNA de Rho. Cette étude suggère donc que YaeO prévient la transcription en bloquant les sites d'union de l'RNA.

CsrA est impliquée dans le contrôle posttranscriptionnel de plusieurs gènes dans le métabolisme de carbohydrates de bactéries. Il fut démontré que CsrA prévient la traduction en bloquant l'accès aux sites de greffe de ribosome. Les études sur le processus de greffe, avec le structure de CsrA, suggèrent un modèle pour la reconnaissance des ARNm cibles.

La dernière section se concentre sur la facteur d'initiation aIF2 $\beta$ . Cette protéine est essentielle pour la synthèse de protéines. La structure de aIF2 $\beta$  fut déterminée et il prouvé que le zinc est nécessaire pour assurer l'intégrité structurelle du cette facteur de traduction. La fonction de aIF2 $\beta$  fut rationalisée en terme de la structure et la information disponible pour le facteurs aIF2 $\alpha$  et aIF2 $\gamma$ .

## Contributions of Authors

A version of Chapter 2, "Solution Structure of YaeO: an inhibitor of Transcription factor Rho" will be submitted for publication. Demetra Elias cloned the Rho130 fragment used in the NMR titrations. The rest of the work described in this chapter was carried out by myself.

Chapter 3 is a modified version of a manuscript accepted for publication in the Journal of Bacteriology. Yan Li performed the diffusion and HPLC experiments. Ekaterina Pomerantseva measured the residual dipolar couplings. Dr. Michael J. Osborne implemented the pulse sequences used for the structure determination. I purified CsrA and determined the initial conditions for determining the solution structure. I also performed the NMR titrations, assignment of NMR signals and calculated the structures.

Chapter 4 is a modified version of the following publications:

- Gutiérrez P, Coillet-Matillon S, Arrowsmith C, Gehring, K. 2002. Zinc is required for structural stability of the C-terminus of archaeal translation initiation factor aIF2 $\beta$ . FEBS Letters, 517:155-158
- Gutiérrez P, Osbourne M, Siddiqui N, Trempe JF, Arrowsmith C, Gehring K. 2004. Structure of translation initiation factor aIF2 $\beta$  from *Methanobacterium thermoautotrophicum*: Implications for translation initiation. Prot Sci 13:659-667.

All of the NMR experiments, analysis and structure calculations described in the manuscript were carried out by myself. I also purified aIF2 $\beta$  and prepared the samples for NMR studies. Dr. Cheryl Arrowsmith provided us with the aIF2 $\beta$  clone. Stéphane Coillet-Matillon collaborated in the initial stages of the characterization of aIF2 $\beta$ . Jean-François Trempe prepared the filamentous phage used in residual dipolar coupling experiments. Dr. Michael J. Osborne provided

assistance for most of the NMR experiments. Nadeem Siddiqui assisted me with varied molecular biology techniques.

## Table of contents

Abstract.....	i
Résumé.....	ii
Contributions of Authors .....	iii
Table of contents.....	v
List of Tables and Figures.....	viii
Acknowledgements.....	x
Abbreviations.....	xi
Chapter 1: Introduction.....	1
1.1 Transcriptional control.....	3
1.1.1 Transcriptional regulation in prokaryotes.....	5
1.1.2 Transcription regulation in eukaryotes .....	15
1.1.3 Transcriptional regulation in archaea.....	20
1.2 Some mechanisms of post-transcriptional control.....	21
1.2.1 Regulatory RNAs.....	21
1.2.2 Capping and polyadenylation.....	24
1.2.3 Splicing .....	25
1.2.4 Regulation of RNA Longevity.....	26
1.2.5 RNA Editing .....	28
1.3 Regulation at the translational level.....	28
1.3.1 Overview of prokaryotic translation .....	29
1.3.2 Eukaryotic translation .....	31
1.3.3 Archaeal translation .....	33
1.4 Scope of the thesis. ....	36
Chapter 2: Structure of YaeO, a Rho-specific inhibitor of transcription termination.....	37
2.1 Abstract.....	37

2.2 Introduction.....	37
2.3 Results and discussion .....	41
2.3.1 Structure determination of YaeO .....	41
2.3.2 Comparative analysis of YaeO .....	42
2.3.3 Interactions between YaeO and Rho.....	49
2.3.5 Discussion.....	54
2.4 Materials and methods.....	56
2.4.1 Sample preparation .....	56
2.4.2 NMR spectroscopy.....	56
2.4.3 Analysis and structure calculations.....	57
2.4.4 Ligand titration.....	57
2.4.5 $\beta$ -galactosidase assay .....	57
2.4.6 Docking.....	58
Chapter 3: Solution structure of the carbon storage regulator protein CsrA from <i>E. coli</i> .....	59
3.1 Abstract.....	59
3.2 Introduction.....	59
3.3 Results and discussion .....	64
3.3.1 Structure determination of the CsrA dimer.....	64
3.3.2 RNA binding of CsrA .....	73
3.3.3 Discussion .....	74
3.4 Materials and methods.....	78
3.4.1 Sample preparation .....	78
3.4.2 Gel filtration.....	78
3.4.3 NMR spectroscopy.....	78
3.4.4 Structure calculations.....	79
3.4.5 NMR titrations .....	80
Chapter 4: Structural studies on archaeal translation initiation factor aIF2 $\beta$ .....	81
4.1 Abstract.....	81
4.2 Introduction.....	81

4.3 Results.....	83
4.3.1 Zinc is required for structural stability of the C-terminus of aIF2 $\beta$ .....	83
4.3.2 Metal binding properties of aIF2 $\beta$ .....	86
4.3.3 Solution structure of aIF2 $\beta$ .....	89
4.3.4. Relaxation and dynamics .....	96
4.3.5 Comparative analysis of the Mt_aIF2 $\beta$ structure.....	96
4.3.6 Discussion .....	97
4.4 Materials and methods .....	106
4.4.1 Protein expression and purification .....	106
4.4.2 Mass spectrometry .....	106
4.4.3 Cobalt visible spectrophotometry .....	106
4.4.4 NMR spectroscopy.....	107
4.4.5 Analysis and structure calculations.....	108
Chapter 5: Structural genomics of gene regulation.....	109
5.1 YaeO .....	109
5.2. CsrA .....	110
5.3 aIF2 $\beta$ .....	111
References.....	113

## List of Tables and Figures

Figure 1.1 Information transfer in prokaryotes and eukaryotes.....	2
Figure 1.2 Overall view of prokaryotic transcription .....	4
Figure 1.3 Transcriptional control of the <i>lac</i> operon .....	7
Figure 1.4 Regulation of the <i>trp</i> operon in <i>E. coli</i> .....	9
Figure 1.5 Rho-dependent transcription termination.....	13
Figure 1.6 Eukaryotic gene structure.....	18
Figure 1.7 Models of sRNA function .....	23
Figure 1.8 Effects of alternative splicing on gene expression .....	27
Figure 1.9 Mechanism of initiation of protein synthesis in eukaryotes.....	35
Figure 2.1 Chromosomal region around the <i>yaeO</i> gene .....	39
Figure 2.2 Effect of <i>yaeO</i> expression on <i>rho</i> gene transcription.....	40
Figure 2.3 <sup>15</sup> N-HSQC and secondary chemical shifts of YaeO.....	43
Figure 2.4 Sequence conservation of the ROF family of proteins.....	44
Figure 2.5 NMR ensemble of YaeO .....	45
Figure 2.6 Charge distribution and sequence conservation of YaeO.....	47
Figure 2.7 Proteins structurally related to YaeO .....	48
Figure 2.8 NMR chemical shift changes upon YaeO-Rho130 binding.....	50
Figure 2.9 YaeO and (dC) <sub>9</sub> binding to Rho130 .....	51
Figure 2.10 Docking model of the YaeO-Rho130 complex.....	53
Figure 2.11 Proposed mechanism of inhibition by YaeO.....	55
Figure 3.1 Effects of CsrA on carbohydrate metabolism .....	61
Figure 3.2 Model of CsrA-mediated regulation of <i>E. coli glgC</i> .....	62
Figure 3.3 Sequence conservation of CsrA.....	63
Figure 3.4 CsrA is dimer in solution.....	65
Table 3.1 Diffusion coefficient (Ds) values at 298 K.....	66
Figure 3.5 <sup>15</sup> N-HSQC of CsrA.....	68
Figure 3.6 Intermolecular interactions in CsrA .....	69
Figure 3.7 NMR ensemble of CsrA.....	70
Figure 3.8 Structure of CsrA.....	72
Figure 3.9 Surface properties of CsrA .....	76

Figure 3.10 Proposed binding of CsrA to target mRNAs.....	77
Figure 4.1 Proteolytic sensitivity of aIF2 $\beta$ in the absence of Zn <sup>+2</sup> .....	84
Figure 4.2 ESI-MS mass spectrum of aIF2 $\beta$ .....	85
Figure 4.3 Cobalt coordination by aIF2 $\beta$ .....	87
Figure 4.4 Zn <sup>+2</sup> binding to the C-terminus of aIF2 $\beta$ .....	88
Figure 4.5 Sequence alignment of proteins related to archaeal IF2 $\beta$ .....	90
Figure 4.6 <sup>1</sup> H- <sup>15</sup> N heteronuclear NOE, NOE constraints and RMSD statistics for Mt_aIF2 $\beta$ .....	93
Figure 4.7 NMR ensemble of Mt_aIF2 $\beta$ .....	94
Figure 4.8 Structure of Mt_aIF2 $\beta$ .....	95
Table 4.2 Protein structures similar to aIF2 $\beta$ .....	99
Figure 4.9 Proteins structurally related to Mt_aIF2 $\beta$ .....	100
Figure 4.10 Electrostatic surface plot of Mt_aIF2 $\beta$ .....	101
Figure 4.11 Mapping of biochemical data to the Mt_aIF2 $\beta$ structure.....	104
Figure 4.12 Proposed function of aIF2 $\beta$ in translation initiation.....	105

## **Acknowledgements**

First, I would like to acknowledge my supervisor, Dr. Kalle Gehring, who provided me with all the tools necessary to pursue my Ph.D. research. I would also like to acknowledge the following people, without whose help this work would not have been possible. Tara Sprules and Guennadi Kozlov were essential during my first years in the lab especially for their patience in teaching how to acquire and process NMR data. Dr. Irena Ekiel who was always willing to help me out with NMR experiments. To Marta Bachetti, Michael Osborne and Nadeem Siddiqui for all their invaluable help, stimulating discussions and encouragement. Also Demetra Elias, Alexej Denisov, Jean-François Trempe, Tudor Moldovenau, Yan Li, Ekaterina Pomerantseva, Qian Liu and all the people that helped me out with different aspects of my research. Last, but not least, I thank Margarita, Stephanie and all my relatives and friends whose support was extremely important in keeping my spirit alive.

## Abbreviations

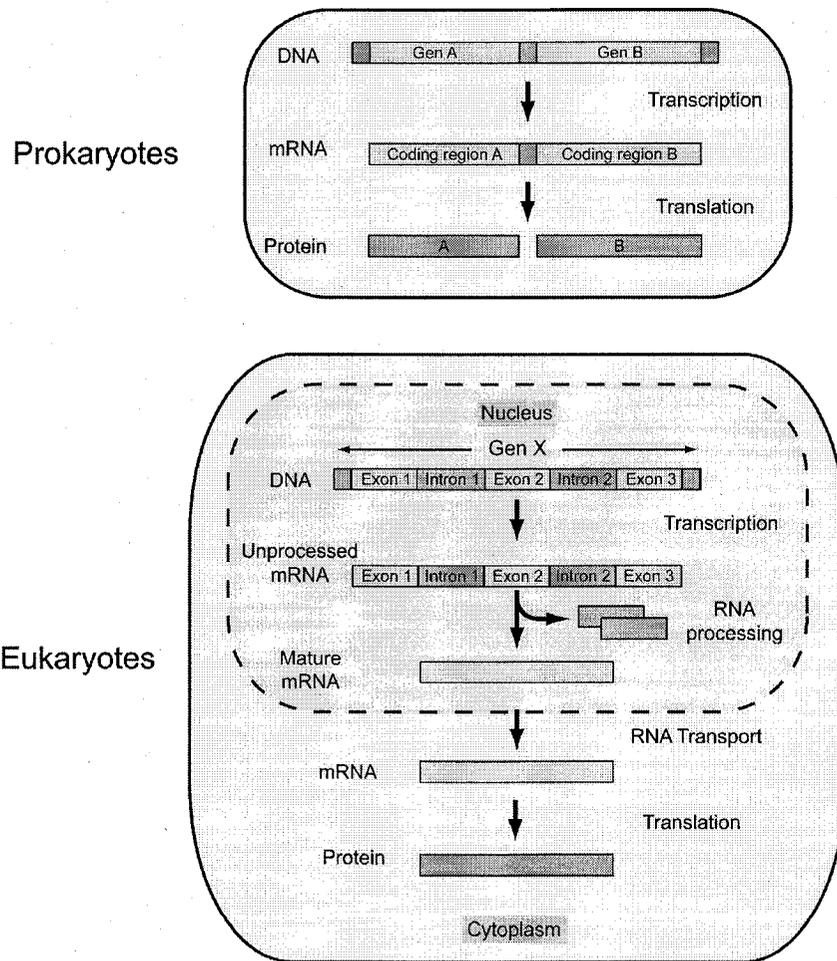
a. a.	Amino acid
ARIA	Ambiguous Restraints in Iterative Assignment
BCM	Bicyclomycin
b.p.	Base pairs
CNS	Crystallography and NMR System
DNA	Deoxyribonucleic acid
DSS	2,2'-dimethyl-2-silapentane-5-sulfonate
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
HMQC	Heteronuclear single quantum coherence
HSQC	Heteronuclear multiple quantum coherence
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser enhancement spectroscopy
nt	Nucleotide
RNAP	RNA polymerase
RDC	Residual dipolar coupling
SD	Shine-Dalgarno
TOCSY	Total correlation spectroscopy
His	Histidine (histidine tag)
Met-tRNA <sub>i</sub>	Methionyl initiator tRNA
mRNA	Messenger RNA
Poly(A)	Polyadenylated
RNA	Ribonucleic acid
RNAP	RNA polymerase
Rut	Rho utilization site
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
snRNA	Small nuclear RNA

TAF	TATA associated factor
TBP	TATA binding protein
tRNA	Transfer RNA
UTR	Untranslated region
UV	Ultraviolet

## Chapter 1: Introduction

In order for a cell to function properly, the levels at which each gene are expressed must be carefully regulated. Generally, proteins needed for cellular proliferation are provided in amounts appropriate for the fastest possible growth in a given medium while other proteins needed for protection, repair or growth under special circumstances are not made until required. In most species, highly sophisticated global regulatory networks modulate the expression of genes in response to environmental and physiological demands (Gottesman, 1984). The first studies on gene regulation began with the work of Jacob and Monod on the regulation of lactose metabolism. Their research led to the celebrated operon model, and introduced concepts such as operon, regulator gene and transcriptional repression, providing a framework that influenced all subsequent thought and research in the field (Jacob and Monod, 1961). This model has been elaborated extensively and today hundreds to thousands of diverse genetic regulatory circuits have been described and characterized experimentally (Wall *et al*, 2004).

Virtually every step from transcription to protein degradation can be exploited by the cell as a target for regulation. In bacteria, genes that encode proteins necessary to perform a coordinated function are normally clustered into operons and their mRNAs are polycistronic. In these organisms, transcription is coupled to the translation machinery producing proteins as soon as the corresponding mRNA starts being synthesized. This explains why a great majority of bacterial genes are regulated at the level of transcription initiation. In eukaryotes, the situation is much more complex as they possess a nuclear membrane, which prevents the simultaneous transcription and translation that occurs in prokaryotes. Furthermore, eukaryotic transcripts must be processed before they can be translated (Figure 1.1).



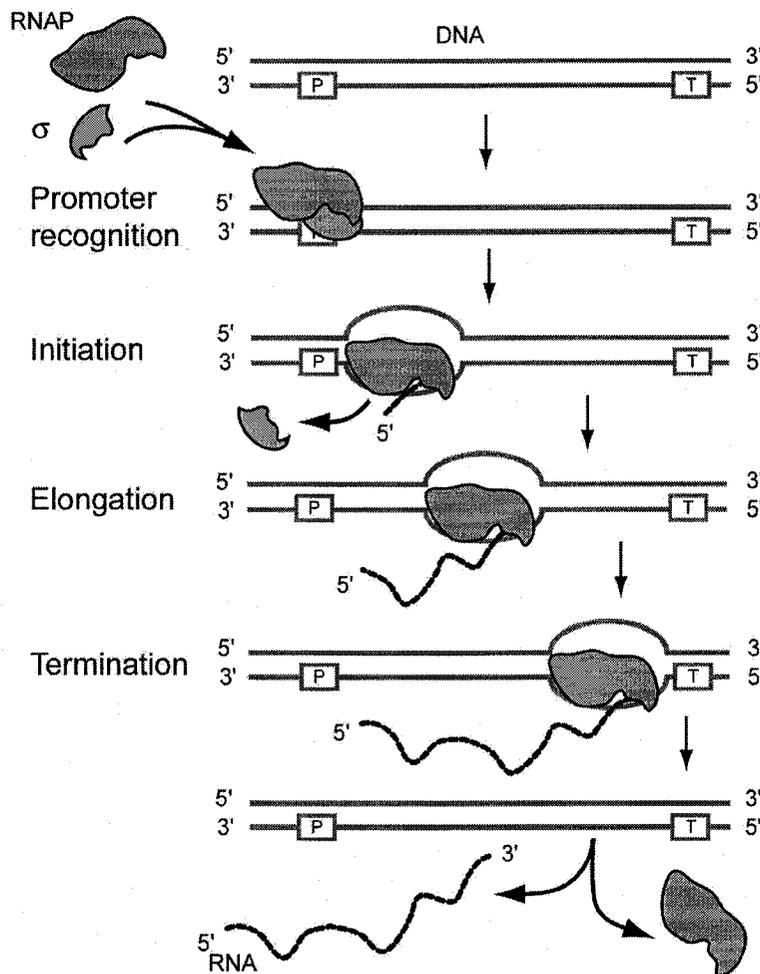
**Figure 1.1 Information transfer in prokaryotes and eukaryotes**

In prokaryotes, a single mRNA molecule contains several protein coding regions and translation is coupled to the transcription process. In eukaryotes, mRNAs code for a single protein and contain noncoding regions or introns that need to be excised before transport to the cytoplasm. Mature mRNA needs to be transported across the nuclear membrane where translation takes place separately from transcription.

In the following sections, I examine some of the main regulatory checkpoints and mechanisms of gene expression in both bacteria and eukaryotes. However, due to the extensive range of strategies for gene control, I will limit this introduction to some transcriptional, post transcriptional and translational regulation systems, as they are the most relevant for the discussion of subsequent chapters.

### **1.1 Transcriptional control**

Transcription is the process where a RNA molecule is synthesized using DNA as a template. It can be divided into three steps: initiation, elongation and termination (Figure 1.2). Initiation is the sequence of events that leads to the assembly of the transcription apparatus on the right position of the DNA to be transcribed. Control of transcript initiation is the most important mechanism for determining whether or not most genes are expressed and the extent to which mRNA produced. Regulation at this stage often involves the prevention of the transcription apparatus to bind its target DNA. These effects are mediated in large part through the activation or repression of mRNA transcript initiation by DNA-binding proteins, sigma factors and/or signal transduction systems. During the elongation phase, ribonucleotides are polymerized into a RNA molecule complementary to the template DNA. The elongation step can be regulated by a very diverse set of protein factors and small molecules that can alter the activity of the RNA polymerase. Termination is related to the release of the transcription machinery and can be modulated by local mRNA structure and specific termination or antitermination factors.



**Figure 1.2 Overall view of prokaryotic transcription**

The initiation step requires the recognition of a promoter sequence where the RNA polymerase can bind. Recognition of the promoter sequence (P) requires the binding sigma factors to the core RNA polymerase. In order to synthesize the RNA molecule, a small region of the DNA is melted forming a transcription bubble. During elongation, RNA polymerase moves along the DNA while adding nucleotides complementary to the template. Transcription terminates at short DNA sequences recognized as termination signals (T) using either a Rho-dependent or Rho-independent mechanism.

## 1.1.1 Transcriptional regulation in prokaryotes

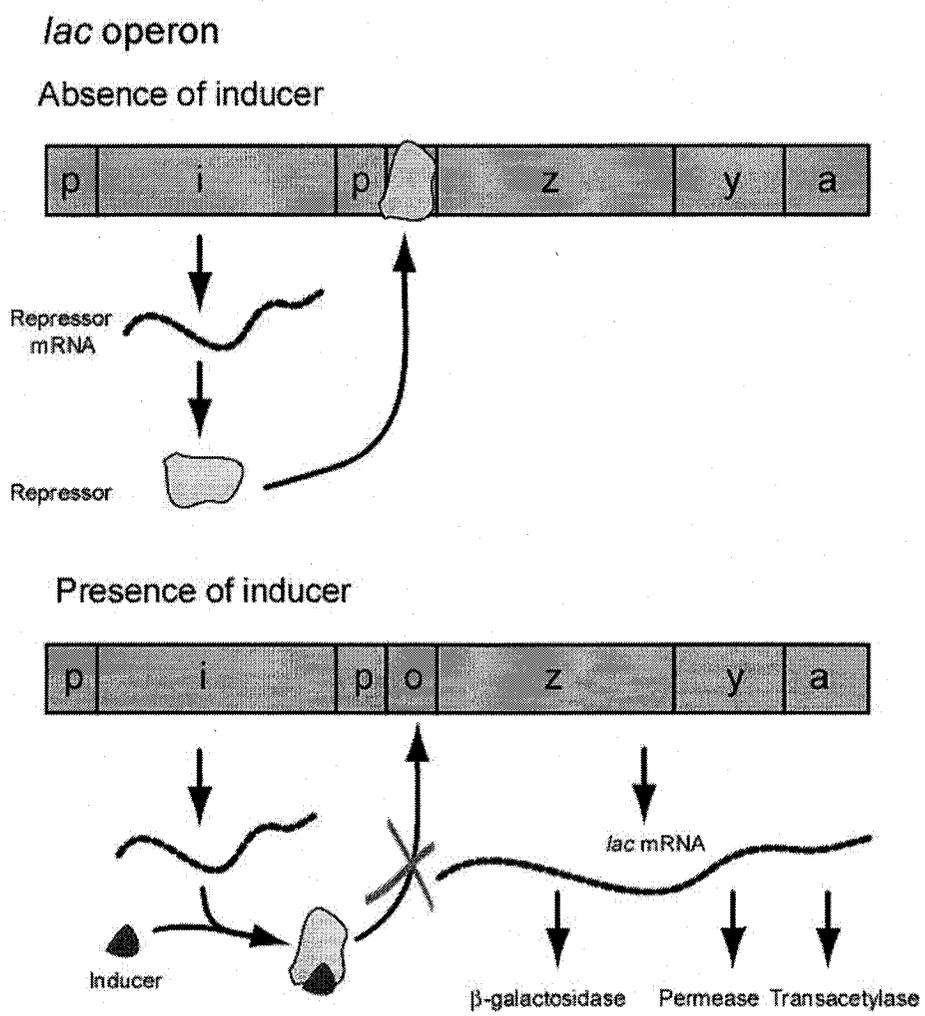
### 1.1.1.1 Prokaryotic transcription initiation

Transcription initiation is the predominant step for control of gene expression in prokaryotes. This step requires the interaction of RNA polymerase with promoter DNA and the formation of an open complex, in which the duplex DNA is unwound (deHaseth *et al*, 1998). Several sequences influencing promoter recognition by the RNA polymerase (RNAP) have been studied intensively (Busby and Ebright, 1994). The two main sequence elements are located 35 and 10 nucleotides upstream of the initiation site. Promoter -10 element, also known as Pribnow box, has the consensus TATAAT and is recognized by domain 2 of the RNAP  $\sigma$  subunit. The -35 element is recognized by domain 4 of the RNAP  $\sigma$  subunit and has the consensus TTGACA (deHaseth *et al*, 1998). Two other important promoters are the extended -10 element and the UP element. The extended -10 element is a ~4 bp motif located immediately upstream of the Pribnow box and is recognized by domain 3 of the RNAP  $\sigma$  subunit (Murakami *et al*, 2002). The UP element is a 20 bp sequence located upstream of the -35 promoter and recognized by the C-terminal domains of RNAP  $\alpha$  subunits (Ross *et al*, 2001). The relative contributions of each of these elements differs from promoter to promoter and their role appears to be docking of the RNA polymerase such that it is competent for open-complex formation (deHaseth *et al*, 1998).

Recognition of a given promoter by the RNA polymerase is regulated by interactions with accessory proteins that can act in a positive or negative manner. The most common of them are the accessory sigma factors that compete with the main  $\sigma$ ,  $\sigma^{70}$ . Alternative  $\sigma$  factors are widely distributed in bacteria and are an efficient mechanism to respond to specific stresses enabling RNAP to transcribe from particular sequence elements (Ishihama, 2000; Maeda *et al*, 2000). The activity of sigma factors can also be controlled by anti-sigma factor that sequester them away from the RNA polymerase (Hughes and Mathee, 1998). Additional transcription factors bind to sequences adjacent to the promoter elements, or operators and can also regulate the accessibility of promoter regions by the RNA polymerase. Repressor proteins prevent binding of the RNA polymerase to the

promoter or interfere with post-recruitment steps in transcription initiation (Muller-Hill, 1998). When a transcription factor binds to a promoter, it can activate or repress transcription initiation of specific genes (Perez-Rueda and Collado-Vides, 2000). Accordingly, bacterial operons are classified as inducible or repressible.

Inducible operons are characteristic of gene ensembles necessary for the utilization of energy and are also known as catabolite regulated operons. In an inducible system, binding to the operator sequence is inhibited by an inducer, normally a molecule at the start of the metabolic pathway governed by the enzymes encoded by the operon. The classic example of inducible operon is the *lac* system, which regulates the metabolism of  $\beta$ -galactosides such as lactose (Figure 1.3). This system is composed of the structural genes *lacZ*, *lacY* and *lacA*, encoding the enzymes  $\beta$ -galactosidase, lactose permease and lactose transacetylase, respectively. Transcription begins at a promoter (*lacP*) upstream of *lacZ* and ends at a terminator beyond *lacA*. The operon is under control of the adjacent *lacI* gene, which encodes the lactose repressor. In the absence of allolactose, the inducer of the *lac* operon, the repressor binds to the *lac* operator (*lacO*) and prevents RNA polymerase from transcribing the operon. However, in the presence of allolactose the repressor can no longer bind to *lacO* allowing the RNA polymerase to bind to *lacP* and initiate transcription (Schlax *et al*, 1995).



**Figure 1.3** Transcriptional control of the *lac* operon

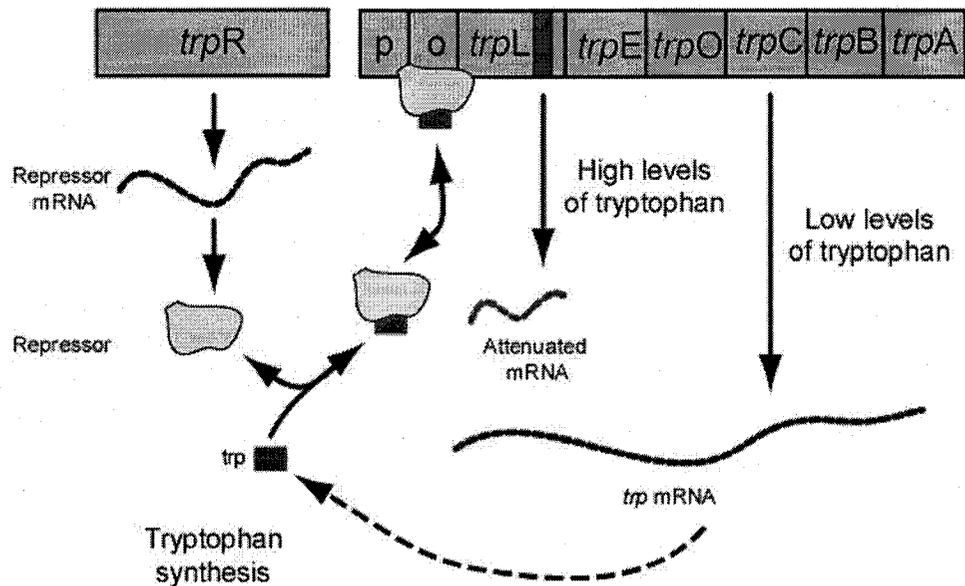
In the absence of the inducer, the repressor protein binds to the *lac* operator and prevents transcription of the corresponding mRNA. However, when lactose is present the *lac* repressor binds to it undergoing a conformational change that prevents it from binding the operator sequence. Under these conditions, transcription from the *lac* operon occurs, allowing the expression of the *lacZ*, *lacY* and *lacZ* genes.

In a repressible system, the repressor only binds to the operator if it is bound to a co-repressor, which is normally the end product of the regulated metabolic pathway. Repressible operons are organized in much the same way as inducible operons, as the structural genes are under the control of a promoter and operator, and there is a gene encoding a repressor. An example of this kind of regulation is the tryptophan operon, *trp* (Figure 1.4). In contrast to the *lac* operon, the *trp* repressor (encoded by the unlinked *trpR*) only binds to the operator when tryptophan is present. Tryptophan acts as a co-repressor as it binds to the repressor and causes a conformational change that allows it join the operator. In the absence of tryptophan, the repressor will not bind to the operator and transcription occurs, allowing the production of the enzymes for its synthesis (Elf *et al*, 2001; Yanofsky, 2001).

#### **1.1.1.2 Prokaryotic transcription elongation**

Ternary complexes of RNA polymerase with its DNA template and nascent transcript are central intermediates in transcription. RNA polymerases are processive enzymes, therefore if the nascent transcript is released prematurely RNAP is unable to rebind the RNA and complete its transcription. Several processes that affect the progression of RNA polymerase in ternary complexes have been discovered. These reactions can be signaled intrinsically, by nucleic acids and the RNA polymerase, or extrinsically, by other regulatory factors (Uptain *et al*, 1997). These factors can affect important transcriptional steps, and therefore play a central role in regulation of gene expression.

### Structure of the *trp* operon



**Figure 1.4 Regulation of the *trp* operon in *E. coli***

Tryptophan acts as a negative regulator of the *trp* operon. Binding of this amino acid to the repressor protein increases its affinity for the operator sequence, inhibiting the transcription of the *trp* genes. A second regulatory element, *trpL*, can further regulate transcription of *trp*. The *trpL* mRNA encodes a non-functional leader peptide rich in tryptophan. When cells have adequate levels of tryptophan-charged tRNA, the leader peptide is synthesized, inducing the formation of a transcription terminator in the leader transcript. However, when cells are deficient in charged tRNA, the ribosome translating *trpL* stalls at one of these tryptophan codons stimulating the formation of an antiterminator structure that prevents folding of the competing hairpin

Throughout the elongation phase, the RNA polymerase can encounter several blocks that include transcriptional pauses, transcriptional arrest and transcript termination (Kassavetis and Chamberlin, 1981). Transcriptional pause and arrest signals can be the result of RNAP's interaction with sequences in the nascent mRNA transcript, in the DNA template, or in response to DNA binding proteins that hinder the progression of ternary complexes. Transcriptional pausing is a temporary impediment to elongation that was first characterized from the study of *E. coli* RNA polymerases complexes synchronously initiated with bacteriophage T7 DNA *in vitro* (Kassavetis and Chamberlin, 1981; Levin and Chamberlin, 1987). Since then, many other RNA polymerases have been shown to pause in the absence of ancillary factors. Pausing is considered as a prerequisite for transcript termination and is an important event in many regulatory processes such as the attenuation of the amino acid biosynthetic operons of enteric bacteria (Landick and Yanofsky, 1987) or the Q-mediated antitermination of phage lambda (Roberts, 1993). However, not all pauses lead to termination (Uptain *et al*, 1997).

Transcriptional arrest can block subsequently initiated RNA polymerases causing the repression of RNA synthesis. Intrinsic blocks to transcription elongation do not require ancillary factors and result from the interaction between the engaged RNAP, the nascent transcript and the DNA template. The best-studied systems of transcriptional arrest are the attenuation mechanisms in enteric bacteria, which involve transcription termination at a site located between the promoter and the structural genes of the operon. For operons regulated in this manner, transcription will terminate depending on the position of the ribosome on a specific short peptide-coding region in the transcript. Many operons concerned with amino acid synthesis and utilization are transcriptionally regulated in this way. Again, the classical example involves the tryptophan operon of *E. coli*. In this system, the leader RNA segment preceding the antiterminator contains a fourteen residue-coding region, *trpL*, which includes two tandem tryptophan codons. When cells have adequate levels of tryptophan-charged tRNA, the leader

peptide is synthesized, inducing the formation of a transcription terminator in the leader transcript. However, when cells are deficient in charged tRNA, the ribosome translating *trpL* stalls at one of these tryptophan codons stimulating the formation of an antiterminator structure that prevents folding of the competing hairpin (Elf *et al*, 2001).

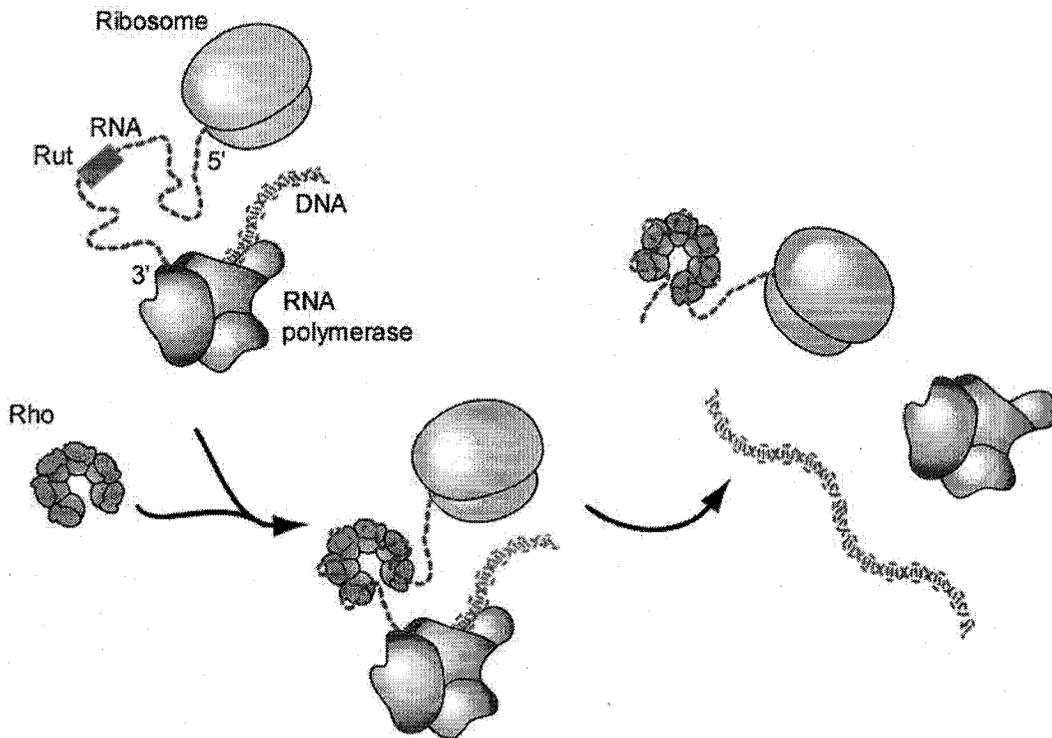
### 1.1.1.3 Prokaryotic transcription termination

Orderly expression of genes depends on the ability of the transcription complex to terminate elongation and release the transcript at the end of the gene. In bacteria, transcription termination sites are determined by specific DNA sequences involved in the formation of hairpin structures that disrupt the RNA polymerase ternary complex, sometimes with help of termination factor Rho. For this reason, bacterial termination is divided in Rho-dependent and Rho-independent mechanisms (Richardson and Greenblatt, 1996).

Rho independent termination sequences have two characteristic features: a series of U residues in the transcribed RNA and, preceding these, a GC-rich self-complementary region with several intervening nucleotides (Rosenberg and Court, 1979; Brendel *et al*, 1986; d'Aubenton Carafa *et al*, 1990). When such a self-complementary region is synthesized in the growing RNA chain, a stem-loop is formed (Ryan and Chamberlin, 1983; Yang and Gardner, 1989; Cheng *et al*, 1991). This structure contributes to the termination process by weakening the interaction between the nascent RNA and the exit region of RNAP (Arndt and Chamberlin, 1990). The base pairs between the U residues at the 3' end of the nascent RNA chain and the A residue in the template DNA strand are more unstable than other types of Watson-Crick base pairs (Martin and Tinoco, 1980). This last feature is also one of the differentiating factors between intrinsic terminators and pause sites (Richardson and Greenblatt, 1996).

Rho-dependent termination was first recognized during *in vitro* studies of transcription of  $\lambda$ -phage DNA. In the presence Rho, transcription from  $\lambda$  DNA yields two short RNA transcripts designated  $t_L$  and  $t_R$ . In absence of Rho, however, transcripts made from  $P_L$  and  $P_R$  are many thousands of nucleotides

long. The termination factor Rho is a RNA-binding protein of 47 kDa that assembles into a hexameric ring structure typical of RNA/DNA helicases (Skordalakes and Berger, 2003). The binding of Rho to its target sequence induces an ATP-dependent translocation in the 3' end of the mRNA that can eventually unwind the RNA-DNA hybrid at the active site of the RNA polymerase (Brennan *et al*, 1987; Geiselmann *et al*, 1993; Platt, 1994; Richardson, 1996). Rho-dependent transcription termination depends on whether Rho moves sufficiently fast to catch up with the RNA polymerase complex. Rho dependent terminators extend over 150-200 base pairs of DNA and consist of a transcription stop point (*tsp*) and a Rho utilization sequence (*rut*) (Lau *et al*, 1982; Lau *et al*, 1984; Richardson, 1990; Zalatan *et al*, 1993). The sequences that specify a Rho-dependent terminator are very diverse and do not conform to a single consensus. However, there are some features that are common to most Rho utilization regions. *rut* sites are about 85 nucleotides long and are the target for Rho binding in the nascent transcript. These Rho-binding sites consist of stretches that are not likely to pair with other segments of the RNA and have a higher than average proportion of C residues (Morgan *et al*, 1985). Rho termination is limited to sections of the mRNA that do not code for proteins, as translating ribosomes will block access to the mRNA (Richardson *et al*, 1975). The ability of Rho to efficiently stop transcription at certain terminators depends on the presence of another factor called NusG (Li *et al*, 1992; Sullivan and Gottesman, 1992). NusG helps Rho to overcome certain kinetic deficiencies (Nehrke and Platt, 1994) and has been shown to interact with the RNA polymerase. (Li *et al*, 1992; Li *et al*, 1993).



**Figure 1.5 Rho-dependent transcription termination**

Rho-dependent termination requires the transcription factor Rho to bind to a *rut* sequence in the target mRNA. Rho then interacts with the RNA polymerase causing it to release the DNA template, thus preventing expression of downstream genes.

Mechanisms that prevent the termination of transcription can be divided as processive and nonprocessive (Richardson and Greenblatt, 1996). Processive mechanisms depend on the activity of specific proteins that modulate the RNA polymerase and alter its response to termination signals. Nonprocessive mechanisms, on the other hand, have only local effects on termination at particular terminators. These terminators are normally located near the beginning of an operon and function as transcriptional attenuators. Examples of nonprocessive antitermination are too numerous to mention in detail. Some of the best characterized systems include the *trp* and *his* operons (Kolter and Yanofsky, 1982), the *tna* operon encoding the tryptophanase enzyme (Stewart and Yanofsky, 1985; Gish and Yanofsky, 1993), the *bgl* operon involved in the catabolism of  $\beta$ -glucosides (Amster-Choder and Wright, 1992; Amster-Choder and Wright, 1993) and the *rpsJ* operon encoding ribosomal proteins S10 and L4 (Zengel and Lindahl, 1990; Sha *et al.*, 1995)

Several processive antitermination mechanisms have been studied. The classic example of this type of mechanism comes from bacteriophage  $\lambda$ . In  $\lambda$ , the transition between the infection stages is dependent on read through of a series of transcriptional terminators, to allow synthesis of transcripts encoding new sets of gene products (Greenblatt *et al.*, 1993; Friedman and Court, 1995). Read-through of these terminators requires binding of the product of one of the early expressed genes, the N protein, to sites on the nascent transcripts (nut sites) (Franklin, 1974). N then forms a complex of host-encoded proteins (NusA, NusB, NusG, ribosomal protein S10) that convert the RNA polymerase into a form resistant to either intrinsic or Rho-dependent transcriptional terminators. These proteins are thought to facilitate the formation and stabilization of elongation complexes by interacting directly with the polymerase or other accessory factors. NusA, NusB, NusG and ribosomal protein S10 are also involved in antitermination of other bacterial genes like the ribosomal RNA operons (Sharrock *et al.*, 1985; Nodwell and Greenblatt, 1991) which are regulated by an antiterminator element, known as

boxA, that renders the RNA polymerase insensitive to Rho-dependent terminators (Li *et al*, 1984; Berg *et al*, 1989).

Another well-known antitermination factor is the  $\lambda$  Q protein which promotes read through of terminators at long distances downstream from their transcription start site (Roberts, 1993). Rather than operate through nascent transcript binding or alternative RNA secondary structures, the Q protein binds a DNA sequence, known as the *qut* site, upstream of the late promoter and exerts its effect on the transcribing RNA polymerase (Yarnell and Roberts, 1992). The Q protein does not require any other bacterial host factor and its activity seems to result from direct contact with the RNA polymerase (Yang *et al*, 1989).

Other proteins, such as Psu and YaeO, can specifically inhibit the activity of Rho (Isaksen *et al*, 1992; Pichoff *et al*, 1998). Psu is a virion protein from the satellite bacteriophage P4 that also suppresses Rho-dependent transcription termination in operons of its helper bacteriophage P2 (Linderoth and Calendar, 1991). Psu works only at Rho-dependent terminators and its activity is not operon specific. The expression of Psu is sufficient to inhibit termination in P2 late genes, plasmid operons and the host chromosome. However, it is still unknown whether Psu inhibits an enzymatic activity of Rho or its interaction with RNA, ATP or RNA polymerase. The second Rho inhibitor, YaeO, was discovered as a suppressor of cell division inhibition by a MalE-MinE fusion protein in *E. coli* (Pichoff *et al*, 1998). It was shown that YaeO interacts with Rho and that its expression reduces termination at Rho-dependent termination sites (Pichoff *et al*, 1998). Studies on the structure of YaeO and its inhibition mechanism are the subject of Chapter 2 of this thesis.

## **1.1.2 Transcription regulation in eukaryotes**

### **1.1.2.1 Eukaryotic transcription initiation**

In eukaryotes, transcription is carried out by three different RNA polymerases: RNAP I, RNAP II and RNAP III. RNAP I is in charge of synthesizing the ribosomal RNAs, except for the 5S species (Paule and White, 2000). RNAP II synthesizes mRNA and some small nuclear RNAs (snRNAs)

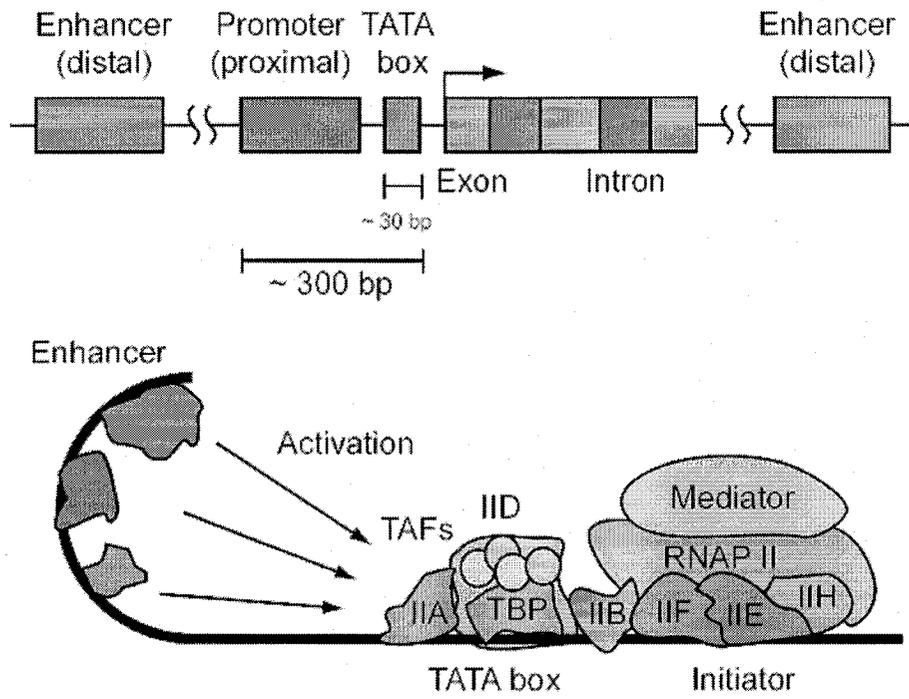
involved in RNA splicing (Hahn, 2004). RNAP III synthesizes the 5S rRNA and the tRNAs (Paule and White, 2000). The most complex mechanisms of transcriptional regulation involve the genes transcribed by RNAP II as they consist of basal promoters and different types of transcriptional regulatory domains. The following text will focus on the regulation of RNAP II.

Transcription by RNAP II requires the formation of a preinitiation complex consisting of RNAP II and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIIF and TFIIF (Warren, 2002). TFIID is a multiprotein complex consisting of the TATA-box binding protein (TBP) at its core and the TBP-associated factors (TAFs) (Zawel and Reinberg, 1995). The regulation of transcription involves the interaction between GTFs that bind to the core promoter element and sequence specific trans-acting factors that bind to promoter elements and modulate the relative efficiency of transcription initiation by activation or repression (Roeder, 2005). Positive or negative cofactors may also modulate transcription by direct interaction with GTFs. DNA packaging, the presence of histones and CpG methylation also have important consequences in the ability of transcriptional factors and RNA polymerases to find access to specific genes and to activate their transcription (Roeder, 2005).

The core promoters of RNAP II comprise DNA sequence motifs within -40 to +40 nucleotides relative to the RNA start site and include the TATA box, the TFIIB recognition element (BRE), the initiator (Inr) and the downstream promoter element (DPE). In the appropriate combinations, these elements are sufficient to direct transcription initiation by the basal RNAP II transcriptional machinery (Blackwood and Kadonaga, 1998). The TATA or Goldberg-Hogness box has the consensus sequence TATA(T/A)A(T/A) and is located 20 to 30 bases upstream of the transcriptional start site. The TATA box is probably the most important promoter. Transcriptional regulation via this sequence element occurs through TATA binding proteins or TBPs (Zawel and Reinberg, 1995). Upstream of the core promoter, there are typically multiple recognition sites for a subgroup of sequence specific DNA-binding transcription factors, which include Sp1, the

CCAAT-binding protein transcription factor (CTF) and the CCAAT-box-binding factor (CBF) (Blackwood and Kadonaga, 1998).

Eukaryotic transcription factors can also recognize other regulatory sequences specific to each mRNA such as enhancers and promoter proximal elements. These elements are distinct from the core promoter elements because they have no apparent role in basal transcription. Enhancers are usually 100-200 base pairs long and are predominantly located upstream (5') of the transcription initiation site. However, there are some cases where they can be located much further upstream or even downstream of the target gene. For example the wing margin enhancer of the *Drosophila cut* locus resides 85 kb upstream of its promoter (Jack *et al*, 1991) whereas the T-cell receptor  $\alpha$ -chain gene enhancer is located about 69 kb downstream of the promoter (Blackwood and Kadonaga, 1998). Another case is the immunoglobulin H $\mu$  core enhancer which lies within the second introns of the transcription unit (Staudt and Lenardo, 1991). Some genes can also be controlled by more than one enhancer region, as in the case of the even-skipped gene in *Drosophila* (Arnosti *et al*, 1996). Promoter-proximal elements, on the other hand, are roughly within the first 200 base pairs upstream of the cap site and lose their influence when moved further from the promoter (McKnight and Kingsbury, 1982). Different combinations of transcription factors can exert differential regulatory effects upon transcriptional initiation and this is one the main mechanism for determining cell-type specificity (Atchison, 1988).



**Figure 1.6 Eukaryotic gene structure**

In contrast to the simple organization of prokaryotic genomes, eukaryotic genes are organized in a more complex manner. Within most eukaryotic genes there are noncoding introns alternated with coding regions or exons. Genes transcribed by RNA polymerase II require at least five accessory protein factors. Eukaryotic transcription is regulated by additional control sequences such as promoter-proximal elements and enhancers as well as the basal promoter elements.

### **1.1.2.2 Eukaryotic transcription elongation**

The genes that are regulated during the elongation process include genes involved in housekeeping functions, cell cycle, development, differentiation and stress response genes as well as oncogenes (Arndt and Kane, 2003). A diverse collection of proteins, including TFIIF, ELL and Elongin, are capable of suppressing transient pausing by RNAP II (Uptain *et al*, 1997; Shilatifard, 1998; Conaway and Conaway, 1999). Several additional elongation factors that suppress pausing have been discovered such as the Cockayne syndrome B protein (Selby and Sancar, 1997). The Tat-Sf1 (Li and Green, 1998) and CA150 proteins (Sune and Garcia-Blanco, 1999) were originally identified by their ability to promote elongation from the HIV-1 LTR and the ELL complex (Shilatifard, 1998).

In addition to elongation factors that interact directly with RNAP II, there exist other proteins that promote elongation by modifying chromatin structure. This level of regulation was clearly demonstrated in studies performed in the *hsp70* gene of *D. melanogaster* (Rougvie and Lis, 1988). Elongation can be also modulated by proteins that modify nucleosomes. Proteins like FACT can interact with nucleosomes and H2A-H2B dimers and may function by promoting nucleosome disassembly during transcription (Belotserkovskaya *et al*, 2003), Swi-Snf and Chd1 which remodel nucleosomes (Hartzog *et al*, 2002) and proteins like Gcn5 and Elp3, which acetylate or methylate histones (Wittschieben *et al*, 2000) (Hampsey and Reinberg, 2003). Three transcription elongation factors that have emerged as playing critical roles are the Spt4, Spt5, and Spt6 proteins (Winston *et al*, 1984). These factors modulate the structure of chromatin and can control transcription in important regulatory and developmental roles. Spt5 and Spt6 are bound within the open reading frames of heat shock (*hsp*) genes after heat shock induction and co-localize with actively transcribing RNA polymerase II (Winston *et al*, 1984).

### **1.1.2.3 Eukaryotic transcription termination**

In eukaryotes, there are three different mechanisms of transcription termination, one for each polymerase. Transcription by RNA polymerase I

terminates at an 18 nt terminator site located approximately 1000 nt downstream of the end of the coding sequence (Grummt *et al*, 1986). Recognition of this site requires a DNA-binding protein known as TTF-I in mice or Reb1p in yeast (Reeder and Lang, 1997). TTF-I then recruits a releasing factor, which catalyses formation of the 3'-end. An exonuclease may trim the 3' end to produce the mature 3'-end. RNAP II uses a terminator region but it is not known what features define this region or how it effects termination (Proudfoot, 1989). Much of the difficulty in resolving this issue is due to the fact that class II transcripts are processed at the 3' end with the addition of a poly(A) tail, which replaces the true 3' end of the transcript (Connelly and Manley, 1988). It is believed that the presence of the cleavage polyadenylation specificity factor (CPSF) and the cleavage stimulation factor (CStF) in association with the C-terminal domain of the large subunit of RNAP II may play a role in regulating termination (Park *et al*, 2004). Transcription by RNAP III is terminated in a manner reminiscent of that in prokaryotes. A small run of U's in a GC-rich region is required as the termination signal (Bogenhagen and Brown, 1981). However, the run of U's is shorter than in prokaryotes and the GC-rich region does not need to adopt any kind of hairpin structure (Bogenhagen and Brown, 1981).

### **1.1.3 Transcriptional regulation in archaea**

Even though, archaeal transcriptional regulators resemble those of bacteria, transcription in archaea is much more like that observed in eukaryotes. Archaea contain a single RNA polymerase of 10-14 subunits, most of them homologous to subunits of eukaryal polymerases (Soppa, 1999). Archaea also possess two general transcription factors, TBP and TFB, which are required for accurate and efficient transcription and are homologues of the eukaryal TBP and TFIIB, respectively (Bell *et al*, 2001). There also exist an archaeal homologue of eukaryotic transcription factor TFIIE called TFE (Bell *et al*, 2001). Archaea have histones, so regulation by histone binding is a possibility as in eukarya (Reeve *et al*, 1997). However, as in bacteria, polycistronic mRNAs are common and sequencing has revealed the presence of genes homologous to bacterial

transcriptional regulators like the *asnC-lrp* family of helix-turn-helix transcriptional regulators (Leigh, 1999). Repression of gene expression invoking the bacterial paradigm was first demonstrated with the control of lytic growth of *Halobacterium* phage  $\phi$ H (Ken and Hackett, 1991; Stolt and Zillig, 1992). Other examples include the negative regulation of the *nif* gene by the nitrogen source in *Methanococcus maripaludis* and *Methanosarcina barkeri* (Cohen-Kupiec *et al*, 1997; Chien *et al*, 1998). Activation of gene expression using transcription factors with eukaryal features occurs in the synthesis of gas vesicles in *Halobacterium salinarum* and *Haloferax mediterranei* (Kruger *et al*, 1998).

## **1.2 Some mechanisms of post-transcriptional control**

Post-transcriptional regulatory mechanisms are also critical determinants of genetic expression. For example, attenuation via antisense RNA transcripts and inhibition of translation are well-recognized approaches for post-transcriptional control (Landick and Yanofsky, 1987; Delihas, 1995). Other genes can also be regulated by specific protein factors targeting, in general, sequences in the untranslated region of the mRNA. The general mechanism for these factors is the inhibition of ribosome binding to target mRNAs followed by degradation by specific RNases. In eukaryotes, post transcriptional regulation is more commonly achieved by altering the RNA transcripts in a very complex set of RNA processing mechanisms and regulatory checkpoints such as the addition of a 5' cap and a 3' poly (A) tail, and the removal of introns. Once a functional mRNA is produced, it must be transported from the nucleus to the cytoplasm where it can be translated into protein. Additionally, the information content of the mRNA can be altered via editing mechanisms that change the original message encoded by the DNA.

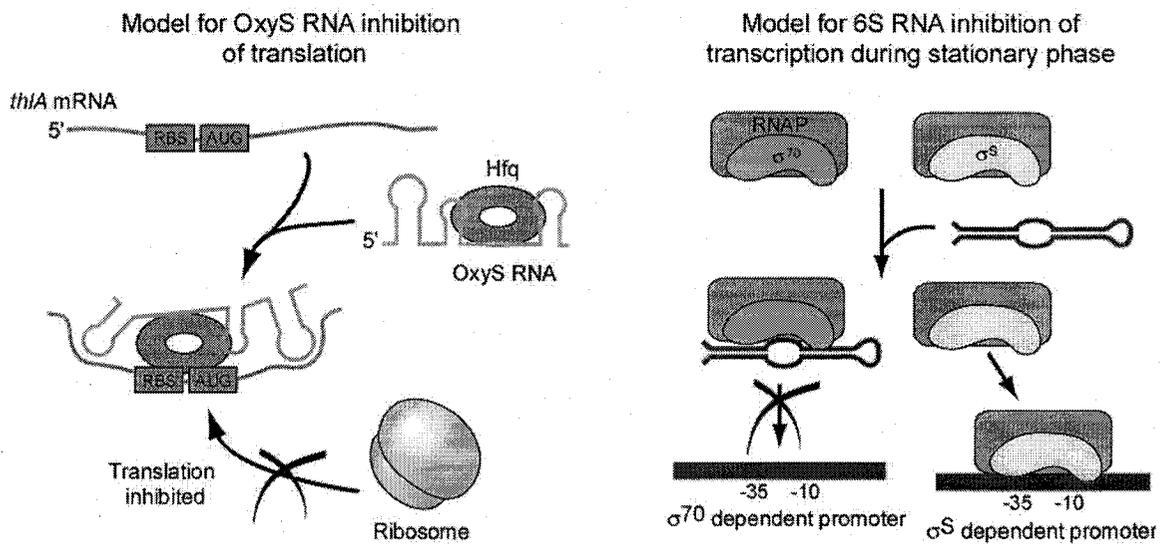
### **1.2.1 Regulatory RNAs**

Bacteria use small RNAs to fine-tune their physiology and adapt to rapidly changing environments. Genome screening has identified around 50 small RNAs in *Escherichia coli* and estimates suggest that the total number might be on the

order of a hundred (Wassarman *et al*, 1999). In general, these are antisense RNAs that regulate only one mRNA target. These antisense RNAs are common in plasmid replication control systems and in bacteriophage immunity systems. This can be seen in plasmid R1, where the small antisense RNA CopA inhibits the translation of the replication protein CopT (Malmgren *et al*, 1997).

Small RNAs can also act as signal transducers and coordinate gene expression in response to a particular condition by targeting global regulators of gene expression. One example of this is the conserved regulatory element *RyhB* involved in iron metabolism (Masse and Gottesman, 2002). In the event of Fe limitation, *ryhB* is strongly expressed causing a dramatic decrease in the level of several mRNAs encoding iron-containing proteins or proteins involved in the intra-cellular storage of iron (Masse and Gottesman, 2002). Another example is Spot 42, a stable RNA, present at about 200 copies per cell that is negatively regulated by the cAMP receptor protein complex. Spot42 has been shown to inhibit translation of *galk* by base-pairing near its ribosome-binding site and appears to regulate the *sucABCD* operon, which encodes enzymes involved in the TCA cycle (Moller *et al*, 2002).

Small RNAs can also act as positive regulators, as in the regulation of the stress/stationary phase sigma factor RpoS ( $\sigma^S$ ) in *E. coli*. Three small RNAs regulate the expression of RpoS: OxyS, DsrA and RprA (Masse *et al*, 2003). Translational initiation of *rpoS* is repressed by a secondary structure in the upstream message that occludes the Shine-Dalgarno (SD) sequence (Altuvia *et al*, 1998). Both DsrA and RprA RNA have complementary sequences that allow them to pair with the upstream leader region of *rpoS* and relieve the inhibition by freeing the SD site. In contrast, OxyS represses *rpoS* translation by modulating the activity of the Hfq replicase as shown in Figure 1.7 (Zhang *et al*, 1998).



**Figure 1.7 Models of sRNA function**

Many small RNAs act as antisense regulators of target mRNAs and generally require the Hfq protein to stimulate pairing. Other sRNAs in bacteria can alter protein activity through direct interaction. For example, the 6S RNA binds to the RNA polymerase containing the  $\sigma^{70}$  subunit, decreasing transcription of  $\sigma^{70}$  dependent promoters.

Other small RNA regulators, like the 6S RNA and CsrB, act by binding a specific target protein, inhibiting or changing its activity. 6S RNA interacts directly with the RNA polymerase regulating transcription from  $\sigma^{70}$  directed RNAP (Wassarman and Storz, 2000). Another example is CsrB, a 366 nucleotide RNA that acts as antagonist of the Carbon storage regulator A (CsrA) (Romeo, 1998). CsrA is a small translational regulatory protein of 61 amino acids, which is involved in the positive regulation of flagella synthesis, acetate metabolism and glycolysis (Baker *et al*, 2002). Additionally, CsrA acts as a negative regulator of biofilm formation, glycogen biosynthesis and catabolism, and gluconeogenesis in *E. coli* (Sabnis *et al*, 1995; Wei *et al*, 2000; Wei *et al*, 2001). CsrA acts by binding to sequences close to the ribosome-binding site of the glycogen metabolism genes and inhibits the initiation of translation. CsrB contains 18 CsrA binding sequences and inhibits CsrA activity by sequestering it. A second small RNA antagonist of CsrA, CsrC has been described recently (Weilbacher *et al*, 2003). The CsrA/CsrB paradigm has been found in a wide variety of bacteria, and is involved in regulation of invasion and virulence in plants and animals. The structural studies performed on CsrA and its binding to CsrB are the subject of Chapter 3 of this thesis.

### **1.2.2 Capping and polyadenylation**

Two features differentiating eukaryotic mRNAs from prokaryotic messengers are the addition of a 3' poly (A) tail and a 5' cap. The poly (A) tail is implicated in mRNA stability, mRNA transport and translation initiation (Gallie, 1998; Hall, 2002). The role of the poly(A) tail in translation is mediated through the poly(A) binding protein (PABP), which interacts with the cap binding complex (Sachs and Varani, 2000). This complex is important for recruiting the ribosome to the mRNA for translation initiation. The poly(A) tail is added post transcriptionally in the nucleus to most mRNA with the exception of those coding for histones (Munroe and Jacobson, 1990; Wickens, 1990; Hall, 2002). Polyadenylation requires the site-specific cleavage of the 3' end of the mRNA 10-30 nucleotides downstream of a conserved hexanucleotide sequence, AAUAAA.

Then, a specific poly(A) polymerase catalyzes the addition of adenosine residues to this cleaved 3' end.

The cap structure (m<sup>7</sup>GpppN, where N is any nucleotide) is obtained by the addition of 7-methylguanosine to the 5' end of the mRNA via an inverted 5', 5'-triphosphate link (Shatkin, 1976). This structure plays a critical role in ribosome recruitment as it interacts with the eukaryotic initiator factor 4E (Sonenberg *et al*, 1978) and is also implicated in pre-mRNA splicing, mRNA export from the nucleus, and mRNA stability (Varani, 1997). Capping is linked to the early stages of transcription initiation and elongation. Elongation factors arrest the transcription complex and help to recruit the capping and splicing machinery to the elongation complex.

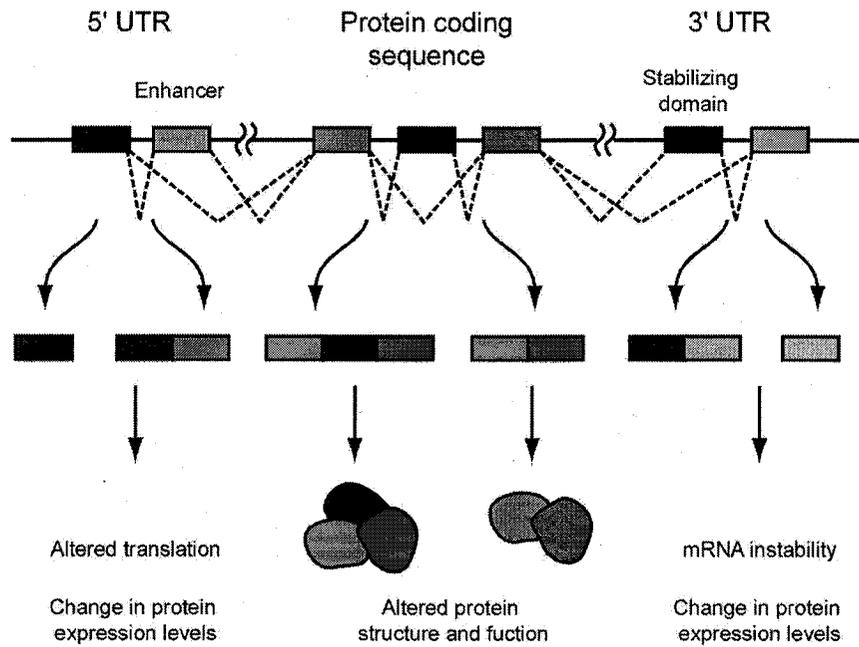
### 1.2.3 Splicing

In eukaryotic organisms, intervening sequences, or introns which interrupt the protein-coding sequences in the primary transcript of many genes are removed by RNA splicing (Jeffreys and Flavell, 1977; Tilghman *et al*, 1978). This is an essential and precisely regulated post-transcriptional process that occurs prior to mRNA translation. RNA splicing takes place in the nucleus and is catalyzed by a complex called the spliceosome formed from the assembly of ribonucleoproteins U1, U2, U5, and U4/U6 snRNPs (Newman, 1998). After assembly of the spliceosome, the reaction occurs in two steps: first, the branch-point A nucleotide in the intron sequence, which is located close to the 3' splice site, attacks the 5' splice site and cleaves it; the cut 5' end of the intron sequence thereby becomes covalently linked to this A nucleotide, forming a branched nucleotide. Secondly, the 3'-OH end of the first exon sequence, which was created in the first step, adds to the beginning of the second exon sequence, cleaving the RNA molecule at the 3' splice site; the two exon sequences are thereby joined to each other and the intron sequence is released as a lariat. The accuracy of RNA splicing is monitored by proof reading mechanisms that are able to target incorrectly spliced mRNA for destruction or that can correct the error (Newman, 1998).

During RNA splicing, individual exons can either be retained in the mature message or targeted for removal to create a diverse array of mRNAs from a single pre-mRNA. This process is known as alternative RNA splicing (Lopez, 1998). Alternative splice events that affect the protein coding region of the mRNA will give rise to proteins which differ in their sequence and therefore in their activities (Figure 1.8). In mammals, for example, the calcitonin gene produces a hormone in one cell type and a neurotransmitter in another, due to exon shuffling (Leff *et al*, 1987). Another example of alternative splicing occurs with the virus SV40 which can direct the synthesis of two proteins, the big T and small t antigens, from the same pre-mRNA (Eul *et al*, 1996). Alternative splicing within the non-coding regions of the RNA can also result in changes in regulatory elements such as translation enhancers or RNA stability domains, which may have a dramatic effect on the level of protein expression.

#### **1.2.4 Regulation of RNA Longevity**

Unlike prokaryotic mRNAs, whose half-lives are all in the range of 1-5 minutes, eukaryotic mRNAs can vary greatly in their stability. This is another important regulatory mechanism as a more stable mRNA will produce more protein before being degraded. The information for mRNA lifespan is predominately found in the 3' UTR. The sequence AUUUA, when found in the 3' UTR, is a signal for early degradation. Multiple copies of this sequence shorten the lifespan of the mRNA. Short-lived mRNAs, which code for proto-oncogene products, cytokines and early response gene products, contain AU-rich elements in their 3' UTR that confer mRNA instability (Chen and Shyu, 1994; Stoecklin *et al*, 1994; Xu *et al*, 1997).



**Figure 1.8 Effects of alternative splicing on gene expression**

Alternative splicing can occur in any region of the nascent mRNA. Insertion or deletion in the 5' UTR contains will affect regulatory regions that control protein expression. Modification of 3' UTR region can have consequences on mRNA stability and therefore protein expression. Alternative splicing within the protein coding sequence results in altered protein structure and function.

### **1.2.5 RNA Editing**

Originally, the term RNA editing was created to describe a phenomenon in which uridine residues are inserted and deleted from mitochondrial RNAs of kinetoplastid protozoa (Benne *et al*, 1986). However, the definition of RNA editing has been expanded to describe numerous cellular processes that result in the modification of RNA resulting in a sequence that differs from that encoded by their original DNA (or RNA) templates. This definition does not include RNA splicing, capping or polyadenylation (Gott and Emeson, 2000). Changes in gene expression attributed to editing have been described in organisms from unicellular protozoa to man, and can affect the mRNAs, tRNAs, and rRNAs present in all cellular compartments (Gott and Emeson, 2000). Creation of new start and stop codons by insertion of uridines or cytidine to uridine (C-to-U) conversions have been observed in trypanosomatid protozoa, plant organelles, and man. Stop codons are also subject to removal by U-to-C changes in plants. Examples of editing events in the middle of genes are varied and include the creation of ORF by nucleotide insertions; frame shifting between alternative ORFs; amino acid substitutions and alterations in splice sites (Nagalla *et al*, 1994; Sharma *et al*, 1994; Skuse *et al*, 1996; Kolakofsky *et al*, 1998; Rueter and Emeson, 1998; Rueter *et al*, 1999).

### **1.3 Regulation at the translational level**

Translation or protein synthesis is a highly regulated process involving hundreds of macromolecules (translation factors, transfer RNAs, mRNA, ribosomal proteins and ribosomal RNAs). As in transcription, translation is divided in three main steps: initiation, elongation and termination. Initiation involves the formation of a complex between the ribosome, the initiator tRNA complex and mRNA. Elongation is where the actual synthesis of the polypeptide chain occurs, by formation of peptide bonds between amino acids. Termination dissociates the translation complex and releases the finished polypeptide chain. Deregulation of translation often leads to cancer, genetic diseases or

developmental defects (Conlon and Raff, 1999; Miron *et al*, 2001). Even though translation can be regulated at many levels, most mechanisms target initiation, which is also the rate-limiting step of protein synthesis (Copeland, 2003). Several viruses can also use the translation regulatory machinery to their advantage in order to facilitate their propagation. Due to the separation between translation and transcription in eukaryotes, regulation at the translational level is a very important regulatory step in higher organisms.

### 1.3.1 Overview of prokaryotic translation

In prokaryotes, the initiation of protein synthesis requires the participation of three initiation factors: IF1, IF2, and IF3. The initiation step also requires the availability of a free 30s ribosomal subunit bound to the anti-association factor IF3 (Hershey and Merrick, 2000). IF1 assists in the IF3 binding to the ribosome and occludes the A site domain of the small ribosomal subunit ensuring that initiation starts in the P site (Carter *et al*, 2001). The start of bacterial protein synthesis also needs an initiator tRNA carrying N-formylmethionine (tRNA<sup>fMet</sup>) which is docked to the small ribosomal subunit with the aid of a small GTP-binding protein, IF2 (Gualerzi and Pon, 1990). Binding of the ribosome to the messenger occurs near the 3' end of the 16S rRNA. At this position, the conserved sequence 3'-UCCUCC-5', also known as Shine-Dalgarno, can base pair with a sequence near the 5' end of each mRNA (Steitz and Jakes, 1975). This pairing aligns the message correctly for the start of translation and allows the 50S subunit to bind to the 30S initiation complex. Formation of the 70S ribosome, promotes the alignment of the AUG initiator codon with the P site. Hydrolysis of the GTP carried by IF2 causes the release of IF2-GDP, P<sub>i</sub> and IF1, allowing the elongation of the peptide chain to proceed.

Protein elongation involves three tRNA binding sites on the 70S ribosome called the P site (peptidyl), the A site (aminoacyl), and the E site (exit). The nascent polypeptide chain is attached to a tRNA in the P site and the A and E sites are empty. A charged tRNA is delivered to the A site in a complex with the protein elongation factor EF-Tu (or EF-1A), which also carries GTP. The loaded

tRNA must have the correct anticodon to base pair with the mRNA codon that is positioned at the A site. Once the appropriate charged tRNA is deposited into the A site, the GTP is hydrolyzed and the EF-Tu-GDP is released. The GTP form of EF-Tu is regenerated with the aid of exchange factor EF-Ts (or EF-1B) (Proud, 2000).

Upon release of EF-Tu-GDP, the polypeptide chain attached to the tRNA in the P site is transferred to the amino group of the amino acid carried by the A-site tRNA. This process is catalyzed by the peptidyltransferase complex, a ribozyme composed of rRNA and ribosomal proteins. Once the peptide bond is formed the uncharged tRNA remaining in the P site is transferred to the E site while the tRNA with the nascent polypeptide is moved to the P site (Proud, 2000). At the same time, the ribosome moves the mRNA by three nucleotides in the 3' direction, placing a new codon adjacent to the empty A site. This step requires the protein factor EF-G (or EF-2) with GTP, which is hydrolyzed in the process (Rodnina *et al*, 1997). As tRNA is released from the E site, the empty A site accepts the aminoacyl tRNA corresponding to the next codon and the process is repeated until a termination signal is reached.

The termination step requires the participation of release factors RF1, RF2 and RF3 (Welch *et al*, 2000). The release factor causes the translation complex to fall apart and cleaves the polypeptide from the final tRNA. RF1 binds to the ribosome when UAA or UAG is in the A site while RF2 binds when UAA or UGA is in that position. The release process is stimulated with the aid of RF3 and GTP hydrolysis (Freistroffer *et al*, 1997). The peptidyltransferase complex transfers the C-terminal residue of the polypeptide chain from the P-site tRNA to a water molecule, releasing the polypeptide chain from the ribosome. Once the RF factors, GDP and tRNA are released, the 70S dissociates with the aid of ribosome recycling factor and initiator factors IF3 and IF1 (Karimi *et al*, 1999).

### 1.3.2 Eukaryotic translation

#### 1.3.2.1 Eukaryotic translation initiation

The first step in the initiation process is binding of the small ribosomal subunit (40S) to the mRNA. Since this is often the rate-limiting step in initiation, it is also often the target of regulation (Sachs and Varani, 2000). As in prokaryotes, eukaryotic translation initiation is regulated by a diverse set of protein factors (eIFs). However, initiation requires many more protein factors in eukaryotes than in prokaryotes (11 versus 3). Some of the initiation factors attach to the ribosomal subunits and others to the messenger RNA. The first step consists in the association of eIF2, GTP and Met-tRNA<sub>i</sub>. This ternary complex, together with eIF1, eIF3, eIF5 and eIF1A, binds to the 40S ribosomal subunit, forming a 43S pre-initiation complex. This ensemble recruits the mRNA-eIF4F complex in an ATP-dependent manner (Pain, 1996). Once the 43S complex is bound to the mRNA, it scans the 5' region, in an ATP dependent mechanism, until the proper initiation codon is found. Efficient recognition of the initiation codon depends on the recognition of a RNA signature, reminiscent of the Shine-Dalgarno sequence, known as the Kozak sequence (Kozak, 1987; Kozak, 1999). At this point, eIF5 comes into play by interacting with eIF3 and hydrolyzing the GTP bound to eIF2. This event causes the release of eIF3, eIF2-GDP and eIF1 and prepares the remaining complex for interaction with the 60S ribosomal. The recycling of eIF2 is controlled by the guanine-nucleotide exchange factor eIF2B, a protein that replaces the GDP of eIF2 with GTP. This is one of the major checkpoints in the regulation of protein synthesis (Pain, 1996; Gray and Wickens, 1998; Hershey and Merrick, 2000). The activity of eIF2 is modulated by phosphorylation of serine 51 of the  $\alpha$  subunit. When eIF2 $\alpha$  is phosphorylated, its affinity for eIF2B is increased resulting in the formation of a very stable complex in which the bound GDP cannot be exchanged for GTP, resulting in translational arrest (Hershey and Merrick, 2000). The activity of other translation factors like eIF4B, 4E, and 4G, is also controlled by their phosphorylation status in response to extracellular stimuli (Raught and Gingras, 1999; Raught *et al*, 2000; Gingras *et al*, 2001) (Figure 1.8)

Translation can also be regulated by the 5' and 3' untranslated regions (Hentze *et al*, 1987; Gray *et al*, 1993; Meyuhas and Hornstein, 2000). For example, the mRNA encoding tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) contains an ARE that represses translation (Gueydan *et al*, 1999; Piecyk *et al*, 2000). A large number of mRNAs contain 3' UTR elements that regulate translation during early embryonic development (Wickens *et al*, 2000; de Moor and Richter, 2001). For example, the formation of the caudal protein gradient in the *Drosophila* embryo is dependent upon a bicoid-response element (BRE) in the 3' UTR of the caudal mRNA. (Wickens *et al*, 2000).

#### 1.3.2.2 Eukaryotic translation elongation

The elongation phase of protein synthesis consists of the addition of amino acids to the C-terminus of the growing polypeptide chain. The first step involves the entrance of the ternary complex formed by EF1A, GTP and aa-tRNA to the A-site of the ribosome. Once the proper base-pairing between the mRNA codon and the tRNA anticodon is made, GTP is hydrolyzed and eEF1A-GDP is released (Merrick and Nyborg, 2000). After this happens, the peptidyl transferase center of the ribosome catalyzes the formation of a peptide bond between the incoming amino acid and the peptide in the P site (Moazed and Noller, 1989). The peptidyl-tRNA is translocated to the P-site in a GTP dependent manner and with the involvement of eEF2 (Rodnina *et al*, 1997). After translocation, eEF2-GDP and the uncharged tRNA are released and the ribosome is ready to accommodate the next aminoacyl-tRNA in the empty A-site.

Regulation of elongation cycle targets primarily the elongation factors eEF1A, EF2 and the exchange factor eEF1B which recycles eEF1A-GDP back to the GTP-bound form (Proud, 2000). These three proteins have been shown to undergo phosphorylation and this may be a mechanism for modulation of elongation *in vivo*. Phosphorylation of mammalian eEF2 inhibits its activity by impairing its interaction with ribosomes (Carlberg *et al*, 1990). The protein kinase responsible for the phosphorylation of eEF2 is the eEF2 kinase and its activity entirely depends on the concentration of  $\text{Ca}^{++}$  and cAMP. It has been postulated

that the cAMP dependence helps to inhibit protein synthesis during periods when cellular energy is required for other processes (Proud, 2000). On the other hand, the activity of the eEF1A/B complex may be enhanced by insulin or phorbol esters through phosphorylation by kinases that target three polypeptide components of this complex (eEF1A, eEF1B $\alpha$  and  $\beta$ ) (Venema *et al*, 1991). eEF1A is a substrate for protein kinase C (Venema *et al*, 1991) while eEF1B can be phosphorylated by casein kinase-2 (CK-2) (Janssen *et al*, 1988) and the cell-cycle regulated kinase, p34<sup>cdc2</sup> (Belle *et al*, 1989).

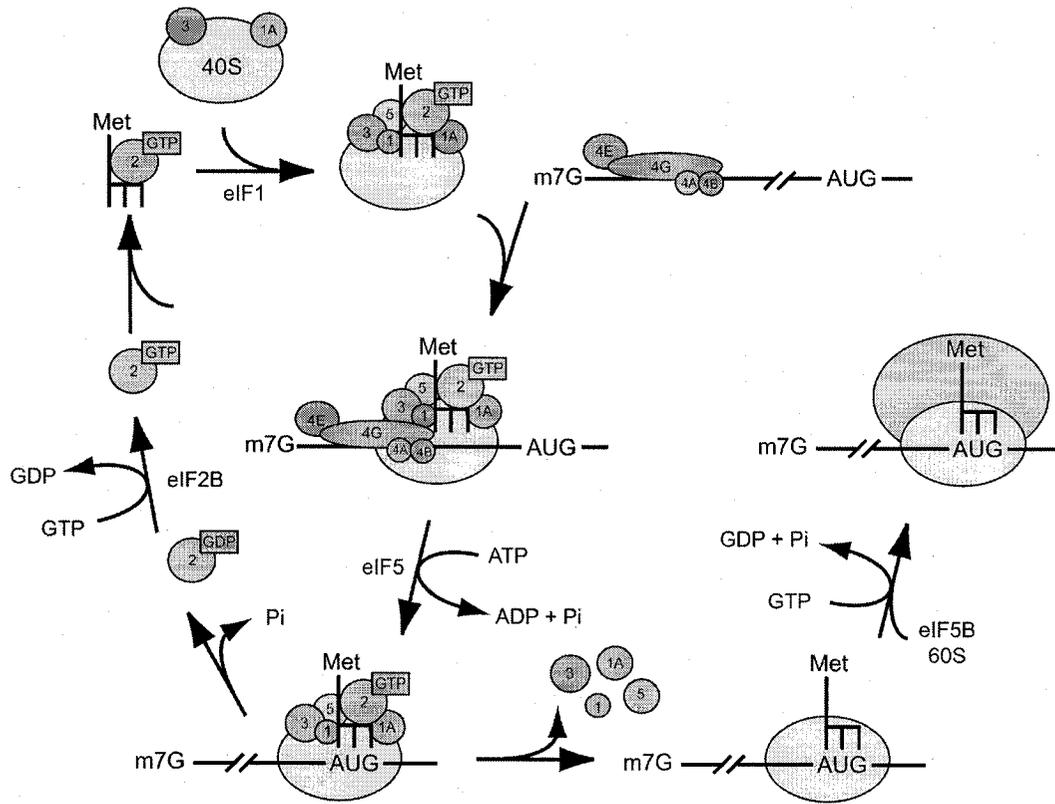
### **1.3.2.3 Eukaryotic translation termination**

In contrast to prokaryotic termination, eukaryotic chain termination requires only one protein factor, eRF1, to recognize all three stop codons (UAA, UAG, and UGA) (Welch *et al*, 2000). eRF1 interacts with eRF3 and together they catalyze the hydrolysis of the peptidyl-tRNA bond and GTP, and release the completed polypeptide chain from the ribosome (Zhouravleva *et al*, 1995). The efficiency of translation termination can be affected by the sequences surrounding the termination codon (Bonetti *et al*, 1995). Trans-acting factors can also modulate translation termination efficiency like the UPF genes (Leeds *et al*, 1992). The efficiency of translation termination can act as a control point to regulate gene expression. An example of this is the headcase gene, *hdc*, in *Drosophila* which has an ORF interrupted by an internal UAA stop codon. Translational readthrough of the internal stop codon is necessary as the long form of the protein is required for *hdc* function (Steneberg *et al*, 1998). Regulated translation termination can also control the gene expression in certain viruses such as the human cytomegalovirus UL4 (Alderete *et al*, 1999).

### **1.3.3 Archaeal translation**

As in transcription, archaeal protein synthesis shares features of both prokaryotes and eukaryotes. Analyses on archaeal mRNAs, revealed that Archaea frequently uses a Shine-Dalgarno sequence 3-10 nt upstream of the start codon (Dennis, 1997). On the other hand, most ribosomal proteins, translation factors

and aminoacyl-tRNA synthetases are similar to their eukaryotic counterparts. Another interesting facet of the archaeal translation apparatus is the large number of posttranslational modifications to the tRNAs and rRNAs, especially among thermophilic Archaea. The elongation and termination steps are achieved with the aid of proteins homologous to elongation factors eEF-1 $\alpha$ , eEF-2 and release factor eRF (Bell and Jackson, 1998).



**Figure 1.9 Mechanism of initiation of protein synthesis in eukaryotes**

See text for details.

#### **1.4 Scope of the thesis.**

The main goal of this thesis is to understand the underlying mechanism of different regulatory systems using nuclear magnetic resonance (NMR) techniques. The systems we studied include proteins involved in the control of gene expression at the transcriptional, post-transcriptional and translational levels. As no structural data existed for any of these proteins, we determined the solution structure of each one and studied the interactions responsible for their function. The structure of these proteins reveal how each one of them can regulate the specific processes where they intervene.

Chapter 2 describes the structural studies I performed on YaeO, a protein involved in the regulation of Rho-dependent transcription termination in bacteria. This project reveals the first structure of a Rho inhibitor. Additionally, we demonstrated that YaeO binds to the primary RNA binding domain of Rho. This study suggests that YaeO inhibits transcription termination by blocking RNA binding to Rho.

In Chapter 3, I investigated the solution structure and RNA interactions of the Carbon storage regulator protein A, CsrA, from *E. coli*. This protein is involved in the post-transcriptional control of genes in carbohydrate metabolism in bacteria. It has been shown that CsrA blocks ribosome binding to its target mRNAs. The binding studies in this section, together with the structure, reveal the potential RNA binding region of CsrA. A model for CsrA recognition of its target mRNAs is also proposed.

Chapter 4 focuses on the archaeal translation initiation factor IF2 $\beta$ . This protein is a key regulator of the overall protein synthesis and its high sequence similarity to its eukaryotic counterpart can help the understanding of the initiation mechanism in higher organisms. The structure of aIF2 $\beta$  was solved and the importance of the zinc ion for the structural integrity of its C-terminus was demonstrated. The function of aIF2 $\beta$  in translation initiation was rationalized in terms of our structure and available structural data of translation factors aIF2 $\alpha$  and aIF2 $\gamma$ .

## **Chapter 2: Structure of YaeO, a Rho-specific inhibitor of transcription termination**

### **2.1 Abstract**

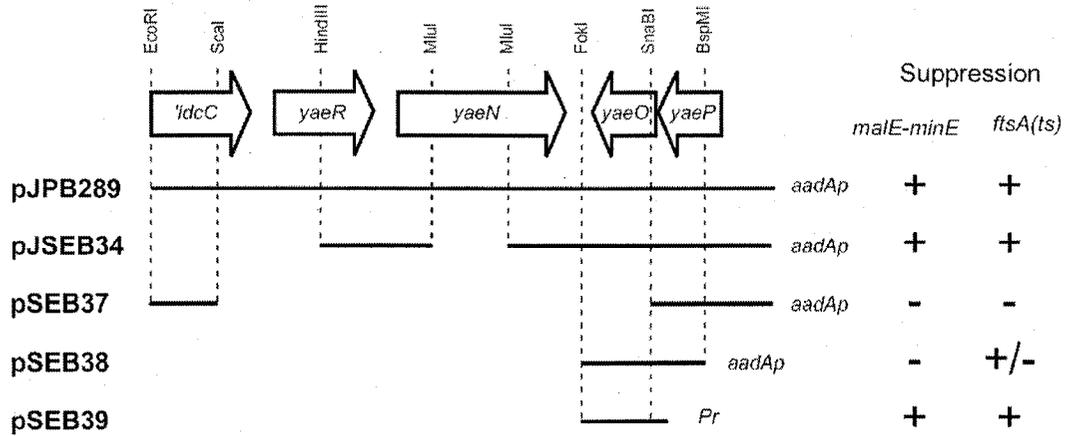
Rho-dependent transcription termination is an essential process for the regulation of bacterial gene expression. Thus far, only two Rho-specific inhibitors of bacterial transcription termination have been described: the *psu* protein from the satellite bacteriophage P4 and YaeO. Here, we report the solution structure of YaeO, which represents the first structure a Rho-specific inhibitor of transcription termination. YaeO is an acidic protein composed of an N-terminal helix and a seven-stranded beta sandwich. NMR chemical shift perturbation experiments revealed that YaeO binds proximal to the primary nucleic acid binding site of Rho. Based on the NMR titration data, a docked model of the YaeO-Rho complex was calculated. These results suggest that YaeO binds outside the Rho hexamer, potentially acting as a competitive inhibitor of RNA binding. In addition, a mechanism detailing the regulation of Rho-dependent transcription termination by YaeO is discussed.

### **2.2 Introduction**

Transcription termination is the process where a nascent RNA is released from its complex with RNA polymerase and DNA template. In bacteria, two main mechanisms of transcription termination have been described. These mechanisms, commonly referred to as Rho-independent and Rho-dependent termination, are essential for the regulation of bacterial gene expression (Richardson and Greenblatt, 1996). Rho-independent termination occurs at a GC-rich self-complementarity region that forms a stem-loop structure believed to cause the RNA polymerase to pause, allowing the release of the RNA (Rosenberg and Court, 1979; Brendel *et al*, 1986). Rho-dependent termination, on the other hand, requires the presence of a hexameric helicase, Rho (Brown *et al*, 1981; Opperman and Richardson, 1994). Rho is an essential transcription factor that binds nucleic acids at specific termination sites (*rut*) and translocates along the RNA until it

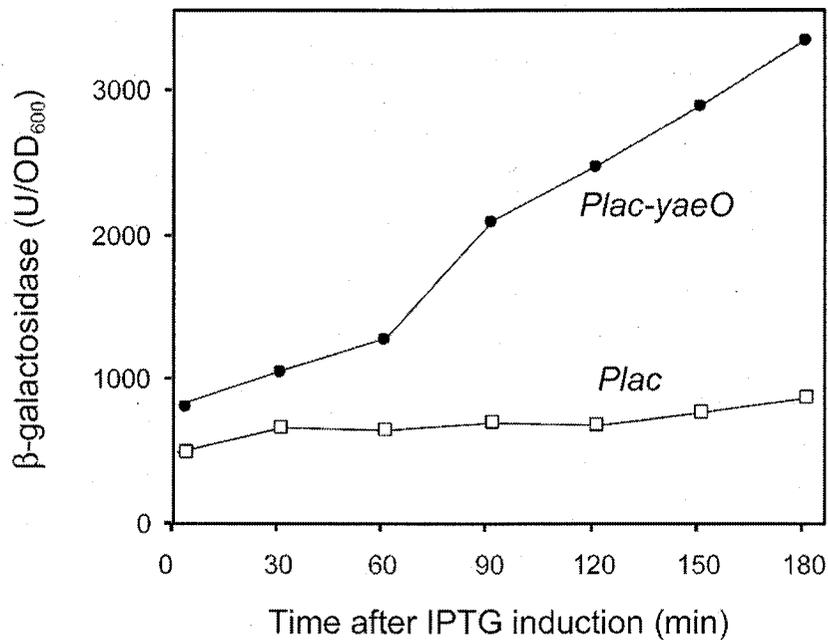
reaches the transcription complex (Geiselmann *et al*, 1993; Platt, 1994; Richardson, 1996). There, it facilitates termination by unwinding RNA/DNA heteroduplexes upon hydrolysis of ATP (Brennan *et al*, 1987).

Currently, only two Rho-specific inhibitors of transcription have been reported. The first to be described is a 21.3 kDa protein encoded by gene *psu* of the satellite bacteriophage P4 (Linderoth and Calendar, 1991). *Psu* interferes with transcription in phage, plasmid and bacterial operons, and its activity does not depend on sequences in the transcript. *In vitro*, protein *Psu* causes efficient read through of Rho-dependent terminators lambda tR1 and TIS2 in a manner that seems to be insensitive to NusG (Linderoth *et al*, 1997). Whether *Psu* inhibits an enzymatic activity of Rho or the interaction of Rho with RNA, ATP, NusG or RNA polymerase is unknown. The second inhibitor is the product of the gene *yaeO*, which has been shown to reduce termination in the Rho-dependent bacteriophage terminator tL1 and upstream the autogenously regulated gene *rho* (Pichoff *et al*, 1998) (Figures 2.1 and 2.2). Additionally, over expression of *YaeO* causes the pleiotropic suppression of some temperature-sensitive mutations. *YaeO* is a 9 kDa acidic protein that binds tightly to Rho but the exact nature of this interaction was unknown (Pichoff *et al*, 1998). Here, the solution structure of *YaeO* is reported, which represents the first structure of a Rho-specific inhibitor of transcription termination. Additionally, the binding surface of the Rho-*YaeO* complex was mapped for both proteins and a mechanism for *YaeO*-mediated regulation suggested.



**Figure 2.1 Chromosomal region around the *yaeO* gene**

*YaeO* was identified as a suppressor of Mal-MinE lethality in *E. coli*. Sequencing of different suppressor plasmids revealed that the target gene was located between *ldcC* to the first codons of *yaeO*. Further subcloning indicated that the suppressor is contained in an operon comprising genes *yaeP* and *yaeO* but only the latter was responsible for the suppression activity. *yaeO* also suppressed some temperature-sensitive mutations in division gene *ftsA*. This figure is based on the work of Sébastien Pichoff and collaborators (Pichoff *et al*, 1998).



**Figure 2.2 Effect of *yaeO* expression on *rho* gene transcription**

Strains JS219/pJPB314/pUC19 (*lacZp*) and JS219/pJPB314/pSEB41 (*lacZp-yaeO*) were grown at 37 °C in LB medium supplemented with 2% glucose. Cultures were maintained below OD<sub>600</sub> = 1 by dilution. The inhibition of Rho function was tested by measuring the expression of gene *rho* in a pSC101-derived plasmid in which most of the *rho* coding region was replaced by *lacZ*. When pSEB41, a derivative of pUC19 carrying the *lacZp-yaeO* fusion was induced, P<sub>*rho-t<sub>rho</sub>-lacZ*</sub> expression increased about four fold 3 hours after addition of IPTG, indicating inhibition of Rho function.

## 2.3 Results and discussion

### 2.3.1 Structure determination of YaeO

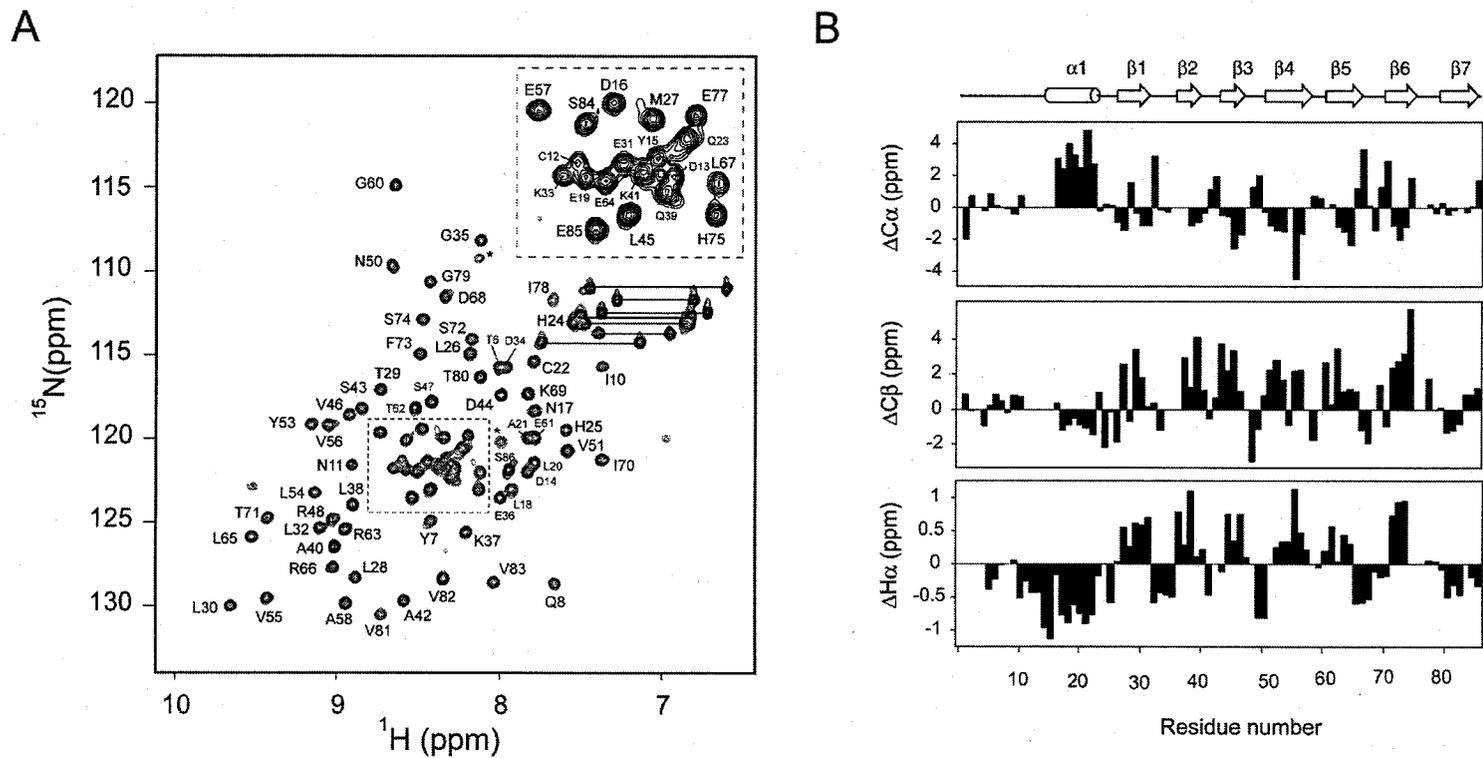
YaeO was produced in *E. coli* as an N-terminal His-tag fusion protein and purified by affinity chromatography. The His-tag was not removed for structural studies as its presence did not perturb the structure of the protein, as evidenced by comparison of  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of the cleaved and uncleaved protein. The protein was labeled uniformly with  $^{15}\text{N}$  or with  $^{15}\text{N}$  and  $^{13}\text{C}$  for NMR analysis. Backbone resonance assignments were obtained with standard triple resonance NMR experiments. The overall structure of YaeO is well defined by NMR data, except for residues 1-8 and 84-86, which are not structured. These residues have almost no long range NOEs show  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOEs close to zero, which are indicative of high mobility in solution (Fig. 2.5). From secondary chemical shift analysis, we deduced that YaeO is composed of one alpha helix and a seven stranded  $\beta$ -sandwich (Fig. 2.3) (Wishart and Sykes, 1994).

YaeO is a member of a family of Rho-dependent transcription termination proteins (Rof). Rof proteins are present in several species of Gram-negative and Gram-positive bacteria, indicating a widespread and conserved function. The location of secondary structure elements relative to the primary sequence reveals that helix  $\alpha 1$  is the best-conserved secondary structure element (Fig. 2.4). Variations exist in the  $\beta 3$ - $\beta 4$  and  $\beta 6$ - $\beta 7$  loops, possibly reflecting differences in their function. Highly conserved residues cluster at helix  $\alpha 1$  or form the hydrophobic core of YaeO. Residues C12, D16, E19, C22 and E52 are completely conserved. The side chains of these residues are located in the exposed surface of YaeO suggesting that their conservation might reflect their biological function. Figure 2.5 shows a stereo superposition of the ten lowest energy NMR structures. The helical region comprises residues 14-23. The two faces of the sandwich are formed by antiparallel beta strands  $\beta 2/\beta 1/\beta 6/\beta 7$  and  $\beta 3/\beta 4/\beta 5$ , respectively. Beta strands one to seven comprise residues 26-32, 36-42, 44-48, 51-58, 61-65, 71-75 and 79-83. Multiple non-polar residues pack between both strands to form a structural core, which includes L28, L30, L54, V56, I70, F73, I78 and V81. The

folded regions of YaeO have random mean square deviations (rmsd) of 0.53 Å for backbone atoms and 1.08 Å for all heavy atoms (Table 2.1). The electrostatic surface representation of YaeO shows that  $\alpha$ 1 together with  $\beta$ 1,  $\beta$ 6, and  $\beta$ 7 form a highly negatively charged patch due to residues D13, D16, E19 and E31 (Fig 2.6).

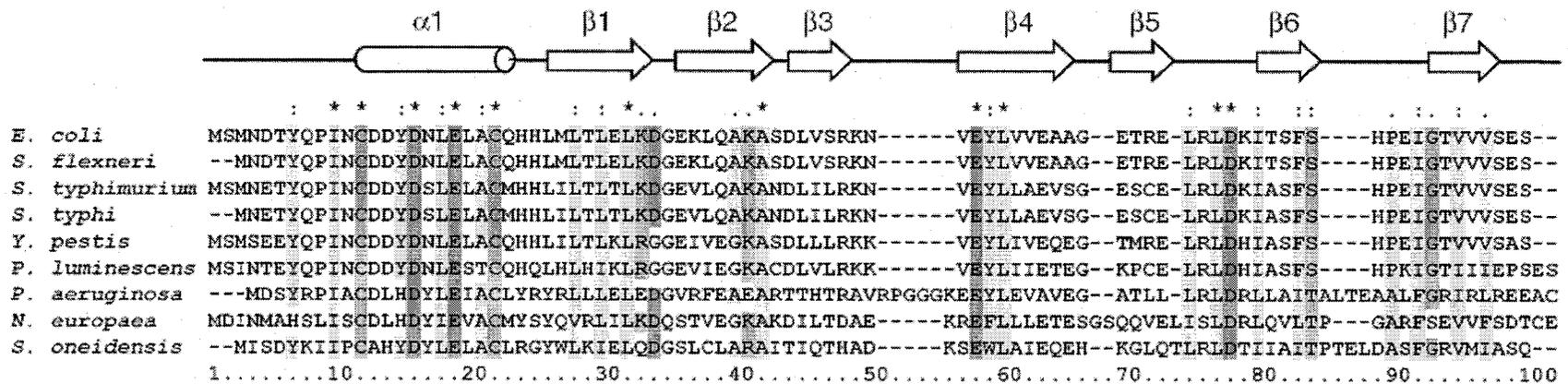
### 2.3.2 Comparative analysis of YaeO

Structure based searches of the Protein Data Bank with the program DALI revealed twelve similar structures with a Z-score larger than 2.0. The most prominent of these are the pleiotropic translational regulator Hfq (1QK1), the heptameric archaeal Sm protein (1I8F), the small nuclear ribonucleoprotein sm d3 fragment (1D3B) and the small ribonucleoprotein sm (1OU8). Z-scores were 4.4, 3.7 3.3, and 3.0 respectively with rms deviations ranging from 2.7 to 3.3 Å for matching C $\alpha$  atoms. These results reveal that, in spite of the lack of recognizable sequence similarity, the fold of YaeO is topologically similar to that of the RNA binding domain of small ribonucleoproteins (Sm-fold). The most important difference between the Sm-fold and YaeO is the presence of an additional  $\beta$ -strand ( $\beta$ 7) in YaeO. It is important to note that, even though most of the hits are transcriptional regulators or involved in mRNA processing, nucleic acid titrations performed on YaeO did not reveal any observable interaction.



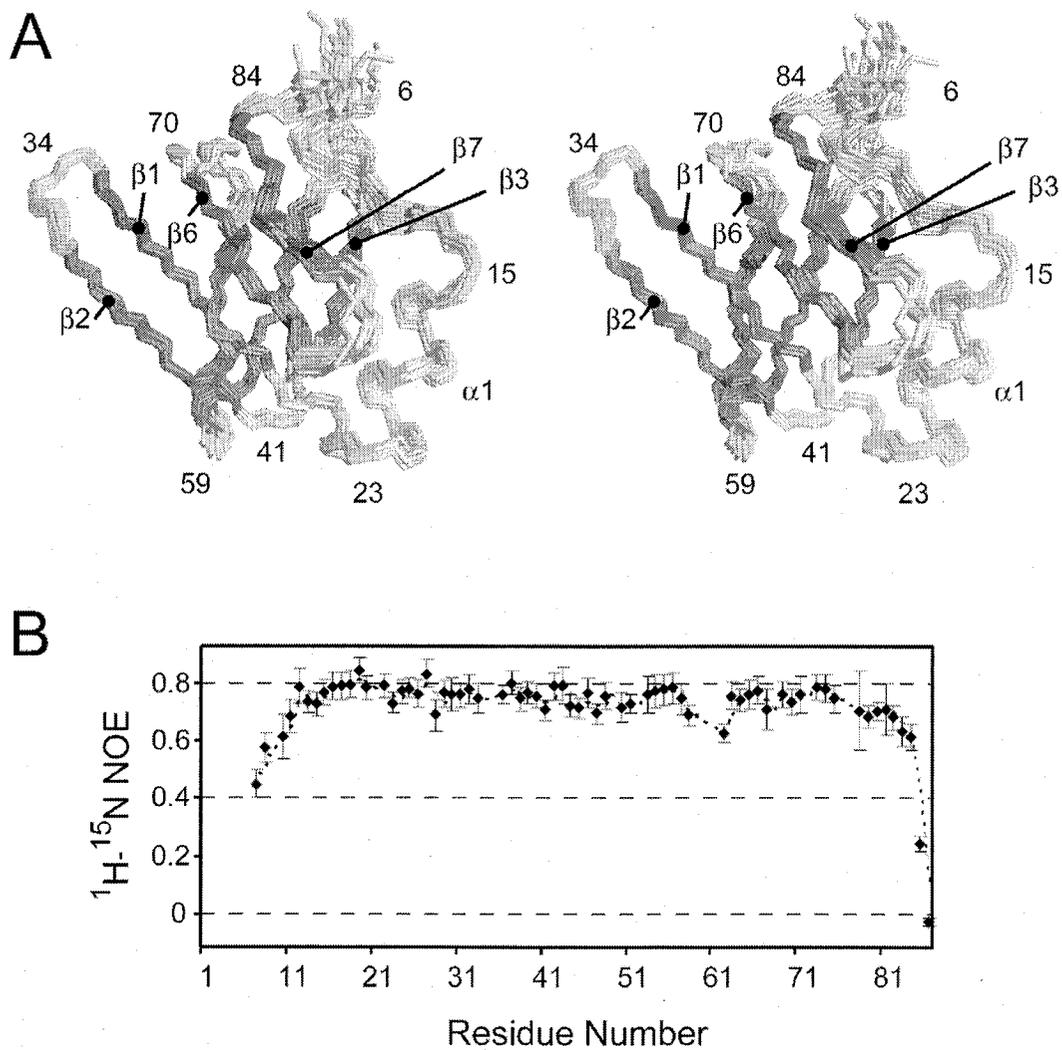
**Figure 2.3**  $^{15}\text{N}$ -HSQC and secondary chemical shifts of YaeO

(A) The backbone amide crosspeaks are labelled with the corresponding amino acid. The unlabelled crosspeaks belong to residues in the His-tag and the glutamine and asparagine sidechains. (B) Plots of the difference between YaeO  $\text{C}\alpha$ ,  $\text{C}\beta$  and  $\text{H}\alpha$  chemical shifts and random coil values.



**Figure 2.4 Sequence conservation of the ROF family of proteins**

Sequences are labeled with their corresponding Swiss-Prot accession code as follows: *Escherichia coli*, Q8X8W7; *Salmonella typhimurium*, Q8ZRN4; *Salmonella typhi*, Q8Z995; *Yersinia pestis*, Q8ZH48; *Photobacterium luminescens*, Q7N8M9; *Pseudomonas aeruginosa*, Q9I519; *Nitrosomas europaea*, Q82XR7; *Shewanella oneidensis*, Q8EJW7; *Vibrio cholerae*, Q9KLO2; *Vibrio vulnificus*, Q8D5E4; *Vibrio parahaemolyticus*, Q87JZ9. Completely conserved acidic, basic, polar and hydrophobic residues are colored red, blue, green and gray, respectively. Glycine residues are yellow. The location of secondary structure elements is shown on top



**Figure 2.5 NMR ensemble of YaeO**

(A) Stereo view showing the ensemble of NMR-derived structures of YaeO, with strands in purple and helices in green. The 10 lowest energy structures are superimposed to the mean. Numbering indicates the amino acid number. (B)  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOE values for backbone amides.

**Table 2.1. Structural statistics for 20 selected conformers for YaeO****Constraints used for structure calculation**

Intraresidue NOEs	(n=0)	270
Sequential range NOEs	(n=1)	253
Medium range NOEs	(n=2,3,4)	68
Long range NOEs	(n>4)	159
Dihedral angle constraints		107
Hydrogen bonds		33
<sup>15</sup> N- <sup>1</sup> H residual dipolar couplings		61
Total number of constraints		951

**Final energies (kcal/mol)**

E <sub>total</sub>	265.67 ± 4.57
E <sub>bond</sub>	8.76 ± 0.55
E <sub>angle</sub>	96.70 ± 2.65
E <sub>improper</sub>	19.52 ± 1.02
E <sub>vdw</sub>	80.94 ± 3.39
E <sub>noe</sub>	30.95 ± 1.68
E <sub>dihedral</sub>	9.02 ± 0.61
E <sub>sani</sub>	19.75 ± 1.53

**RMS deviation from idealized covalent geometry**

Bonds (Å)	0.0025 ± 0.00001
Angles (°)	0.5081 ± 0.0072
Improper (°)	0.4390 ± 0.01134

**Average RMS difference to mean structure (Å)**

Backbone atoms	0.40 ± 0.12
All heavy (non-hydrogen) atoms	1.04 ± 0.10
All atoms	1.22 ± 0.08

**RMS deviation from NMR restraints**

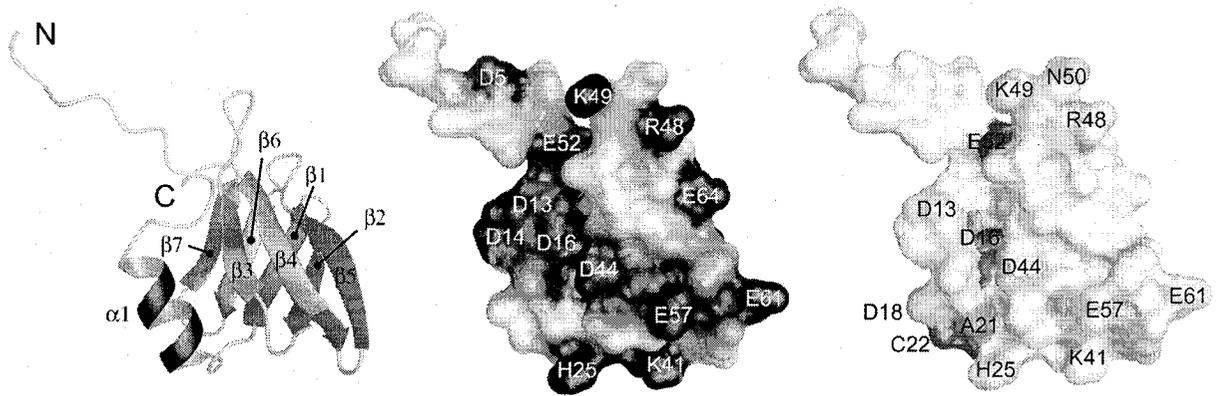
Distance restraints (Å)	0.0241 ± 0.0007
Dihedral angle restraints (°)	0.8315 ± 0.0281

**Average Ramachandran statistics (%)**

Residues in most favored regions	87.1
Residues in additional allowed regions	10.8
Residues in generously allowed regions	2.1
Residues in disallowed regions	0.0

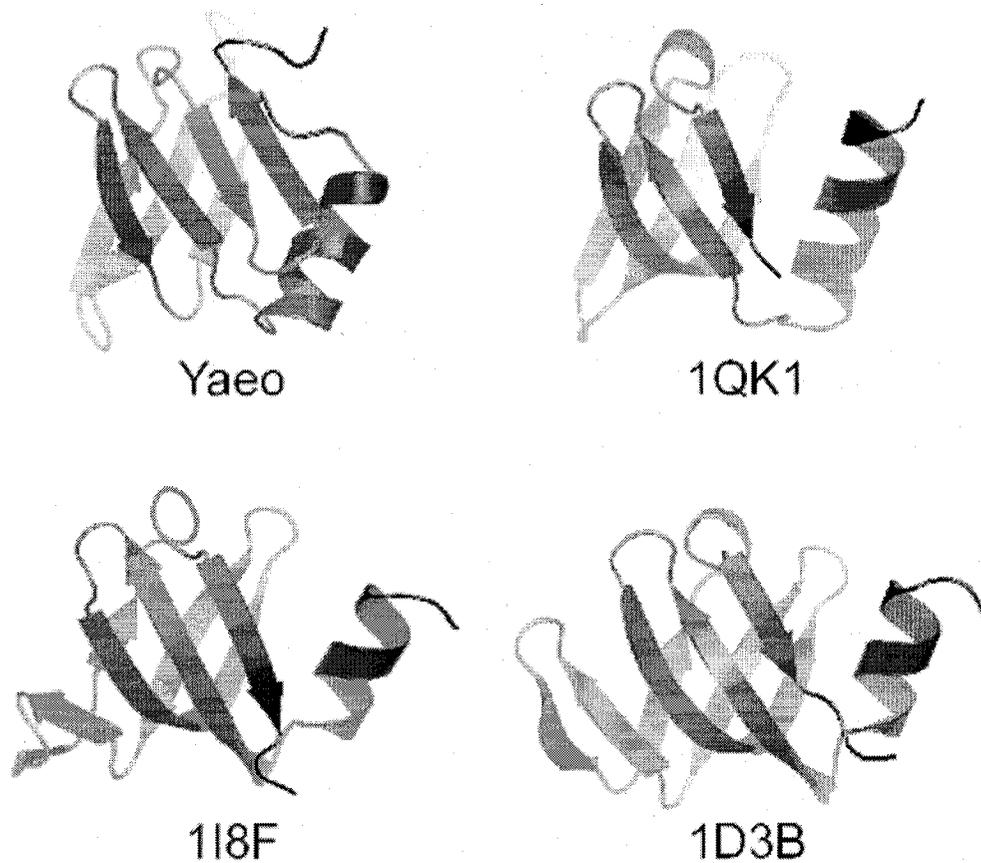
**Analysis of residual dipolar couplings**

RMSD (Hz)	1.390 ± 0.035
Q-factor	0.138 ± 0.003



**Figure 2.6 Charge distribution and sequence conservation of YaeO**

Ribbon display and surface plots of the YaeO structure. The surface electrostatic potential (middle) was calculated using MOLMOL (Koradi *et al.*, 1996). red, white and blue correspond to negative, neutral and positive potential, respectively. The degree of sequence conservation was derived from Clustal-X (Left). White to green defines the range for non conserved to 100% conserved residues.



**Figure 2.7 Proteins structurally related to YaeO**

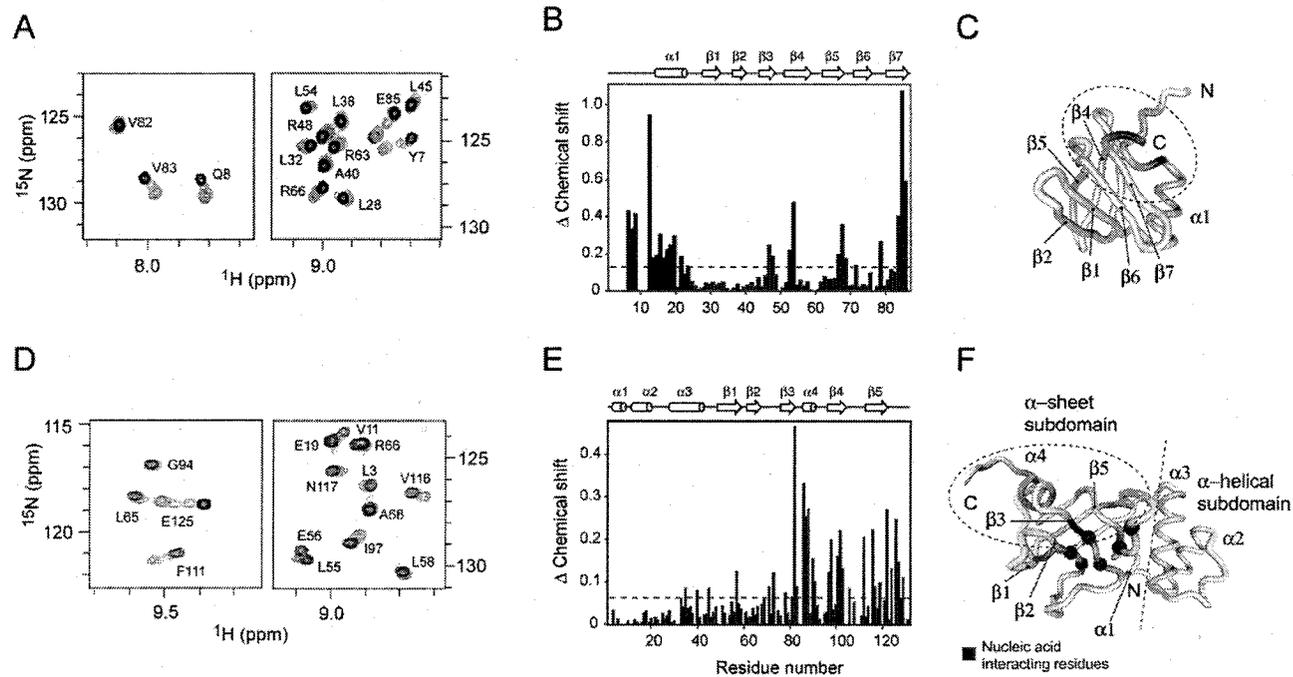
According to the program DALI (Holm and Sander, 1998), YaeO shares significant structural similarity ( $Z$ -score  $> 2.0$ ) proteins containing the Sm fold. PDB codes correspond to the pleiotropic translational regulator Hfq (1QK1), the heptameric archaeal Sm protein (1I8F) and the small nuclear ribonucleoprotein sm d3 fragment (1D3B).  $Z$ -scores were 4.4, 3.7 and 3.3 respectively with rms deviations ranging from 2.7 to 3.3 Å for matching  $C\alpha$  atoms.

### 2.3.3 Interactions between YaeO and Rho

In order to understand the mechanism of transcription termination inhibition by YaeO, NMR experiments were performed to observe the interaction of YaeO with Rho *in vitro*. A truncated version of Rho, Rho130, from *E. coli* was constructed. This fragment corresponds to the primary RNA binding site of Rho (residues 1-130) and has been shown to be a good model of Rho-oligonucleotide interactions (Briercheck *et al*, 1998). We performed an NMR titration by recording a series of  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -labeled YaeO as a function of unlabeled Rho130 concentration. Complex formation occurred, as evidenced by chemical shift perturbations in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra (Fig. 2.8). Comparison of the bound and free spectra allowed us to map the binding site of Rho on YaeO. The largest chemical shifts occurred in the N and C-termini, helix  $\alpha$ 1 and strands  $\beta$ 3,  $\beta$ 4,  $\beta$ 5 and  $\beta$ 7. These regions localize to one edge of the  $\beta$ -sandwich with clustered acidic residues. These results suggest that the unfolded N and C-termini of YaeO become structured upon binding to Rho.

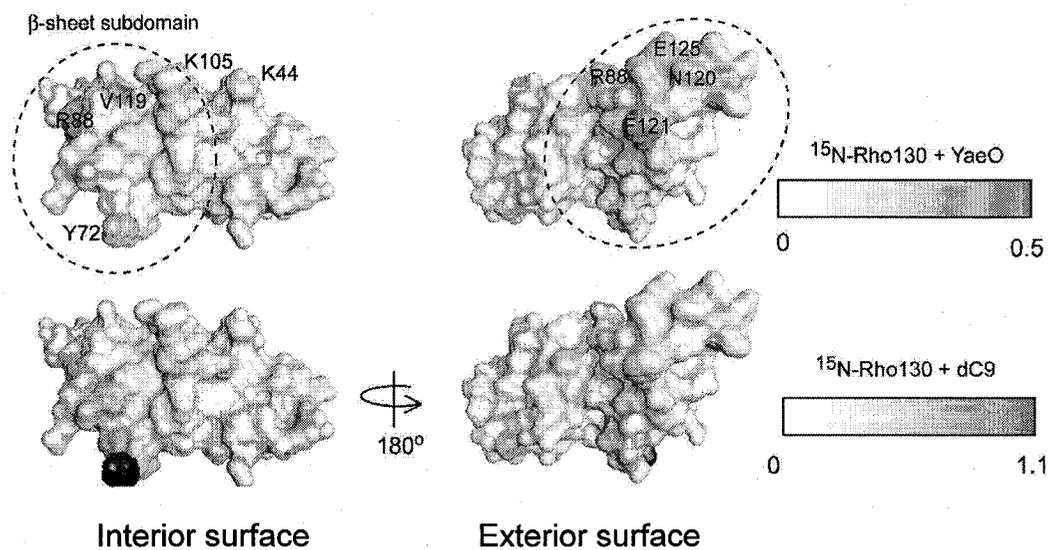
As the structure of Rho130 has been solved by NMR (Briercheck *et al*, 1998), we decided to map the interaction with YaeO on Rho130. Rho130 was uniformly  $^{15}\text{N}$ -labeled and the  $^1\text{H}$ - $^{15}\text{N}$  HSQC peaks were assigned using the chemical shifts for the previous determined NMR structure, kindly provided by Dr. G. Rule (Briercheck *et al*, 1996). Rho130 is composed of an  $\alpha$ -helical (residues 1-47) and a  $\beta$ -sheet sub-domain (residues 48-130). Chemical shift perturbation analysis showed that YaeO binds primarily to strands  $\beta$ 3,  $\beta$ 4,  $\beta$ 5 and helix 4 of the  $\beta$ -sheet sub-domain. Minor shifts were also observed for helix 3 and strands 1 and 2.

Structural evidence has shown that RNA binds to the N-terminal domain of Rho mainly by contacts with F64, R66, D78, Y80, E108 and Y110 (Skordalakes and Berger, 2003). These residues are neighbors of the YaeO interacting region of Rho suggesting a mechanism of transcription termination inhibition by blocking the RNA binding site of Rho (Figure 2.9).



**Figure 2.8 NMR chemical shift changes upon YaeO-Rho130 binding**

(A)  $^{15}\text{N}$ -HSQCs of YaeO during titration with Rho130. Contours are colored according to the YaeO-Rho130 ratio (1:0, blue; 2:1, green; 1:1, orange and 1:2, red). (B) Chemical shift changes from the titration in A. (C) Mapping of chemical shift changes onto the structure of YaeO. (D)  $^{15}\text{N}$ -HSQCs of Rho130 during titration with YaeO. Contours colored according to the Rho130-YaeO ratio (1:0, blue; 3:1, green; 3:2, orange and 1:1, red). (E) Chemical shift changes from the titration in D. (F) Mapping of chemical shift changes onto the structure of Rho130.

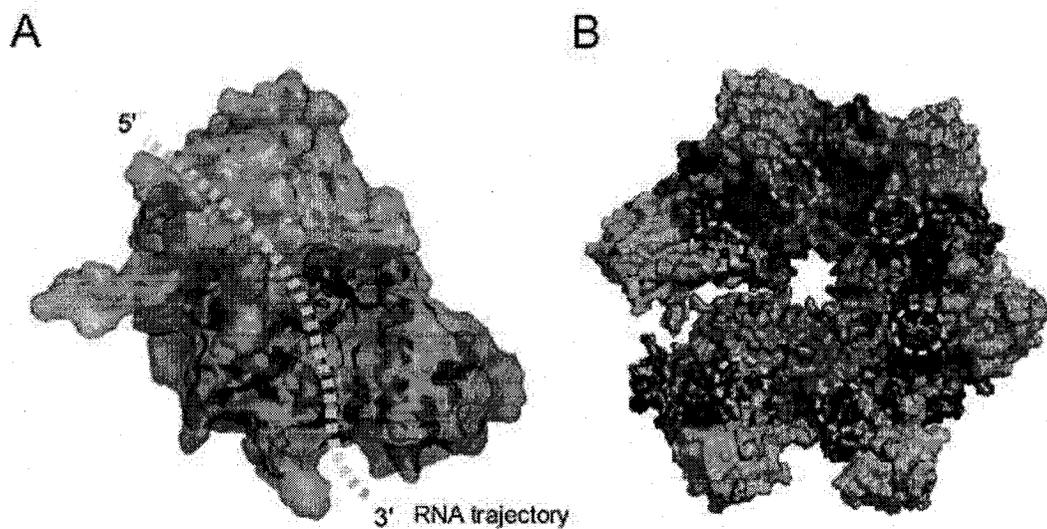


**Figure 2.9 YaeO and (dC)<sub>9</sub> binding to Rho130**

YaeO binds primarily to the external surface of the  $\beta$ -sheet subdomain of Rho130. This site is proximal to the (dC)<sub>9</sub> binding site but does not overlap with it. This suggests that YaeO inhibition is due to blocking of the RNA exit from the primary RNA binding pocket.

#### 2.3.4 Model of the Rho-YaeO complex

As NMR titration data for the YaeO-Rho interaction was available for both proteins, a model of the complex was built using high ambiguity driven protein docking (HADDOCK) (Dominguez *et al.*, 2003). Ambiguous interaction restraints were derived from the NMR titration data by selecting residues with both the biggest chemical shifts and solvent accessibility. Unfolded residues were allowed to move freely during the docking protocol. The model with the lowest intermolecular energy after the last stage of refinement in Cartesian space with explicit solvent was selected. The interaction between YaeO and Rho seems to be facilitated by charge complementarity and burial of  $\sim 2558 \text{ \AA}^2$  of surface area. This is consistent with *in vitro* binding results that show the YaeO-Rho interaction is salt dependent and can be disrupted at high ionic strength (0.4 M KCl). (Pichoff *et al.*, 1998). The docking model suggests seven potential salt bridges: D5:R87, D44:K100, D44:K115, E:52:R88, E57:R102 and E64:K105. The model is compatible with the hexameric structure Rho as there are no clashes when the YaeO-Rho130 model is superposed to the hexameric complex.



**Figure 2.10 Docking model of the YaeO-Rho130 complex**

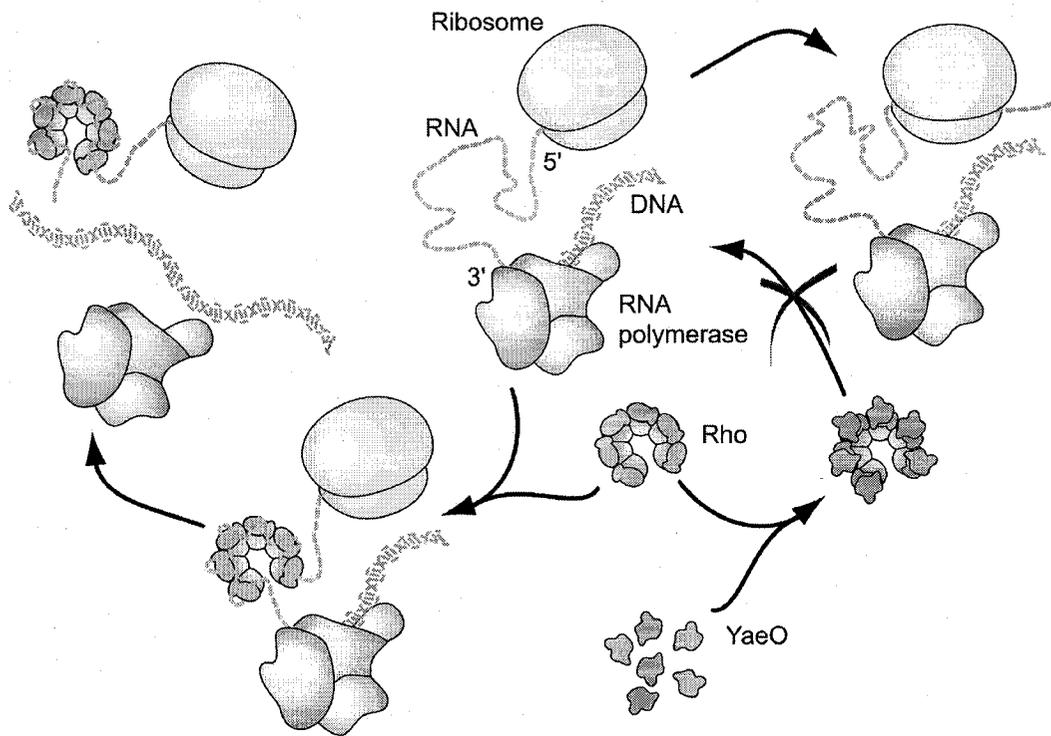
(A) Lowest energy YaeO-Rho130 complex calculated with the program HADDOCK. Chemical shift perturbation data shown in Figure 2.8 was used for the docking. YaeO is colored green and Rho130 is magenta. (B) Model of YaeO bound to hexameric Rho and RNA (yellow). The ATP binding domain is shown in blue

### 2.3.5 Discussion

Our data demonstrate the YaeO binds to the transcription factor Rho and provide a structural basis for its inhibitory effects on transcription termination. It is likely that YaeO binds to the Rho hexamer in a 1:1 monomer-to-monomer ratio. YaeO inhibition likely results from a reduced affinity of Rho for RNA when YaeO is bound. YaeO probably acts as a competitive inhibitor of RNA binding. Previous studies have suggested that the pathway of RNA binding to hexameric Rho consist of four steps: PR1, PR2, PR3 and PR4 (Kim and Patel, 2001). State PR1 is formed when the RNA binds to primary binding site of Rho. Bound RNA then fills the continuous binding sites in the crown to form PR2. The third step consists of the opening of the ring leading to passage of the RNA through the central channel to form PR3. Finally, PR4 is formed when the ring closes rendering Rho competent in ATPase and translocation activities. Our data suggest that YaeO may act by inhibiting the formation of the PR1 and/or PR2 states.

Little is known about the way *yaeO* expression is regulated, however, previous studies determined that high expression of YaeO is required for efficient inhibition of transcription termination. This is not surprising as Rho represents approximately 0.3% of the total protein content in *E. coli*. More studies need to be done in order to understand the mechanism of Rho-inhibition by YaeO and whether its action is in concert with other auxiliary factors such as NusG or NusA.

Rho is the target of bicyclomycin (BCM), an antibiotic biosynthetically derived from leucine and isoleucine. Bicyclomycin has been shown to possess antimicrobial activity against Gram-negative bacteria such as *Escherichia coli*, *Shigella*, and *Salmonella* and gram positives as *Micrococcus luteus* (Magyar *et al*, 1996). BCM has been used for the treatment of nonspecific diarrhea in humans and bacterial diarrhea in calves and swine. This study shows a new possibility for the rational design of antibiotics targeting Rho based on the knowledge of the structure of YaeO.



**Figure 2.11 Proposed mechanism of inhibition by YaeO**

YaeO may inhibit transcription termination by binding to each RNA binding domain in the Rho hexamer. This will prevent Rho to bind to the termination signal in the mRNA allowing transcription of downstream genes.

## 2.4 Materials and methods

### 2.4.1 Sample preparation

The gene *yaeO* from *E. coli* K12 was subcloned into pET15b (Novagen, Inc., Madison, WI) and expressed in *E. coli* BL21 as an oligo-histidine (His tag) fusion protein of 106 residues. Cells were grown at 37 °C to an OD<sub>600</sub> of 0.8 and induced with 1 mM IPTG. Afterwards, the temperature was reduced to 30°C and the cells were allowed to express the protein for 3 hours before harvesting. The media used were either LB or M9 minimal media containing <sup>15</sup>N ammonium chloride and/or <sup>13</sup>C-glucose (Cambridge Isotopes Laboratory, Andover, MA). YaeO was purified by affinity chromatography on Ni<sup>2+</sup>-loaded chelating sepharose (Amersham Pharmacia Biotech, Piscataway, NJ). NMR samples were ~2 mM protein in 50 mM phosphate buffer (1 mM NaN<sub>3</sub>, 2 mM DTT, pH 7.0). Rho130 was cloned, expressed and purified in a similar fashion.

### 2.4.2 NMR spectroscopy

NMR experiments were recorded at 303 K on a Bruker Avance 600 MHz spectrometer. Backbone and side-chain assignments of YaeO were determined using HNCACB, CBCA(CO)NH, <sup>15</sup>N-edited TOCSY and <sup>13</sup>C-edited TOCSY. NOE data for the structure determination were obtained from homonuclear NOESY, <sup>15</sup>N-edited or <sup>13</sup>C-edited 3D NOESY experiments. Evaluation of spectra and manual assignments were completed with XEASY (Bartels *et al*, 1995). IPAP-HSQC experiments for measuring <sup>15</sup>N-<sup>1</sup>H dipolar couplings were recorded on an isotropic medium and a sample containing 18 mg/ml Pfl phage (Hansen *et al*, 1998; Ottiger *et al*, 1998). <sup>15</sup>N-<sup>1</sup>H heteronuclear NOE data were measured by taking the ratio of peak intensities from experiments performed with and without <sup>1</sup>H presaturation. Hydrogen bond constraints were introduced to secondary structure regions as determined by chemical shift analysis, and characteristic NOE patterns.  $\phi$  and  $\psi$  dihedral restraints were obtained using the TALOS (Cornilescu *et al*, 1999). All NMR spectra were processed using either XWINNMR version

2.5 or 3.1 (Bruker Biospin) or GIFA (Malliavin *et al*, 1998). Evaluation of spectra and manual assignments were completed with XEASY (Bartels *et al*, 1995).

#### **2.4.3 Analysis and structure calculations**

CNS 1.1 software (Brunger *et al*, 1998) was used to generate an initial fold of YaeO with a basic set of NOEs acquired manually from manual assignments of 3D <sup>15</sup>N-edited NOESY and 2D homonuclear NOE spectra including dihedral angle and hydrogen bond constraints (Wüthrich, 1986). These calculations generated a fold that was used as a model template for automated assignments by ARIA1.1 (Nilges *et al*, 1997). The final structure of YaeO was calculated using the constraints in Table 2.1 and collected from the experiments described above. In the final round of calculations, CNS 1.1 was extended to incorporate RDC restraints for further refinement. The axial and rhombic components of the alignment tensor were defined from a histogram of measured RDCs (Clare *et al*, 1998) and optimized by a grid search method (Clare *et al*, 1998). Twenty structures were selected based on the lowest overall energy and least violations to represent final structures. PROCHECK-NMR was used to generate Ramachandran plots to check the protein's stereochemical geometry (Laskowski *et al*, 1993). The coordinates of YaeO have been deposited in the RCSB under PDB code 1SG5.

#### **2.4.4 Ligand titration**

Ligand titration experiments were performed by recording a series of <sup>15</sup>N-HSQC spectra on uniformly <sup>15</sup>N-labeled YaeO or Rho130 (~2 mM) in the presence of different amounts of ligand in the 1-2.0 mM range. The protein sample and stock solutions of the ligands were all prepared for NMR as described above.

#### **2.4.5 $\beta$ -galactosidase assay**

Strains JS219/pJPB314 (*lacZp*) and JS219/pJPB314/pSEB41 (*lacZp-yaeO*) were grown at 37 °C in Luria broth supplemented with 0.2% glucose.

Samples were taken every 30 minutes for measurements of  $\beta$ -galactosidase activity. Cultures were maintained below  $OD_{600}=1$  by dilution.  $\beta$ -galactosidase assays were performed as described by Miller (Miller, 1972). Strains and plasmids were kindly provided by Dr. J-P. Bouché (CNRS, Toulouse Cédex, France).

#### **2.4.6 Docking**

Docking of YaeO and Rho130 was done using HADDOCK (Dominguez *et al*, 2003). Ambiguous interaction restraints (AIRs) were defined based for residues with chemical shifts perturbation above the average and at least 40% solvent exposed. Mobile regions were determined based on heteronuclear NOE data. 200 structures were calculated during the first step of rigid body energy minimization followed by 50 structures of semirigid simulated annealing in torsion angle space. The best model was selected based on the RMSD at the interface and the intermolecular energy.

## Chapter 3: Solution structure of the carbon storage regulator protein CsrA from *E. coli*

### 3.1 Abstract

The carbon storage regulator A (CsrA) is a protein responsible for the repression of a variety of stationary-phase-genes in bacteria. In this work we describe a NMR-based structure of the CsrA dimer and its RNA-binding properties. CsrA is a dimer of two identical subunits, each composed of five strands, a small  $\alpha$ -helix and a flexible C-terminus. NMR titration experiments suggest that the  $\beta$ 1-  $\beta$ 2 and  $\beta$ 3-  $\beta$ 4 loops and the C-terminal helix are important elements in RNA binding. Even though the  $\beta$ 3-  $\beta$ 4 loop contains a highly conserved RNA binding motif, GxxG, typical of KH domains, our structure excludes CsrA from being a member of this protein family, as was previously suggested. We also propose a model for the recognition of mRNAs downregulated by CsrA.

### 3.2 Introduction

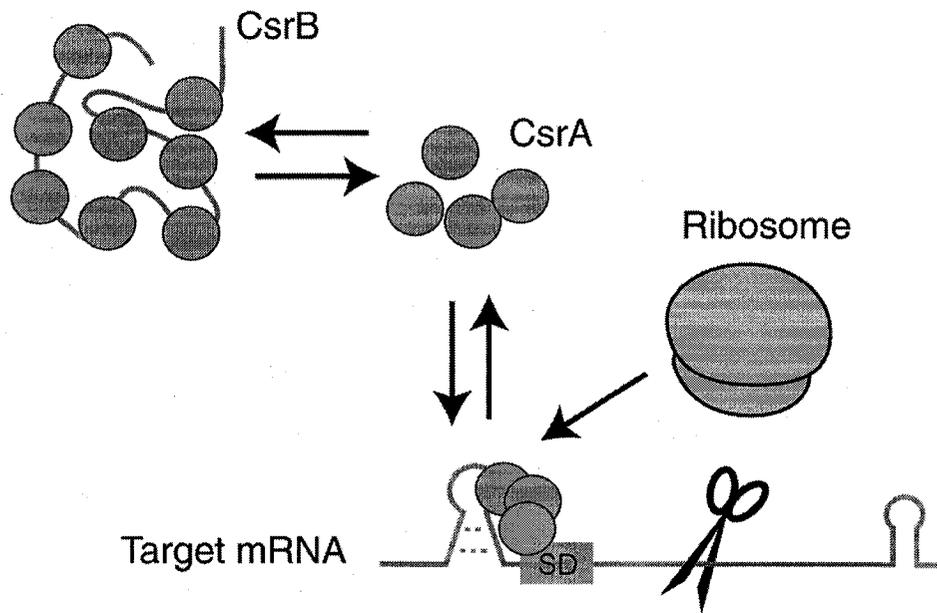
The carbon storage regulator A (CsrA) is a central component of the global regulatory system Csr, which is responsible for the repression of a variety of stationary-phase-genes (Romeo, 1998). CsrA negatively regulates gluconeogenesis, glycogen biosynthesis and catabolism, and biofilm formation (Romeo *et al*, 1993; Sabnis *et al*, 1995; Jackson *et al*, 2002). Additionally, CsrA can activate glycolysis, acetate metabolism and flagellum biosynthesis (Sabnis *et al*, 1995; Wei *et al*, 2000; Wei *et al*, 2001). CsrA acts post-transcriptionally by repressing gene expression of essential enzymes in the carbohydrate metabolism like ADP-glucose pyrophosphorylase (*glgC*), glycogen synthase (*glgA*), glycogen branching enzyme (*glgB*), and glycogen phosphorylase (*glgP*) (Figure 3.1). CsrA destabilizes target mRNAs by binding in a region within -18 and +31 nucleotides of the coding region, which includes the ribosome-binding site (Liu *et al*, 1995). This prevents translation of the corresponding mRNA and promotes its degradation by endogenous RNases. As a result, a decrease in the intracellular

levels of the glycogen biosynthetic enzymes and decreased synthesis of intracellular glycogen is observed.

Intracellular levels of CsrA are regulated by two untranslated RNA molecules, CsrB and CsrC, that act as antagonists by sequestering CsrA and preventing its binding to target mRNAs (Liu and Romeo, 1997; Gudapaty *et al*, 2001; Weilbacher *et al*, 2003). CsrA binding to both CsrB and CsrC seems to be mediated by a highly repetitive sequence element, 5'-CAGGA(U,C,A)G-3', located in the loops of predicted CsrB/C hairpins (Liu and Romeo, 1997; Weilbacher *et al*, 2003). It has been proposed that CsrA exists in equilibrium between CsrB/C and CsrA-regulated mRNAs, implying that CsrB/C levels are a key determinant of CsrA activity in the cell (Figure 3.2).

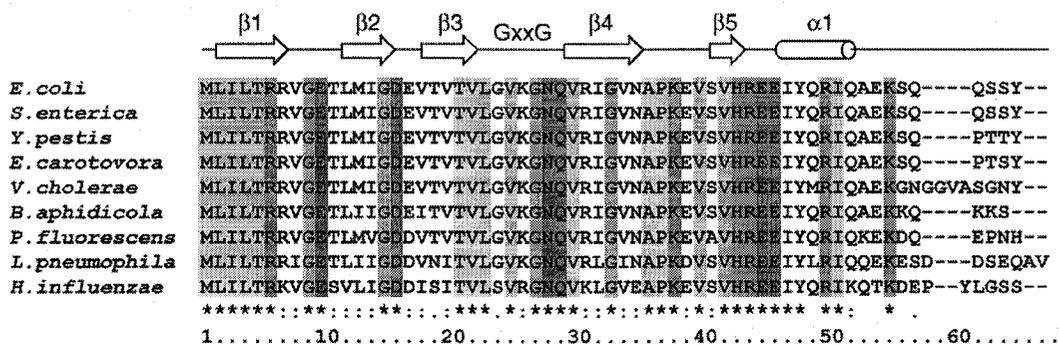
CsrA homologues have been recognized for important roles in the regulation of stationary phase gene expression in other bacterial species (Figure 3.3). The CsrA homologue (RsmA) of *Erwinia* species regulates a variety of genes involved in soft-rot disease of higher plants (Cui *et al*, 1995; Cui *et al*, 1999). *csrA* and *csrB* in *Salmonella enterica* serovar *Typhimurium* regulate genes involved in epithelial cell invasion by this species (Altier *et al*, 2000). In *Pseudomonas aeruginosa*, the Csr (Rms) system controls the quorum sensing systems Las and Rhl, which regulate several of its virulence factors.





**Figure 3.2 Model of CsrA-mediated regulation of *E. coli glgC***

The RNA CsrB functions as an antagonist of CsrA action. CsrA inhibits translation of *glgC* by binding to two positions within the *glgCAP* leader. The upstream site is contained within a short RNA hairpin and CsrA while the second site overlaps the *glgC* SD sequence. The absence of translation allows endonucleolytic cleavage, resulting in rapid degradation of the message (Figure adapted from (Baker, Morozov et al. 2002))



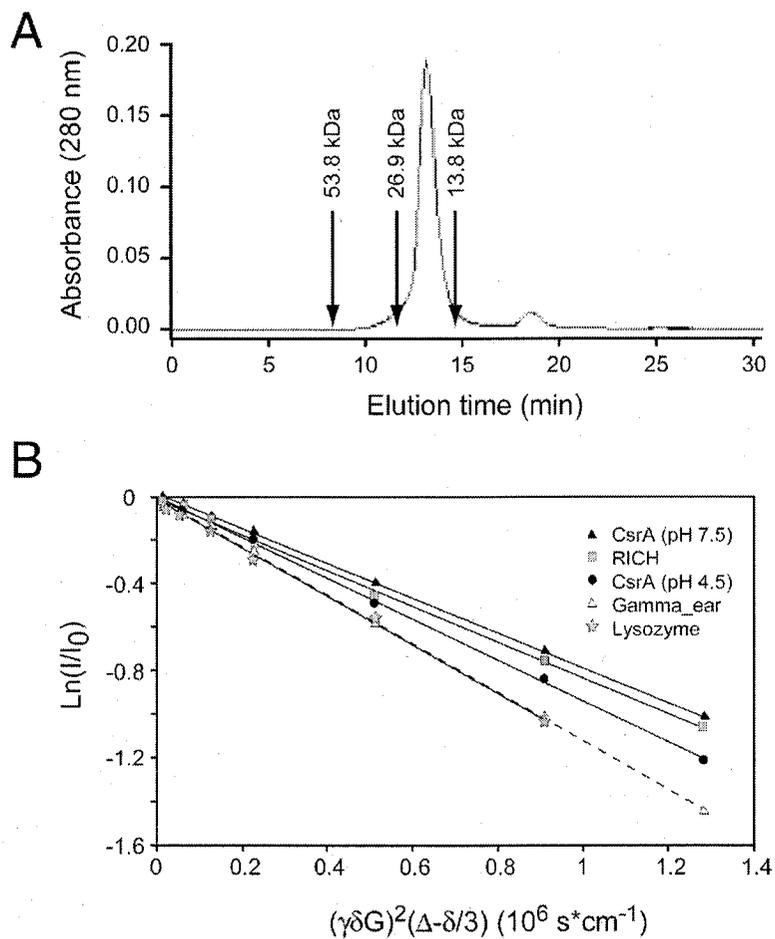
**Figure 3.3 Sequence conservation of CsrA**

The CsrA sequence from *Escherichia coli* K12 (gi161306043) is aligned with homologous proteins from *Salmonella enterica* (gi:16766132), *Yersinia pestis* (gi:16123457), *Erwinia carotovora* (gi:50122288) *Vibrio Cholerae* (gi:9654977), *Buchnera aphidicola* (gi:21672665), *Pseudomonas fluorescens* (gi:38489882), *Legionella pneumophila* (gi:54296805) and *Haemophilus influenzae* (gi:16272754). Completely conserved acidic, basic, polar and hydrophobic residues are colored red, blue, green and gray, respectively. Glycine residues are yellow. The location of secondary structure elements is shown on top.

### 3.3 Results and discussion

#### 3.3.1 Structure determination of the CsrA dimer

Mass spectrometry of cross-linked CsrA demonstrated that CsrA exists in solution as a dimer of identical subunits (Dubey *et al*, 2003). Our preliminary work showed that CsrA aggregates at physiological pH at concentrations above 0.1mM. At pH 7.5, size exclusion chromatography showed the presence of three peaks with apparent molecular weights of 18, 36 and 54 kDa compatible with the formation of dimers, tetramers and hexamers (data not shown). Size exclusion chromatography showed that at low pH (~4.5), CsrA does not aggregate into higher order forms but remains as a dimer (Fig. 3.4). Gel filtration data was also confirmed by NMR self-diffusion experiments (Fig. 3.4 and Table 3.1), (Ekiel *et al*, 1997). At pH 4.5 CsrA gave a diffusion coefficient of  $0.93 \times 10^6 \text{ cm}^2/\text{second}$ , in agreement with the formation of a dimer at low pH (apparent molecular weight of 18.6 kDa). The apparent molecular weight for the aggregate at pH 7.5 is ~29.3 kDa. NMR experiments for determining the solution structure of CsrA were performed at pH 4.5.



**Figure 3.4 CsrA is dimer in solution**

(A) Gel filtration chromatogram of CsrA. Protein standards are: regeneration induced CNP homolog (RICH), 53.8 kDa; RICH in 1 mM DTT (26.9 kDa) and gamma-adaptin ear protein (13.8 kDa) as indicated in the chromatogram. CsrA eluted as a ~18 kDa protein, consistent with a dimeric form. (B) PFG-self diffusion experiments for CsrA. The slopes in the plot are proportional to the diffusion coefficient (Ds).

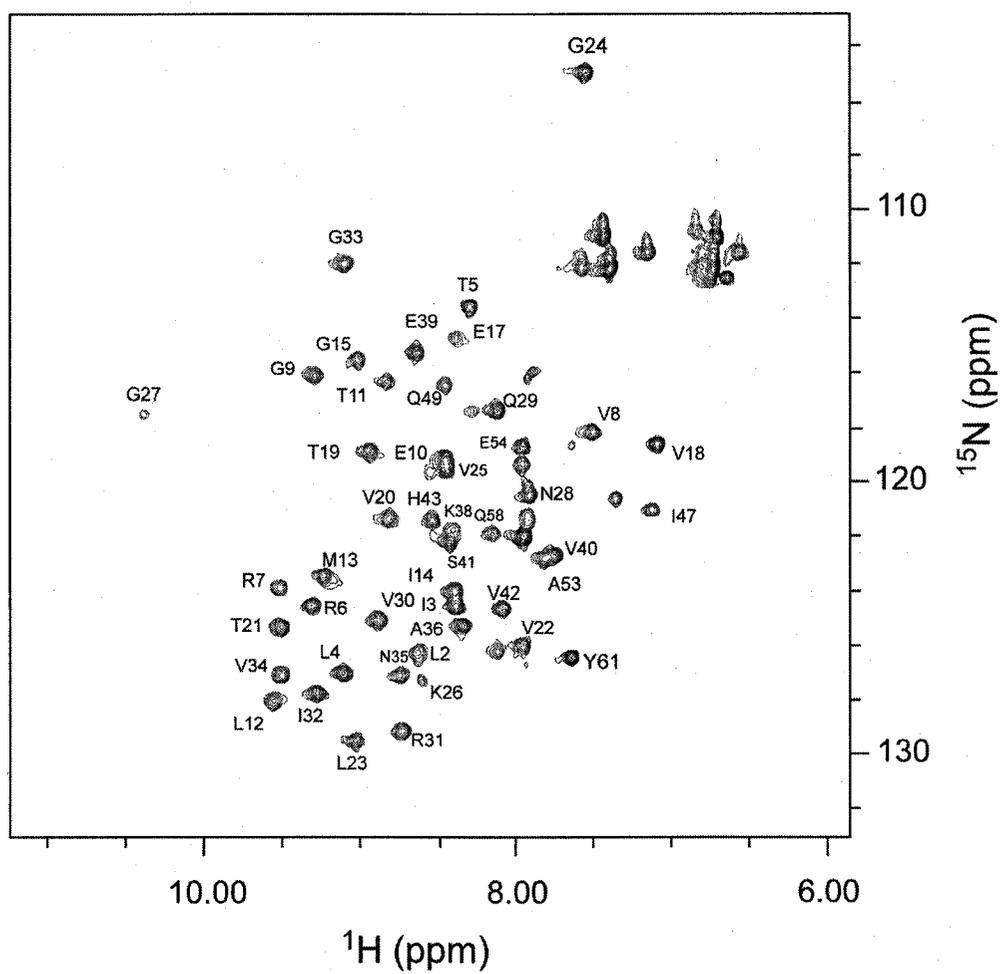
**Table 3.1 Diffusion coefficient (Ds) values at 298 K**

<b>Protein</b>	<b>Mw (kDa)</b>	<b>Diffusion coefficient (<math>D_{sRICH}/D_s</math>)<sup>3</sup> (<math>10^6 \text{ cm}^2/\text{second}</math>)</b>	
CsrA (pH 4.5) <sup>(1)</sup>	18.0	0.93	0.88
CsrA (pH 7.5)	-	0.80	1.02
Gamma_ear	13.9	1.11	0.74
Lysozyme	14.3	1.12	0.73
RICH	26.9	0.82	1.00

<sup>(1)</sup> Molecular weight of the CsrA dimer

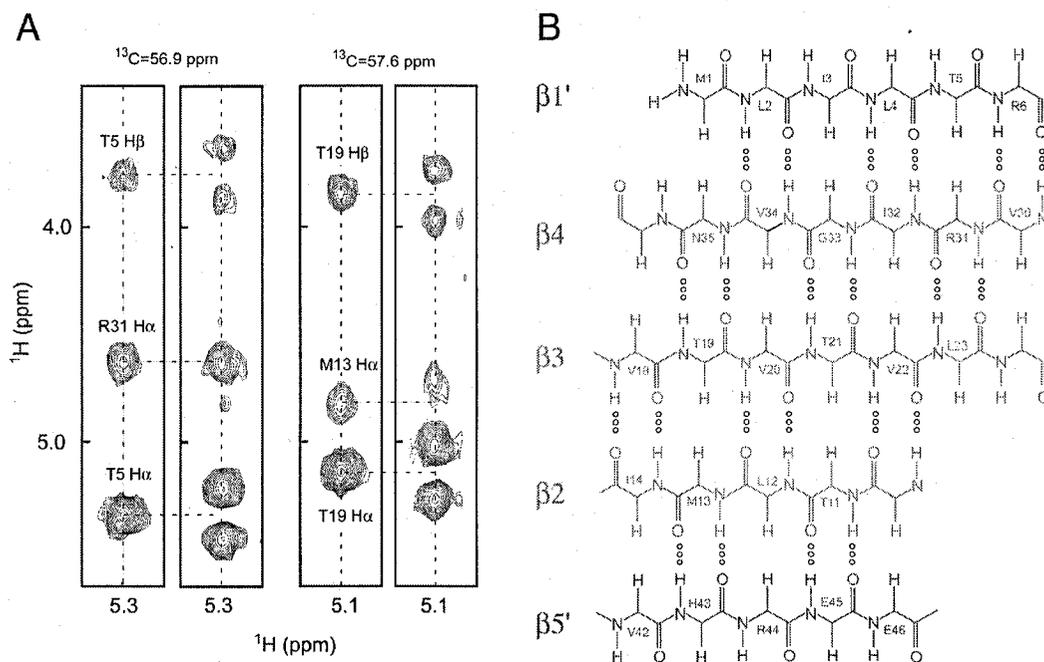
CsrA was uniformly labeled with  $^{15}\text{N}$  or doubly labeled with  $^{15}\text{N}$  and  $^{13}\text{C}$  for NMR analysis. Backbone resonance assignments were obtained with standard triple resonance NMR experiments (Figure 3.5). Chemical shift indexes of  $\text{C}\alpha$ ,  $\text{C}\beta$  and  $\text{H}\alpha$  (Wishart and Sykes, 1994) and the analysis of sequential and short-range NOE connectivities involving NH,  $\text{H}\alpha$ , and  $\text{H}\beta$  protons indicate that the CsrA monomer is composed of five  $\beta$ -strands and a short  $\alpha$ -helix. The unstructured C-terminus is unfolded as shown by measurement of backbone  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOEs. Analysis of  $^{13}\text{C}$ -edited NOESY experiments recorded in conjunction with and without carbon decoupling on a 1:1 mixture of  $^{13}\text{C}/^{15}\text{N}$ -labeled/nonlabeled CsrA, allowed us to determine intermolecular NOEs and the hydrogen bond network defining the CsrA dimer (Fig.3.6).

Even though 95% of backbone and side chain resonances were unambiguously determined, assignment of NOE cross-peaks was challenging and ambiguous at several points. For instance, the core region of the protein was found to be rich in valine residues (~20 %) with proton and carbon nuclei resonating within a narrow chemical shift range. The dimeric nature of CsrA contributed further to this ambiguity. However, the high content of antiparallel  $\beta$ -sheet within CsrA allowed the structure of CsrA to be defined using relatively sparse NMR-derived restraints (Fig. 3.7 and Table 3.2).



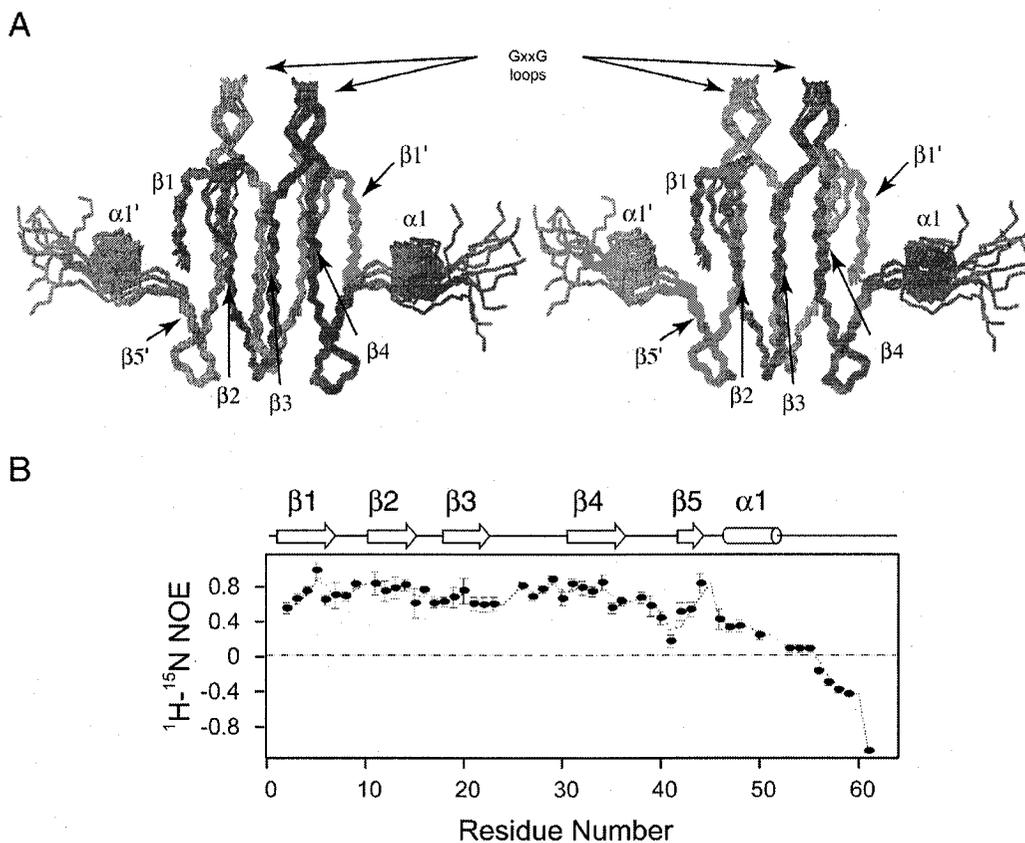
**Figure 3.5**  $^{15}\text{N}$ -HSQC of CsrA

The backbone amide crosspeaks are labelled with the corresponding amino acid. The unlabelled crosspeaks belong to glutamine and asparagine sidechains. The spectrum was taken at pH 7.0 and 303 K.



**Figure 3.6 Intermolecular interactions in CsrA**

(A) Representative 2D strips of  $^{13}\text{C}$  edited NOESY experiments with and without carbon decoupling in a 1:1 sample of  $^{13}\text{C}/^{15}\text{N}$ -labeled/nonlabeled protein. Peaks from the carbon-coupled experiment are shown to the left for both sets of strips. NOEs resulting from intermolecular interactions appear as singlet in both experiments, while intramolecular NOEs are doublets in the uncoupled experiment. (B) Schematic representation of the hydrogen bond network in CsrA. Red and magenta represent different CsrA molecules.



**Figure 3.7 NMR ensemble of CsrA**

(A) Stereo view of the backbone atoms for residues 1-55 of 10 selected conformers. Adjacent subunits of the dimer are colored in magenta and blue respectively. (B) Values of  $\{^1\text{H}\}$ - $^{15}\text{N}$  heteronuclear NOEs for backbone amides of CsrA.

**Table 3.2 Structural statistics for 20 selected conformers of CsrA****Constraints used for structure calculation**

Intraresidue NOEs	(n=0)	254
Sequential range NOEs	(n=1)	74
Medium range NOEs	(n=2,3,4)	22
Long range NOEs	(n>4)	68
Intermolecular NOEs		78
Dihedral angle constraints		110
Hydrogen bonds		46
<sup>15</sup> H- <sup>1</sup> H residual dipolar couplings		104
Total number of constraints		710

**Final energies (kcal/mol)**

E <sub>total</sub>	163.03 ± 4.52
E <sub>bond</sub>	5.18 ± 0.51
E <sub>angle</sub>	62.89 ± 1.99
E <sub>improper</sub>	5.85 ± 0.93
E <sub>VdW</sub>	2.10 ± 0.70
E <sub>noe</sub>	13.70 ± 1.59
E <sub>dihedral</sub>	2.10 ± 0.70
E <sub>sani</sub>	34.82 ± 5.32

**Deviations from idealized geometry**

Bonds (Å)	0.0016 ± 0.0001
Angles (°)	0.3401 ± 0.0054
Impropers (°)	0.1991 ± 0.0157

**RMS deviation from experimental restraints**

Distance restraints (Å)	0.0183 ± 0.0011
Dihedral angle restraints (°)	0.4144 ± 0.0671

**RMS deviations of the 20 structures from the mean coordinates (Å)**

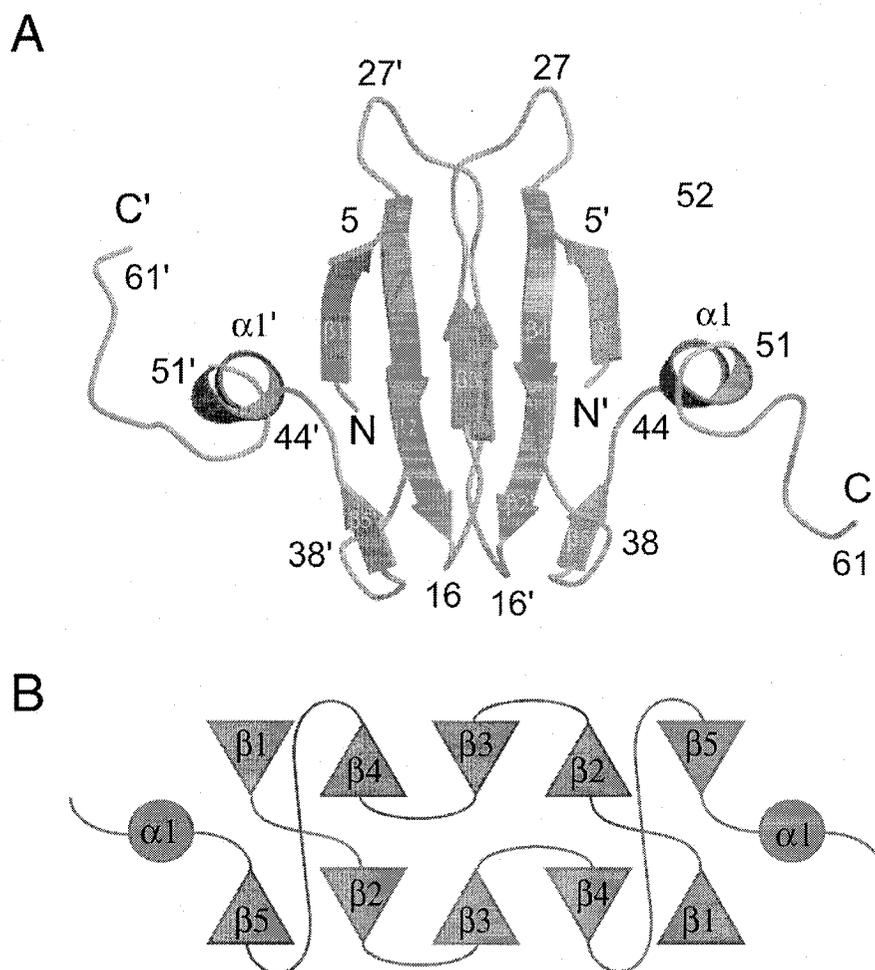
Backbone atoms	0.5422 ± 0.1919
All heavy (non-hydrogen) atoms	1.1434 ± 0.2313
All atoms	1.4072 ± 0.1919

**Average Ramachandran statistics for structured regions (%)**

Residues in most favored regions	82.57
Residues in additional allowed regions	16.36
Residues in generously allowed regions	1.07
Residues in disallowed regions	0.00

**Analysis of residual dipolar couplings**

RMSD (Hz)	1.028 ± 0.085
Q-factor	0.131 ± 0.011



**Figure 3.8 Structure of CsrA**

(A) Ribbon depiction of CsrA. (B) Topology diagram of the CsrA structure showing the way  $\beta$ -sheets are connected in the dimer.

Each CsrA monomer is composed of five strands,  $\beta$ 1-  $\beta$ 5, corresponding to residues 2-6, 10-15, 18-23, 30-35 and 41-43. Residues 46-50 fold into a short  $\alpha$ -helix followed by an unstructured C-terminus (residues 51-61). In the dimer, strands  $\beta$ 1 and  $\beta$ 5 of one monomer hydrogen bond to  $\beta$ 4' and  $\beta$ 2' of the other monomer, forming a mixed antiparallel  $\beta$ -sheet (Fig. 3.8). Packing of these two mixed  $\beta$ -sheets forms the core of CsrA.

In spite of the low sequence similarity, it was proposed that CsrA was a member of the KH domain family, a group that comprises a diverse series of RNA-binding proteins (Liu *et al*, 1995). The characteristic signature of this protein family is the presence of a  $\sim$ 30 amino acid segment that expands around a conserved GxxG core sequence (where x is any amino acid, with a preference for basic residues) (Adinolfi *et al*, 1999). In CsrA, the GxxG motif has the sequence GVKG (residues 24-27) and locates in the loop connecting strands  $\beta$ 3 and  $\beta$ 4. Our structure proves that CsrA is not a member of the KH family of proteins, which have a characteristic  $\beta\alpha\alpha\beta\alpha$  topology (Musco *et al*, 1996; Musco *et al*, 1997) and differs from that of CsrA. However, it is still possible for the GxxG sequence in CsrA to be involved in the recognition of the GGA triplet present in all CsrA binding sites (Baker *et al*, 2002).

### 3.3.2 RNA binding of CsrA

Charged residues in CsrA are grouped into well-defined clusters on the protein surface (Fig. 3.9). The main basic patch comprises residues R6, R7, K26, R31 and the backbone amides of N28 and Q29, defining a putative RNA binding site. Residues E10, E45, E46 and D16, E17, E39 give rise to well defined acidic patches located on the side and bottom of the CsrA molecule. Electrostatic interactions between these basic and acidic patches may explain CsrA aggregation at high concentrations.

In order to map the RNA binding surface of CsrA, we obtained HSQC spectra of the  $^{15}$ N-enriched protein in the presence of different target RNAs. Based on the CAP leader mRNA sequence and the CsrB consensus, three

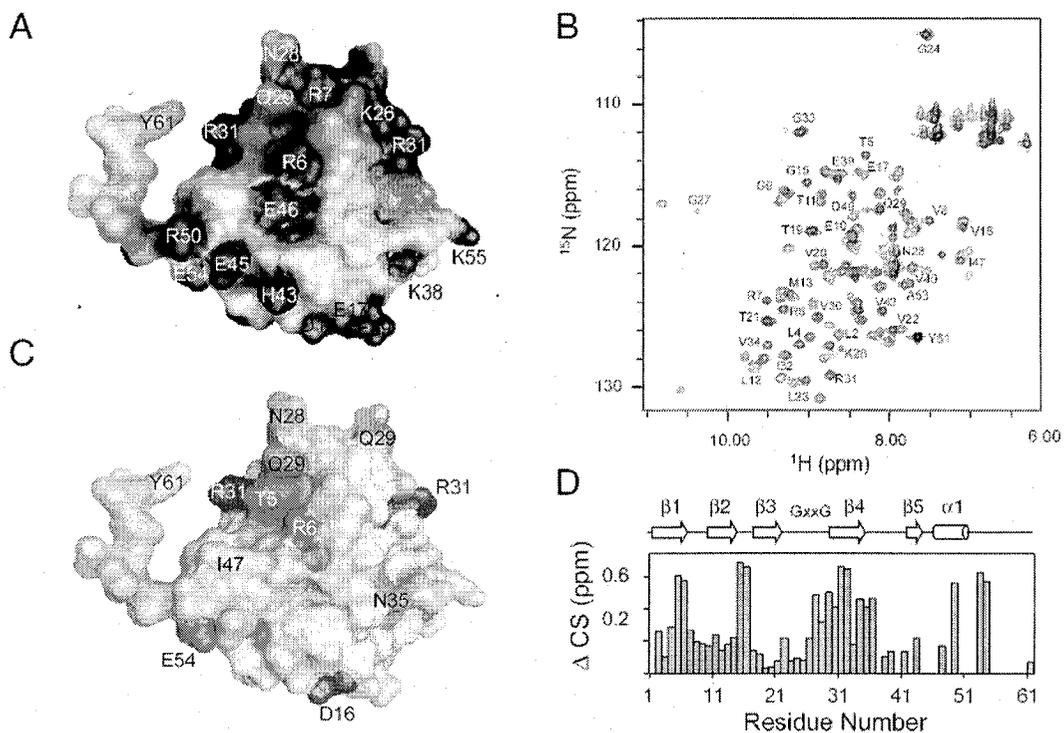
different RNAs containing the GGA sequence were designed as follows: glgC25, (ACCUGCACACGGAUUGUGUGGUUC); glgC15 (CACACGGAUUGUGUG) and the CsrB consensus (CAGGAUG). As low pH could alter the protonation state of the nucleotide bases, thereby affecting RNA recognition, binding experiments were performed at pH 7.5. Titrations with both the glgC25 and glgC15 hairpin showed high affinity binding as indicated by the slow exchange in the NMR time-scale between bound and unbound forms. Binding of glgC25 and glgC15 affect almost all the CsrA amide signals, suggesting a large conformational change and/or major protein-RNA interactions upon RNA binding (Fig. 3.9). However, the CsrB consensus sequence caused no chemical shift perturbations. Contrary to glgC25 and glgC15, which are predicted to form hairpin structures, the CsrB consensus is expected to be single stranded (Baker *et al*, 2002). It is possible that CsrA affinity for RNA is greatly reduced when not in the duplex form. These results are consistent with previous reports showing that CsrA binds to the GGC sequence with higher affinity if it is part of a hairpin loop (Baker *et al*, 2002). Titration data mapped into the CsrA structure (Fig. 3.9), suggests that the loops connecting  $\beta$ 1- $\beta$ 2,  $\beta$ 3- $\beta$ 4 (GxxG motif), strand  $\beta$ 4 and the C-terminus are the regions responsible for RNA binding. In CsrA, conserved and surface-exposed residues are probably the ones involved in recognizing the GGA signature. Candidate residues are R6, R7, E10, N28, Q29, V30 and R31.

### 3.3.3 Discussion

Toeprint analyses have been performed to identify the position of bound CsrA in target mRNAs (Baker *et al*, 2002; Dubey *et al*, 2003). In the case of the *glgCAP* transcript, RNA digestion and gel mobility assays were performed on a 134 nucleotides untranslated leader containing the CsrA binding site. Binding of CsrA protects both the single stranded glgC Shine-Dalgarno (SD) sequence and the *glgCAP* hairpin further upstream from cleavage by RNase T1 and  $Pb^{2+}$  (Baker *et al*, 2002). Structural changes seem to occur in the hairpin RNA as CsrA binding enhances the cleavage of the sequence in the stem loop protected in the unbound form.

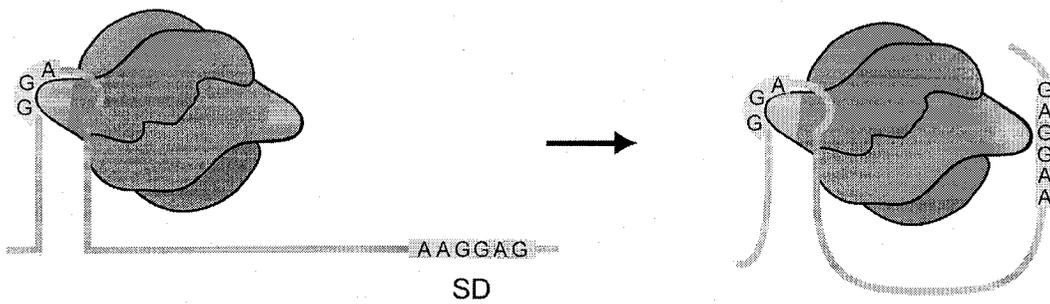
In light of our structural data we postulate that CsrA dimer presents its GxxG motifs to simultaneously recognize both GGA sequences in the Shine-Dalgarno and the upstream hairpin loop. Since SD sequences are present in most of the bacterial mRNAs CsrA potentially requires a secondary signal to recognize the correct transcripts to be regulated. The upstream hairpin loop in glgCAP transcript may act as an allosteric activator for CsrA binding to the downstream SD sequence. CsrA seems to bind the GGA sequence with higher affinity when present as part of a hairpin loop than in single stranded sequences. This is supported by experiments where the SD-CsrA interaction was not sufficiently strong to disrupt the complex formed by reverse transcriptase (Baker *et al*, 2002). Our titration experiments substantiated that CsrA binds preferentially to hairpin loop structures. In addition, the GGA sequence in the hairpin loop affects the affinity of CsrA binding to the SD (Baker *et al*, 2002). Furthermore, conformational changes seem to occur upon binding to RNA as shown by our own data and foot printing studies and RNA structure mapping that demonstrate that the base of the glgCAP leader RNA hairpin is disrupted when CsrA is bound (Baker *et al*, 2002). The binding affinity is probably higher for the hairpin due to the reduced conformational entropy associated with this structure.

We propose that CsrA binding to its target mRNAs may occur in two steps. First, the CsrA dimer recognizes the hairpin loop upstream the SD and binds to the GGA sequence using one of its GxxG loops. At this point it is possible that both RNA and CsrA experience a conformational change that increases the CsrA affinity for the GGA sequence in the single stranded Shine-Dalgarno. CsrA can then bind to the GGA sequence through the second GxxG motif therefore preventing transcription and making the RNA more susceptible for degradation (Fig. 3.10).



**Figure 3.9 Surface properties of CsrA**

(A) Surface potential of the CsrA structure. Blue and red colors indicate positive and negative electrostatic potential, respectively. (B) Superposition of  $^{15}\text{N}$ -HSQC spectra of CsrA in the absence (blue) and presence (red) of the glg15 RNA (5'-CACACGGAUUGUGUG-3'). (C) Mapping of chemical shift changes from B onto the CsrA structure. Chemical shift changes that could not be measured are colored gray (D) Measured chemical shift changes versus residue number from the RNA titration, calculated as  $[(\Delta\text{H})^2 + (0.2 \cdot \Delta\text{N})^2]^{1/2}$ . Secondary structural elements are shown on top. Blank spaces represent shifts that could not be traced with certainty.



**Figure 3.10 Proposed binding of CsrA to target mRNAs**

CsrA binds to the GGA sequence in hairpin region upstream the Shine-Dalgarno consensus through one its GxxG motifs (yellow). This interaction promotes a conformational change in both RNA and CsrA that increases the affinity for the GGA sequence in the SD sequence downstream (right).

### **3.4 Materials and methods**

#### **3.4.1 Sample preparation**

Gene *csrA* from *E. coli* K12 was subcloned into pET15b (Novagen, Inc., Madison, WI) and expressed in *E. coli* BL21 as an oligo-histidine (His tag) fusion protein of 9 kDa. Cells were grown at 37°C to an OD600 of 0.8 and induced with 1 mM IPTG. Afterwards, the temperature was reduced to 30°C and the cells were allowed to express the protein for 3 hours before harvesting. The media used were either LB or M9 minimal medium containing <sup>15</sup>N ammonium chloride and/or <sup>13</sup>C-glucose (Cambridge Isotopes Laboratory, Andover, MA). CsrA was purified by affinity chromatography on Ni<sup>2+</sup>-loaded chelating Sepharose (Amersham Pharmacia Biotech, Piscataway, NJ). NMR samples were ~1 mM protein in 50 mM sodium acetate buffer, 300 mM NaCl at pH 4.5. For preparation of <sup>13</sup>C/<sup>15</sup>N-labeled/unlabeled protein samples, equal amounts of purified <sup>13</sup>C/<sup>15</sup>N-labeled and unlabeled protein were mixed in the presence of 6 M urea overnight. Urea was removed by extensive dialysis against NMR buffer.

#### **3.4.2 Gel filtration**

The oligomeric state of CsrA was determined using gel filtration (Hiload 16/60 Superdex 75, Pharmacia Biotech). Regeneration induced CNP homolog (RICH), 53.8 kDa; RICH in 1 mM DTT (26.9 kDa) and gamma-ear protein (13.8 kDa) were used as standards. Samples were run with a flow rate of 1 ml/min at room temperature in NMR buffer as described above. CsrA eluted from the column at a predicted molecular mass of ~18 kDa as expected for a dimer.

#### **3.4.3 NMR spectroscopy**

NMR experiments were recorded at 303 K on a Bruker Avance 600 MHz spectrometer. Backbone and side-chain assignments of CsrA were determined using HNCACB, CBCA(CO)NH, edited <sup>15</sup>N/<sup>13</sup>C-TOCSY-HMQC, <sup>13</sup>C-HCCH-TOCSY and <sup>13</sup>C-(h)CCH-TOCSY experiments. NOE data for the structure determination were obtained from homonuclear NOESY, <sup>15</sup>N-edited or <sup>13</sup>C-edited 3D NOESY experiments. Backbone assignments at pH 7.5 were determined using

HNCA and CBCA(CO)NH experiments. The intermolecular NOEs were detected using a filter-edited 3D NOESY spectrum and a pair of identical  $^{13}\text{C}$ -edited 3D NOESY with/without decoupling in the indirect  $^1\text{H}$  dimension (100-ms mixing time). A  $^{13}\text{C}/^{15}\text{N}$ -labeled/nonlabeled protein sample (1:1) was used for these experiments.  $^1\text{H}$ - $^{15}\text{N}$  residual dipolar coupling constants were measured from comparison of IPAP-HSQC experiments recorded on CSRA with and without 2.5% C12E5/hexanol (Rückert and Otting, 2000) For the measurement of dipolar couplings we used 50 mM sodium acetate buffer at pH 4.5 and 0.5 mM CSRA. All NMR spectra were processed using either XWINNMR version 2.5 or 3.1 (Bruker Biospin) or NMRPipe (Delaglio *et al*, 1995). Evaluation of spectra and manual assignments were completed with NMRView (Johnson and Blevins, 1994). Pulse field gradient (PFG) self-diffusion experiments were done according to the method reported by Ekiel (Ekiel *et al*, 1997).

#### 3.4.4 Structure calculations.

CNS 1.1 software (Brunger *et al*, 1998) was used to generate an initial fold of CsrA with a basic set of 122 NOEs manually assigned from NOESY spectra (104 intramolecular and sequential NOEs). Hydrogen bond constraints were introduced to secondary structure regions as determined by chemical shift analysis, characteristic NOE patterns and analysis of amide exchange rates. Dihedral restraints  $\psi$  and  $\phi$  were obtained using the TALOS program (Cornilescu *et al*, 1999). These calculations generated a fold that was used as a model template for automated assignments by ARIA1.1 (Nilges *et al*, 1997). The final structure of CsrA was calculated with a total set of 710 constraints collected from the experiments described earlier. Non-crystallographic symmetry restraints (NCS) were used to keep both subunits in the dimer with the same conformation. In the final round of calculations, CNS 1.1 was extended to incorporate RDC restraints for further refinement using the torsion angle space. The axial and rhombic components were defined from a histogram of measured RDCs (Clare *et al*, 1998) and optimized by a grid search method (Clare *et al*, 1998). Twenty structures were selected based on the lowest overall energy and least violations to

represent final structures. PROCHECK-NMR was used to generate Ramachandran plots to check the protein's stereochemical geometry (Laskowski *et al*, 1993). A summary of the structural statistics for CsrA is shown in Table 1. The coordinates of CsrA have been deposited in the RCSB under PDB code 1Y00 and the NMR assignments under BMRB accession 11855.

#### **3.4.5 NMR titrations**

RNA titrations were performed by recording a series of  $^{15}\text{N}$ -HSQC spectra on uniformly  $^{15}\text{N}$ -labeled CsrA (~0.7 mM CsrA) in the presence of different amounts of ligand concentrations in the 0-2.0 mM range. As high concentrations of imidazole improve the solubility of CsrA at physiological pH, the protein sample and RNA stock solutions were prepared in 500 mM deuterated imidazole, 300 mM NaCl at pH 7.5. The RNA sequences used were 5'-ACCUGCACACGGAUUGUGUGGUUC-3' (glg25), 5'-CACACGGAUUGUGUG-3' (glg15) and 5'-CAGGAUG-3' (CsrB consensus sequence) and were synthesized in the Core DNA & Protein Services, University of Calgary.

## Chapter 4: Structural studies on archaeal translation initiation factor aIF2 $\beta$

### 4.1 Abstract

aIF2 $\beta$  is the archaeal homolog of eIF2 $\beta$ , a member of the eIF2 heterotrimeric complex, implicated in the delivery of Met-tRNA<sub>i</sub><sup>Met</sup> to the 40S ribosomal subunit. We have determined the solution structure of the intact  $\beta$  subunit of aIF2 from *Methanobacterium thermoautotrophicum* and showed the importance of zinc for the stability of the C-terminus. aIF2 $\beta$  is composed of an unfolded N-terminus, a mixed  $\alpha/\beta$  core domain and a C-terminal zinc finger. NMR data shows the two folded domains display restricted mobility with respect to each other. Analysis of the aIF2 $\beta$  structure docked to tRNA allowed for the identification of a putative binding site for the  $\beta$  subunit in the ternary translation complex. Based on structural similarity and biochemical data a role for the different secondary structure elements is suggested.

### 4.2 Introduction

Archaeal translation initiation factor aIF2 is a heterotrimeric protein, consisting of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits, with high sequence similarity to their eukaryotic counterparts (eIF2). eIF2 plays a critical role in the initiation of protein synthesis by forming a ternary complex with GTP and the aminoacylated initiator methionyl-tRNA (Met-tRNA<sub>i</sub><sup>Met</sup>). This complex binds to the small ribosomal subunit and with the aid of other translation factors scans from the 5' end of mRNA (Bell and Jackson, 1998). Upon recognition of the initiation codon, GTP is hydrolyzed and the eIF2-GDP complex is released. This leads to assembly of the 80S ribosome at the initiation codon and the start of protein elongation. The recycling of eIF2 between successive rounds of translation requires an additional protein factor, the guanine nucleotide exchange factor IF2B, which catalyses the exchange of GDP bound to eIF2 for GTP (Kimball, 1999; Pestova and Hellen, 2000).

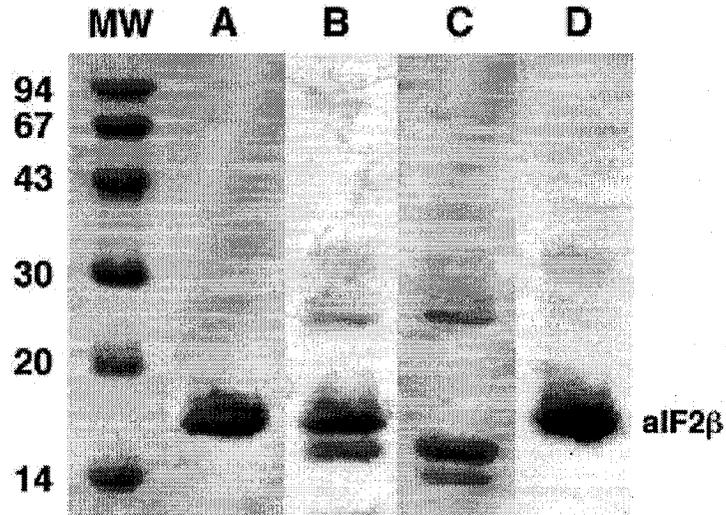
Distinct functions have been observed for each subunit of eIF2. The  $\alpha$  subunit is a global regulator of protein synthesis in eukaryotes. Phosphorylation of eIF2 $\alpha$  regulates the exchange rate of GDP to GTP in a/eIF2, altering its availability for translation initiation through the inhibition of Met-tRNA<sub>i</sub><sup>Met</sup> binding (Pain, 1996). The  $\gamma$  subunit is responsible for GTP binding and its similarity to EF-Tu (~27% identity, ~50% similarity), allowed for the identification of the Met-tRNA<sub>i</sub><sup>Met</sup> binding region (Schmitt *et al*, 2002). The  $\beta$  subunit of eIF2, is implicated in a variety of interactions with other translation factors. For example, its N-terminus binds to eIF5, the GTPase activating factor (GAP) for eIF2, and to the  $\epsilon$  subunit of the exchange factor eIF2B (Asano *et al*, 1999). This region has also been shown to bind RNA *in vitro* through three lysine repeats (Laurino *et al*, 1999). The C-terminal region of eIF2 $\beta$  contains another potential RNA binding motif. Mutations in this C<sub>2</sub>-C<sub>2</sub> zinc finger result in spontaneous GTPase activity and alter the correct recognition of the AUG codon (Huang *et al*, 1997). The  $\beta$  subunit has also been implicated in binding to the  $\delta$  subunit of eIF2B and to crosslink GTP and Met-tRNA<sub>i</sub><sup>Met</sup> (Bommer and Kurzhalia, 1989; Gaspar *et al*, 1994). Archaeal aIF2 $\beta$  has ~50% similarity and ~30% identity to the C-terminal half of eukaryotic IF2 $\beta$  but lacks of N-terminal polylysine tracts, which are ubiquitous in eukaryotes (Laurino *et al*, 1999; Thompson *et al*, 2000). a/eIF2 $\beta$  also shares a high degree of sequence similarity with eIF5, an eukaryotic initiation factor important for stimulating hydrolysis of GTP by the ternary complex.

Solution structures of the N and C-terminal domains of aIF2 $\beta$  from *M. jannaschii* (Mj\_aIF2 $\beta$ ) have been determined (Cho and Hoffman, 2002). Here, we present the solution structure of the intact archaeal translation initiation factor 2 $\beta$  from *Methanobacterium thermoautotrophicum*, including the interdomain linker, absent in the Mj\_aIF2 $\beta$  structure. Based on comparison with structurally similar proteins and previous known biochemical data, we propose roles for the different regions of a/eIF2 $\beta$  in translation initiation. Additionally, we demonstrate the importance of zinc in the stability of the C-terminal domain.

## 4.3 Results

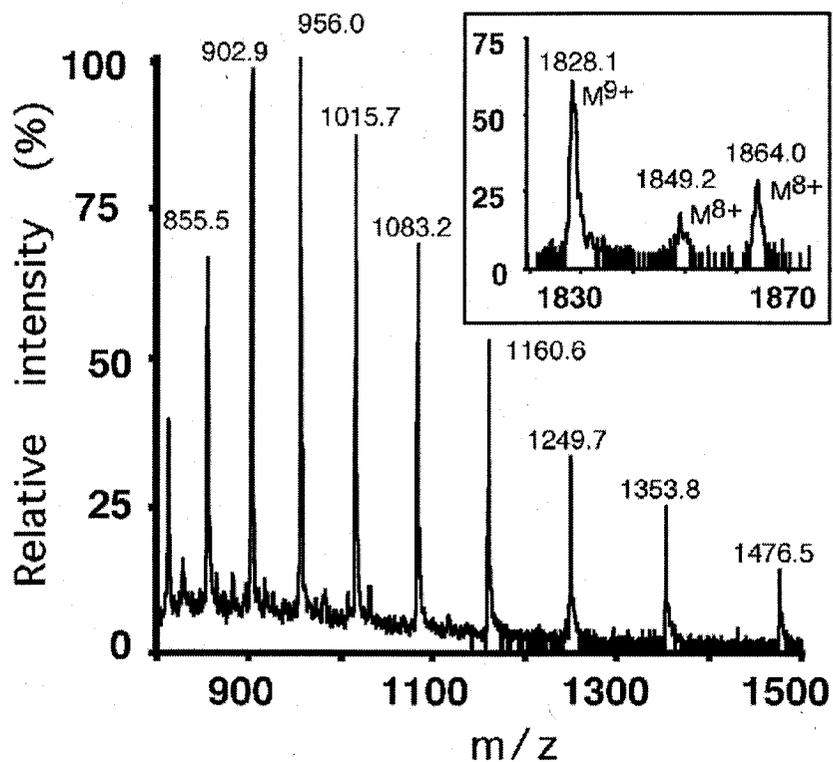
### 4.3.1 Zinc is required for structural stability of the C-terminus of aIF2 $\beta$

The first evidence that zinc is important for the stability of aIF2 $\beta$  came from differences in the size of aIF2 $\beta$  when produced in *E. coli* grown in Luria broth (LB) and minimal salts (M9) media. SDS-PAGE of the His-tag fusion protein from an LB culture showed the presence of a single band with the expected molecular size of 16.2 kDa (Figure 4.1). In contrast, an additional band smaller by ~20 residues, was observed for protein purified from M9 cultures. This pointed to nutrient deficiency as a probable cause. Cleavage of the N-terminal histidine tag showed the same pattern as the uncleaved form, demonstrating that the degradation was at the C-terminus. As the predicted zinc finger motif for aIF2 $\beta$  is also located at the C-terminus, this suggested that zinc was possibly the missing nutrient. Supplementation with 50  $\mu$ M ZnCl<sub>2</sub> allowed us to obtain full-length aIF2 $\beta$  from minimal medium. Mass spectrometry revealed cleavage sites at L121 and L120 (Figure 4.2). As proteolytic digestion of proteins in absence of their metal ligand has been observed previously for other metalloproteins (Bicknell *et al.*, 1985) we concluded that the absence of zinc causes the C-terminus of aIF2 $\beta$  to unfold, making it susceptible to cleavage by endogenous proteases.



**Figure 4.1 Proteolytic sensitivity of aIF2 $\beta$  in the absence of Zn<sup>+2</sup>**

(A) SDS gel of aIF2 $\beta$  purified from LB culture. The protein migrates as a single band and has the expected size of 16.2 kDa. (B) Protein purified from M9 minimal medium revealing a proteolytic fragment about 20 residues smaller than full-length aIF2 $\beta$ . (C) Cleavage of the N-terminal His-tag demonstrating that the heterogeneity is C-terminal. (D) aIF2 $\beta$  purified from M9 minimal medium supplemented with 50  $\mu$ M ZnCl<sub>2</sub> showing the enhanced the stability of aIF2 $\beta$  against endogenous *E. coli* proteases.



**Figure 4.2 ESI-MS mass spectrum of aIF2 $\beta$**

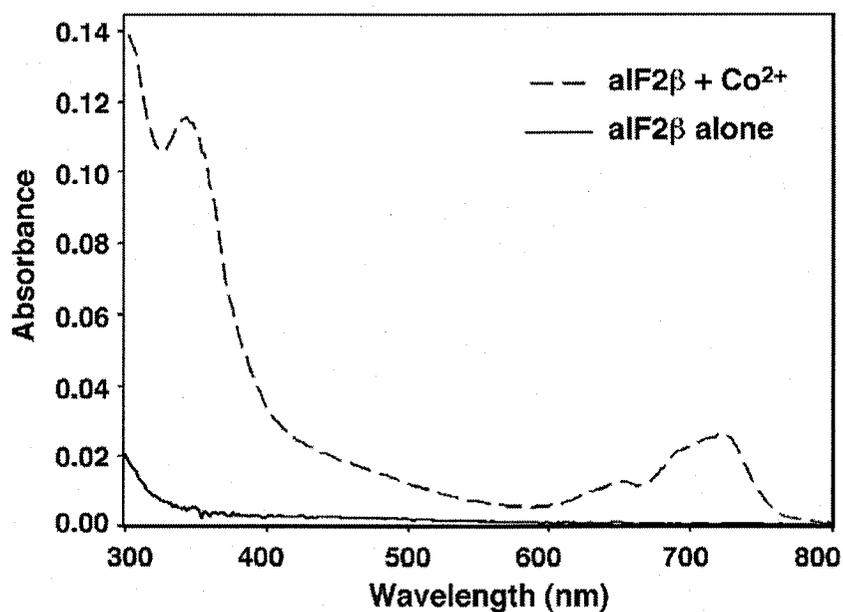
aIF2 $\beta$  grown in LB shows a single form of 16234.30 Da. Insert: Spectrum of protein grown in M9 medium (without zinc) shows the presence of fragments cleaved after residues L120 and L121 (M<sup>8+</sup> peaks at 1849 m/z and 1864 m/z) in addition to the full length product (M<sup>9+</sup> peak at 1828 m/z).

### 4.3.2 Metal binding properties of aIF2 $\beta$

To test the metal binding properties of aIF2 $\beta$ , visible spectra of aIF2 $\beta$  in the presence and absence of Co<sup>2+</sup> were recorded. Cobalt substitution for zinc has been commonly used to address the coordination environment of structural zinc sites in proteins (Maret and Vallee, 1993). Figure 4.3 shows the visible spectrum of a cobalt-aIF2 $\beta$  adduct. The spectrum contains an absorbance maximum at 730 nm, which is characteristic of tetrahedral coordination of Co<sup>2+</sup> (Bertini and Luchinat, 1984).

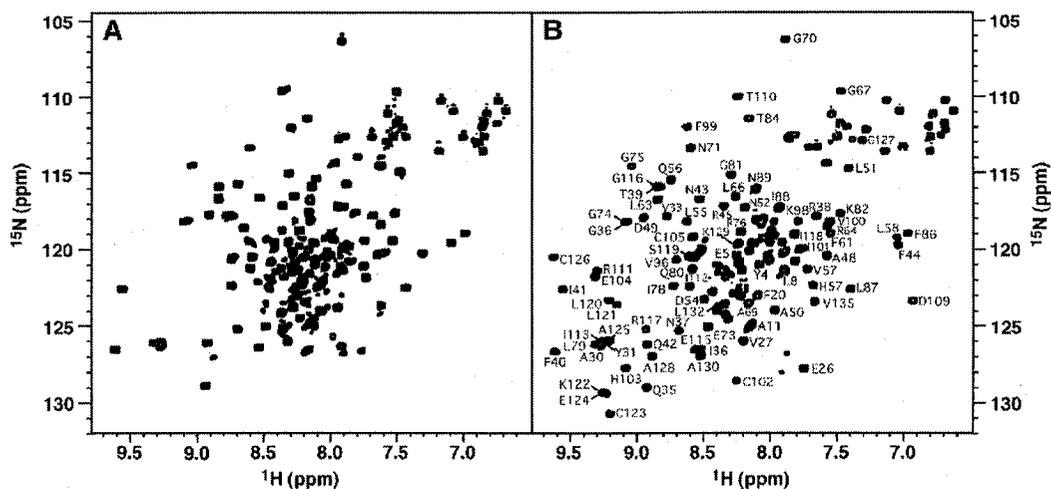
The importance of zinc for folding of the C-terminus was clearly seen in <sup>15</sup>N-<sup>1</sup>H HSQC spectra of aIF2 $\beta$  with and without zinc (Figure 4.4). Upon addition of ZnCl<sub>2</sub> to the protein, several amide resonances shifted away from the overlapped amide resonances between 7.5 and 8.5 ppm in the proton dimension. These spectral changes indicate acquisition of additional folded structure in the presence of zinc. The dependence of zinc binding on the oxidation state of the cysteines was tested by recording <sup>15</sup>N-<sup>1</sup>H HSQC spectra in the absence of DTT. No spectral changes were observed upon addition of zinc to oxidized aIF2 $\beta$  (data not shown). The assignments of the backbone resonances of the zinc bound aIF2 $\beta$  reveal that most of the zinc-shifted peaks are from the putative C-terminal C<sub>2</sub>-C<sub>2</sub> zinc-binding motif.

Zinc is essential for the structural stability of known zinc fingers, so it was surprising that given the high conservation of the cysteines in aIF2 $\beta$  and eIF2 $\beta$ , the importance of this metal for IF2 $\beta$  structure or function has not been previously demonstrated. Several mutagenesis studies have shown the importance of residues at the C-terminus for the function of yeast eIF2 $\beta$  (Donahue *et al*, 1988). This region is strongly conserved and we believe that these residues play an important role in the specific recognition of the Met-tRNA<sub>i</sub><sup>Met</sup>-AUG duplex. If true, the zinc finger would also constitute an important structural element signal for the release of the initiation complex.



**Figure 4.3 Cobalt coordination by aIF2β**

In the absence of CoCl<sub>2</sub>, aIF2β (45 μM) is colorless, (continuous line). Addition of 80 μM CoCl<sub>2</sub> led to an absorption peak at 730 nm, which is distinctive of tetrahedrally complexed cobalt (dashed line). Further additions of CoCl<sub>2</sub> caused no additional changes in the absorption spectrum.



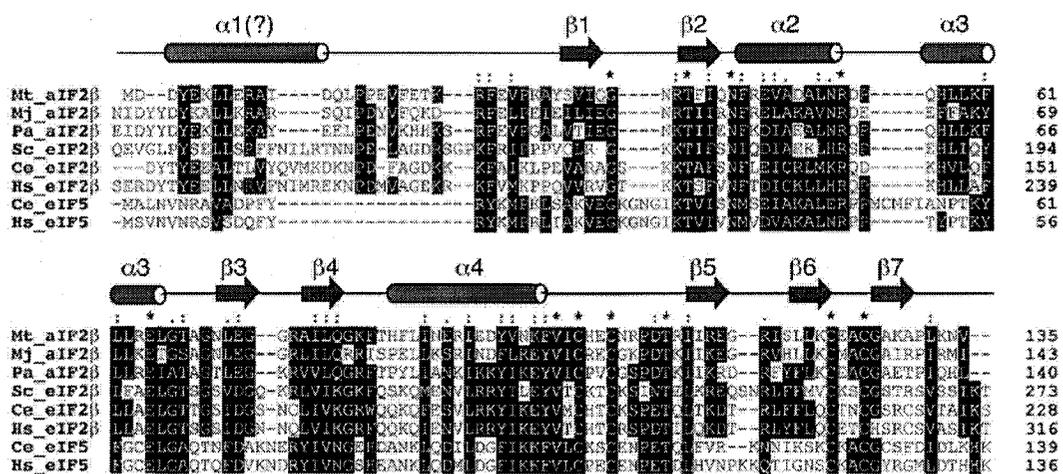
**Figure 4.4**  $\text{Zn}^{+2}$  binding to the C-terminus of aIF2 $\beta$

(A)  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum in the absence of  $\text{Zn}^{+2}$ , showing random coil amide chemical shifts or missing signals for most of the C-terminal residues. (B) Spectrum in the presence of  $\text{Zn}^{+2}$ , showing greater dispersion with additional downfield peaks characteristic of  $\beta$ -strands.

### 4.3.3 Solution structure of aIF2 $\beta$

The solution structure of aIF2 $\beta$  was determined by heteronuclear multidimensional NMR spectroscopy and calculated using standard molecular dynamics protocols. The protein used for structural studies included all 135 residues from Mt\_aIF2 $\beta$  and an additional three residues from the purification tag. Backbone,  $^1\text{H}$ ,  $^{15}\text{N}$ , and  $^{13}\text{C}$  assignments for all residues (excluding K29 and F86) and >95% of the side chain protons were obtained. Regular secondary structure was determined from the chemical shift index (Wishart and Sykes, 1994) and confirmed by observation of characteristic NOEs. The position of secondary structure elements relative to the sequence is shown on Figure 4.5.

In contrast to the *M. jannaschii* protein (Cho and Hoffman, 2002), the *Methanobacterium* protein is stable over long periods of time and the unfolded N-terminus is proteolytically stable. This allowed us to obtain a higher number of distance restraints than the homologous *M. jannaschii* structure plus additional orientation restraints based on  $^1\text{H}$ - $^{15}\text{N}$  residual dipolar couplings (RDCs) measured in Pfl phage (Table 4.1).  $\{^1\text{H}\}$ - $^{15}\text{N}$  NOEs indicate that under our conditions residues 1-30 are unfolded (Figure 4.6). The Mt\_aIF2 $\beta$  structure can be divided in three regions: an unfolded N-terminus, a core domain and a C-terminal zinc finger domain. The core and zinc finger domains have a backbone RMSDs to the mean of 0.54 and 0.63 Å respectively (Table 4.1). The zinc finger is mobile with respect to the core domain, as evidenced by the lack of long range NOEs between these two regions. An average orientation of both domains was obtained by the use of residual dipolar couplings, giving a backbone RMSD to the mean of 0.76 Å for residues 30 to 130 (Figure 4.7).



**Figure 4.5 Sequence alignment of proteins related to archaeal IF2β**

*M. thermoautotrophicum* (Mt\_aIF2β), *M. jannaschii* (Mj\_aIF2β), *P. abyssi* (Pa\_aIF2β); eukaryotic eIF2β from *S. cerevisiae* (Sc\_eIF2β), *H. sapiens* (Hs\_eIF2β), *C. elegans* (Ce\_eIF2β); and N-terminal eIF5 from *H. sapiens* (Hs\_eIF5) and *C. elegans* (Ce\_eIF5). Secondary structure elements are shown on top. Helix α1 is hypothetical and is proposed based on secondary structure predictions and chemical shift index. Swissprot ID numbers for the sequences shown are O27797, Q57562, O58312, PO9064, P20042, Q21230, P55010 and Q22918.

The core domain, (residues 30-98) is composed of three  $\alpha$ -helices ( $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ ) packed against an antiparallel four-stranded  $\beta$ -sheet ( $\beta 1$ ,  $\beta 2$ ,  $\beta 3$  and  $\beta 4$ ), in  $\alpha\beta\beta\alpha\beta\beta\alpha$  topology (Figure 4.8). The  $\beta$  sheet region is formed by residues 32-35, 38-41, 71-73 and 77-80, and the helical bundle comprises residues 44-51, 55-64 and 87-98. The zinc finger domain (residues 99-135) is composed of three antiparallel  $\beta$ -strands ( $\beta 5$ ,  $\beta 6$  and  $\beta 7$ ) encompassing residues 112-116, 120-124 and 128-130 respectively. Helix  $\alpha 4$ , links the core domain and the zinc finger. The absence of significant chemical shift changes in the core domain upon folding of the C-terminus (Cho and Hoffman, 2002; Gutierrez *et al*, 2002), the lack of NOEs between the two domains and their different RDC alignment tensors suggest that there are minimal interactions between the core domain and the zinc finger. Superposition of the core and zinc finger domains of Mj\_aIF2 $\beta$  to Mt\_aIF2 $\beta$  gives a backbone RMSDs of 2.24 Å. The most divergent parts are  $\beta 3$ ,  $\beta 4$  and the loop connecting  $\alpha 4$  and  $\beta 5$ . Only the coordinates for the separate domains of Mj\_aIF2 $\beta$  are available.

**Table 4.1 Structural statistics for 20 selected conformers for aIF2 $\beta$** **Constraints used for structure calculation**

Intraresidue NOEs	(n=0)	385
Sequential range NOEs	(n=1)	155
Medium range NOEs	(n=2,3,4)	76
Long range NOEs	(n>4)	91
Dihedral angle constraints		144
Hydrogen bonds		31
<sup>15</sup> N- <sup>1</sup> H residual dipolar couplings		44
Total number of constraints		835

**Average RMS difference to mean structure (Å)**

Residues	30-130	30-98	98-130
Backbone atoms	0.76±0.12	0.54±0.11	0.63±0.16
All heavy (non-hydrogen) atoms	1.41±0.09	1.24±0.09	1.45±0.14

**Average energy values (kcal mole<sup>-1</sup>)**

E <sub>total</sub>	284.06±25.20
E <sub>bond</sub>	9.19±0.87
E <sub>angle</sub>	89.18±4.87
E <sub>improper</sub>	9.50±1.31
E <sub>VdW</sub>	149.90±23.96
E <sub>NOE</sub>	12.46±2.15
E <sub>dihedral</sub>	5.90±1.17
E <sub>sani</sub>	6.06±1.88

**RMS deviation from idealized covalent geometry**

Bonds (Å)	0.0022±0.0001
Angles (°)	0.380±0.010
Improper (°)	0.232±0.016

**RMS deviation from NMR restraints**

Distance restraints (Å)	0.018±0.001
Dihedral angle restraints (°)	0.575±0.058

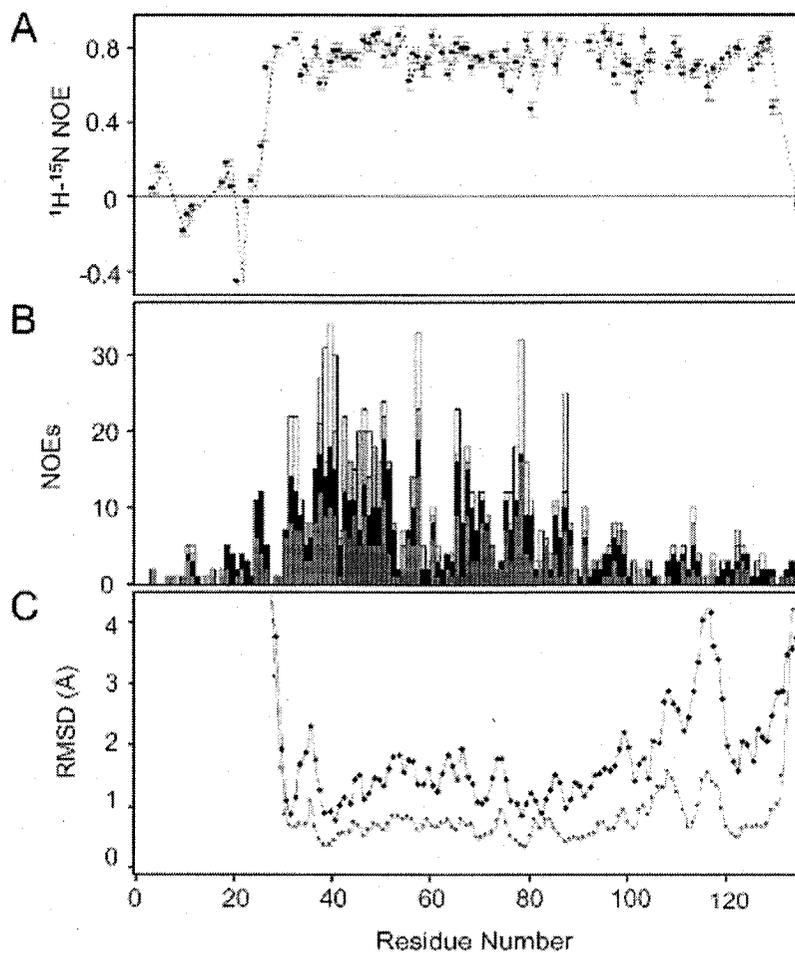
**Average Ramachandran statistics of folded regions (%)**

Residues in most favored regions	78.35
Residues in additional allowed regions	21.28
Residues in generously allowed regions	0.35
Residues in disallowed regions	0.0

**Analysis of residual dipolar couplings\***

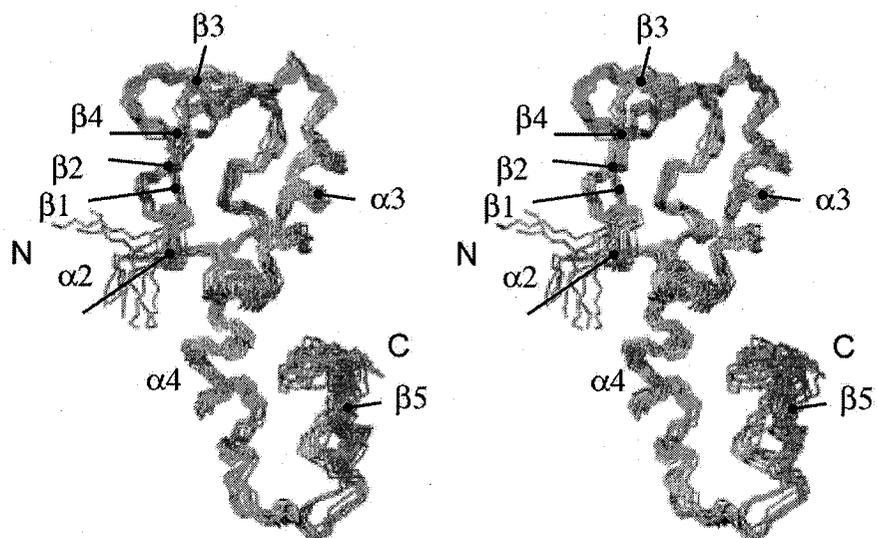
	<b>Core domain</b>	<b>Zinc finger</b>
RMSD (Hz)	0.808±0.055	0.726±0.135
Q-factor	0.057±0.039	0.072±0.013

\*Values are quoted with respect to separate alignment tensors defining each domain.



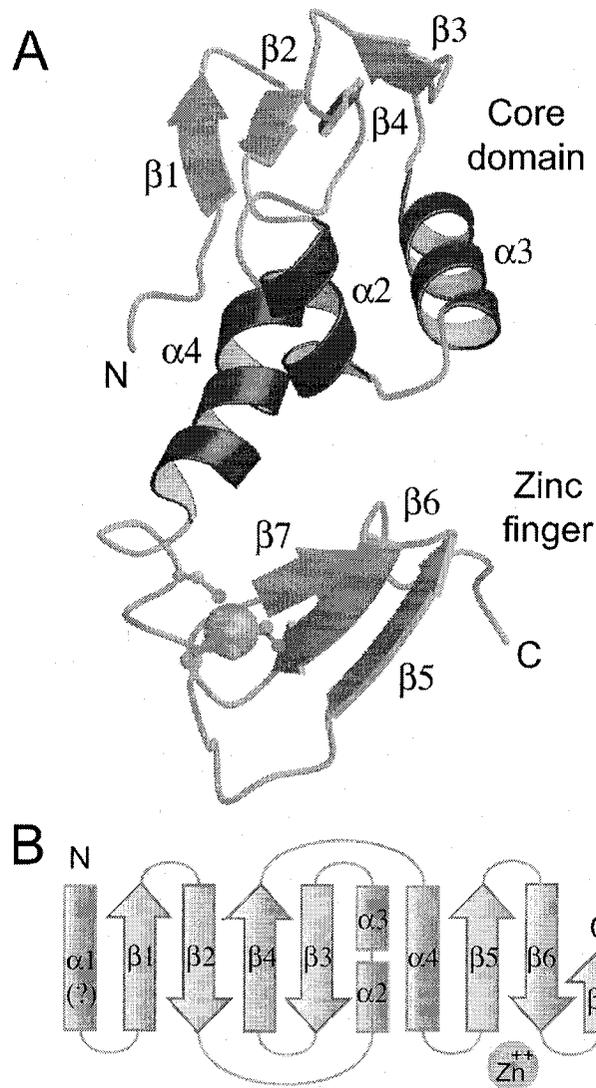
**Figure 4.6  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear NOE, NOE constraints and RMSD statistics for Mt\_aIF2 $\beta$**

$^1\text{H}$ - $^{15}\text{N}$  heteronuclear NOE values obtained at a proton frequency of 500 MHz. (B) A summary of all unambiguous NOEs: intraresidue, sequential, medium and long-range NOEs are shown in dark gray, black, light gray and white, respectively. (C) A comparison of the average backbone RMSD ( $\text{\AA}$ ) per residue of the 20 lowest energy structures calculated without (diamonds) and with (circles)  $^1\text{H}$ - $^{15}\text{N}$  residual dipolar couplings. The fit was calculated for residues 30 to 130.



**Figure 4.7 NMR ensemble of Mt\_aIF2 $\beta$**

The RMSD to the mean structure for backbone atoms in the folded region (30-130) is  $0.76 \pm 0.12$  Å. RMSD values for the core domain (30-98) and the zinc finger (98-130) are  $0.54 \pm 0.11$  and  $0.63 \pm 0.16$  Å, respectively.



**Figure 4.8 Structure of Mt\_aIF2 $\beta$**

Ribbon representation of Mt\_aIF2 $\beta$ , excluding the unfolded N-terminus. The order of the  $\beta$  strands and  $\alpha$  helices is indicated and the sphere represents the zinc ion. (B) Topology diagram of Mt\_aIF2 $\beta$  showing the connectivity between the secondary structure elements.  $\beta$  strands and  $\alpha$  helices are colored purple and green respectively.

#### 4.3.4. Relaxation and dynamics

<sup>15</sup>N-relaxation data (steady state {<sup>1</sup>H}-<sup>15</sup>N NOE, R<sub>1</sub>, R<sub>2</sub>) were recorded to gain an insight into the motional properties of aIF2 $\beta$  (Figure 4.6 and data not shown). The unstructured N-terminus exhibits intermediate NOE values for residues belonging to helix  $\alpha$ 1 consistent with the presence of a partially populated helix in this region (Eliezer *et al*, 2000). The structured core and zinc finger domains exhibit a trimmed weighted correlation time of ~8.45 ns which is in excellent agreement with the predicted value for a 135 residue protein using the Stokes-Einstein equation (8.5 ns at 303K) indicating that these domains do not tumble completely freely of each other. However, the lower than expected NOE value in the structured domains (trimmed mean 0.73) indicates that some mobility on the ps-ns timescales is present. This was confirmed by the poor fits of the R<sub>2</sub>/R<sub>1</sub> ratios to non-isotropic rotational diffusion models (Lee *et al*, 1997; Osborne and Wright, 2001), despite the high degree of anisotropy predicted from hydrodynamics calculations ( $D_{\parallel}/D_{\perp} = 1.43$ ). A possible source for these motions can be restricted inter-domain motions, which would be consistent with the lack of observable NOE contacts between the domains and the large RMSD prior to refinement with residual dipolar couplings.

#### 4.3.5 Comparative analysis of the Mt\_aIF2 $\beta$ structure

The structural classification databases Dali/FSSP (Holm and Sander, 1998) and SCOP (Hubbard *et al*, 1997) were used for comparative analysis of the Mt\_aIF2 $\beta$  structure. The coordinates were compared with known structures using the Dali and SSM search tools. Fifteen proteins with Z scores higher than 2.0 showed similarity to either the core domain or the zinc finger (Table 4.2). All the structures related to the core domain are nucleic acid binding proteins with a helix-turn-helix (HTH) structural motif. This motif is composed of an alpha helix, a linking or turn region and a second alpha helix (recognition helix) involved in sequence specific nucleic acid interactions. The most closely related structures are: Elk-1 (Mo *et al*, 2000), heat shock transcription factor (Damberger *et al*, 1994), GABP $\alpha$  (Batchelor *et al*, 1998) SAP-1 (Mo *et al*, 1998), PU.1 (Pio *et al*,

1996), Mu repressor (Wojciak *et al*, 2001), Mu transposase (Clubb *et al*, 1994), Ribosomal protein L11 (Markus *et al*, 1997) and hRFX1 (Gajiwala *et al*, 2000). Six of these proteins have been solved with their cognate DNA. This interesting result suggests a role for  $\alpha 2$ ,  $\alpha 3$ ,  $\beta 3$  and  $\beta 4$  aIF2 $\beta$  in binding to nucleic acids (Figure 4.9).

Six proteins structurally related to the zinc finger domain were also identified: zinc finger domain from yeast transcription factor SW15 (Neuhaus *et al*, 1992), tramtrack protein (Fairall *et al*, 1993), ribosomal protein L36 (Hard *et al*, 2000), erythroid transcription factor GATA-1 (Omichinski *et al*, 1993), transcription elongation factor TFIIB (Zhu *et al*, 1996) and transcription elongation factor SII (Qian *et al*, 1993). As with the core domain, these proteins are involved in the recognition of sequence specific double-stranded nucleic acids. Taken together, these results suggest that a/eIF2 $\beta$  is involved in the recognition of nucleic acids.

#### 4.3.6 Discussion

The role of different regions of a/eIF2 $\beta$  in translation initiation can be deduced based on our structure and established biochemical facts. The central portion of eukaryotic IF2 $\beta$  (equivalent to the unfolded N-terminus of Mt\_aIF2 $\beta$ ) is necessary for the interaction with eIF2 $\gamma$  as shown by immunoprecipitation, yeast two-hybrid and GST pull-down assays (Thompson *et al*, 2000; Hashimoto *et al*, 2002).  $C\alpha$ ,  $C\beta$  and  $H\alpha$  secondary chemical shifts suggest the presence of a partially folded alpha helix ( $\alpha 1$ ) at the N-terminus (Gutierrez *et al*, 2002). Secondary structure predictions show that this region could form a highly amphipathic helix which could interact with a hydrophobic patch on aIF2 $\beta$ . Mutations in this region of aIF2 $\beta$  affect the hydrolysis of GTP by the  $\gamma$  subunit (Hashimoto *et al*, 2002). Surface potential analysis of the aIF2 $\beta$  structure reveals a conserved hydrophobic patch formed by  $\beta 6$  and  $\alpha 6$  (residues 175-179 and 188-197). The loop connecting these elements is involved in GTP binding. This region could constitute a binding site for the N-terminus of aIF2 $\beta$ .

Analysis of the surface potential of aIF2 $\beta$  reveals the presence of clustered basic residues typical of RNA binding proteins. Figure 4.10 shows the surface potential, as calculated with the program MOLMOL (Koradi *et al*, 1996). Negative patches originate from the charges on residues E46, D49, E65, E73, E90, E93, D94, E104, D109 and E115. The observed distribution of basic residues suggests a putative interaction site comprised by R53, H57, K60, R64 and R76 in helices  $\alpha$ 2,  $\alpha$ 3 and strands  $\beta$ 3 and  $\beta$ 4. R76 and H57 are conserved throughout most of the a/eIF2 $\beta$  sequences and R53 is completely conserved. The closely related structures obtained from the DALI search suggest a RNA binding region in the core domain of aIF2 $\beta$ . The key residue for this function is probably R53, located in the loop connecting  $\alpha$ 2 and  $\alpha$ 3. From the sequence alignments, it seems that at least another basic residue has to be present in helix  $\alpha$ 3 at positions equivalent to 57, 60 or 64 of Mt\_aIF2 $\beta$ . Unfortunately, no functional studies of mutants in this region have been done.

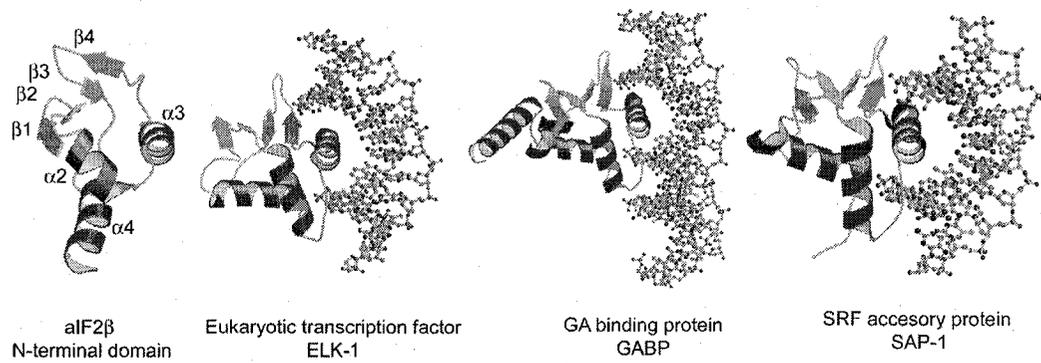
Several mutagenesis studies have shown the importance of C-terminal residues for the function of eIF2 $\beta$  in yeast (Donahue *et al*, 1988). Of particular interest, are the non-conservative substitutions R248T, R253I, R253S, L254P, V268F and V268G (corresponding to positions R114, R117, I118 and L132 of the archaeal protein) that are located at the tip of the zinc finger and allow translation to initiate at UUG codons instead of AUG. As zinc finger domains are normally associated with the recognition of sequence specific double stranded nucleic acids, the C-terminus of aIF2 $\beta$  is likely to constitute a second RNA binding region.

**Table 4.2 Protein structures similar to aIF2 $\beta$**

Name and description	PDB	Z <sup>a</sup>	Secondary structure elements <sup>b</sup>									
			$\beta$ 1	$\beta$ 2	$\alpha$ 2	$\alpha$ 3	$\beta$ 3	$\beta$ 4	$\alpha$ 4	$\beta$ 5	$\beta$ 6	$\beta$ 7
<b>Core Domain</b>												
Elk-1/DNA	1dux	3.8		X	X	X	X	X	X			
Heat shock transcription factor	3hsf	3.6	X	X	X	X	X	X	X			
Mouse GAP/Domain	1awc	3.5	X	X	X	X	X	X	X			
PU.1 ETS domain	1pue	3.1	X	X	X	X	X	X	X			
DNA binding protein SAP-1	1bc7	3.2	X	X	X	X	X	X	X			
Mu bacteriophage repressor	1g4d	2.7			X	X				X		
Mu transposase	1tns	2.3			X	X			X	X		
RNA binding domain of L11	1fow	2.2		X	X	X	X	X				
Transcription factor hrfx 1	1dp7	2.1		X	X	X	X	X	X			
<b>Zinc Finger</b>												
SW15 zinc finger domain	1ncs	4.0										
Tramtrack protein	2drp	3.6								X	X	X
Ribosomal protein L36	1dfe	3.3								X	X	X
N-terminal TFIIIB	1pft	2.9								X	X	X
Transcription factor GATA-1	1gau	2.6								X	X	X
Transcription factor SII	1tfi	2.3								X	X	X

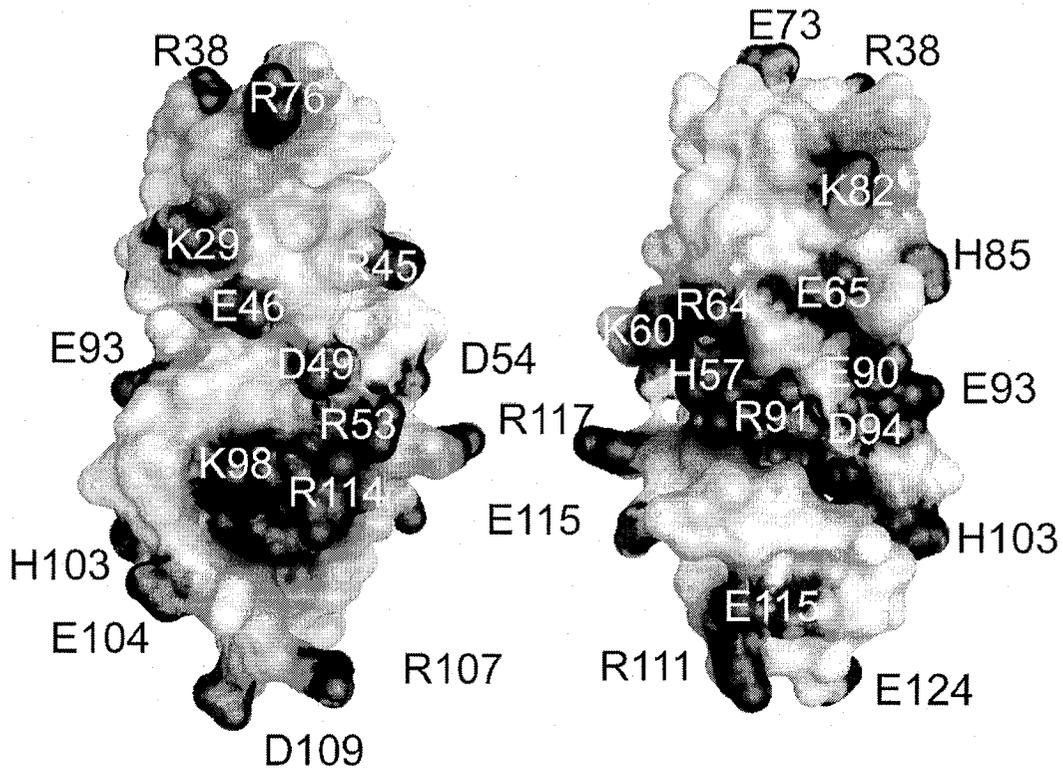
<sup>a</sup> Corresponding Dali Z score

<sup>b</sup> X refers whether the corresponding secondary structure is present in the structure under comparison. References are cited in the text.



**Figure 4.9 Proteins structurally related to Mt\_aIF2β**

ELK-1, GABP and SAP-1 are shown with their cognate nucleic acid, suggesting a role for the HTH motif of aIF2β in nucleic acid recognition.



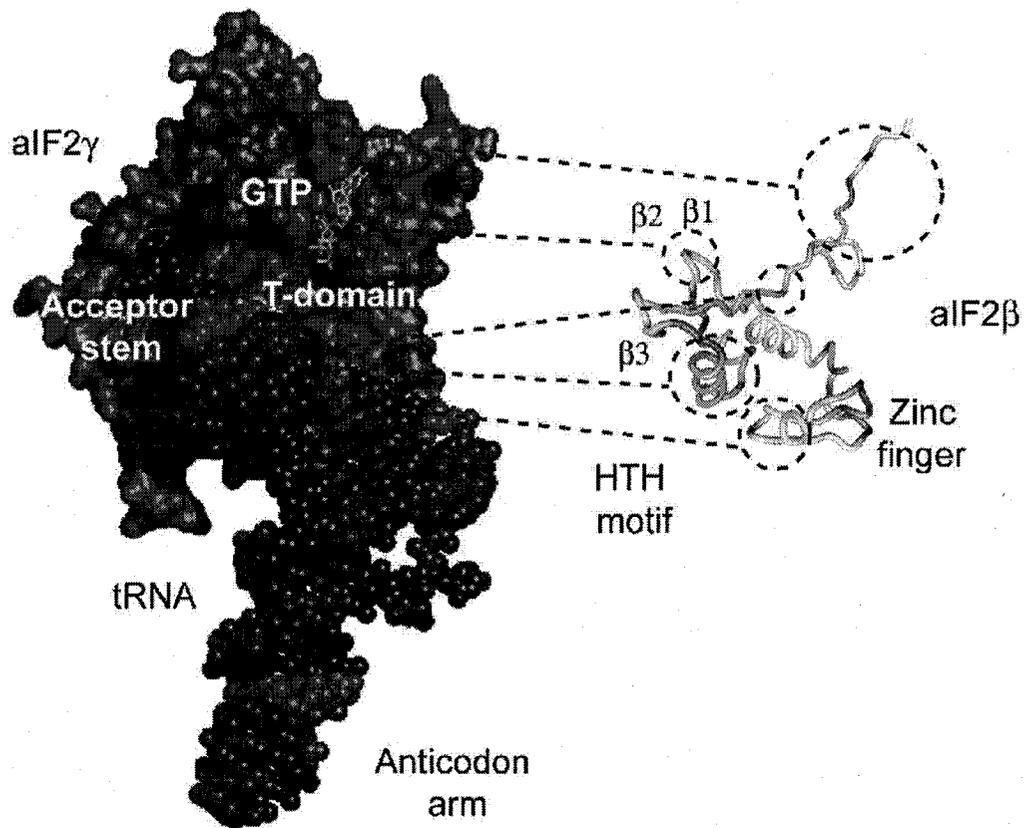
**Figure 4.10 Electrostatic surface plot of Mt\_aIF2 $\beta$**

Acidic and basic residues are colored red and blue, respectively. Surface generated with MOLMOL. Residues 1-28 were excluded for clarity. Both images correspond to 180° rotations.

Based on the aIF2 $\gamma$  structure from *P. abyssi* docked to tRNA (Schmitt *et al.*, 2002), a model for aIF2 $\beta$  in the ternary initiation complex can be hypothesized (Figure 4.11). Assuming that the acceptor stem of tRNA<sub>i</sub> is recognized by the  $\gamma$ -subunit, an interaction of aIF2 $\beta$  with the T-domain is proposed as the size of aIF2 $\beta$  rules out a direct involvement in the recognition of the codon-anticodon interaction. However, there is a clear connection between the recognition of the initiation site (AUG) and the rate of GTP hydrolysis. Affinity labelling with GTP analogs has suggested that eIF2 $\beta$  is in close proximity to the guanine base and ribose moieties of GTP in a region that maps to strands  $\beta$ 1 and  $\beta$ 2 (Bommer and Kurzchalia, 1989; Bommer *et al.*, 1991). At this position, eIF5 presents a well-conserved GNG insertion at the loop connecting these two strands, which may be related to the GAP activities of aIF2 $\beta$ , eIF2 $\beta$  and eIF5. It is possible that upon recognition of the initiation codon, some major structural changes occur in the preinitiation complex that trigger the hydrolysis of GTP. This is supported by early work where conformational changes in the tertiary structure of tRNA upon formation of the codon-anticodon interaction were detected (Schwarz *et al.*, 1976; Robertson *et al.*, 1977; Moller *et al.*, 1979). A similar event may occur upon recognition of the initiation codon, where aIF2 $\beta$  could act as an element signal that stimulates GTP hydrolysis (Figure 4.12).

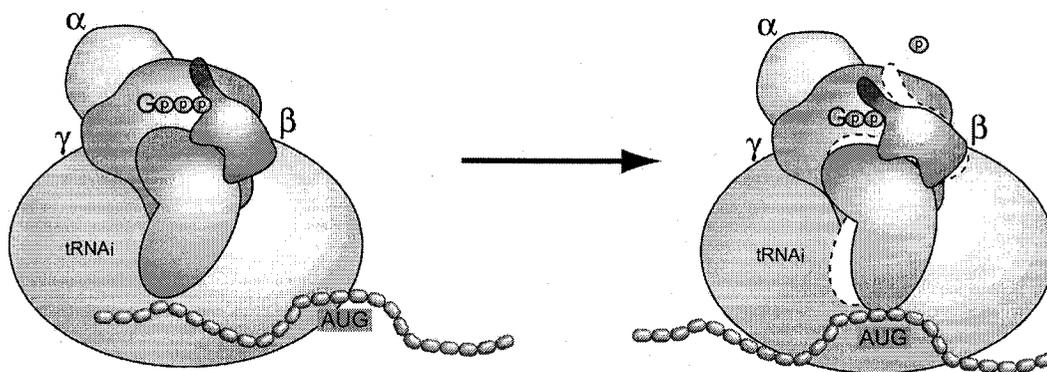
Initiator tRNAs have several unique sequence and structural characteristics that distinguish them from elongator tRNAs. For example, the A1:U72 base pair at the end of the acceptor stem and the three consecutive G:C base pairs in the anticodon stem (G29:C41, G30:C40, G31:C39). Initiators also lack the T $\psi$ C sequence in loop IV of the T-domain containing A54 in place of T54 (of the T $\psi$ C sequence) and A60 instead of pyrimidine-60 (Sprinzl *et al.*, 1998; Hinnebusch, 2000). A54, A50:U64, U51:A63 in the T-region are critical discriminating features as mutating these residues together with A1:U72 can confer elongator function *in vitro* (Drabkin and RajBhandary, 1998). These sequences are potential targets for aIF2 $\beta$  recognition and further studies using

mutagenesis should elucidate the tRNA<sub>i</sub> binding properties of aIF2 $\beta$ . Most studies on a/eIF2 $\beta$  have focused on the N-terminal region or the zinc finger; however, further investigations on the HTH motif and the  $\beta$ 1- $\beta$ 2 turn will provide insight in the role of a/eIF2 $\beta$  in the tRNA<sub>i</sub> recognition and GTP hydrolysis. The proposed model provides a framework for understanding the processes regulated by a/eIF2 in translation initiation.



**Figure 4.11 Mapping of biochemical data to the Mt\_aIF2 $\beta$  structure**

Biochemical data for a/eIF2 $\beta$  mapped to a model of *P. abyssi* aIF2 $\gamma$  (green, PDB identifier 1KK1) bound to tRNA (magenta, PDB identifier 1B23). A location for aIF2 $\beta$  in the ternary complex is suggested. Regions in aIF2 $\beta$  involved in GTP, tRNA<sub>i</sub>, and aIF2 $\gamma$  interactions are circled and connected to their corresponding ligand (see text for details). Positions that differentiate initiator tRNA from elongator tRNA are colored in blue. aIF2 $\beta$  has been colored based on sequence conservation where red represents highly conserved residues.



**Figure 4.12 Proposed function of aIF2β in translation initiation**

Upon recognition of the initiation codon, aIF2β could act as an element signal that stimulates GTP hydrolysis (See text for details)

## **4.4 Materials and methods**

### **4.4.1 Protein expression and purification**

Initiation factor aIF2 $\beta$  from *Methanobacterium thermoautotrophicum* (gene MTH1769) was subcloned into pET15b (Novagen, Inc., Madison, WI) and expressed in *E. coli* BL21 as an oligo-histidine (His-tag) fusion protein of 161 residues. Cells were grown at 37°C to an OD600 of 0.8 and induced with 1 mM IPTG. Afterwards, the temperature was reduced to 30°C and the cells were allowed to express the protein for 3 hours before harvesting. The media used were either LB or M9 minimal media containing <sup>15</sup>N-ammonium chloride and/or <sup>13</sup>C-glucose (Cambridge Isotopes Laboratory, Andover, MA) and 50  $\mu$ M ZnCl<sub>2</sub>. aIF2 $\beta$  was purified by heat denaturation of endogenous *E. coli* proteins and affinity chromatography on Ni<sup>2+</sup>-loaded chelating sepharose (Amersham Pharmacia Biotech, Piscataway, NJ). The N-terminal His-tag was cleaved from aIF2 $\beta$  by treatment for 24 h at room temperature with thrombin (Amersham Pharmacia Biotech) at 1 unit per mg fusion protein. Benzamidine sepharose was used to remove thrombin.

### **4.4.2 Mass spectrometry**

ESI/MS analyses were performed utilizing a Perkin Elmer, API III spectrometer. aIF2 $\beta$  was purified as above and dialyzed extensively against 10% acetic acid. For protein grown in M9 medium, 1 Da per nitrogen was subtracted from the measured mass to account for the <sup>15</sup>N isotope enrichment.

### **4.4.3 Cobalt visible spectrophotometry**

Ultraviolet and visible light spectra of cobalt-aIF2 $\beta$  complexes were obtained on a Cary 100 UV-visible spectrophotometer at 20°C. The spectrum was scanned at 600 nm/min from 800 to 200 nm with a 1-cm path length. Data points were collected each 2 nm. The sample was 0.6 mg/ml proteins in 50 mM phosphate buffer, 150 mM NaCl, 2 mM dithiothreitol (DTT) at pH 6.0. CoCl<sub>2</sub> was added from a 1 mM stock solution to final concentrations of 80, 160 and 240  $\mu$ M.

The protein was grown on LB medium and purified as described. Samples was dialyzed extensively against buffer containing 1 mM EDTA, to remove bound zinc and dialyzed again against buffer to get rid of EDTA.

#### 4.4.4 NMR spectroscopy

All NMR experiments were recorded at 310 K using standard double and triple resonance techniques on  $^{15}\text{N}$  or  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled samples (Bax and Grzesiek, 1993). All of the experiments were done on a Bruker DRX500 and Varian INOVA 800 MHz spectrometers. The following experiments were recorded and evaluated: (1) for backbone assignments: HNCACB and CBCACONH (Grzesiek *et al*, 1992; Constantine *et al*, 1993); (2) for side-chain and NOE assignments: from  $^{15}\text{N}$ -TOCSY,  $^{15}\text{N}$ -edited NOESY, 2D homonuclear NOESY in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$ ; (3) for dihedral angle restraints:  $^3\text{J}_{\text{HN-H}\alpha}$  coupling constants were obtained from HNHA experiment (Kuboniwa *et al*, 1994); (4) for  $^{15}\text{N}$ - $^1\text{H}$  dipolar couplings: an IPAP-HSQC experiment on an isotropic medium and on a sample containing 18 mg/ml Pfl phage (Hansen *et al*, 1998; Ottiger *et al*, 1998); (5) for backbone dynamics:  $^{15}\text{N}$ - $^1\text{H}$  heteronuclear NOE data were measured by taking the ratio of peak intensities from experiments performed with and without  $^1\text{H}$  presaturation. Hydrogen bond constraints were introduced to secondary structure regions as determined by chemical shift analysis, HNHA experiments and characteristic NOE patterns. Hydrogen bonds were defined as a restraint from the carbonyl oxygen to the amide hydrogen and nitrogen, using a standard length of 1.8 Å and 2.8 Å respectively. Additional  $\psi$  and  $\phi$  dihedral restraints were obtained using TALOS (Cornilescu *et al*, 1999). All NMR spectra were processed using either XWINNMR version 2.5 or 3.1 (Bruker Biospin) or GIFA (Malliavin *et al*, 1998). Evaluation of spectra and manual assignments were completed with XEASY (Bartels *et al*, 1995). NMR samples were ~1.0 mM protein in 50 mM Bis-tris buffer, 0.30 M NaCl, 50  $\mu\text{M}$   $\text{ZnCl}_2$ , 1 mM DTT, 0.02% (w/v)  $\text{NaN}_3$  at pH 6.0.

#### 4.4.5 Analysis and structure calculations

CNS 1.1 software (Brunger *et al*, 1998) was used to generate an initial fold of aIF2 $\beta$  with a basic set of NOEs acquired from manual assignments of 3D  $^{15}\text{N}$ -edited NOESY and 2D homonuclear NOE spectra including dihedral angle and hydrogen bond constraints (Wüthrich, 1986). These calculations generated a fold that was used as a model template for automated assignments by ARIA1.1 (Nilges *et al*, 1997). The final structure of aIF2 $\beta$  was calculated with a total set of 835 constraints (Table 4.1) collected from the experiments described earlier. In the final round of calculations, CNS 1.1 was extended to incorporate RDC restraints for further refinement, using the torsion angle space. The axial and rhombic components of the alignment tensor were defined from a histogram of measured RDCs (Clare *et al*, 1998) and optimized by a grid search method (Clare *et al*, 1998). Refinement of the whole protein using a single alignment tensor resulted in poor fits, which may reflect interdomain motion. We therefore proceeded to refine the structure using two separate alignment tensors to define each well-structured domain. The Twenty structures were selected based on the lowest overall energy and least violations to represent final structures. PROCHECK was used to generate Ramachandran plots to check the protein's stereochemical geometry (Laskowski *et al*, 1993). The coordinates of aIF2 $\beta$  have been deposited in the RCSB under PDB code 1NEE and the NMR assignments under BMRB accession 4385.

## Chapter 5: Structural genomics of gene regulation

The proteins studied in this thesis are part of two separate structural genomics projects. YaeO and CsrA were studied under the Montreal-Kingston structural genomics initiative which focuses on genes from both *E. coli* K-12 and O157:H7 (Matte *et al.*, 2003). Translation initiation factor aIF2 $\beta$  was part of the *Methanobacterium thermoautotrophicum* structural genomics project (Yee *et al.*, 2002).

Structural genomics efforts have three main goals: (1) To determine a large number of protein structures to complement the expanding databases of genome sequences; (2) the generation of biological insights into how proteins function at the atomic level and (3) to help assign and verify functions for these proteins in combination with other biochemical approaches. The structures derived from these projects will also yield valuable clues to the rules for predicting protein folding. For this thesis, proteins were chosen based on their implication in regulating gene expression and the absence of a known structure. In the following sections, I outline the main contributions to knowledge derived from each of the proteins studied in this thesis. In each case, structural studies have revealed new functional aspects of the protein studied and allowed the proposal of models explaining their function.

### 5.1 YaeO

No structures with significant sequence similarity to YaeO were found using BLAST searches and only one annotation regarding YaeO's function in the cell was published (Pichoff *et al.*, 1998). Our main objective was to solve the solution structure of YaeO and confirm its involvement in the modulation of bacterial transcription termination. This is also the first structural determination of a Rho-specific inhibitor of transcription termination. The YaeO structure reveals that its fold is topologically similar to that of the RNA binding domain of small ribonucleoproteins (Sm-fold). The most important difference between the Sm-fold

and YaeO is the presence of an additional  $\beta$ -strand ( $\beta 7$ ) in YaeO. Sm proteins tend to aggregate into higher order complexes by the interaction of  $\beta 1$  and  $\beta 6$  of adjacent monomers (Kambach *et al*, 1999; Schumacher *et al*, 2002). In YaeO, this oligomerization is probably prevented by the presence of an additional strand  $\beta 7$ . Interestingly, Sm protein can function as pleiotropic regulators and can associate with exported snRNAs at short single-stranded regions. It would be expected that YaeO, having a similar fold to Sm protein would bind to nucleic acids as well; a hypothesis compatible with an inhibitory effect in the formation of the Rho-RNA complex. However our binding studies suggest that this is not the case and is probably a result of the highly acidic nature of YaeO. Whether the similarities between YaeO and Sm proteins is a coincidence or the result of an evolutionary relationship remains an open question.

Rho is unique to prokaryotes and essential for the viability of many bacterial species and as a consequence it is an attractive target for drug development. The use of antibiotics targeting Rho dates back from the isolation of bicyclomycin in 1972 from *Streptomyces sapporonensis* and *Streptomyces aizumenses* (Kamiya *et al*, 1972; Miyamura *et al*, 1972). Currently, BCM is commercially available under the name Bicozamycin and has been used for the treatment of nonspecific diarrhea. BCM is a reversible, noncompetitive inhibitor of ATP turnover that exerts its activity by binding at the interface of adjacent C-terminal domains of Rho. BCM also slows Rho's movement along RNA and acts as a mixed-type inhibitor for RNA binding at the secondary site (Skordalakes *et al*, 2005). In the cell, BCM attenuates the ability of Rho to reach and dissociate the RNA polymerase from its DNA template, yielding unnaturally long RNA transcripts. Our studies on YaeO reveal an additional mechanism for the inhibition of Rho activity and set the stage for future efforts to develop additional antibiotics targeting Rho.

## 5.2. CsrA

The carbon storage regulator A, CsrA, has received a lot of attention during the past decade as it is involved in a very wide range of physiological

processes in *E. coli*. Some of them include the regulation of gluconeogenesis, glycogen biosynthesis and catabolism, and biofilm formation (Romeo *et al*, 1993; Sabnis *et al*, 1995; Jackson *et al*, 2002); the activation of glycolysis, acetate metabolism and flagellum biosynthesis (Sabnis *et al*, 1995; Wei *et al*, 2000; Wei *et al*, 2001) and quorum sensing systems.

No structure was available for a CsrA homolog at the start of this project. However, CsrA was annotated as being a member of the KH-domain protein family. One of the most interesting observations derived from our work is the demonstration that CsrA protein is not a member of the KH protein family. This misleading annotation is the result of poor sequence analysis (Liu *et al*, 1995). Indeed, CsrA constitutes a novel fold that adds to the increasing list of RNA binding motifs such as the RRM, KH, OB and Sm folds. CsrA is an intimately associated dimer held together by the formation of a mixed  $\beta$ -sheet between  $\beta 1'$ - $\beta 4$  and  $\beta 2$ - $\beta 5'$ . This structural arrangement also raises interesting questions regarding the mechanism of folding, as the proper structure of CsrA can only be acquired after a second CsrA molecule has been translated. After submission of our paper coordinates for the X-ray structure of the *P. aeruginosa* homolog were released. The two structures are very similar with an RMSD of 2.8 Å for backbone atoms.

Functionally, our studies revealed insights into the RNA binding region of CsrA. According to our data, the GxxG motif is implicated in the recognition of target RNA sequences. The dimeric structure of CsrA allows it to possess two identical RNA binding motifs to simultaneously recognize both GGA sequences in the Shine-Dalgarno and the upstream hairpin loop that differentiate target mRNAs. We proposed that CsrA first recognizes the hairpin loop upstream the SD experiencing a conformational change that increases its CsrA affinity for the GGA sequence in the single stranded Shine-Dalgarno sequence.

### 5.3 $\alpha$ IF2 $\beta$

Structural data for the three subunits of the  $\alpha$ /eIF2 complex has been elusive for many years. For this reason, we decided to undertake structural studies

of the archaeal homolog of the  $\beta$  subunit of eIF2. At the time, insights into the function of a/eIF2 came from the examination of the subunit's primary sequences and genetic and biochemical analyses (Hershey and Merrick, 2000). While subunits  $\alpha$  and  $\gamma$  are known for their regulatory and tRNA and GTP binding functions, respectively, the role for the beta subunit was not very clear.

Our first contribution was the demonstration that zinc is an important structural element of aIF2 $\beta$ . The fact that a/eIF2 $\beta$  contains a C<sub>2</sub>-C<sub>2</sub> zinc finger-like sequence has been known for a long time but attempts to demonstrate zinc binding had failed and it was assumed that this ion was not present or required. Our work showed that in the absence of zinc the C-terminus is unfolded and is the target for endogenous proteases (Gutierrez *et al*, 2002). Our solution structure was also the first for the intact aIF2 $\beta$  (before our publication, the solution structures for the independent domains of the *M. janaschii* homolog was reported (Cho and Hoffman, 2002)) Most studies on a/eIF2 $\beta$  have focused on the N-terminal region or the zinc finger; however we observed that the core domain has a fold similar to the helix-turn-helix motif. This is a signature fold of nucleic acid binding proteins and shows that a/eIF2 $\beta$  has another potential RNA binding region (Gutierrez *et al*, 2004).

Our proposal is that upon recognition of the initiation codon, major structural changes occur in the preinitiation complex (a/eIF2-GTP-tRNA<sub>i</sub>) that trigger the hydrolysis of GTP. During this, process aIF2 $\beta$  acts as the element signal that links the recognition of the initiation site, *via* its RNA binding motifs, with the GTPase activity of aIF2 $\gamma$ .

## References

- Adinolfi, S., C. Bagni, M. A. Castiglione Morelli, F. Fraternali, G. Musco and A. Pastore** (1999). Novel RNA-binding motif: the KH module. *Biopolymers* **51**(2): 153-64.
- Alderete, J. P., S. Jarrahan and A. P. Geballe** (1999). Translational effects of mutations and polymorphisms in a repressive upstream open reading frame of the human cytomegalovirus UL4 gene. *J Virol* **73**(10): 8330-7.
- Altier, C., M. Suyemoto and S. D. Lawhon** (2000). Regulation of *Salmonella enterica* serovar *typhimurium* invasion genes by *cstA*. *Infect Immun* **68**(12): 6790-7.
- Altuvia, S., A. Zhang, L. Argaman, A. Tiwari and G. Storz** (1998). The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *Embo J* **17**(20): 6069-75.
- Amster-Choder, O. and A. Wright** (1992). Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. *Science* **257**(5075): 1395-8.
- Amster-Choder, O. and A. Wright** (1993). Transcriptional regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. *J Cell Biochem* **51**(1): 83-90.
- Arndt, K. M. and M. J. Chamberlin** (1990). RNA chain elongation by *Escherichia coli* RNA polymerase. Factors affecting the stability of elongating ternary complexes. *J Mol Biol* **213**(1): 79-108.
- Arndt, K. M. and C. M. Kane** (2003). Running with RNA polymerase: eukaryotic transcript elongation. *Trends Genet* **19**(10): 543-50.
- Arnosti, D. N., S. Barolo, M. Levine and S. Small** (1996). The *eve stripe 2* enhancer employs multiple modes of transcriptional synergy. *Development* **122**(1): 205-14.

- Asano, K., T. Krishnamoorthy, L. Phan, G. D. Pavitt and A. G. Hinnebusch** (1999). Conserved bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *Embo J* **18**(6): 1673-88.
- Atchison, M. L.** (1988). Enhancers: mechanisms of action and cell specificity. *Annu Rev Cell Biol* **4**: 127-53.
- Baker, C. S., I. Morozov, K. Suzuki, T. Romeo and P. Babitzke** (2002). CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol Microbiol* **44**(6): 1599-610.
- Bartels, C., M. Billeter, P. Gunter and K. Wutrich** (1995). The program XEASY for computer supported NMR spectral analysis of biological molecules. *J Biomol NMR* **6**: 1-10.
- Bartels, C., T. Xia, M. Billeter, P. Güntert and K. Wüthrich** (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **6**: 1-10.
- Batchelor, H. A., D. E. Piper, F. C. de la Brousse, S. L. McKnight and C. Wolberger** (1998). The structure of GABP $\alpha/\beta$ : An ETS Domain-Ankyrin Repeat Heterodimer Bound to DNA. *Science* **279**: 1037-41.
- Bax, A. and S. Grzesiek** (1993). Methodological advances in protein NMR. *Acc. Chem. Res* **26**: 131-138.
- Bell, S. D. and S. P. Jackson** (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**(6): 222-8.
- Bell, S. D., C. P. Magill and S. P. Jackson** (2001). Basal and regulated transcription in Archaea. *Biochem Soc Trans* **29**(Pt 4): 392-5.
- Belle, R., J. Derancourt, R. Poulhe, J. P. Capony, R. Ozon and O. Mulner-Lorillon** (1989). A purified complex from *Xenopus* oocytes contains a p47 protein, an in vivo substrate of MPF, and a p30 protein respectively

homologous to elongation factors EF-1 gamma and EF-1 beta. *FEBS Lett* **255**(1): 101-4.

**Belotserkovskaya, R., S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky and D. Reinberg** (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**(5636): 1090-3.

**Benne, R., J. Van den Burg, J. P. Brakenhoff, P. Sloof, J. H. Van Boom and M. C. Tromp** (1986). Major transcript of the frameshifted *coxII* gene from *trypanosome* mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**(6): 819-26.

**Berg, K. L., C. Squires and C. L. Squires** (1989). Ribosomal RNA operon anti-termination. Function of leader and spacer region box B-box A sequences and their conservation in diverse micro-organisms. *J Mol Biol* **209**(3): 345-58.

**Bertini, I. and C. Luchinat** (1984). High spin cobalt(II) as a probe for the investigation of metalloproteins. *Adv Inorg Biochem* **6**: 71-111.

**Bicknell, R., A. Schaeffer, D. S. Auld, J. F. Riordan, R. Monnanni and I. Bertini** (1985). Protease susceptibility of zinc- and apo-carboxypeptidase A. *Biochem Biophys Res Commun* **133**(2): 787-93.

**Blackwood, E. M. and J. T. Kadonaga** (1998). Going the distance: a current view of enhancer action. *Science* **281**(5373): 61-3.

**Bogenhagen, D. F. and D. D. Brown** (1981). Nucleotide sequences in *Xenopus* 5S DNA required for transcription termination. *Cell* **24**(1): 261-70.

**Bommer, U. A., R. Kraft, T. V. Kurzchalia, N. T. Price and C. G. Proud** (1991). Amino acid sequence analysis of the beta- and gamma-subunits of eukaryotic initiation factor eIF-2. Identification of regions interacting with GTP. *Biochim Biophys Acta* **1079**(3): 308-15.

**Bommer, U. A. and T. V. Kurzchalia** (1989). GTP interacts through its ribose and phosphate moieties with different subunits of the eukaryotic initiation factor eIF-2. *FEBS Lett* **244**(2): 323-7.

- Bonetti, B., L. Fu, J. Moon and D. M. Bedwell** (1995). The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol* **251**(3): 334-45.
- Brendel, V., G. H. Hamm and E. N. Trifonov** (1986). Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. *J Biomol Struct Dyn* **3**(4): 705-23.
- Brennan, C. A., A. J. Dombroski and T. Platt** (1987). Transcription termination factor rho is an RNA-DNA helicase. *Cell* **48**(6): 945-52.
- Briercheck, D. M., T. J. Allison, J. P. Richardson, J. F. Ellena, T. C. Wood and G. S. Rule** (1996).  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  resonance assignments and secondary structure determination of the RNA-binding domain of E.coli rho protein. *J Biomol NMR* **8**(4): 429-44.
- Briercheck, D. M., T. C. Wood, T. J. Allison, J. P. Richardson and G. S. Rule** (1998). The NMR structure of the RNA binding domain of E. coli rho factor suggests possible RNA-protein interactions. *Nat Struct Biol* **5**(5): 393-9.
- Brown, S., E. R. Brickman and J. Beckwith** (1981). Blue ghosts: a new method for isolating amber mutants defective in essential genes of *Escherichia coli*. *J Bacteriol* **146**(1): 422-5.
- Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren** (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**(Pt 5): 905-21.
- Busby, S. and R. H. Ebright** (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**(5): 743-6.

- Carlberg, U., A. Nilsson and O. Nygard** (1990). Functional properties of phosphorylated elongation factor 2. *Eur J Biochem* **191**(3): 639-45.
- Carter, A. P., W. M. Clemons, Jr., D. E. Brodersen, R. J. Morgan-Warren, T. Hartsch, B. T. Wimberly and V. Ramakrishnan** (2001). Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**(5503): 498-501.
- Chen, C. Y. and A. B. Shyu** (1994). Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol Cell Biol* **14**(12): 8471-82.
- Cheng, S. W., E. C. Lynch, K. R. Leason, D. L. Court, B. A. Shapiro and D. I. Friedman** (1991). Functional importance of sequence in the stem-loop of a transcription terminator. *Science* **254**(5035): 1205-7.
- Chien, Y., J. D. Helmann and S. H. Zinder** (1998). Interactions between the promoter regions of nitrogenase structural genes (nifHDK2) and DNA-binding proteins from N<sub>2</sub>- and ammonium-grown cells of the archaeon *Methanosarcina barkeri* 227. *J Bacteriol* **180**(10): 2723-8.
- Cho, S. and D. W. Hoffman** (2002). Structure of the beta Subunit of Translation Initiation Factor 2 from the Archaeon *Methanococcus jannaschii*: A Representative of the eIF2beta/eIF5 Family of Proteins. *Biochemistry* **41**(18): 5730-42.
- Clore, G. M., A. M. Gronenborn and A. Bax** (1998). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* **133**(1): 216-21.
- Clore, G. M., A. M. Gronenborn and N. Tjandra** (1998). Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* **131**(1): 159-62.

- Clubb, R. T., J. G. Omichinski, H. Savilahti, K. Mizuuchi, A. M. Gronenborn and G. M. Clore** (1994). A novel class of winged helix-turn-helix protein: the DNA-binding domain of Mu transposase. *Structure* **2**(11): 1041-8.
- Cohen-Kupiec, R., C. Blank and J. A. Leigh** (1997). Transcriptional regulation in Archaea: in vivo demonstration of a repressor binding site in a methanogen. *Proc Natl Acad Sci U S A* **94**(4): 1316-20.
- Conaway, J. W. and R. C. Conaway** (1999). Transcription elongation and human disease. *Annu Rev Biochem* **68**: 301-19.
- Conlon, I. and M. Raff** (1999). Size control in animal development. *Cell* **96**(2): 235-44.
- Connelly, S. and J. L. Manley** (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**(4): 440-52.
- Constantine, K. L., V. Goldfarb, M. Wittekind, M. S. Friedrichs, J. Anthony, S. C. Ng and L. Mueller** (1993). Aliphatic <sup>1</sup>H and <sup>13</sup>C resonance assignments for the 26-10 antibody VL domain derived from heteronuclear multidimensional NMR spectroscopy. *J Biomol NMR* **3**(1): 41-54.
- Copeland, P. R.** (2003). Regulation of gene expression by stop codon recoding: selenocysteine. *Gene* **312**: 17-25.
- Cornilescu, G., F. Delaglio and A. Bax** (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**(3): 289-302.
- Cui, Y., A. Chatterjee, Y. Liu, C. K. Dumenyo and A. K. Chatterjee** (1995). Identification of a global repressor gene, *rsmA*, of *Erwinia carotovora* subsp. *carotovora* that controls extracellular enzymes, N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in soft-rotting *Erwinia* spp. *J Bacteriol* **177**(17): 5108-15.

- Cui, Y., A. Mukherjee, C. K. Dumenyo, Y. Liu and A. K. Chatterjee** (1999). rsmC of the soft-rotting bacterium *Erwinia carotovora* subsp. *carotovora* negatively controls extracellular enzyme and harpin(Ecc) production and virulence by modulating levels of regulatory RNA (rsmB) and RNA-binding protein (RsmA). *J Bacteriol* **181**(19): 6042-52.
- Damberger, F. F., J. G. Pelton, C. J. Harrison, H. C. Nelson and D. E. Wemmer** (1994). Solution structure of the DNA-binding domain of the heat shock transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy. *Protein Sci* **3**(10): 1806-21.
- d'Aubenton Carafa, Y., E. Brody and C. Thermes** (1990). Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J Mol Biol* **216**(4): 835-58.
- de Moor, C. H. and J. D. Richter** (2001). Translational control in vertebrate development. *Int Rev Cytol* **203**: 567-608.
- deHaseth, P. L., M. L. Zupancic and M. T. Record, Jr.** (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol* **180**(12): 3019-25.
- Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax** (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**(3): 277-93.
- Delihias, N.** (1995). Regulation of gene expression by trans-encoded antisense RNAs. *Mol Microbiol* **15**(3): 411-4.
- Dennis, P. P.** (1997). Ancient ciphers: translation in Archaea. *Cell* **89**(7): 1007-10.
- Dominguez, C., R. Boelens and A. M. Bonvin** (2003). HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**(7): 1731-7.

- Donahue, T. F., A. M. Cigan, E. K. Pabich and B. C. Valavicius** (1988). Mutations at a Zn(II) finger motif in the yeast eIF-2 beta gene alter ribosomal start-site selection during the scanning process. *Cell* **54**(5): 621-32.
- Drabkin, H. J. and U. L. RajBhandary** (1998). Initiation of protein synthesis in mammalian cells with codons other than AUG and amino acids other than methionine. *Mol Cell Biol* **18**(9): 5140-7.
- Dubey, A. K., C. S. Baker, K. Suzuki, A. D. Jones, P. Pandit, T. Romeo and P. Babitzke** (2003). CsrA regulates translation of the *Escherichia coli* carbon starvation gene, *cstA*, by blocking ribosome access to the *cstA* transcript. *J Bacteriol* **185**(15): 4450-60.
- Ekiel, I., M. Abrahamson, D. B. Fulton, P. Lindahl, A. C. Storer, W. Levadoux, M. Lafrance, S. Labelle, Y. Pomerleau, D. Groleau, L. LeSauter and K. Gehring** (1997). NMR structural studies of human cystatin C dimers and monomers. *J Mol Biol* **271**(2): 266-77.
- Elf, J., O. G. Berg and M. Ehrenberg** (2001). Comparison of repressor and transcriptional attenuator systems for control of amino acid biosynthetic operons. *J Mol Biol* **313**(5): 941-54.
- Eliezer, D., J. Chung, H. J. Dyson and P. E. Wright** (2000). Native and non-native secondary structure and dynamics in the pH 4 intermediate of apomyoglobin. *Biochemistry* **39**(11): 2894-901.
- Eul, J., M. Graessmann and A. Graessmann** (1996). Trans-splicing and alternative-tandem-cis-splicing: two ways by which mammalian cells generate a truncated SV40 T-antigen. *Nucleic Acids Res* **24**(9): 1653-61.
- Fairall, L., J. W. Schwabe, L. Chapman, J. T. Finch and D. Rhodes** (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**(6454): 483-7.

- Franklin, N. C.** (1974). Altered reading of genetic signals fused to the *N* operon of bacteriophage lambda: genetic evidence for modification of polymerase by the protein product of the *N* gene. *J Mol Biol* **89**(1): 33-48.
- Freistroffer, D. V., M. Y. Pavlov, J. MacDougall, R. H. Buckingham and M. Ehrenberg** (1997). Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *Embo J* **16**(13): 4126-33.
- Friedman, D. I. and D. L. Court** (1995). Transcription antitermination: the lambda paradigm updated. *Mol Microbiol* **18**(2): 191-200.
- Gajiwala, K. A., H. Chen, F. Cornille, B. P. Roques, W. Reith, B. Mach and S. K. Burley** (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* **403**: 916-21.
- Gallie, D. R.** (1998). A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation. *Gene* **216**(1): 1-11.
- Gaspar, N. J., T. G. Kinzy, B. J. Scherer, M. Humbelin, J. W. Hershey and W. C. Merrick** (1994). Translation initiation factor eIF-2. Cloning and expression of the human cDNA encoding the gamma-subunit. *J Biol Chem* **269**(5): 3415-22.
- Geiselman, J., Y. Wang, S. E. Seifried and P. H. von Hippel** (1993). A physical model for the translocation and helicase activities of *Escherichia coli* transcription termination protein Rho. *Proc Natl Acad Sci U S A* **90**(16): 7754-8.
- Gingras, A. C., B. Raught and N. Sonenberg** (2001). Regulation of translation initiation by FRAP/mTOR. *Genes Dev* **15**(7): 807-26.
- Gish, K. and C. Yanofsky** (1993). Inhibition of expression of the tryptophanase operon in *Escherichia coli* by extrachromosomal copies of the *tna* leader region. *J Bacteriol* **175**(11): 3380-7.

- Gott, J. M. and R. B. Emeson** (2000). Functions and mechanisms of RNA editing. *Annu Rev Genet* **34**: 499-531.
- Gottesman, S.** (1984). Bacterial regulation: global regulatory networks. *Annu Rev Genet* **18**: 415-41.
- Gray, N. K., S. Quick, B. Goossen, A. Constable, H. Hirling, L. C. Kuhn and M. W. Hentze** (1993). Recombinant iron-regulatory factor functions as an iron-responsive-element-binding protein, a translational repressor and an aconitase. A functional assay for translational repression and direct demonstration of the iron switch. *Eur J Biochem* **218**(2): 657-67.
- Gray, N. K. and M. Wickens** (1998). Control of translation initiation in animals. *Annu Rev Cell Dev Biol* **14**: 399-458.
- Greenblatt, J., J. R. Nodwell and S. W. Mason** (1993). Transcriptional antitermination. *Nature* **364**(6436): 401-6.
- Grummt, I., H. Rosenbauer, I. Niedermeyer, U. Maier and A. Ohrlein** (1986). A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. *Cell* **45**(6): 837-46.
- Grzesiek, S., H. Dobeli, R. Gentz, G. Garotta, A. M. Labhardt and A. Bax** (1992).  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* **31**(35): 8180-90.
- Gualerzi, C. O. and C. L. Pon** (1990). Initiation of mRNA translation in prokaryotes. *Biochemistry* **29**(25): 5881-9.
- Gudapaty, S., K. Suzuki, X. Wang, P. Babitzke and T. Romeo** (2001). Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. *J Bacteriol* **183**(20): 6017-27.
- Gueydan, C., L. Droogmans, P. Chalon, G. Huez, D. Caput and V. Krays** (1999). Identification of TIAR as a protein binding to the translational

- regulatory AU-rich element of tumor necrosis factor alpha mRNA. *J Biol Chem* **274**(4): 2322-6.
- Gutierrez, P., S. Coillet-Matillon, C. Arrowsmith and K. Gehring** (2002). Zinc is required for structural stability of the C-terminus of archaeal translation initiation factor aIF2beta. *FEBS Lett* **517**(1-3): 155-8.
- Gutierrez, P., M. J. Osborne, N. Siddiqui, J. F. Trempe, C. Arrowsmith and K. Gehring** (2004). Structure of the archaeal translation initiation factor aIF2 beta from *Methanobacterium thermoautotrophicum*: implications for translation initiation. *Protein Sci* **13**(3): 659-67.
- Hahn, S.** (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* **11**(5): 394-403.
- Hall, T. M.** (2002). Poly(A) tail synthesis and regulation: recent structural insights. *Curr Opin Struct Biol* **12**(1): 82-8.
- Hampsey, M. and D. Reinberg** (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**(4): 429-32.
- Hansen, M. R., L. Mueller and A. Pardi** (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5**(12): 1065-74.
- Hard, T., A. Rak, P. Allard, L. Kloo and M. Garber** (2000). The solution structure of ribosomal protein L36 from *Thermus thermophilus* reveals a zinc-ribbon-like fold. *J Mol Biol* **296**(1): 169-80.
- Hartzog, G. A., J. L. Speer and D. L. Lindstrom** (2002). Transcript elongation on a nucleoprotein template. *Biochim Biophys Acta* **1577**(2): 276-86.
- Hashimoto, N. N., L. S. Carnevalli and B. A. Castilho** (2002). Translation initiation at non-AUG codons mediated by weakened association of eukaryotic initiation factor (eIF) 2 subunits. *Biochem J* **367**(Pt 2): 359-68.
- Hentze, M. W., S. W. Caughman, T. A. Rouault, J. G. Barriocanal, A. Dancis, J. B. Harford and R. D. Klausner** (1987). Identification of the

iron-responsive element for the translational regulation of human ferritin mRNA. *Science* **238**(4833): 1570-3.

**Hershey, J. W. B. and W. C. Merrick** (2000). The pathway and mechanism of initiation of protein synthesis. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 33-88.

**Hinnebusch, A.** (2000). Mechanism and regulation of initiator methionyl-tRNA binding to ribosomes. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. Mathews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 185-243.

**Holm, L. and C. Sander** (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**(1): 316-9.

**Huang, H. K., H. Yoon, E. M. Hannig and T. F. Donahue** (1997). GTP hydrolysis controls stringent selection of the AUG start codon during translation initiation in *Saccharomyces cerevisiae*. *Genes Dev* **11**(18): 2396-413.

**Hubbard, T. J., A. G. Murzin, S. E. Brenner and C. Chothia** (1997). SCOP: a structural classification of proteins database. *Nucleic Acids Res* **25**(1): 236-9.

**Hughes, K. T. and K. Mathee** (1998). The anti-sigma factors. *Annu Rev Microbiol* **52**: 231-86.

**Isaksen, M. L., S. T. Rishovd, R. Calendar and B. H. Lindqvist** (1992). The polarity suppression factor of bacteriophage P4 is also a decoration protein of the P4 capsid. *Virology* **188**(2): 831-9.

**Ishihama, A.** (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**: 499-518.

**Jack, J., D. Dorsett, Y. Delotto and S. Liu** (1991). Expression of the *cut* locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development* **113**(3): 735-47.

- Jackson, D. W., K. Suzuki, L. Oakford, J. W. Simecka, M. E. Hart and T. Romeo** (2002). Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. *J Bacteriol* **184**(1): 290-301.
- Jacob, F. and J. Monod** (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318-56.
- Janssen, G. M., G. D. Maessen, R. Amons and W. Moller** (1988). Phosphorylation of elongation factor 1 beta by an endogenous kinase affects its catalytic nucleotide exchange activity. *J Biol Chem* **263**(23): 11063-6.
- Jeffreys, A. J. and R. A. Flavell** (1977). The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell* **12**(4): 1097-108.
- Johnson, B. A. and R. A. Blevins** (1994). NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* **4**: 603-614.
- Kambach, C., S. Walke, R. Young, J. M. Avis, E. de la Fortelle, V. A. Raker, R. Luhrmann, J. Li and K. Nagai** (1999). Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**(3): 375-87.
- Kamiya, T., S. Maeno, M. Hashimoto and Y. Mine** (1972). Bicyclomycin, a new antibiotic. II. Structural elucidation and acyl derivatives. *J Antibiot (Tokyo)* **25**(10): 576-81.
- Karimi, R., M. Y. Pavlov, R. H. Buckingham and M. Ehrenberg** (1999). Novel roles for classical factors at the interface between translation termination and initiation. *Mol Cell* **3**(5): 601-9.
- Kassavetis, G. A. and M. J. Chamberlin** (1981). Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *J Biol Chem* **256**(6): 2777-86.
- Ken, R. and N. R. Hackett** (1991). Halobacterium halobium strains lysogenic for phage phi H contain a protein resembling coliphage repressors. *J Bacteriol* **173**(3): 955-60.

- Kim, D. E. and S. S. Patel** (2001). The kinetic pathway of RNA binding to the *Escherichia coli* transcription termination factor Rho. *J Biol Chem* **276**(17): 13902-10.
- Kimball, S. R.** (1999). Eukaryotic initiation factor eIF2. *Int J Biochem Cell Biol* **31**(1): 25-9.
- Kolakofsky, D., T. Pelet, D. Garcin, S. Hausmann, J. Curran and L. Roux** (1998). Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. *J Virol* **72**(2): 891-9.
- Kolter, R. and C. Yanofsky** (1982). Attenuation in amino acid biosynthetic operons. *Annu Rev Genet* **16**: 113-34.
- Koradi, R., M. Billeter and K. Wuthrich** (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**(1): 51-5, 29-32.
- Kozak, M.** (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**(20): 8125-48.
- Kozak, M.** (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**(2): 187-208.
- Kruger, K., T. Hermann, V. Armbruster and F. Pfeifer** (1998). The transcriptional activator GvpE for the halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. *J Mol Biol* **279**(4): 761-71.
- Kuboniwa, H., S. Grzesiek, F. Delaglio and A. Bax** (1994). Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J Biomol NMR* **4**(6): 871-8.
- Landick, R. and C. Yanofsky** (1987). Transcription attenuation. *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. F. C. Neidhardt. Washington, DC, American Society for Microbiology. **2**: 1276-301.

- Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton** (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**: 283-291.
- Lau, L. F., J. W. Roberts and R. Wu** (1982). Transcription terminates at lambda tR1 in three clusters. *Proc Natl Acad Sci U S A* **79**(20): 6171-5.
- Lau, L. F., J. W. Roberts, R. Wu, F. Georges and S. A. Narang** (1984). A potential stem-loop structure and the sequence CAAUCAA in the transcript are insufficient to signal rho-dependent transcription termination at lambda tR1. *Nucleic Acids Res* **12**(2): 1287-99.
- Laurino, J. P., G. M. Thompson, E. Pacheco and B. A. Castilho** (1999). The beta subunit of eukaryotic translation initiation factor 2 binds mRNA through the lysine repeats and a region comprising the C<sub>2</sub>-C<sub>2</sub> motif. *Mol Cell Biol* **19**(1): 173-81.
- Lee, L. K., M. Rance, W. J. Chazin and A. G. Palmer, 3rd** (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13C alpha nuclear spin relaxation. *J Biomol NMR* **9**(3): 287-98.
- Leeds, P., J. M. Wood, B. S. Lee and M. R. Culbertson** (1992). Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol Cell Biol* **12**(5): 2165-77.
- Leff, S. E., R. M. Evans and M. G. Rosenfeld** (1987). Splice commitment dictates neuron-specific alternative RNA processing in *calcitonin/CGRP* gene expression. *Cell* **48**(3): 517-24.
- Leigh, J. A.** (1999). Transcriptional regulation in Archaea. *Curr Opin Microbiol* **2**(2): 131-4.
- Levin, J. R. and M. J. Chamberlin** (1987). Mapping and characterization of transcriptional pause sites in the early genetic region of bacteriophage T7. *J Mol Biol* **196**(1): 61-84.
- Li, J., R. Horwitz, S. McCracken and J. Greenblatt** (1992). NusG, a new *Escherichia coli* elongation factor involved in transcriptional

- antitermination by the N protein of phage lambda. *J Biol Chem* **267**(9): 6012-9.
- Li, J., S. W. Mason and J. Greenblatt** (1993). Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev* **7**(1): 161-72.
- Li, S. C., C. L. Squires and C. Squires** (1984). Antitermination of *E. coli* rRNA transcription is caused by a control region segment containing lambda nut-like sequences. *Cell* **38**(3): 851-60.
- Li, X. Y. and M. R. Green** (1998). The HIV-1 Tat cellular coactivator Tat-SF1 is a general transcription elongation factor. *Genes Dev* **12**(19): 2992-6.
- Linderoth, N. A. and R. L. Calendar** (1991). The Psi protein of bacteriophage P4 is an antitermination factor for rho-dependent transcription termination. *J Bacteriol* **173**(21): 6722-31.
- Linderoth, N. A., G. Tang and R. Calendar** (1997). *In vivo* and *in vitro* evidence for an anti-Rho activity induced by the phage P4 polarity suppressor protein Psi. *Virology* **227**(1): 131-41.
- Liu, M. Y. and T. Romeo** (1997). The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J Bacteriol* **179**(14): 4639-42.
- Liu, M. Y., H. Yang and T. Romeo** (1995). The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* **177**(10): 2663-72.
- Lopez, A. J.** (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32**: 279-305.
- Maeda, H., N. Fujita and A. Ishihama** (2000). Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res* **28**(18): 3497-503.

- Magyar, A., X. Zhang, H. Kohn and W. R. Widger** (1996). The antibiotic bicyclomycin affects the secondary RNA binding site of *Escherichia coli* transcription termination factor Rho. *J Biol Chem* **271**(41): 25369-74.
- Malliavin, T. E., J. L. Pons and M. A. Delsuc** (1998). An NMR assignment module implemented in the Gifa NMR processing program. *Bioinformatics* **14**(7): 624-31.
- Malmgren, C., E. G. Wagner, C. Ehresmann, B. Ehresmann and P. Romby** (1997). Antisense RNA control of plasmid R1 replication. The dominant product of the antisense RNA-mRNA binding is not a full RNA duplex. *J Biol Chem* **272**(19): 12508-12.
- Maret, W. and B. L. Vallee** (1993). Cobalt as probe and label of proteins. *Methods Enzymol* **226**: 52-71.
- Markus, M. A., A. P. Hinck, S. Huang, D. E. Draper and D. A. Torchia** (1997). High resolution solution structure of ribosomal protein L11-C76, a helical protein with a flexible loop that becomes structured upon binding to RNA. *Nat Struct Biol* **4**(1): 70-7.
- Martin, F. H. and I. Tinoco, Jr.** (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Res* **8**(10): 2295-9.
- Masse, E. and S. Gottesman** (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* **99**(7): 4620-5.
- Masse, E., N. Majdalani and S. Gottesman** (2003). Regulatory roles for small RNAs in bacteria. *Curr Opin Microbiol* **6**(2): 120-4.
- Matte, A., J. Sivaraman, I. Ekiel, K. Gehring, Z. Jia and M. Cygler** (2003). Contribution of structural genomics to understanding the biology of *Escherichia coli*. *J Bacteriol* **185**(14): 3994-4002.
- McKnight, S. L. and R. Kingsbury** (1982). Transcriptional control signals of a eukaryotic protein-coding gene. *Science* **217**(4557): 316-24.

- Merrick, W. C. and J. Nyborg** (2000). The protein biosynthesis and elongation cycle. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 89-125.
- Meyuhas, O. and E. Hornstein** (2000). Translational control of TOP mRNAs. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 671-694.
- Miller, J. H.** (1972). *Experiments in molecular genetics*. [Cold Spring Harbor, N.Y.], Cold Spring Harbor Laboratory.
- Miron, M., J. Verdu, P. E. Lachance, M. J. Birnbaum, P. F. Lasko and N. Sonenberg** (2001). The translational inhibitor 4E-BP is an effector of PI(3)K/Akt signalling and cell growth in *Drosophila*. *Nat Cell Biol* 3(6): 596-601.
- Miyamura, S., N. Ogasawara, H. Otsuka, S. Niwayama and H. Tanaka** (1972). Antibiotic no. 5879, a new water-soluble antibiotic against gram-negative bacteria. *J Antibiot (Tokyo)* 25(10): 610-2.
- Mo, Y., B. Vaessen, K. Johnston and R. Marmorstein** (1998). Structures of SAP-1 bound to DNA targets from the *E74* and *c-fos* promoters: insights into DNA sequence discrimination by Ets proteins. *Mol Cell* 2(2): 201-12.
- Mo, Y., B. Vaessen, K. Johnston and R. Marmorstein** (2000). Structure of the elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. *Nat Struct Biol* 7(4): 292-7.
- Moazed, D. and H. F. Noller** (1989). Intermediate states in the movement of transfer RNA in the ribosome. *Nature* 342(6246): 142-8.
- Moller, A., U. Wild, D. Riesner and H. G. Gassen** (1979). Evidence from ultraviolet absorbance measurements for a codon-induced conformational change in lysine tRNA from *Escherichia coli*. *Proc Natl Acad Sci U S A* 76(7): 3266-70.

- Moller, T., T. Franch, C. Udesen, K. Gerdes and P. Valentin-Hansen (2002).** Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev* 16(13): 1696-706.
- Morgan, W. D., D. G. Bear, B. L. Litchman and P. H. von Hippel (1985).** RNA sequence and secondary structure requirements for rho-dependent transcription termination. *Nucleic Acids Res* 13(10): 3739-54.
- Muller-Hill, B. (1998).** Some repressors of bacterial transcription. *Curr Opin Microbiol* 1(2): 145-51.
- Munroe, D. and A. Jacobson (1990).** Tales of poly(A): a review. *Gene* 91(2): 151-8.
- Murakami, K. S., S. Masuda, E. A. Campbell, O. Muzzin and S. A. Darst (2002).** Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296(5571): 1285-90.
- Musco, G., A. Kharrat, G. Stier, F. Fraternali, T. J. Gibson, M. Nilges and A. Pastore (1997).** The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat Struct Biol* 4(9): 712-6.
- Musco, G., G. Stier, C. Joseph, M. A. Castiglione Morelli, M. Nilges, T. J. Gibson and A. Pastore (1996).** Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* 85(2): 237-45.
- Nagalla, S. R., B. J. Barry and E. R. Spindel (1994).** Cloning of complementary DNAs encoding the amphibian bombesin-like peptides Phe8 and Leu8 phyllolitorin from *Phyllomedusa sauvagei*: potential role of U to C RNA editing in generating neuropeptide diversity. *Mol Endocrinol* 8(8): 943-51.
- Nehrke, K. W. and T. Platt (1994).** A quaternary transcription termination complex. Reciprocal stabilization by Rho factor and NusG protein. *J Mol Biol* 243(5): 830-9.

- Neuhaus, D., Y. Nakaseko, J. W. Schwabe and A. Klug** (1992). Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional  $^1\text{H}$  nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet. *J Mol Biol* **228**(2): 637-51.
- Newman, A.** (1998). RNA splicing. *Curr Biol* **8**(25): R903-5.
- Nilges, M., M. J. Macias, S. I. O'Donoghue and H. Oschkinat** (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. *J Mol Biol* **269**(3): 408-22.
- Nodwell, J. R. and J. Greenblatt** (1991). The *nut* site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. *Genes Dev* **5**(11): 2141-51.
- Omichinski, J. G., G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S. J. Stahl and A. M. Gronenborn** (1993). NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261**(5120): 438-46.
- Opperman, T. and J. P. Richardson** (1994). Phylogenetic analysis of sequences from diverse bacteria with homology to the *Escherichia coli rho* gene. *J Bacteriol* **176**(16): 5033-43.
- Osborne, M. J. and P. E. Wright** (2001). Anisotropic rotational diffusion in model-free analysis for a ternary DHFR complex. *J Biomol NMR* **19**(3): 209-30.
- Ottiger, M., F. Delaglio and A. Bax** (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* **131**(2): 373-8.
- Pain, V. M.** (1996). Initiation of protein synthesis in eukaryotic cells. *Eur J Biochem* **236**(3): 747-71.
- Park, N. J., D. C. Tsao and H. G. Martinson** (2004). The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by

- truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Mol Cell Biol* **24**(10): 4092-103.
- Paule, M. R. and R. J. White** (2000). SURVEY AND SUMMARY Transcription by RNA polymerases I and III. *Nucl. Acids Res.* **28**(6): 1283-1298.
- Perez-Rueda, E. and J. Collado-Vides** (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res* **28**(8): 1838-47.
- Pestova, T. V. and C. U. Hellen** (2000). The structure and function of initiation factors in eukaryotic protein synthesis. *Cell Mol Life Sci* **57**(4): 651-74.
- Pichoff, S., L. Alibaud, A. Guedant, M. P. Castanie and J. P. Bouche** (1998). An *Escherichia coli* gene (*yaeO*) suppresses temperature-sensitive mutations in essential genes by modulating Rho-dependent transcription termination. *Mol Microbiol* **29**(3): 859-69.
- Piecyk, M., S. Wax, A. R. Beck, N. Kedersha, M. Gupta, B. Maritim, S. Chen, C. Gueydan, V. Krusys, M. Streuli and P. Anderson** (2000). TIA-1 is a translational silencer that selectively regulates the expression of TNF-alpha. *Embo J* **19**(15): 4154-63.
- Pio, F., R. Kodandapani, C. Z. Ni, W. Shepard, M. Klemsz, S. R. McKercher, R. A. Maki and K. R. Ely** (1996). New insights on DNA recognition by ets proteins from the crystal structure of the PU.1 ETS domain-DNA complex. *J Biol Chem* **271**(38): 23329-37.
- Platt, T.** (1994). Rho and RNA: models for recognition and response. *Mol Microbiol* **11**(6): 983-90.
- Proud, C.** (2000). Control of the elongation phase of protein synthesis. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 719-39.

- Proudfoot, N. J.** (1989). How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**(3): 105-10.
- Qian, X., S. N. Gozani, H. Yoon, C. J. Jeon, K. Agarwal and M. A. Weiss** (1993). Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* **32**(38): 9944-59.
- Raught, B. and A. C. Gingras** (1999). eIF4E activity is regulated at multiple levels. *Int J Biochem Cell Biol* **31**(1): 43-57.
- Raught, B., A. C. Gingras, S. P. Gygi, H. Imataka, S. Morino, A. Gradi, R. Aebersold and N. Sonenberg** (2000). Serum-stimulated, rapamycin-sensitive phosphorylation sites in the eukaryotic translation initiation factor 4GI. *Embo J* **19**(3): 434-44.
- Reeder, R. H. and W. H. Lang** (1997). Terminating transcription in eukaryotes: lessons learned from RNA polymerase I. *Trends Biochem Sci* **22**(12): 473-7.
- Reeve, J. N., K. Sandman and C. J. Daniels** (1997). Archaeal histones, nucleosomes, and transcription initiation. *Cell* **89**(7): 999-1002.
- Richardson, J. P.** (1990). Rho-dependent transcription termination. *Biochim Biophys Acta* **1048**(2-3): 127-38.
- Richardson, J. P.** (1996). Structural organization of transcription termination factor Rho. *J Biol Chem* **271**(3): 1251-4.
- Richardson, J. P. and J. Greenblatt** (1996). Control of RNA chain elongation and termination. *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. F. C. Neidhardt. Washington, DC, American Society for Microbiology Press: 822-848.
- Richardson, J. P., C. Grimley and C. Lowery** (1975). Transcription termination factor Rho activity is altered in *Escherichia coli* with *suA* gene mutations. *Proc Natl Acad Sci U S A* **72**(5): 1725-8.

- Roberts, J. W.** (1993). RNA and protein elements of *E. coli* and lambda transcription antitermination complexes. *Cell* **72**(5): 653-5.
- Robertson, J. M., M. Kahan, W. Wintermeyer and H. G. Zachau** (1977). Interactions of yeast tRNA<sup>Phe</sup> with ribosomes from yeast and *Escherichia coli*. A fluorescence spectroscopic study. *Eur J Biochem* **72**(1): 117-25.
- Rodnina, M. V., A. Savelsbergh, V. I. Katunin and W. Wintermeyer** (1997). Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* **385**(6611): 37-41.
- Roeder, R. G.** (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett* **579**(4): 909-15.
- Romeo, T.** (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol Microbiol* **29**(6): 1321-30.
- Romeo, T., M. Gong, M. Y. Liu and A. M. Brun-Zinkernagel** (1993). Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* **175**(15): 4744-55.
- Rosenberg, M. and D. Court** (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu Rev Genet* **13**: 319-53.
- Ross, W., A. Ernst and R. L. Gourse** (2001). Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**(5): 491-506.
- Rougvie, A. E. and J. T. Lis** (1988). The RNA polymerase II molecule at the 5' end of the uninduced *hsp70* gene of *D. melanogaster* is transcriptionally engaged. *Cell* **54**(6): 795-804.
- Rückert, M. and G. Otting** (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* **122**: 7793-97.

- Rueter, S. M., T. R. Dawson and R. B. Emeson** (1999). Regulation of alternative splicing by RNA editing. *Nature* **399**(6731): 75-80.
- Rueter, S. M. and R. B. Emeson** (1998). Adenosine-to-Inosine conversion in mRNA. *Modification and Editing of RNA*. H. Grosjean and R. Benne. Washington, DC, ASM Press: 343-61.
- Ryan, T. and M. J. Chamberlin** (1983). Transcription analyses with heteroduplex trp attenuator templates indicate that the transcript stem and loop structure serves as the termination signal. *J Biol Chem* **258**(8): 4690-3.
- Sabnis, N. A., H. Yang and T. Romeo** (1995). Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. *J Biol Chem* **270**(49): 29096-104.
- Sachs, A. B. and G. Varani** (2000). Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat Struct Biol* **7**(5): 356-61.
- Schlax, P. J., M. W. Capp and M. T. Record, Jr.** (1995). Inhibition of transcription initiation by *lac* repressor. *J Mol Biol* **245**(4): 331-50.
- Schmitt, E., S. Blanquet and Y. Mechulam** (2002). The large subunit of initiation factor aIF2 is a close structural homologue of elongation factors. *Embo J* **21**(7): 1821-32.
- Schumacher, M. A., R. F. Pearson, T. Moller, P. Valentin-Hansen and R. G. Brennan** (2002). Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *Embo J* **21**(13): 3546-56.
- Schwarz, U., H. M. Menzel and H. G. Gassen** (1976). Codon-dependent rearrangement of the three-dimensional structure of phenylalanine tRNA, exposing the T-psi-C-G sequence for binding to the 50S ribosomal subunit. *Biochemistry* **15**(11): 2484-90.

- Selby, C. P. and A. Sancar** (1997). Cockayne syndrome group B protein enhances elongation by RNA polymerase II. *Proc Natl Acad Sci U S A* **94**(21): 11205-9.
- Sha, Y., L. Lindahl and J. M. Zengel** (1995). Role of NusA in L4-mediated attenuation control of the S10 r-protein operon of *Escherichia coli*. *J Mol Biol* **245**(5): 474-85.
- Sharma, P. M., M. Bowman, S. L. Madden, F. J. Rauscher, 3rd and S. Sukumar** (1994). RNA editing in the Wilms' tumor susceptibility gene, *WT1*. *Genes Dev* **8**(6): 720-31.
- Sharrock, R. A., R. L. Gourse and M. Nomura** (1985). Defective antitermination of rRNA transcription and derepression of rRNA and tRNA synthesis in the nusB5 mutant of *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**(16): 5275-9.
- Shatkin, A. J.** (1976). Capping of eucaryotic mRNAs. *Cell* **9**(4 PT 2): 645-53.
- Shilatifard, A.** (1998). Factors regulating the transcriptional elongation activity of RNA polymerase II. *Faseb J* **12**(14): 1437-46.
- Shilatifard, A.** (1998). Identification and purification of the Holo-ELL complex. Evidence for the presence of ELL-associated proteins that suppress the transcriptional inhibitory activity of ELL. *J Biol Chem* **273**(18): 11212-7.
- Skordalakes, E. and J. M. Berger** (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* **114**(1): 135-46.
- Skordalakes, E., A. P. Brogan, B. S. Park, H. Kohn and J. M. Berger** (2005). Structural mechanism of inhibition of the Rho transcription termination factor by the antibiotic bicyclomycin. *Structure (Camb)* **13**(1): 99-109.
- Skuse, G. R., A. J. Cappione, M. Sowden, L. J. Metheny and H. C. Smith** (1996). The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res* **24**(3): 478-85.

- Sonenberg, N., M. A. Morgan, W. C. Merrick and A. J. Shatkin (1978).** A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci U S A* **75**(10): 4843-7.
- Soppa, J. (1999).** Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31**(5): 1295-305.
- Sprinzl, M., C. Horn, M. Brown, A. Ioudovitch and S. Steinberg (1998).** Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res* **26**(1): 148-53.
- Staudt, L. M. and M. J. Lenardo (1991).** Immunoglobulin gene transcription. *Annu Rev Immunol* **9**: 373-98.
- Steitz, J. A. and K. Jakes (1975).** How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc Natl Acad Sci U S A* **72**(12): 4734-8.
- Steneberg, P., C. Englund, J. Kronhamn, T. A. Weaver and C. Samakovlis (1998).** Translational readthrough in the *hdc* mRNA generates a novel branching inhibitor in the drosophila trachea. *Genes Dev* **12**(7): 956-67.
- Stewart, V. and C. Yanofsky (1985).** Evidence for transcription antitermination control of tryptophanase operon expression in *Escherichia coli* K-12. *J Bacteriol* **164**(2): 731-40.
- Stoecklin, G., S. Hahn and C. Moroni (1994).** Functional hierarchy of AUUUA motifs in mediating rapid interleukin-3 mRNA decay. *J Biol Chem* **269**(46): 28591-7.
- Stolt, P. and W. Zillig (1992).** *In vivo* studies on the effects of immunity genes on early lytic transcription in the *Halobacterium salinarium* phage phi H. *Mol Gen Genet* **235**(2-3): 197-204.
- Sullivan, S. L. and M. E. Gottesman (1992).** Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell* **68**(5): 989-94.

- Sune, C. and M. A. Garcia-Blanco** (1999). Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol Cell Biol* **19**(7): 4719-28.
- Thompson, G. M., E. Pacheco, E. O. Melo and B. A. Castilho** (2000). Conserved sequences in the beta subunit of archaeal and eukaryal translation initiation factor 2 (eIF2), absent from eIF5, mediate interaction with eIF2gamma. *Biochem J* **347 Pt 3**: 703-9.
- Tilghman, S. M., P. J. Curtis, D. C. Tiemeier, P. Leder and C. Weissmann** (1978). The intervening sequence of a mouse beta-globin gene is transcribed within the 15S beta-globin mRNA precursor. *Proc Natl Acad Sci USA* **75**(3): 1309-13.
- Uptain, S. M., C. M. Kane and M. J. Chamberlin** (1997). Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* **66**: 117-72.
- Varani, G.** (1997). A cap for all occasions. *Structure* **5**(7): 855-8.
- Venema, R. C., H. I. Peters and J. A. Traugh** (1991). Phosphorylation of elongation factor 1 (EF-1) and valyl-tRNA synthetase by protein kinase C and stimulation of EF-1 activity. *J Biol Chem* **266**(19): 12574-80.
- Venema, R. C., H. I. Peters and J. A. Traugh** (1991). Phosphorylation of valyl-tRNA synthetase and elongation factor 1 in response to phorbol esters is associated with stimulation of both activities. *J Biol Chem* **266**(18): 11993-8.
- Wall, M. E., W. S. Hlavacek and M. A. Savageau** (2004). Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5**(1): 34-42.
- Warren, A. J.** (2002). Eukaryotic transcription factors. *Curr Opin Struct Biol* **12**(1): 107-14.
- Wassarman, K. M. and G. Storz** (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* **101**(6): 613-23.

- Wassarman, K. M., A. Zhang and G. Storz** (1999). Small RNAs in *Escherichia coli*. *Trends Microbiol* **7**(1): 37-45.
- Wei, B., S. Shin, D. LaPorte, A. J. Wolfe and T. Romeo** (2000). Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate. *J Bacteriol* **182**(6): 1632-40.
- Wei, B. L., A. M. Brun-Zinkernagel, J. W. Simecka, B. M. Pruss, P. Babitzke and T. Romeo** (2001). Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol Microbiol* **40**(1): 245-56.
- Weilbacher, T., K. Suzuki, A. K. Dubey, X. Wang, S. Gudapaty, I. Morozov, C. S. Baker, D. Georgellis, P. Babitzke and T. Romeo** (2003). A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* **48**(3): 657-70.
- Welch, E. M., W. Wang and S. W. Peltz** (2000). Translation termination: it's not the end of the story. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 467-485.
- Wickens, M.** (1990). How the messenger got its tail: addition of poly(A) in the nucleus. *Trends Biochem Sci* **15**(7): 277-81.
- Wickens, M., E. B. Goodwin, J. Kimble, S. Strickland and M. Hentze** (2000). Translational control of developmental decisions. *Translational control of gene expression*. N. Sonenberg, J. W. B. Hershey and M. B. Matthews. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 295-370.
- Winston, F., D. T. Chaleff, B. Valent and G. R. Fink** (1984). Mutations affecting Ty-mediated expression of the *HIS4* gene of *Saccharomyces cerevisiae*. *Genetics* **107**(2): 179-97.
- Wishart, D. S. and B. D. Sykes** (1994). The <sup>13</sup>C chemical-shift index: a simple method for the identification of protein secondary structure using <sup>13</sup>C chemical-shift data. *J Biomol NMR* **4**(2): 171-80.

- Wittschieben, B. O., J. Fellows, W. Du, D. J. Stillman and J. Q. Svejstrup** (2000). Overlapping roles for the histone acetyltransferase activities of SAGA and elongator *in vivo*. *Embo J* **19**(12): 3060-8.
- Wojciak, J. M., J. Iwahara and R. T. Clubb** (2001). The Mu repressor-DNA complex contains an immobilized 'wing' within the minor groove. *Nat Struct Biol* **8**(1): 84-90.
- Wüthrich, K.** (1986). *NMR of proteins and nucleic acids*. New York, Wiley-Interscience.
- Xu, N., C. Y. Chen and A. B. Shyu** (1997). Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay. *Mol Cell Biol* **17**(8): 4611-21.
- Yang, M. T. and J. F. Gardner** (1989). Transcription termination directed by heteroduplex thr attenuator templates. Evidence that the transcript stem and loop structure is the termination signal. *J Biol Chem* **264**(5): 2634-9.
- Yang, X. J., J. A. Goliger and J. W. Roberts** (1989). Specificity and mechanism of antitermination by Q proteins of bacteriophages lambda and 82. *J Mol Biol* **210**(3): 453-60.
- Yanofsky, C.** (2001). Advancing our knowledge in biochemistry, genetics, and microbiology through studies on tryptophan metabolism. *Annu Rev Biochem* **70**: 1-37.
- Yarnell, W. S. and J. W. Roberts** (1992). The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. *Cell* **69**(7): 1181-9.
- Yee, A., X. Chang, A. Pineda-Lucena, B. Wu, A. Semesi, B. Le, T. Ramelot, G. M. Lee, S. Bhattacharyya, P. Gutierrez, A. Denisov, C. H. Lee, J. R. Cort, G. Kozlov, J. Liao, G. Finak, L. Chen, D. Wishart, W. Lee, L. P. McIntosh, K. Gehring, M. A. Kennedy, A. M. Edwards and C. H.**

- Arrowsmith** (2002). An NMR approach to structural proteomics. *Proc Natl Acad Sci U S A* **99**(4): 1825-30.
- Zalatan, F., J. Galloway-Salvo and T. Platt** (1993). Deletion analysis of the *Escherichia coli* rho-dependent transcription terminator trp t'. *J Biol Chem* **268**(23): 17051-6.
- Zawel, L. and D. Reinberg** (1995). Common themes in assembly and function of eukaryotic transcription complexes. *Annu Rev Biochem* **64**: 533-61.
- Zengel, J. M. and L. Lindahl** (1990). Ribosomal protein L4 stimulates *in vitro* termination of transcription at a NusA-dependent terminator in the *S10* operon leader. *Proc Natl Acad Sci U S A* **87**(7): 2675-9.
- Zhang, A., S. Altuvia, A. Tiwari, L. Argaman, R. Hengge-Aronis and G. Storz** (1998). The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. *Embo J* **17**(20): 6061-8.
- Zhouravleva, G., L. Frolova, X. Le Goff, R. Le Guellec, S. Inge-Vechtomov, L. Kisselev and M. Philippe** (1995). Termination of translation in eukaryotes is governed by two interacting polypeptide chain release factors, eRF1 and eRF3. *Embo J* **14**(16): 4065-72.
- Zhu, W., Q. Zeng, C. M. Colangelo, M. Lewis, M. F. Summers and R. A. Scott** (1996). The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat Struct Biol* **3**(2): 122-4.