EXPRESSION AND CHARACTERISATION OF A GENE ENCODING RbpD, AN RNA-BINDING PROTEIN IN Anabaena SP. STRAIN PCC 7120

ROBIN LEE TREMBLAY

Expression and Characterisation of a Gene Encoding RbpD, an RNA-Binding Protein in *Anabaena* sp. strain PCC 7120

by

Robin Lee Tremblay

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Department of Biochemistry/Faculty of Science
Memorial University of Newfoundland

January 2000

St. John's Newfoundland

Abstract

The RNA-binding protein RbpD, from the cyanobacterium *Anabaena sp.* strain PCC 7120 was expressed in *Escherichia coli* and successfully purified using the IMPACT I system (New England Biolabs). The *rbpD* gene was cloned into the pCYB1 expression vector by using polymerase chain reaction to introduce NdeI and SapI restriction sites at the 5' end 3' ends of the gene respectively. The 3'-end mutagenesis also changed the stop codon into a cysteine codon. The resulting gene encoded a fusion protein consisting of RbpD, the *Saccharomyces cerevisiae* VMA intein and a chitin binding domain. Expression of the fusion protein was observed in *E. coli* strain MC1061 but Western blot analysis using an intein-directed antibody indicated that significant *in vivo* intein-directed splicing of the fusion protein occurred. We were unable to eliminate this problem; no fusion protein expression was observed in 8 other *E. coli* strains tested. Wild-type RbpD was purified following binding of the fusion protein to a chitin column and overnight cleavage in the presence of a reducing agent, dithiothreitol. A number of modifications to the manufacturer's purification protocol were found to be necessary for successful purification. The NaCl concentration in the cleavage and elution buffer was increased from 50 mM to 500 mM to eliminate problems of RbpD solubility. An increase in the dithiothreitol concentration of the cleavage buffer from 30 mM to 50 mM was required for full cleavage.

A modified form of RbpD containing an hexa-histidine tag in loop 3 of the RNA recognition motif, RbpD1, was also successfully purified. The *rbpD1* gene, previously

constructed by Cynthia Slade, was cloned into the pTRC99A expression vector. An NcoI site was introduced at the 5'-end of the gene using site-directed mutagenesis. This modification also changed the second codon of the gene from serine to alanine. The RbpD1 protein was expressed in *E. coli* strain BL21(DE3)pLysS following induction with IPTG and purified using a nickel-NTA agarose affinity column. The protein was eluted with 100 mM imidazole and appeared to be pure upon analysis using polyacrylamide gel electrophoresis.

The RNA-binding activity of RbpD and RbpD1 were first determined using Sepharose-4B-, polyacrylhydrazido-agarose-, or agarose-bound RNA homopolymers. Both proteins bound strongly to poly(U), less strongly to poly(G), weakly to poly(A), and not at all to poly(C). This pattern is consistent with that observed for other cyanobacterial RNA-binding proteins. There was no apparent difference in the binding affinities of RbpD and RbpD1 indicating that the presence of the 6x-histidine tag had no effect. Experiments to detect binding between RbpD and a conserved sequence element in the 5'- untranslated region of *rbpD* using both electrophoretic mobility shift assays and nitrocellulose filter binding were unsuccessful. Similarly, attempts to detect binding between RbpD and size-fractionated radioactively labelled poly(U) by electrophoretic mobility shift assays were unsuccessful. Two SELEX experiments were also unsuccessful. In both cases, no increase in specific binding over background was detected through four rounds of selection.

This thesis is dedicated to my parents, Kerry and Lee Tremblay
who have always supported me in anything I have put my mind to.

I would like to thank Dr. Martin Mulligan, my supervisor on this project, for all the help he has provided, and also for the emotional support. I'd also like to thank Tom Belbin for question answering, Kerry Tremblay for editing the writing in this thesis (and learning lots of biochemistry while doing so!), Lee Tremblay for computer support, and Dr. John Brosnan who has gone above and beyond the call of duty in his role as Head of the Department. I also thank Dr. Margaret Brosnan and Dr. David Heeley who were on my advisory committee.

I would like to thank my fiancé Jason Churchill for being there through all the stress.

Table of Contents

Section			Page
Abstract			ii
Acknowledgements			iv
Table of Contents			v
List of Tables			viii
List of Figures			ix
List of Abbreviations and Symbols			xii
CHAPTER 1: INTRODUCTION			
1.1	Cyanobacteria		2
1.2	*Anabaena* sp. strain PCC 7120		4
	1.2.1	General	4
	1.2.2	Heterocysts	4
	1.2.3	Mechanism of Heterocyst Formation	7
1.3.	RNA-Binding Proteins		10
	1.3.1	RNP-Type RNA-Binding Proteins	10
	1.3.2	RNA-Binding Proteins in Cyanobacteria	13
	1.3.3	RbpD in *Anabaena* 7120	15
	1.3.4	U1A	18
	1.3.5	Heterogeneous Nuclear Ribonucleoproteins	21
	1.3.6	Glycine Loops	24
	1.3.7	Evolutionary Trends in RNA-Binding Proteins	26
1.4	Cold-Shock Proteins		28
1.5	Inteins		30
	1.5.1	General	30
	1.5.2	Rationale for the Placement of Inteins	31
	1.5.3	Mechanism of Intein Excision	32
	1.5.4	Inteins in Cyanobacteria	33
	1.5.5	Inteins for Protein Expression Systems	36

Section		Page
1.6	SELEX	37
1.7	Aims	43
CHAPTER 2: MATERIALS AND METHODS		
2.1	Materials	45
2.2	Culturing and Cloning	45
	2.2.1 Media	45
	2.2.2 Competent Cells	50
	2.2.3 Transformations	50
	2.2.4 Isolation of Plasmid DNA	51
	2.2.5 Cloning of DNA	52
2.3	Sequencing of DNA	53
2.4	Detection of Proteins	55
2.5	Expression of RbpD	56
	2.5.1 Cloning of *rbpD* Into pCYB1	56
	2.5.2 Expression and Purification of RbpD	57
	2.5.3 Western Blotting of Fusion Protein	60
2.6	Expression of RbpD1	61
	2.6.1 Site Directed Mutagenesis of *rbpD1a*	61
	2.6.2 Expression of RbpD1	65
2.7	Storage of proteins	66
2.8	Binding Experiments	66
	2.8.1 Polymer Binding	66
	2.8.2 Binding of 5'-Untranslated Region – DNA	67
	2.8.3 Binding of 5'-Untranslated Region – RNA	68
	2.8.4 Binding of Poly(U)	70
2.9	SELEX	71
CHAPTER 3: RESULTS AND DISCUSSION – PROTEIN EXPRESSION		
3.1	Expression of RbpD	76
	3.1.1 Cloning of *rbpD* Into pCYB1	76

Section **Page**

3.1.2 Expression of the Fusion Protein in *E. coli* MC1061 79
3.1.3 Expression of the Fusion Protein in *E. coli* BL21(DE3)pLysS 84
3.1.4 Western Blotting 92
3.1.5 Expression of RbpD in Other Strains of *E. coli* 96
3.1.6 Purification of RbpD 99

3.2 Expression and Purification of RbpD1 102
3.2.1 Background 102
3.2.2 Site-Directed Mutagenesis 110
3.2.3 Affects of the NcoI Site in *rbpD1* 113
3.2.4 Expression and Purification of RbpD1 114

CHAPTER 4: RESULTS AND DISCUSSION – CHARACTERISATION OF RbpD

4.1 Binding to Agarose-Bound Polymers 123

4.2 Binding to 5' – Untranslated Region of *rbpD* – DNA 127

4.3 Binding to 5' – Untranslated Region of *rbpD* – RNA 131

4.4 Binding of RbpD to Poly(U) 139

4.5 SELEX 140

CHAPTER 5: GENERAL DISCUSION

5.1 General Discussion 151

5.2 Future Work 156

References 158

Appendix 1: Cloning and Sequencing of *nifH from *Chlorogloeopsis sp.* PCC 6192** 174

List of Tables

Table		Page
2.1	Plasmids used in this work	46
2.2	Bacterial Strains used in this work	48
2.3	Oligonucleotides synthesised in this work	54

List of Figures

Figure		Page
Chapter 1		
1.1	Photo of *Anabaena* 7120	6
1.2	U1A RNA-Recognition motif bound to an RNA hairpin	12
1.3	Nucleotide sequence of the *rbpD* gene from *Anabaena* 7120 and its inferred amino acid sequence	17
1.4	Schematic diagram of the mechanism of intein splicing	35
1.5	Schematic diagram illustrating the theory of SELEX	40
Chapter 2		
2.1	Protocol for expression and purification of RbpD	59
2.2	Schematic diagram of the Altered Sites II mutagenesis procedure	63
Chapter 3		
3.1	Schematic drawing of the splicing mechanism of the IMPACT I system	78
3.2	Cloning *rbpD* in pCYB1	81
3.3	pRLT1 digested with NdeI and KpnI	83
3.4	Induction of RbpD fusion protein expression in *E. coli* MC1061	86
3.5	Expression of the RbpD fusion protein from pRLT1 in *E. coli* strain MC1061	88
3.6	Expression of the RbpD fusion protein in *E. coli* BL21(DE3)pLysS	91
3.7	Western blot analysis of the RbpD fusion protein in *E. coli* strain MC1061	94

Figure		Page
3.8	Expression of the RbpD fusion protein from pRLT1 in eight strains of *E. coli*	98
3.9	RbpD elution with 0.1% Triton-X100 in Cleavage Buffer	101
3.10	Expression and elution of RbpD	104
3.11	Illustration of RbpD showing the sites of attempted 6x His-tag insertion	107
3.12	Picture of U1A RRM1 bound to an RNA ligand	109
3.13	Mutagenesis of *rbpD1a*	112
3.14	pRLT3 digested with NcoI and HindIII	116
3.15	RbpD1 expression under induction of 1 mM and 2 mM IPTG	118
3.16	Expression and purification of RbpD1	120
Chapter 4:		
4.1	Binding of RNA-binding proteins to RNA homopolymers	125
4.2	Conserved portion of the 5' – untranslated region from *rbpD*	129
4.3	Binding of RbpD to the 5' – untranslated region in the DNA form	133
4.4	Transcription of the 5' – untranslated region of *rbpD*	135
4.5	Binding of RbpD to the 5' – untranslated region in the RNA form	138
4.6	Poly(U) binding experiment	142
4.7	SELEX – Experiment #1	146
4.8	SELEX – Experiment #2	149
Appendix 1:		
A.1	Sequence of the *Chlorogloeopsis* 6912 *nifH** gene	179

Figure		Page
A.2	Comparison between NifH* sequences of *Chlorogloeopsis* 6912 and *Anabaena* 7120 and the NifH sequence of *Anabaena* 7120	181

List of Abbreviations

BCIP	5-bromo-4-chloro-3-indolyl phosphate
bZLM	Leucine zipper motif in hnRNP C
cAMP	Cyclic AMP
CAPS	3-[cyclohexylamino]-1-propanesulfonic acid
CBD	Chitin-binding domain
CIRP	Cold induced RNA-binding protein
dNTP	Deoxynucleotide triphosphate
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminotetracetic acid
EMSA	Electrophoretic mobility shift assay
g	Gravitational force
His-tag	protein containing an insertion of six adjacent histidine residues
hnRNP	Heterogeneous nuclear ribonucleoprotein
IMPACT I	Intein Mediated Purification with an Affinity Chitin-binding Tag (New England Biolabs)
IPTG	Isopropyl β-D-thiogalactopyranoside
MCS	Multiple cloning site
MOPS	3-[N-morpholino]propanesulfonic acid
NBT	Nitro blue tetrazolium

NMR	Nuclear magnetic resonance
NTP	Nucleotide triphosphate
PABP	Yeast polyadenylate binding protein
PCR	Polymerase chain reaction
PMSF	Phenylmethylsulfonyl fluoride
Poly(A)	Polyadenylic acid
Poly(C)	Polycytidylic acid
Poly(G)	Polyguanylic acid
Poly(U)	Polyuridylic acid
PVDF	Polyvinylidine difluoride
RBP	RNA-binding protein
RNP	Ribonucleoprotein
RRM	RNA-recognition motif
RT-PCR	Reverse transcription polymerase chain reaction
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SELEX	Systematic Evolution of Ligands by EXponential enrichment
snRNP	Small nuclear ribonucleoprotein
ssDNA	Singe stranded DNA
TE	Tris-EDTA
U1A	Human U1A protein from the U1 spliceosomal complex
UTR	Untranslated region
VMA	Vacuolar ATPase subunit from *Saccharomyces cerevisiae*

Chapter 1:

Introduction

1.1 Cyanobacteria

The cyanobacteria are ancient. Fossil records are fragmented, but the earliest unicellular and filamentous cyanobacteria found to date are from sedimentary rocks formed 3500 million years ago. The heterocystous and branching forms of this phylum developed later, after the Precambrian era. These fossils are morphologically very similar to the modern day cyanobacteria (Wilmot, 1994).

Based on sequence analysis from 16S RNA, cyanobacteria form one of eleven major phyla of eubacteria. However, very little is known about the taxonomy of the cyanobacteria themselves. As of 1989, over 50% of the cyanobacterial strains in collections did not correspond to the diagnosis of the taxa to which they were supposed to belong (Komarek and Anagnostidis, 1989). On the basis of DNA-DNA hybridisation studies, even *Anabaena* sp. strain PCC 7120, the organism of study in this thesis, should properly be renamed *Nostoc* sp. strain PCC 7120.

Rippka *et al.* (1979) first classified the cyanobacterial strains in the Pasteur Culture Collection into five sections based on their morphology and method of cell division. By this method of classification, Sections I and II correspond to unicellular organisms that multiply by binary or multiple fission. Section III is made up of filamentous strains which have only vegetative cells. Sections IV and V correspond to the heterocyst and akinete forming strains which differ by whether the cells divide in one plane or multiple planes.

From the molecular biology point of view however, the reclassification of the five cyanobacterial sections by partial sequencing of 16S rRNA by Giovannoni *et al.* (1988) is more useful. This classification demonstrates that Sections I and III are not really true lines of descent, but are made up of a number different lineages as well as some lineages that were mixed together. Sections II, IV and V however, are made up of coherent phylogenetic clusters. The clusters contain a number of short branches diverging at the base of the evolutionary tree, which probably represents the period when oxygen concentrations in the air and oceans were rising during the Precambrian era, allowing the colonisation of new biotopes. This led to a great divergence in the cyanobacterial strains (Wilmot, 1994).

Cyanobacteria are important globally as a source of fixed nitrogen. Economically they are important both as a source of fixed nitrogen in "green" fertilisers and in supplementing reserves of fixed nitrogen in Asian rice fields (Rice *et al.*, 1982). Additionally, as the population of the world grows, nitrogen-fixing cyanobacteria are finding more use in sustainable agriculture due to the expense and environmental costs of making ammonia chemically. Replacing the fixed nitrogen made naturally by nitrogen fixing bacteria in legumes alone would take 288 tonnes of fuel and close to $30 billion U.S. annually (Vance, 1997). The idea of transgenic crop plants that contain nitrogen-fixing nodules is thus becoming more attractive (van Kammen, 1997; Martinez-Romero *et al.*, 1997).

1.2 *Anabaena* sp. strain PCC 7120

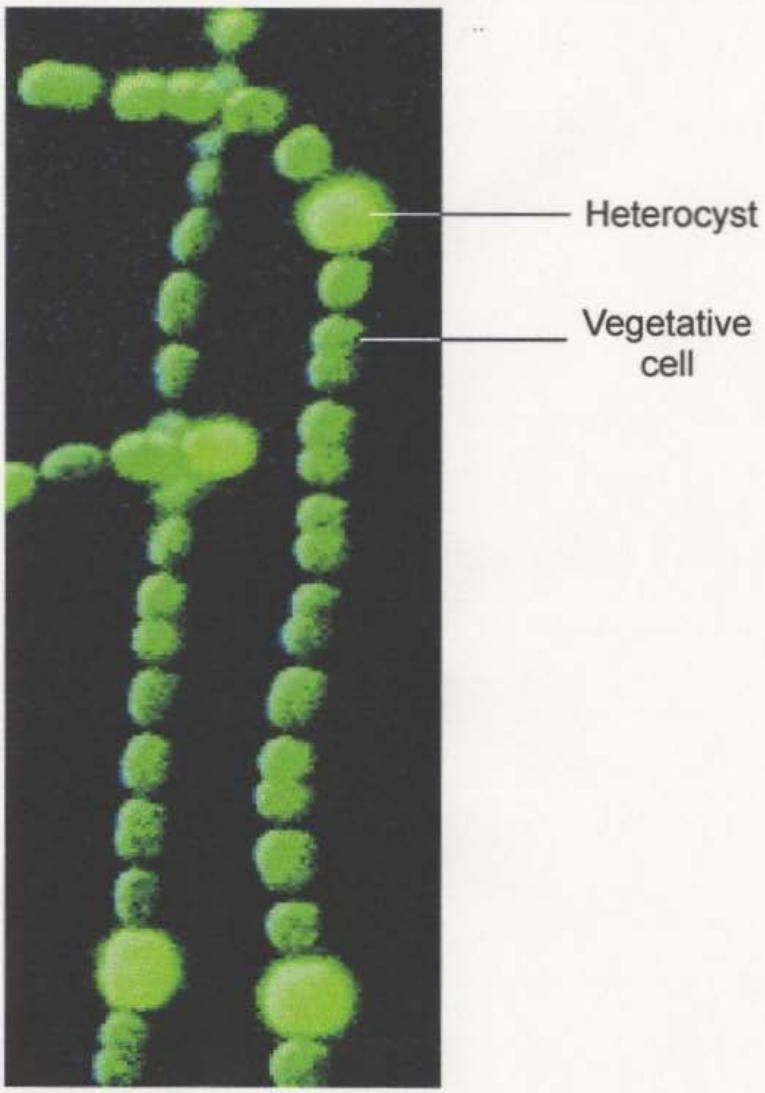
1.2.1 General

Anabaena sp. strain PCC 7120 (*Anabaena* 7120) (Figure 1.1) is a filamentous, heterocyst-forming cyanobacterium. It has a relatively large genomic size at 6.4 million base-pairs (Bancroft *et al.*, 1989). Unlike most prokaryotes, the cells within the *Anabaena* 7120 filament are able to communicate with one another, partly in order to facilitate the intercellular exchange of metabolites between the heterocysts and vegetative cells (Wolk *et al.*, 1993). Because of the flexibility allowed by intercellular communication, certain cells are able to differentiate into anaerobic, nitrogen fixing heterocysts when a fixed source of nitrogen is not available (Haselkorn, 1998; Yoon and Golden, 1998; Wolk 1989).

1.2.2 Heterocysts

In the time before oxygen existed in the atmosphere, around 2.2 billion years ago, micro-organisms lived anaerobically, utilising energy from the sunlight, carbon and nitrogen from the atmosphere. Over time, the use of water as a reductant polluted the atmosphere with oxygen; this transition was detrimental to nitrogen fixation because of the negative reducing potential required for the fixation process (Zhou *et al.*, 1998b). Some micro-organisms adapted to the presence of oxygen by modifying a portion of their cells into nitrogen fixation specialists (Giovannoni *et al.*, 1988). In doing so, these cells became unable to perform photosynthesis and lost the ability to produce oxygen (Wolk, 1996; Wolk *et al.*, 1994; Thiel, 1993). The specialised cells are called heterocysts.

Figure 1.1: Photo of *Anabaena* 7120. This species grows as a filament, and experiences some communication between cells. Under conditions where a source of fixed nitrogen is lacking, every tenth cell will undergo a series of reactions to become an anaerobic, nitrogen-fixing heterocyst. The heterocysts in this picture are expressing green fluorescent protein under the control of a promoter active in the heterocyst. (Photo by Bill Buikema from http://www.thewebpros.com/ccb/science.html)


Heterocyst
Vegetative cell

Because the van der Waal's radii of N and O are so similar (15 Å and 14 Å respectively), it was not possible in nature to permit nitrogen entry into the heterocyst without also permitting entry to oxygen, which would destroy the nitrogen fixation reaction. The heterocyst cells therefore possess a barrier made of a layer of polysaccharide, surrounding a layer of glycolipid, which in turn surrounds a wall layer corresponding to the vegetative cell wall. The layers effectively block the entrance of both the oxygen and nitrogen (Wolk, 1996; Wolk *et al.*, 1988). Different species of the cyanobacteria permit entry of nitrogen (and oxygen) by having 1, 2, or 3 pores between the heterocyst and the adjacent vegetative cells. The oxygen that passes through is reduced to water to prevent it from destroying the nitrogen fixation enzymes (Wolk *et al.*, 1994; Buikema and Haselkorn, 1991a).

Transcriptional regulation of heterocyst development is important in the cell. In *Anabaena cylindrica*, 15 to 25% of the genomic space is reserved for genes that are expressed only in heterocysts (Wolk *et al.*, 1994).

1.2.3 Mechanism of Heterocyst formation

Much of the study into the differentiation of heterocysts has been done with *Anabaena* 7120. In this species, active heterocysts develop in the filament within 24 hours of nitrogen step-down. Heterocysts develop about every tenth cell along the filament (Wilcox *et al.*, 1973a; 1973b; Wolk *et al.*, 1974).

When the fixed source of nitrogen is removed from a culture of *Anabaena* 7120, an unidentified group of genes begins to respond. Between 2 and 3.5 hours after nitrogen

step-down, *hetR* becomes active. This coincides with the appearance of proheterocysts along the filament (Cai and Wolk, 1997; Buikema and Haselkorn, 1991b).

Four to ten hours after the step-down, other genes begin to be active. A gene for proheterocyst maturation, *devA*, becomes active, as does *hepA* which is responsible for the deposition of the heterocyst envelope (Wolk, 1996). If *hepA* is mutated then the heterocysts which are formed tend to be permeable to O_2 and lack a cohesive polysaccharide layer (Wolk *et al.*, 1993). The gene for *hetM*, which is required for the synthesis of heterocyst glycolipids is also induced (Cai and Wolk, 1997).

By 18 hours after nitrogen step-down, XisA has removed an 11 kb fragment from the *nifHDK* operon. This allows expression of the dinitrogenase reductase (a dimer encoded by *nifH*) and dinitrogenase (a tetramer encoded by *nifD* and *nifK*). These enzymes are responsible for fixing nitrogen. Twenty-four hours after the removal of a source of fixed nitrogen, the heterocysts are fixing nitrogen for the filament (Wolk *et al.*, 1994).

Fogg (1949) postulated that the pattern of heterocyst formation in the filaments is based on a competition between a diffusible inhibitor and a non-diffusible activator. The inhibitor should be produced quickly and migrate to other cells to inhibit the activator. The activator should be produced more slowly, and have the ability to activate a positive feedback loop to initiate the process of differentiation. It should also activate synthesis of the inhibitor. The inhibitor should diffuse through the vegetative cells (perhaps being degraded by the cells), setting up a gradient of inhibitor which would decrease as the number of cells form the existing heterocyst increased. At the same time, the vegetative

cells should have made the activator substance. When a cell is reached that has less than a threshold level of inhibitor, the activator could set up the positive feedback loop, inducing both cell differentiation and production of more inhibitors so that other cells do not also start the differentiation process. (Zhou *et al.*, 1998a; Black *et al.*, 1993; Wilcox *et al.*, 1973a; 1973b).

HetR is theorised to be the potential activator (Buikema and Haselkorn, 1991a; 1991b). The expression of *hetR* from a multi-copy plasmid causes the development of heterocysts even in the presence of a fixed nitrogen source (Cai and Wolk, 1997; Black and Wolk, 1994; Liang *et al.* 1992). It is one of the earliest known genes to be activated in the differentiation process, and it is thought to activate its own synthesis within the heterocyst (Haselkorn, 1998; Yoon and Golden, 1998; Buikema and Haselkorn, 1991a; 1991b). In immunoblot studies, HetR has been localised to the differentiating heterocysts and proheterocysts (Zhou *et al.*, 1998a).

PatS, a 17 amino acid peptide, is thought to be the inhibitor (Yoon and Golden, 1998). Over-expression of *patS* prevents all heterocyst formation, while inactivating the gene allows heterocysts to form, even in the presence of a fixed nitrogen source. Contiguous heterocysts develop in filaments of *patS* mutants (Elhai, 1999; Haselkorn, 1998; Yoon and Golden, 1998).

Once the pattern of heterocyst differentiation has been determined, the progress of the differentiation relies on several interlocking steps. For example, development of the mature heterocyst can only be halted by the addition of transcriptional inhibitors such as rifampicin, fluorouracil or proflavin, 1 to 2 or 8 to 10 hours after nitrogen step-down.

The first time slot corresponds to the period in which HetR is becoming active. Inactivation of the transcription of *hetR* prevents the formation of proheterocysts and thus the creation of mature heterocysts. Once the proheterocyst is formed, the chemicals have no effect on the formation of the mature heterocyst unless they are added 8 to 10 hours after nitrogen step-down. This period of time corresponds to the transcription of the envelope biosynthesis genes (Bradley and Carr, 1977). These results strongly indicate a cascade events of that rely on the initial step and then on other intermediate steps rather than on environmental pressures.

1.3 RNA-Binding Proteins

1.3.1 RNP-Type RNA-Binding Proteins

RNA binding proteins are ubiquitous in eukaryotes. The most common type of RNA-binding proteins are those containing the RNP domain, or ribonucleoprotein domain. These proteins are characterised by an RNA-recognition motif (RRM) that is made up of 70-90 residues that participate in RNA binding (Figure 1.2). In the protein, the RRM forms a domain containing two α-helices which are arranged across the back of four anti-parallel β-strands. Two sequences in the RRM are more highly conserved than the rest of the domain and are thought to interact most directly with the bound RNA. These are RNP1, made up of eight amino acids, and RNP2, which is made of six amino acids (Nagai *et al.*, 1995; Hanano *et al.*, 1996; Li and Sugiura, 1991). The RNP1 and RNP2 sequences lie on the third and first β-strands respectively.

Figure 1.2: U1A RNA-Recognition motif bound to an RNA hairpin. The diagram shows the first RRM in U1A bound to an RNA hairpin. The RRM secondary structure is composed of a βαββαβ motif. RNP1 lies on β-strand three of the RRM (green), and RNP2 lies on β-strand one (cyan). The most variable part of the RRM lies in loop three of the domain. In U1A, this domain protrudes through the RNA hairpin and is essential for the RNA-protein interaction. The diagram was taken from http://www.tulane.edu/~biochem/nolan/lectures/rna/u1a.htm, and is referenced to Oubridge *et al.* (1994).


RNP1
RNP2
Loop 3
RNA
hairpin

Most RRMs contain solvent-exposed tyrosine and phenylalanine side-chains on β-strands 1 and 3. The most conserved of these residues are phenylalanine residues at position 13 on β-strand 1 and positions 54 and 56 on β-strand 3. The geometry of the residues suggests that they are involved in base stacking; it also accounts for the ability of many RRM-containing proteins to be cross-linked with their RNA targets (Kenan *et al.*, 1991). X-ray crystallography studies of the U1A protein RRM with the cognate RNA have confirmed these interactions (Oubridge *et al.*, 1994).

The first RNA-binding protein of the RNP type to be identified was the yeast polyadenylate binding protein which contains four tandem repeats of the RRM (Nagai, 1996). The RNP-domain has since been found in over 200 other proteins. The RNP-type RNA-binding proteins are normally involved in post-transcriptional metabolism. Some of the functions for these proteins in eukaryotic cells are capping, polyadenylation, splicing and alternative splicing, transport of mRNA, localisation of RNA and RNA turnover (Kim and Baker, 1993; Haynes, 1992). Some of the RNP-type proteins are induced in response to cellular stresses, such as cold, both in prokaryotes (Sato, 1995; Maruyama *et al.*, 1999) and eukaryotes (Dunn *et al.*, 1996; Breiteneder *et al.*,1994).

1.3.2 RNA-Binding Proteins in Cyanobacteria

The first paper showing that RNP-type binding proteins might not be strictly eukaryotic came with the paper by Mulligan *et al.* (1994); demonstrating a gene encoding an RRM-containing protein in *Chlorogloeopsis* sp. strain PCC 6912 (*rbpA*). Mulligan *et al.* used Southern blotting with a DNA probe for *rbpA* to demonstrate that heterocyst-

forming cyanobacteria tend to have more copies of these genes than do non-heterocyst-forming unicellular cyanobacteria. Maruyama *et al.* (1999) postulate that the number of RNA-binding protein genes in different cyanobacteria may be related to the size of the genomes. This fits with the theory that the size of the cyanobacterial genomes arose from multiplication of an original genome of 1.2 megabase-pairs (Herdman *et al.*, 1979).

Little is known about RNA-binding proteins in cyanobacteria (Smith, *et al.*, 1996). No binding sequences from RNA have been identified and no binding constants have been measured. Some cyanobacterial RNA-binding proteins seem to be involved in the cold-shock process. RNA-binding proteins containing glycine-rich domains are upregulated when the growing temperature is dropped from 38°C to 22°C. However, those that do not contain the glycine-rich region are not regulated by temperature (Sato, 1995; Maruyama *et al.*, 1999; Chamot *et al.*, 1999). Little is known about this temperature dependence except that under conditions of cold shock, at least in *Anabaena variabilis* M3 and *Anabaena* 7120, RbpA1 seems to act as a repressor for the initial steps in heterocyst formation when nitrate ions are present (Sato and Wada, 1996).

From an evolutionary point of view, the cyanobacterial RNA-binding proteins are very interesting to study. Until this year, the conserved nature of the RRM sequences in chloroplasts and cyanobacteria had researchers convinced that cyanobacterial RNA-binding proteins were the precursor for the chloroplast version of the proteins (Mulligan *et al.*, 1994; Sugita and Sugiura, 1994). But in 1999, Maruyama *et al.* took this further. They showed, using phylogenetic studies, that the chloroplast RNA-binding proteins probably diverged from eukaryotic RNA-binding proteins before the addition of the

glycine-rich region or duplication of the RRMs, rather than arising from the cyanobacterial proteins. Also, until recently, cyanobacteria were the only prokaryotes found to have RNP-type RNA-binding proteins. Most other bacteria seem to contain cold-shock domain proteins, which perform the same job as the RRM-containing RNA-binding proteins in cyanobacteria. However, Maruyama *et al.* found RRM-containing proteins in the genome sequences of *Helicobacter pylori* (Tomb *et al.*, 1997) and *Treponema pallidum*, (Fraser *et al.*, 1998). This suggests the possibility that other bacteria may contain these proteins as well.

1.3.3 RbpD in *Anabaena* 7120

The *rbpD* gene is one of eight genes in *Anabaena* 7120 encoding an RRM-containing protein, and corresponds to the *rbpC* gene in *Anabaena varabilis* M3 (Maruyama *et al.*, 1999). It resides on a 374 base-pair fragment lying between EcoRI and HindIII sites in the genome. The fragment was cloned into a pBR322 vector by Chris Holden (1995) and sequenced. The open reading frame in the fragment encodes a protein that is 110 amino acids long (Figure 1.3). RbpD is interesting because it contains a slightly longer glycine–rich region than the other RRM-containing proteins encoded by *Anabaena*. This region is shown in outlined letters in Figure 1.3. The C-terminal four amino acids also have the sequence "RRSY", whereas most of these proteins have the conserved sequence "RNRY". This gene was chosen as the single gene to study because none of its characteristics have yet to be elucidated by any of the groups working in the area of cyanobacterial RNA-binding proteins.

Figure 1.3: Nucleotide sequence of the *rbpD* gene from *Anabaena* 7120 and its inferred amino acid sequence. The gene resides on a 374 base-pair EcoRI-HindIII fragment. The open reading frame is 110 amino acids long, and the calculated molecular weight of the expressed protein is 12 031 kDa. The RNP2 hexapeptide is highlighted in red and the RNP1 octapeptide is highlighted in blue. The glycine-rich auxiliary domain is written in outline letters. This data is from Holden (1995).

EcoRI

GAATTCGGAGACATTGTATGTCAATTTACATCGGTAACTTATCCTACCAAGTTACAGAGGAAGACCTAAAGCTGG

M S I Y I G N L S Y Q V T E E D L K L A

CCTTCGCAGAGTACGGAAAAGTTAGCCGCGTTCAATTACCAACCGACCGTGAAACTGGCCGTCCTCGTGGGTTTG

F A E Y G K V S R V Q L P T D R E T G R P R G F A

CTTTTGTGGAAATGGAAACAGAAGCTCAAGAAACCGCAGCCATTGAAGCACTGGATGGTGATGAATGGATGGGAC

F V E M E T E A Q E T A A I E A L D G A E W M G R

GTGATTTAAAAGTCAACAAAGCTAAACCCCGTGAAGAAAGAAGTTCTTCTCCTCGTGGTGGCGGCGGTAGTTGGG

D L K V N K A K P R E E R S S S P R G G G G S W G

GTAATAATAACCGTGGTGGTGGCGGCGGTGGTAATCGCCGTAGTTACTAAATCCTATTGCTGAAACGCAAGCTT

HindIII

N N N R G G G G G G N R R S Y*

1.3.4 U1A

The human U1A protein is probably the best known and understood of the RNP-type proteins; interaction of its RRM with stem-loop II of U1 RNA is the basis for much of the understanding of how RRMs and RNA interact. U1 snRNP is made up of U1 RNA (165 nucleotides), three U1 specific protein subunits (A, 70 kDa, and C) and several core Sm proteins that are common to the other spliceosomal snRNPs. The U1 RNA is complementary to the 5'-end of the splice site at its own 5'-end, contributing to the localisation of the splice site. U1 RNA folds into a secondary structure that contains four stem-loop structures (Evans *et al.*, 1993). U1A, a 283 amino acid protein, binds to stem-loop II (Scherly *et al.*, 1989). It is able to regulate its own production by also binding to an internal loop in its 3'-untranslated region, blocking polyadenylation of the pre-mRNA. The internal loop contains 7 nucleotides with a sequence identical to those found in the loop portion of stem-loop II in U1 RNA. Polyadenylation is blocked when U1A (bound to the 3'-untranslated region) interacts directly with the poly(A) polymerase (Gubser and Varani, 1996).

U1A consists of two RRM domains joined by a linker. Only the N-terminal domain is required for binding to U1 RNA. In examining the binding of the first RRM of the protein with hairpin II of U1 RNA by X-ray crystallography, Oubridge *et al.* (1994) found that the RNA loop lies across the β-sheet of the RRM, fitting into the groove between β-strand 2 and β-strand 3. In the structure, the N- and C-terminals of U1A are not well ordered in the free protein, nor is loop 3 of the protein. However, when U1A is bound to the RNA-hairpin, the network of hydrogen bonds at the molecular interface

keeps all the parts of the protein in place. The RNA bases in the loop interact extensively with the side-chains of the RNP1 and RNP2 amino acids (Jessen *et al.*, 1991), and also with the main-chain amide and carbonyl groups in the C-terminal region of the peptide.

Amino acids 2 to 98 in U1A are sufficient to bind to the U1 RNA with full affinity, but residues up to amino acid 114 in the protein help to add specificity to the binding. This specificity in the C-terminal end of the first RRM has been found to be critical to RNA recognition in a number of proteins, including U1A, hnRNP and U1 70K (Allain *et al.*, 1996). Multidimensional heteronuclear NMR of a peptide that included the first 117 amino acids allowed the discovery of a third α-helix (helix C) in the U1A RNP-domain. This helix forms from residues 92-98. In the free protein, helix C interacts with the β-sheet, forming a hydrophobic core involving the residues L44, I58, F56, I93, I94 and M97. The position of the helix prevents exposure of the hydrophobic residues, stabilising the protein. Because the helix blocks part of the binding site however, it must rotate 136° upon the interaction of RNA with the RRM, exposing the binding site (Avis *et al.*, 1996).

The loops between the β-strands are important as well. Loop 3, between β-strands 2 and 3, contains one of the most variable regions of the RRM. If a five amino acid sequence (residues 44 to 48) is moved from this region in U2B" to the corresponding position in U1A, it changes the specificity of U1A so that it binds the U2B" cognate RNA rather than its own (Scherly *et al.*, 1990; Bentley and Keene, 1991; Hoffman *et al.*, 1991). Two loops, one between β-strand 1 and α-helix A, and one between β-strands 2 and 3 (loop 3), have a number of basic amino acids that are essential for binding to RNA.

These include K22, K23, K27, R47, K50 and K52. These two loops form a pair of basic "jaws" that hold the backbone of the RNA in position (Evans *et al.*, 1993; Jessen *et al.*, 1991).

It is not only the protein that experiences conformational changes on interaction with the RNA. Using NMR to examine RNA resonances in complex with U1A, Hall (1994) found that there were significant resonance shifts in both the nucleotide bases and ribose sugars of the RNA. This indicated that the environment of many ribonucleosides was being altered on binding to the protein. Hall concluded that the formation of the RNA-protein complex resulted from a number of interdependent interactions, where the failure to make one contact affected the formation of other contacts.

Surprisingly, the C-terminal RNA-recognition motif in U1A does not appear to bind to RNA (RNA-polymers, snRNAs or random RNA sequences), at least not in its monomeric form. The domain has good sequence and structural conservation with the N-terminal RNA-binding domain and contains all the core hydrophobic amino acids. There are a number of possible reasons for this lack of RNA binding. The C-terminal recognition motif may have the wrong amino acid residues at critical positions. It may lack the necessary electrostatic contacts for binding, or lack the appropriate amino acids displayed in the correct geometric arrangement. The tertiary structure of the domain might inhibit RNA contact. Or, the domain may simply require an auxiliary protein in order to be able to bind to the RNA, as does U2B" (Lu and Hall, 1995).

1.3.5 Heterogeneous nuclear Ribonucleoproteins

The heterogeneous nuclear ribonucleoprotein particles (hnRNPs) are those proteins that bind to hnRNAs, but are not a stable part of some other nuclear complex such as snRNPs (Dreyfuss *et al.*, 1993; Dreyfuss, *et al.*, 1988). At present about thirty hnRNP proteins varying in size from 34 to 120 kDa, have been identified by two-dimensional gel electrophoresis (Krecic and Swanson, 1999). The hnRNPs contain an RNP-type RNA-binding motif and many have demonstrated RNA-binding activity. The two most studied hnRNP proteins are hnRNP C and hnRNP A1. Most is known about the structure and binding of the C protein, which is important because it allows comparisons to be made with the information known about U1A. The A1 protein is less well understood in terms of its structure, but the cyanobacterial RNA-binding proteins are often compared to hnRNP A1 because it contains a glycine-rich region which may affect its protein-protein interactions or RNA-binding interactions (Dreyfuss *et al.*, 1993).

The C protein will be discussed first. The protein is 41-43 kDa and is confined to the nucleus of interphase cells. It consists of two parts, an N-terminal RNA-binding domain and a C-terminal acidic domain that contains a putative NTP-binding region. HnRNP C is often post-translationally modified by phosphorylation. The protein is involved in transcript packaging, splicing, and nuclear retention (Krecik and Swanson, 1999).

The solution structure of the RNA-binding domain of hnRNP C was determined by multidimensional heteronuclear NMR of residues 2-94. The resulting structure was very similar to that of U1A with the exception that the residues in the loop between β-

strands 2 and 3 were missing (this is the variable loop which is involved in determining RNA-binding specificity in U1A) (Wittekind *et al.*, 1992). Photochemical cross-linking and SELEX (see Section 1.6) experiments determined that hnRNP C prefers a poly(U) sequence for binding (5 and 6 Us respectively) (Gorlach *et al.*, 1994; Swanson and Dreyfuss, 1988). The authors suggested that the RRM confers a general RNA binding ability and that the variable regions in the loops, and N-terminals and C-terminals (especially residues 95-104) confer the specificity of interaction. Chemical shifts from the NMR data suggest that the RNA polymer is bound on the β-sheet of the RRM, with the sheet acting as a platform for binding rather than as a binding pocket. This leaves the bound RNA exposed for interaction with other proteins (Dreyfuss *et al.*, 1993; Gorlach *et al.*, 1992).

In the past two years, there has been a suggestion that the RRM of hnRNP C is not the principal portion of the protein involved in RNA binding. Shahied-Milam *et al.* (1998) have shown that hnRNP C binds in a co-operative manner as a tetramer to blocks of RNA that are 700 nucleotides long, and also to snRNAs U1, U2, and U6 as well as to a 116 nucleotide ribosomal RNA transcript. The winner from the SELEX experiment (Gorlach *et al.*, 1994) did not bind as strongly to hnRNP C as these sequences, possibly because of the short oligonucleotide length used in the SELEX experiments. The binding of U1, U2, U6 and the 116 nucleotide RNAs is not mediated by the RRM as originally thought, but by a 28 amino acid leucine zipper motif (bZLM). Using competition studies, Shahied-Milam *et al.* (1998) theorise that the RRM may function as a negative allosteric modulator of hnRNP C-protein interactions.

Shahied–Milam *et al.* (1998) also made two RRM deletion mutants in domains considered important for RNA-binding. One mutant contained the first 115 amino acids, the RRM domain, and the other contained amino acid residues 119-290, consisting of the bZLM and acidic carboxy-terminus. In the mutant containing only the RRM, there was reduced affinity for target RNA as compared to the wild-type protein. The affinity of the mutant containing the bZLM and acidic domain for the RNA did not lessen as compared to the full length protein, but this mutant did show severe defects in RNA-activated tetramer-formation. Taken with the RNA-binding domain's low affinity for RNA, the authors postulate that the RRM may be important in protein-protein interactions rather than protein-RNA interactions. This does have some precedent in the U2B" snRNP protein, which interacts with U2A snRNP protein through its RRM before being able to bind to its cognate RNA (Hall and Kranz, 1995).

Heterogeneous nuclear ribonucleoprotein A1 (34 kDa) is involved in alternative splicing of pre-mRNA, mRNA transport and telomere biogenesis (Krecik and Swanson, 1999). The protein contains two RNA-binding domains (Merrill *et al.*, 1988), positioned with their β-sheets oriented in opposite directions, and a C-terminal glycine-rich region (Dreyfuss *et al.*, 1993). There are a number of different isoforms of the protein, generated by alternative splicing (Buvoli *et al.*, 1990). The length of loop 3 in the RRMs of hnRNP A1 is varied by alternative splicing. This may be important in determining the specificity of RNA binding. More diversity is created using post-translational modifications such as phosphorylation and arginine methylation.

SELEX experiments demonstrated that hnRNP A1 binds to sequences resembling 5' and 3' splice sites of pre-mRNA. Mutation of the conserved AG found at the 3' splice site severely inhibits the binding of hnRNP A1. If a SELEX experiment was done using the N-terminal and C-terminal RRMs separately from the whole protein, then different RNA sequences were bound in each experiment. This indicated that the binding specificity of the A1 protein is a result of both RNA-binding domains acting as a single RNA-binding composite (Burd and Dreyfuss, 1994). Shamoo *et al.* (1997) used multidimensional heterogeneous nuclear NMR in order to determine the solution structure of hnRNP A1. The two RRMs stack against one another through an antiparallel interaction of helix B in RRM1 and the corresponding helix in RRM2. The glycine-rich region in the C-terminal region of the protein seems to be involved in facilitating co-operative binding of RNA by a hnRNP A1 dimer (Dreyfuss, *et al.*, 1993).

1.3.6 Glycine Loops

Glycine-rich regions are common in many proteins including locicrins, keratins and in the hnRNP proteins. In most of these proteins, glycines compose about 40% of the region and are interspersed with aromatic amino acids (Steinert *et al.*, 1991). In hnRNPs the glycine rich region is thought to be involved in protein-protein interactions, and sometimes in protein-nucleic acid interactions, as with hnRNP A1 (Haynes, 1992).

Plant glycine-rich proteins are very similar to the cyanobacterial RNA-binding proteins. The plant proteins are generally 16-17 kDa in size (comparable to the 12-14 kDa seen in cyanobacterial RNA-binding proteins) and contain a single RRM which is

60-80% identical to all other RRMs studied. The plant glycine-rich auxiliary domain contains about 70% glycines, and 10-15% arginines. It also contains a number of RGG boxes which have previously been defined as RNA-binding domains (Alba and Pagès, 1998). The RNA-binding proteins in cyanobacteria do not have the RGG box, but do tend to have fairly high amounts of charged amino acids, including arginines and asparagines (See Figure 1.3).

The variability in glycine-rich regions does tend to be quite high, even within a specific group of proteins, because there is little evolutionary pressure for sequence conservation, as long as the common structural motif of being able to form glycine loops in the interaction with other proteins is maintained (Steinert *et al.*, 1991).

While the actual function of these proteins is not known, most of the proteins with a single RRM and a glycine-rich region are induced under conditions of cold-shock. In barley, alfalfa, carrot and maize, a number of proteins that contain the RNA-binding domain and glycine-rich region have been found that are produced in response to cold temperatures (Carpenter *et al.*, 1994; Hirose, *et al.*, 1993; Ludevid *et al.*, 1992). Similar proteins in these organisms, and others, are also produced in response to other stresses, including drought, heavy metal stress and wounding (Dunn *et al.*, 1996). The CIRP (Cold Inducible RNA-binding Protein) protein in mouse, which contains the same domains, is also involved in a cold temperature response (Nishiyama, *et al.*, 1997). The significance of the induction is not known since the protein levels in these responses has not been measured (Alba and Pagès, 1998).

1.3.7 Evolutionary Trends in RNA-Binding Proteins

In the far distant past, a cyanobacterium-like organism probably formed an endosymbiotic-type relationship with another micro-organism which eventually led to the rise of the eukaryotic plant cell. The phylogenetic data, from 16S rRNA sequences, support the hypothesis that the plant chloroplasts are one of the cyanobacterial sublines of descent and actually should be included with the cyanobacteria as a holophyletic group (Sugita *et al.*, 1997; Giovannoni *et al.*, 1988). Given the hypothesis that chloroplasts arose from cyanobacteria, the relationship between chloroplast and cyanobacterial RNA-binding proteins may be important to understanding the properties of the cyanobacterial RNA-binding proteins.

In support of this hypothesis is the finding that many of the genes in cyanobacteria and plants are similar. This makes the comparison of any protein found in plants with the corresponding protein found in cyanobacteria (or vice-versa) desirable, to see if the proteins diverged before or after the divergence of prokaryotes and eukaryotes, and also to look at the changes brought about through the millennia. Alba and Pagès (1998) suggest that the RRM from an early cyanobacterium-like organism may have been transferred from the endosymbiont to the nucleus of the early eukaryote. This gene may later have duplicated and fused to other genes, giving rise to nuclear-encoded, chloroplast RNA-binding proteins and glycine-rich proteins. Support for the theory that the RNA-binding proteins descend from a common ancestor comes from evidence of both the plant and cyanobacterial polymer-binding experiments. RNA-binding proteins from both types

of organisms have a strong preference for binding to G and U nucleotide tracts (Ye and Sugiura, 1992; Ludevid *et al,* 1992; Li and Sugiura, 1991).

Topological data for the RNA-binding domain argue that the domain is ancient. The domain forms a globular structure with a conserved three-dimensional shape even between such dissimilar proteins as snRNP U1A and hnRNP C which share only 20% sequence identity. Tellingly though, the identity between these two proteins is found mostly in the hydrophobic residues which interact with the cognate RNAs (Fukami-Kobayashi *et al.*, 1993). Many of the modern day RNP-type RNA-binding proteins arose through a combination of gene duplication and intragenic domain duplication. The phylogenetic data show that some of the RNA-binding proteins with multiple RRMs contain more conservation between the RRMs of the different proteins than between the RRMs of the same protein (as in the poly(A)-binding proteins). This indicates that domain duplication occurred in a common ancestor and then each RRM evolved independently (Birney *et al.*, 1993; Fukami-Kobayashi *et al.*, 1993).

The glycine-rich regions of plant glycine-rich proteins and cyanobacterial RNA-binding proteins arose separately from the RRMs. Because of the similarities between the two types of proteins (described in Section 1.3.6), Alba and Pagès (1998) suggested that the protein, including a single RRM and glycine-rich region, must be an extremely ancient structure originating before the divergence of prokaryotes and eukaryotes. However, phylogenetic studies by Maruyama *et al.* (1999) showed that within eukaryotes, the glycine-rich proteins are monophyletic and completely independent of the cyanobacterial lineage of RNA-binding protein glycine-rich region. The analysis

indicated that the glycine-rich domain was probably added during the diversification of cyanobacterial RNA-binding proteins. The similarity between the two regions is a case of convergent evolution. This indicated that chloroplast RNA-binding proteins are not direct descendants of cyanobacterial RNA-binding proteins, but diverged from other eukaryotic RNA-binding proteins before either the duplication of the RRM or the addition of the glycine-rich domain (Maruyama *et al.*, 1999).

1.4 Cold-Shock Proteins

Cold-shock proteins are also of interest because they demonstrate convergent evolution with RNA-binding proteins. Both protein families contain RNP1 and RNP2 sequences as well as many of the basic and aromatic amino acids that form the RRM. Both cold shock proteins and cyanobacterial RNA-binding proteins also have a glycine rich region (with the exception of RbpB in *Anabaena* 7120 and RbpD in *Anabaena variabilis*) (Maruyama *et al.*, 1999; Graumann and Marahiel, 1996a). However, the remainder of the proteins have little in common with one another in terms of sequence and topology. The differences in the remainder of the proteins indicate that the domains have converged toward the same function rather than having diverged from a single ancestral protein (Graumann and Marahiel, 1996b).

In fact, three families of proteins have been found that contain domains with the RNP1 and RNP2 sequences. These families are the cold shock proteins (Schnuchel, 1993; Landsman, 1992), the RNA-binding proteins, and the bacterial Rho factors. In all three families, the active domains contain the RNP motifs on two spatially conserved β-

strands forming part of an anti-parallel β-sheet (Schindelin *et al.*, 1993). Graumann and Marahiel (1996b) proposed that selection pressure has created a β-strand surface to allow single-stranded oligonucleotide binding, where the basic amino acid residues attract the negative phosphate backbone of the oligonucleotide. The aromatic residues stack with the bases, and there are a number of glycine residues present to allow close association of the protein and oligonucleotide sequence to facilitate specific recognition.

Both RNA-binding proteins from cyanobacteria and the cold shock proteins have experienced some degree of convergent evolution from the functional perspective, as well as from sequence specificity. Proteins in both families are induced during a drop in temperature. There is also some speculation that proteins in both families may operate as molecular chaperones (Maruyama *et al.*, 1999; Sato, 1995). Interestingly, one of the chloroplast RNA-binding proteins also contains a glycine-rich region, and is involved in response to various stresses.

Some authors (Graumann and Marahiel, 1996a; Sato, 1995) have proposed that the cyanobacterial RNA-binding proteins might be the functional equivalent of the bacterial cold shock proteins. Lending credence to this speculation is the fact that no cold shock proteins have been found in cyanobacteria to date, and vice-versa.

The vast majority of the strains tested (Mulligan *et al.*, 1994) have multiple copies of genes encoding RNA-binding proteins in their genomes. Graumann and Marahiel (1996a) hypothesise that cold shock is more likely to occur and may present a greater obstacle to single cell micro-organisms than does heat shock. Having multiple copies of the cold shock genes would allow cold shock adaptation even if several of the genes were

knocked out. This same logic seems to hold for the RNA-binding protein genes in cyanobacteria.

1.5 Inteins

1.5.1 General

Inteins are protein introns that are encoded within the amino acid sequence of other proteins. They have been found in mycobacteria, thermophilic archaebacteria, yeast, chloroplasts (Pietrokovski, 1996), and more recently in bacteriophage (Lazarevic *et al.*, 1998; Derbyshire and Belfort, 1998). A single gene encodes the sequence for two proteins which are transcribed and translated together. The internal protein segment, or intein (Perler *et al.*, 1994), is precisely excised from the N- and C-exteins, usually resulting in a functional protein being released from the precursor protein. The N- and C-exteins ligate together to form a functional protein from the precursor protein (Clarke, 1994; Chong *et al.*, 1996).

Many of the excised inteins act as homing endonucleases that catalyse the lateral transfer of the DNA sequence for the intein into inteinless alleles of the gene. This endonuclease activity is unrelated to the splicing reaction implemented by the intein (Xu, *et al.*, 1994; Clarke, 1994; Chong *et al.*, 1996). While the amino acids required for the splicing of the extein are present on the N- and C-terminal ends of the intein, the endonuclease domain tends to be in the centre of the intein (InBase: http://www.neb.com/neb/inteins/int_intro.html).

1.5.2 Rationale for the Placement of Inteins

There is an ongoing debate about whether inteins provide some sort of evolutionary advantage to the gene, or whether they are molecular parasites that spread by duplicating themselves into inteinless alleles of the genes in which they reside. If the introduction of an intein conferred a selective advantage, then it would account for the persistence of inteins. If the inteins were molecular parasites, it would explain the fast and efficient splicing of the inteins, as this would provide minimal disruption to the infected gene product. Spreading into inteinless alleles would increase the survival chances of the parasite, explaining the endonuclease activity (Pietrokovski, 1996).

In over 70% of the inteins identified, the spliced exteins make up a protein that is involved in DNA metabolism of some sort (Derbyshire and Belfort, 1998; Chong *et al.*, 1996). Knowing that inteins are very efficient at splicing out of the host protein so that they do not disrupt the function of the exteins, the question arises as to why the range of intein-containing genes is so small, remaining mostly in these metabolic genes. Four reasons have been put forward to explain the narrowness of range. The first is that the presence of the intein may confer a selective advantage to the organism by residing in these genes. However, the nature of the advantage has yet to be determined. The second is that certain DNA structures present only in these genes may facilitate entry of the DNA encoding the intein. The third is that the sites coding for intein sequences might not readily tolerate the removal of the DNA because of the chance of improper excision. Fourthly, most of the places in which intein encoding DNA exists are genes which code

duplicated functions, and which may not upset the cell much if the integrity of the expressed protein is affected (Derbyshire and Belfort, 1998).

1.5.3 Mechanism of Intein Excision

Comparison of the splice junctions in inteins shows a number of conserved amino acids which are involved in the splicing reaction. There is a thiol- or hydroxyl-containing Cys, Thr or Ser residue at the upstream and downstream ends of the inteins. Inteins from thermophilic archaebacteria generally contain serine or threonine, while those from the mesophilic bacteria generally contain cysteine. There is also generally a dipeptide motif, His-Asn, at the C-terminus of the intein preceding the thiol/hydroxyl position. The dipeptide is generally preceded by a number of hydrophobic residues (Xu *et al.*, 1994).

Most of the original work on the mechanism of intein splicing was done using the extremely thermophilic archaebacterium *Pyrococcus sp.* GB-D (which grows at 95°C to 104°C (Xu *et al.*, 1994)). Intein-containing genes from this species were cloned into an *E. coli* expression system and the intein precursors isolated. Each of the intermediates could be isolated by carrying out *in vitro* reactions at low temperature (12 to 15°C), and then the splicing reaction could be observed by raising the temperature (Chong *et al.*, 1996). However, if the intein splicing mechanism was to be used as a method to express proteins from mesophilic organisms, the extremes in temperatures required by the thermophilic inteins would be detrimental. Instead, the results from *Pyrococcus sp.* GB-D were used as a guide to deduce the splicing mechanism in mesophilic organisms.

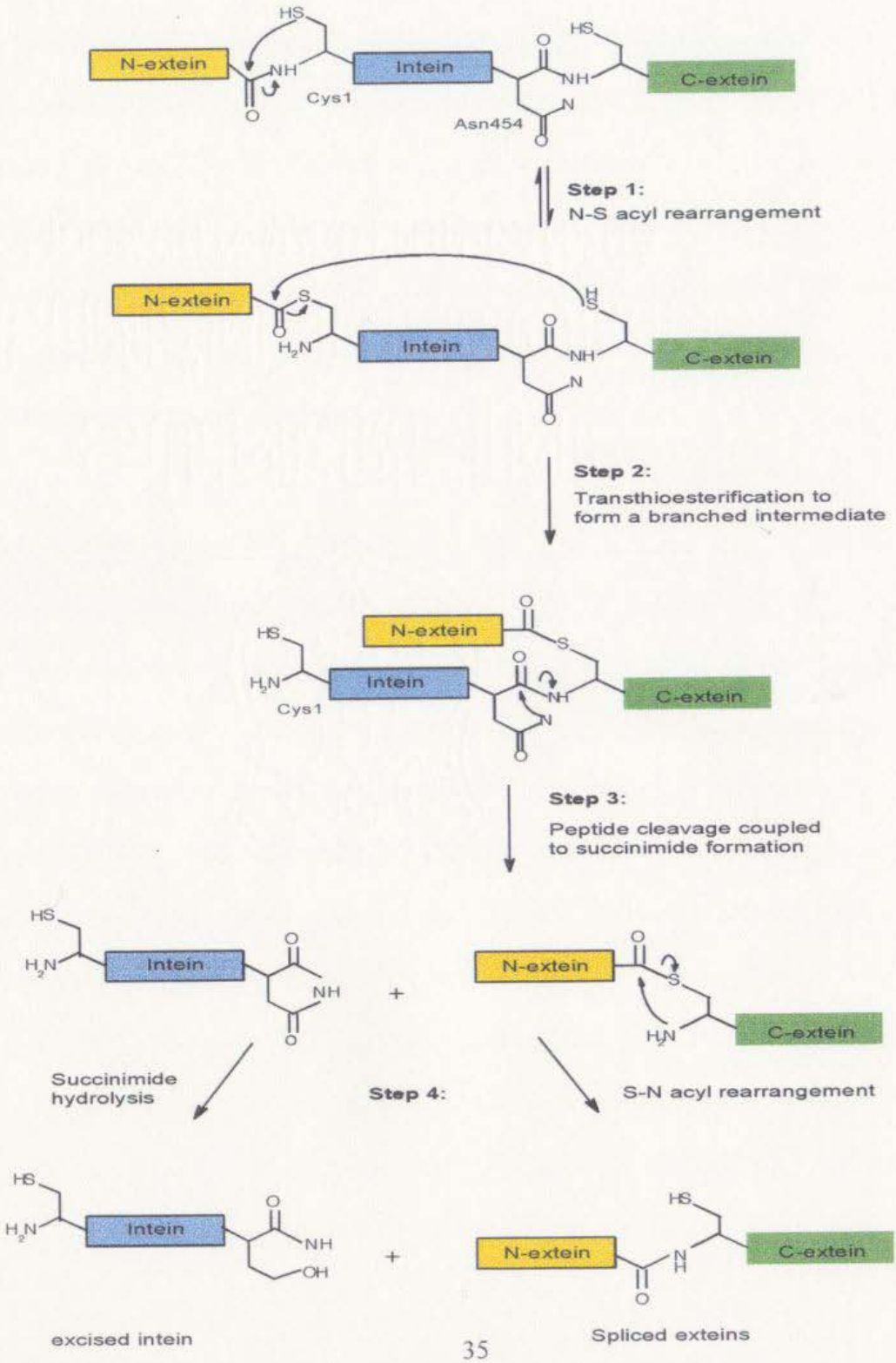
The mechanism of the excision process for a mesophilic intein is shown in Figure 1.4. The first step involves the formation of a thioester intermediate through an N-S acyl rearrangement at Cys1 in the upstream splice junction (Shao *et al.*, 1996). The second step creates a branched intermediate through a transthioesterification reaction. This is followed by a very fast cyclisation of Asn454 (forming succinimide) which allows the intein to excise from the protein in the third step. Step 4 consists of an S-N acyl rearrangement, so that the N- and C-exteins are now joined by a peptide bond; the succinimide also hydrolyses back to Asn454 (Xu *et al.*, 1994).

1.5.4 Inteins in Cyanobacteria

Recently, two inteins have been found in the cyanobacterial strain *Synechocystis sp.* PCC 6803. One is in the *dnaB* gene (Pietrokovski, 1996), and the other is a split intein in the *dnaE* genes (Wu *et al.*, 1998). These are the first inteins to have been found in eubacteria outside of the *Mycobacterium* genes.

The DnaB intein in *Synechocystis* shares the same integration point in the centre of 15 conserved amino acids with the DnaB intein in the chloroplast of the red alga *Porphyra purpurea*. Although there are a number of silent substitutions in the nucleotide sequence of the integration area, the presence of the intein might represent a single evolutionary event. Red alga are thought to have evolved from cyanobacteria 1.25 – 2.1 billion years ago. This implies that inteins are extremely ancient and also that the DnaB inteins have survived in their hosts a very long time (Pietrokovski, 1996).

Figure 1.4: Schematic diagram of the mechanism of intein splicing. The first residue of the intein is designated Cys1; the last residue is Asn 454; positions are those of the *S. cerevisiae* VMA intein. The first step of intein removal involves the formation of a thioester intermediate through an N-S acyl rearrangement at Cys1 in the upstream splice junction. The second step creates a branched intermediate through a transthioesterification reaction with Cys 455. This is followed by a very fast cyclisation of Asn454 (forming succinimide) which allows the intein to excise from the protein in the third step. Step 4 accounts for an S-N acyl rearrangement, so that the N- and C-exteins are now joined by a peptide bond; the succinimide also hydrolyses back to Asn. Mutation of Asn454 to alanine (N454A) can prevent the cyclisation step, thus preventing the release of the intein from the C-terminal extein, a feature used in the isolation of proteins using the IMPACT I system (Chong *et al.*, 1997).


N-extein
Intein
C-extein
Cys1
Asn454
Step 1:
N-S acyl rearrangement
Step 2:
Transthioesterification to
form a branched intermediate
Step 3:
Peptide cleavage coupled
to succinimide formation
Succinimide
hydrolysis
Step 4:
S-N acyl rearrangement
excised intein
Spliced exteins

As for the DnaE intein, it was the first split intein to be found. The N-extein and portion of intein are coded by one strand of DNA while the C–extein and portion of intein are found 745 226 base-pairs away and are coded by the opposite strand of DNA. After translation of the two genes, the proteins undergo a *trans*-splicing reaction so that the two exteins form one protein. This has important implications for intein evolution. The intein probably arose from a continuous intein that lost its continuity during a genomic rearrangement. It also shows that inteins can have *trans*-splicing activity even in biological systems. The split intein may represent a novel mechanism for regulating expression of the DnaE protein (Wu *et al.* 1998).

1.5.5 Inteins for Protein Expression Systems

Purification of recombinant proteins through the use of affinity tags is a convenient and widely used technology. Some examples of this include the maltose-binding protein system, glutathione S-transferase and polyhistidine systems. However, following most of the protein purification procedures, the tag must be cleaved from the larger protein. The proteases used in these procedures are not always specific and may cleave at a secondary site as well as the intended site. Many of the proteases also require elevated temperatures which may affect the stability of the protein. Sometimes cleavage is made impossible by secondary structures in the protein.

Inteins provide an efficient way of avoiding these problems while still providing the advantages of affinity chromatography. The inclusion of an affinity tag in place of the C-extein of the protein allows the purification of the protein (Chong *et al.*, 1997;

Severinov and Muir, 1998). If the protein of interest is fused to the intein in place of the N-extein, then the intein cleaves the larger protein at the correct site without relying on a protease or chemical to do the cleaving. Preventing the cyclisation of Asn454 by mutating it to an alanine (N454A) (Chong *et al.*, 1997), prevents the release of the intein from the C-terminal affinity tag, a feature used in the isolation of proteins using the IMPACT I (Intein Mediated Purification with an Affinity Chitin-binding Tag) system.

RbpD from *Anabaena* 7120 was expressed using the IMPACT I system from New England Biolabs (1997). This system makes use of the intein from the vacuolar ATPase subunit (VMA) of the yeast *Saccharomyces cerivisae*, and incorporates the use of a chitin-binding domain in the C-extein position as an affinity tag (Chong *et al.*, 1997; New England Biolabs, 1997).

1.6 SELEX

In 1990, three labs independently developed ways to simultaneously screen over 10^{15} different nucleic acid sequences in order to select for different functionalities (Klug and Famulok, 1994). "*In vitro* selection" was developed by G.F. Joyce, "*in vitro* evolution" was developed by J.W. Szostak and "SELEX" (Systematic Evolution of Ligands using EXponential enrichment) was developed by L. Gold.

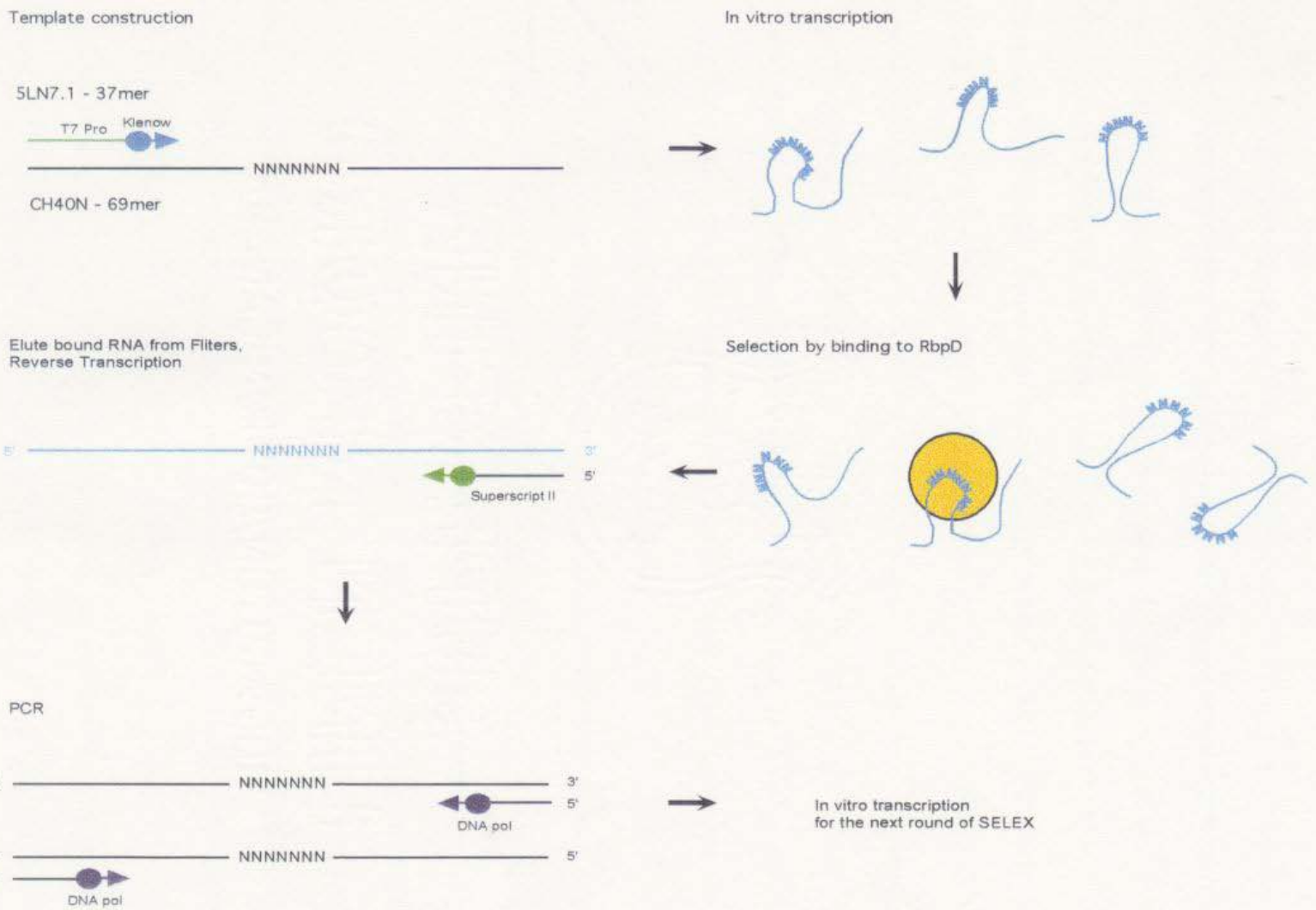
The theory behind the *in vitro* selection techniques is fairly straightforward. The first step of the process is to create a pool of oligonucleotides for selection, using a DNA synthesiser. The oligonucleotides generally contain a fully random sequence flanked by defined primer binding sites. The oligonucleotides and the molecule applying the

selection pressure are incubated together, at which point the active aptamers are isolated. The aptamers are subjected to reverse transcription-polymerase chain reaction (RT-PCR) in order to amplify those sequences that bound. Subsequent rounds of selection allow the section of increasingly specific RNA species (Gold *et al.*, 1993; Keene, 1996; Ouellette and Wright, 1995; Tuerk and Gold, 1990). The system as applied to RbpD is shown in Figure 1.5.

In vitro selection techniques can be used for a number of purposes. They can be used to find sequences of RNA that will bind to certain proteins, and from these sequences the secondary or tertiary structure that the RNA must adopt for binding can be determined. These methods have even been used to select for ribozymes such as RNA-ligase ribozymes and RNA-cleavage ribozymes (Lorsch and Szostak, 1996).

The techniques have found a potential diagnostic use in measuring the levels of therapeutic drugs (Allen *et al.*, 1996; Jenison *et al.*, 1994). Theophyllin is used in patients for the treatment of asthma, bronchitis and emphysema, but it is toxic at higher levels. The traditional way of detecting the drug has been with antibodies, but monoclonal antibodies had the problem of cross-reacting with caffeine and theobromine. *In vitro* selection isolated an RNA sequence that can bind to theophyllin with a 10-fold higher specificity than the antibody (Jenison *et al.*, 1994). As well, the RNA ligands were negatively selected for by running the theophylline aptamers over a column with caffeine attached to the matrix, thus removing the cross-reacting RNAs (Gold *et al.*, 1995). This type of detection might be useful for other drugs as well, or for substances which have little antigenicity (Gold *et al.*, 1993).

Figure 1.5: Schematic diagram illustrating the theory of SELEX. An oligonucleotide (5LN7.1) containing a T7 promoter (T7 Pro) is hybridized to the SELEX template which contains a number of random nucleotides in the middle portion. Klenow DNA polymerase is used to extend the 5LN7.1 oligonucleotide, creating a double stranded template. T7 RNA polymerase is then used to transcribe RNA from the template. The RNA is isolated and subjected to binding by RbpD in a nitrocellulose filter binding assay. Bound RNA is isolated from the filters, and reverse transcribed using SuperscriptII (Gibco BRL). PCR is used to amplify the cDNA containing the bound sequences using the 5LN7.1 and 3LN7.1, oligonuclotides as primers. The amplified fragments are subjected to *in vitro* transcription to initiate the next round of SELEX (Tuerk and Gold, 1990).


Template construction
5LN7.1 - 37mer
T7 Pro
Klenow
NNNNNNN
CH40N - 69mer
In vitro transcription
Selection by binding to RbpD
Elute bound RNA from Fliters,
Reverse Transcription
5'
NNNNNNN
3'
5'
Superscript II
PCR
5'
NNNNNNN
3'
5'
DNA pol
3'
NNNNNNN
5'
5'
DNA pol
In vitro transcription
for the next round of SELEX

In vitro selection has been useful for the generation of ribozymes. This method relies on the change that the ribozyme brings about to identify active aptamers. One example was the isolation of an ATP-dependent kinase ribozyme (Lorsch and Szostak, 1994). In this experiment, a sequence containing two constant regions of RNA, which made up an ATP-binding domain, and three random portions of RNA were used. The molecules also contained two primer-binding sites at the 5'- and 3'-ends. The RNA was incubated with ATP-γ-S and then with activated thiopropyl agarose. Those RNA sequences forming active kinase-ribozymes formed a dithio-linkage through the γ-thiophosphate to which they had attached themselves. Since the active ribozymes were thus attached to the column, the inactive RNA sequences could simply be washed away. β-mercaptoethanol was used to elute the active ribozymes.

One of the more recent advances in SELEX has been the use of genomic SELEX. This method creates the starting library from the organism of interest so that the best aptamers found will correspond to those sequences available to the protein *in vivo* (Gold *et al.*, 1995; 1997). The library consists of sequences derived from the genome (starting from every nucleotide of the genome), flanked by fixed regions to allow PCR amplification (Singer *et al.*, 1997). This adaptation of SELEX allows researchers to be more confident that the sequences they are isolating are recognised *in vivo* by the protein.

One of the problems with *in vitro* selection is that the length of random nucleotides that can be synthesised is limited. However, this problem has been addressed with the use of mutagenic PCR to introduce additional mutations. If manganese ions and different ratios of deoxynucleoside triphosphates are introduced, then Taq polymerase is

forced to mis-replicate the template DNA. In the course of the PCR steps, additional mutations are introduced so that over many rounds of *in vitro* selection, aptameric species appear and then disappear as they are replaced with more efficient RNA sequences (Klug and Famulok, 1994; Lorsch and Szostak, 1996; Szostak and Ellington, 1993).

A second limitation to *in vitro* selection methods is that some sequences are not reverse transcribed well because of high levels of secondary or tertiary structure in the RNA. This introduces some bias to the population of RNA molecules that are available for selection and binding. As a result, some sequences may be out-competed even though they bind the protein strongly. PCR also introduces some bias since some sequences do not replicate as fast as others under the conditions of the PCR. The selection conditions themselves also affect the results. Varying the salt concentration, the buffer composition and even the elution volume can alter the results of the selection (Klug, and Famulok, 1994). Finally, small amounts of ligand require elution of the bound RNA by denaturation. This can lead to artefacts being formed when RNA binds to the matrix, especially if nitrocellulose is used as the matrix. If this happens, then most of the species amplified for the next round will be specific for the filter rather than the molecule or protein that is being targeted (Klug and Famulok, 1994).

In vitro selection processes do have the advantage that they introduce the least bias of any method of selection known (Schneider *et al.*, 1993). A more important advantage to these methods is that the processes, with their multiple rounds of amplification, can be applied to almost any conventional purification scheme. The most common schemes adapted to using the selection and amplification are affinity

chromatography, filter binding, gel mobility shift and immunoprecipitation (Szostak and Ellington, 1993). This makes *in vitro* selection extremely flexible and convenient to apply to RNA-ligand problems. Most importantly though, these methods do not rely on any knowledge about the folding of the RNA. As the structures for more aptamers are solved, more complex structures, which form as a result of non-canonical base pairing, have been found. Secondary structures such as bulges, pseudo-knots and 1-3-2 stacks (stacking interactions between three adjacent nucleotides, where the bend is so sharp that the third nucleotide lies between the first two) have been found (Gold, *et al.*, 1997; Davis *et al.*, 1996).

1.7 Aims

The aim of this M.Sc. project was first to isolate and express both wild-type and His-tagged RbpD, and then start to characterise these proteins. The second goal of this study was to begin to characterise RbpD in order to determination its function in the *Anabaena* 7120 cell. Characterisation studies of the wild-type proteins involved the determination of a binding constant for RbpD, and determining an RNA sequence to which RbpD will bind. We also wanted to determine the effects of adding a tag to RbpD, and if possible, determine the *in vivo* and *in vitro* effects of removing the glycine rich region of RbpD. The examination of RbpD in *Anabaena* 7120 will add to the knowledge of the role of RNA binding proteins in cyanobacteria, and also to the knowledge of why different domains of the protein are required.

Chapter 2:

Materials and Methods

2.1 Materials

All chemicals were purchased from either Sigma Chemicals or Fisher Biotech and were reagent-grade or better. Media components were purchased from Difco. Restriction and modification enzymes were purchased from MBI Fermentas, New England Biolabs, Phamacia-Amersham, Promega or Gibco BRL. The enzymes were used according to the specifications of the manufacturer. Oligonucleotides were synthesised by Gibco BRL, Cortec (Queen's University, Kingston), or Operon Technologies (California).

2.2 Cloning and Culturing Methods

All plasmids used in this work are shown Table 2.1, and all strains of *Escherichia coli* used for cloning are shown in Table 2.2.

2.2.1 Media:

E. coli strains were grown in LB broth (10 g/l Bactotryptone; 5 g/l Yeast Extract; 10 g/l NaCl, pH 7.5). Plates were made by adding 15 g Bacto-Agar (Sambrook *et al.*, 1989). Antibiotics were used in the following concentrations: ampicillin, carbenicillin, and ticarcillin at 100 μg/ml, tetracycline at 15 μg/ml, kanamycin at 100 μg/ml, and chloramphenicol at 50 μg/ml.

Table 2.1 **Plasmids used in this work**

Plasmids	Description	Reference
pUC18	Ap^R	Yannisch-Perron *et al.*, 1985
pUC19	Ap^R, identical to pUC18; the multiple cloning site is reversed.	Yannisch-Perron *et al.*, 1985
pRLT4	pBR322 with 3.1 kb fragment containing *rbpA* and *CgnifH** from *Chlorogloeopsis* 6912	this work
pRLT5	pUC18 with 1.8 EcoRV fragment containing *CgnifH** cloned in the SmaI site	this work
pRLT6	pUC18 with a 1.6 EcoRI - HindIII with *CgnifH** from *Chlorogloeopsis* 6912.	this work
pAnR4.1	pUC18 with a 400 bp insert carrying the *rbpD* gene from *Anabaena* 7120.	Holden, 1995
pCYB1	Ap^R plasmid containing gene for chitin binding domain and intein gene in the same reference frame in which the gene of interest is to be inserted	New England Biolabs 1997
pRLT1	pCYB1 with 0.4 kb *rbpD* NdeI - SapI fragment from *Anabaena* 7120.	this work
pALTER-1	Tet^R Ap^S plasmid can be used to do mutagenesis.	Promega, 1996

Table 2.1 - continued

pCSS2	pUC18 with *rbpD1a* (*rbpD* with six His codons inserted at position 134) inserted EcoRI - HindIII	Slade, 1998
pRLT2	pALTER with *rbpD1a* inserted in the EcoRI - HindIII site.	this work
pRLT3	pTRC99A with *rbpD1* (*rbpD1a* which has been mutated to include an NcoI site) inserted in the NcoI - HindIII site.	this work
pTZ18R	Ap^R plasmid contains a T7 promoter allowing RNA to be transcribed in large quantities. Contains the pUC18 multiple cloning site.	Mead *et al.*, 1986
pRLT7	pTZ18R containing an insert corresponding to the 5'-untranslated region of the *rbpD* gene of *Anabaena* 7120.	this work
pTRC99A	Ap^R This plasmid contains a strong *trc* promoter, and is inducible by IPTG. Contains a *lacZ* RBS and $lacI^q$ repressor.	Pharmacia Amersham, 1994

Table 2.2 **Bacterial Strains Used in This Work**

E. coli Strain	Characteristics	References
MC1061	*hsdR, mcrB, araD139Δ(araABC-leu)7679, ΔlacX74, galU, galK, rpsL, thi*	Meissner *et al.*, 1987
DH2	*recA1, endA1, thi-1, hsdR17(m_k^-, m_k^+), supE44, rel41*	Philippe Bernard
ES1301	*lacZ53, mutS201::Tn5, thyA36, rha-5, metB1, deoC, IN (rrnD - rrnE)*	Siegal *et al.*, 1982
JM109	*endA1, recA1, gyrA96, thi, hsdR17, (m_k^-, m_k^+), relA1, supE44, Δ(lac-proAB), [F', traD36, proA$^+$B$^+$, lacI^{q}ZΔM15]*	Yannisch-Peron, *et al.*, 1985
DH5α	*supE44, ΔlacU169 (ϕ80lacZΔM15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1*	Sambrook *et al.*, 1989
LE392	*e14$^-$(mcrA$^-$), hsdR514, supE44, supF58,lacY1 or Δ(lacIZY)6, galK2, galT22, metB1, trpR55*	Sambrook *et al.*, 1989
HB101	*supE44, ara14, galK2, lacY1 Δ(gpt-proA)62, rpsL20(Str^r), xyl-5, mtl-1, recA13,* Δ(mcrC-mrr) HsdS$^-$(r^-m^-)	Sambrook *et al.*, 1989
XL1Blue	*recA1, endA1, gyrA86, thi-1, hsdR17, supE44, relA1, lac[F' proAB, lacI^{q}ZΔM15, Tn10(tet^R)]*	Bullock *et al.*, 1987

Table 2.2 - continued

B410	*lacI^q, recA1, endA1, thi-1, hsdR17(r_k^-, m_k^-), supE44, relA1*	Philippe Bernard
BL21(DE3)pLysS	F-, *dcm, ompT, hsdS(r_B^-, m_B^-), gal λ(DE3) [pLysS, Cam^r]*	Studier *et al.*, 1990 Weiner *et al.*, 1994
ER2566	F-, κ-, *fhuA2[lon], ompT, lacZ::T7 gene1, gal, sulA11, Δ(mcrC-mrr) 114::IS10R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS), endA1 [dcm]*	New England Biolabs, 1999

2.2.2 Competent Cells:

All competent cells, except for *E. coli* strains ES1301 and JM109, were made from 500 ml cultures grown to an optical density (600 nm) of 0.8-1.0. The cells were then pelletted at 4000 rpm for 10 minutes and resuspended in 250 ml of 0.1 M $CaCl_2$. After a thirty minute incubation on ice, the cells were again pelletted for 10 minutes at 4000 rpm and resuspended in 5 ml 0.1 M $CaCl_2$/15% glycerol. The cells were portioned out in 50 µl fractions and frozen at -70°C until required.

E. coli ES1301 and JM109 competent cells were prepared by a modified RbCl method as described in the Altered Sites II manual (Promega, 1996). Cells were grown in 250 ml cultures of LB media to an optical density (600 nm) of 0.4-0.6, and the cells pelletted by centrifugation for five minutes at 4500 x g. The pellet was then resuspended in 100 ml TFB1 (30 mM potassium acetate; 10 mM $CaCl_2$; 50 mM $MnCl_2$; 100 mM RbCl; 15% glycerol, pH 5.8). The suspension was incubated on ice for five minutes and then pelletted as above. Ten ml of TFB2 (10 mM MOPS, pH 6.5; 75 mM $CaCl_2$; 10 mM RbCl; 15% glycerol, pH 6.5) was used to resuspend the pellet. 100 µl portions were frozen in liquid nitrogen and stored at -70°C.

2.2.3 Transformations:

Transformation of most ampicillin resistant plasmids was carried out by the five minute method (Pope and Kent, 1996). Fifty µl of the competent cells were diluted with 450 µl 0.1 M $CaCl_2$ (except for ES1301 and JM109 competent cells which were used directly). Eighty µl of the diluted cell suspension were then mixed with an appropriate

amount of plasmid (depending on the plasmid concentration) and 0.1 M $CaCl_2$ to bring the volume to 100 µl. The solution was placed on ice for five minutes then plated on an agar plate containing the appropriate drug.

If plasmids created through a ligation reaction were being transformed, or if the plasmid contained tetracycline or kanamycin resistance, then a longer transformation procedure was used (Sambrook *et al.*, 1989). The 100 µl of plasmid and cells was set up the same way as in the five minute method and incubated on ice for 30 minutes, then heat shocked at 42°C for 45 seconds. One ml of LB was then added to the cells and the mixture left at 37°C for 30 minutes. The cells were pelletted in a microfuge for one minute and resuspended in 100 µl LB for plating on a plate with the appropriate drug.

2.2.4 Isolation of Plasmid DNA

Small scale isolation of plasmid DNA was carried out using Iyer's (1994) Merlin MiniPrep method. Cells were grown overnight in a 5 ml LB culture and pelletted by centrifugation. The pellets were then resuspended in 200 µl Cell Resuspension Solution (50 mM Tris-HCl, pH 7.5; 10 mM EDTA; 100 µg/ml RNaseA). The resuspended cells were lysed using Cell Lysis Solution (0.2 M NaOH; 1% SDS). Then the proteins and genomic DNA were precipitated using 200 µl Neutralisation Solution (1.25 M potassium acetate; 1.24 M glacial acetic acid). After the mixture was centrifuged for five minutes, the supernatant was transferred to a new tube containing 1 ml of DNA-binding Resin (Celite resin; 7 M guanidine hydrochloride). The slurry was passed through a Promega mini column and the column then washed with Column Wash Solution (200 mM NaCl;

20 mM Tris-HCl, pH 7.5; 5 mM EDTA; 50% ethanol). The purified plasmid DNA was eluted into a fresh microfuge tube using 60 µl of warm (60°C) TE. Recovery of the plasmid was estimated by restriction digestion and gel electrophoresis.

2.2.5 Cloning of DNA

DNA fragments were cloned after purifying the desired fragment from an agarose gel slice (Heery, 1990). The slice was placed in a punctured 0.5 ml microfuge tube containing a small piece of aquarium filter floss (Levine, 1994) and centrifuged into a 1.5 ml microfuge tube for ten minutes at 6000 rpm. The bottom tube collected 100-200 µl of eluate which was washed with an equal volume of 1:1 phenol:chloroform. The purified fragment of DNA was removed with the aqueous layer, ethanol precipitated and resuspended in TE (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, pH 8.0) (Sambrook *et al.*, 1989).

Ligation reactions were carried out in a 10 µl volume using approximately equal (as determined by gel electrophoresis) amounts of DNA from plasmid digested with the appropriate enzyme(s) and the purified DNA fragment. One µl T4 DNA Ligase (Pharmacia) was added as well as 1 µl of One-Phor-All ligation buffer (Pharmacia). The final volume was reached by adding TE. Ligations were incubated at least six hours at room temperature to ensure complete ligation (Bolivar and Backman, 1979). Transformation was done as described in Section 2.2.3.

2.3 Sequencing of DNA

The Sanger *et al.* (1977) method with the modifications introduced by United States Biolabs (USB) (Sequenase 2.0) was used to sequence DNA cloned in plasmids (Kraft, *et al.*, 1988; Wang, 1988). If the fragment to be sequenced was cloned in a pUC plasmid, then the –40 primer (5' - GTT TTC CCA GTC ACG AC – 3') (Table 2.3) and reverse primer (5' - AAC AGC TAT GAC CAT G – 3') were used to initiate the sequencing. If the fragment was cloned in a pCYB vector, the primers provided by New England Biolabs were utilised (P_{tac} forward primer: 5'- GAG CGG ATA AGA ATT TCA CAC AGG – 3' and the intein reverse primer: 5'- ACC CAT GAC CTT ATT AGG AAC CTC – 3').

Sequencing reactions were carried out by first denaturing the plasmid. Fifty µl miniprep DNA was incubated with 5 µl 2 M NaOH; 2 mM EDTA for five minutes at 37°C. The sample was placed on ice with 17.5 µl water and 17.5 µl 3 M sodium acetate, pH 5.2 to neutralise the solution. The now single-stranded DNA was precipitated with ethanol, and resuspended in 20 µl TE.

In order to anneal the primers to the template, 7 µl of template DNA solution, 2 µl reaction buffer (USB), and 1 µl of the primer were boiled and allowed to cool slowly to room temperature in a beaker of water. Once cooled, 1 µl DTT, 2 µl labelling mixture (USB), 1µl [α-^{35}S] ATP (1250 Ci/mmol) and 2 µl diluted (1:6) Sequenase (USB) were added to the tube. Portions of 3.5 µl were removed from this mix and added to 2.5 µl portions of each of the ddNTPs. The reactions were incubated at 37°C for five minutes, and then 4 µl stop solution was added.

Table 2.3 Oligonucleotides synthesised in this work

Oligonucleotide (company)	Sequence (5' to 3')	Purpose
- 40 primer (USB)	GTTTTCCCAGTCACGAC	sequencing in pUC plasmids
Reverse primer (USB)	AACAGCTATGACCATG	sequencing in pUC plasmids
P_{tac} forward primer (New England Biolabs)	GAGCGGATAAGAATTTCACACAGG	sequencing of pCYB plasmids
Intein reverse primer (New England Biolabs)	ACCCATGACCTTATTAGGAACCTC	sequencing of pCYB plasmids
RbpD(NdeI) (Gibco BRL)	GGAGATCATATGTGTCAA	introduction of NdeI site to *rbpD*
RbpD(SapI) (Gibco BRL)	TTTTTTGCTCTTCTGCAGTAACTAC-GGCGATTACC	introduction of SapI site and changing of stop codon to Cys codon in *rbpD*
RbpD1(NcoI) (Gibco BRL)	GATTCGGAGACTTGTATGTCAATT-TACAT	mutagenesis
RbpDloop (Cortec)	TCTCCGAATTTAACTCTCTACATT-TCCTGAATTCGGAGA	electrophoretic mobility shift assay
RbpDloopR1 (Cortec)	AATTCTCTCCGAATTTAACTCTC-TACATTTCCTGAATTCGGAGAC	transcription of the 5'-UTR of *rbpD*
RbpDloopS1 (Cortec)	TCGCGTCTCCGAATTCAGGAAAT-GTAGAGAGTTAAATTCGGAGAG	transcription of the 5'-UTR of *rbpD*
5LN7.1 (Operon Technologies)	GAAATTAATACGACTCACTATAG-GGAGGACGATGCGG	SELEX experiments
CH40N (Operon Technologies)	TCCCGCTCGTCGTCTG(N_{40})CCGC-ATCGTCCTC	SELEX experiments
3LN7.1 (Operon Technologies)	TCCCGCTCGTCGTCTG	SELEX experiments

The reaction products were separated on a preheated 6% acrylamide gel run at 35 W. After the gels were run, they were fixed in 10% methanol, 10% acetic acid for thirty minutes, then dried onto Whatman No. 3. The gel was then exposed to Kodak XAR5 film for at least one day at room temperature.

2.4 Detection of Proteins

Expressed proteins were separated, unless otherwise stated, using PhastSystem gels (Pharmacia, 1986a,b). Gels consisted of an 8-25% acrylamide gradient or 20% homogeneous separating gel with a 3.5% stacking gel. The proteins were suspended in an equal volume of SDS-loading buffer (187.5 mM Tris-HCl, pH 6.8; 6% SDS; 30% glycerol; 0.03% bromophenol blue) for electrophoresis or in cracking buffer (65 mM Tris-HCl, pH 6.8; 13% glycerol; 1.3% SDS; 0.02% sodium azide) if a cell extract was being electrophoresed. Coomassie staining was done using a PhastSystem dye formulation (Pharmacia, 1986c). The stock stain consisted of one tablet of PhastGel Blue R dissolved in 80 ml H_2O, and 120 ml methanol. The stain was made by mixing one part stock stain solution with one part 20% methanol in water. Gels were stained for at least twenty minutes in a petri dish with gentle shaking and then destained in 30% methanol; 10% acetic acid for approximately one hour. The gel was left in preserving solution (5% glycerol; 10% acetic acid) for at least twenty minutes and allowed to dry before scanning.

Silver staining was performed using the Plus-One silver staining system from Pharmacia-Amersham. The gel was fixed in 40% ethanol, 10% acetic acid for 30 minutes,

then sensitised for 30 minutes in 30% ethanol, 0.125% glutaraldehyde, 0.2% sodium thiosulphate, 6.8% sodium acetate. The gel was then washed three times with water and exposed to the silver reaction solution (0.25% silver nitrate; 0.015% formaldehyde) until the gel turned slightly yellow. Water was used to wash off the excess silver reaction solution. The colour was developed using a solution of 2.5% sodium carbonate; 0.007% formaldehyde until the protein bands were distinct. The reaction was stopped with the addition of 7.3% EDTA for 10 minutes. The gel was preserved in 10% glycerol. All percentages except the ethanol and glycerol are in (w/v).

2.5 Expression of RbpD

2.5.1 Cloning of *rbpD* into pCYB1

The IMPACT I system was chosen for expression and purification of the RbpD protein. The 374 base-pair *rbpD* gene fragment was present on the plasmid AnR4.1 (See Table 2.1). Oligonucleotides were designed to mutate the gene for insertion into the plasmid pCYB1 using PCR. The first oligonucleotide (5' – GGA GAT CAT ATG TGT CAA – 3') (Table 2.3) introduced an NdeI site at the start codon of the *rbpD* gene. The second oligonucleotide (5' – TTT TTT GCT CTT CTG CAG TAA CTA CGG CGA TTA CC – 3') changed the stop codon in the gene to a cysteine codon and also introduced a SapI site to the gene.

PCR was utilised to amplify the *rbpD* gene fragment. The amplified fragment was digested with NdeI and SapI and ligated into pCYB1, which had been digested with the same restriction endonucleases. The resulting plasmid, pRLT1, contained the *rbpD*

gene in the same reference frame as the gene for the vacuolar ATPase subunit from *Saccharomyces cerevisiae* (VMA) intein.

2.5.2 Expression and Purification of RbpD

The protocol for the expression and purification of RbpD is shown in Figure 2.1 and follows the IMPACT I protocol from New England Biolabs, with a number of modifications. *E. coli* MC1061 was transformed with pRLT1. A single colony from the transformed cells was inoculated into a 5 ml overnight LB culture. This overnight culture was used to inoculate a larger culture of between 500 ml and 3 l, which was grown to an optical density (600 nm) of 0.5-0.7. The culture was induced with 1 mM IPTG for 4-8 hours at 37 °C.

The culture was then centrifuged to pellet the cells, and the supernatant decanted. Cells were resuspended in 10-60 ml of column buffer (20 mM Tris-HCl, pH 8.0; 50-500 mM NaCl; 0.1 mM EDTA; 0.1% Triton-X 100) depending on the size of the culture. If the resuspension volume was less than 40 ml, the cells were then sonicated. If the resuspension volume was larger than 40 ml, then the cells were passed through a french press at 12 000 psi to disrupt the membranes. The crude extract was centrifuged for 15 to 30 minutes to clarify the extract and remove cell debris. The clarified extract was loaded slowly onto a chitin column. The column was washed with Cleavage Buffer (20 mM Tris-HCl, pH 8.0; 50-500 mM NaCl; 0.1 mM EDTA; 30-50 mM DTT) and the column was allowed to sit overnight at 4°C with no flow.

Figure 2.1: Protocol for expression and purification of RbpD. The steps of purification are shown on the left side of the figure; buffers and other parameters are shown on the right side of the figure. The components of the buffer that were altered from the manufacturers details are shown in red. These factors were manipulated extensively in the course of trying to purify RbpD.

5 ml overnight culture

↓

Scale up to 500 ml and grow
to O.D. of 0.5-0.7
Induce with 1 mM IPTG for 4-8 hr — 15, 30 or 37°C

↓

Spin down cells, 5000 rpm, 5 min
Resuspend in 10 mL column buffer — 20 mM Tris, pH 8.0
50-500 mM NaCl
0.1 mM EDTA
0.1% Triton-X 100

↓

Sonicate
Freeze thaw
Clarify by centrifuge, 12000 rpm, 15 min

↓

Load on chitin column, 0.5 mL/min — Use 0.5-3 mL column
Wash with 10 column volumes buffer

↓

Pass 3 column volumes cleavage buffer through column quickly — 20 mM Tris pH 8.0
50-500 mM NaCl
0.1 mM EDTA
30-50 mM DTT

↓

Allow to sit overnight in order to
cleave

↓

Wash cleaved protein off column — 20 mM Tris pH 8.0
50-500 mM NaCl
0.1 mM EDTA
0-0.5% Triton-X 100

↓

Strip column — 20 mM NaHepes pH 8.0
500 mM NaCl
1% SDS

RbpD was eluted the following day using Cleavage Buffer without DTT. The column was stripped of the residual intein and chitin-binding domain using Stripping Buffer (20 mM NaHEPES, pH 8.0; 500 mM NaCl; 1% SDS).

2.5.3 Western Blotting of Fusion Protein Expressed From pRLT1

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) was performed using a 15% acrylamide separating gel (Separating Buffer: 9.1% Tris-HCl, pH 6.8; 0.2% SDS (all w/v)) with 4% stacking gel (Stacking Buffer: 1.6% Tris-HCl, pH 6.8; 0.1% SDS; 0.01% sodium azide (all w/v)) on a BioRad minigel system. Duplicate gels were loaded and run at a constant voltage of 150V until the bromophenol blue dye in the loading buffer reached the end of the gel. One gel was stained (10% glacial acetic acid (v/v); 50% methanol (v/v); 0.25% Coomassie Brilliant Blue (w/v)) to visualise all protein present. The second gel was subjected to Western blotting as follows.

The gel was loaded onto a transfer apparatus, submerged in Transfer Buffer (10 mM CAPS; 10% methanol), and electrophoresed at 60 V for at least two hours at 4 °C to transfer all protein to a PVDF membrane. Complete transfer was checked by staining the gel in Coomassie Brilliant Blue to ensure that no protein was still present in the gel. The membrane was immersed in 3% skim milk powder (Carnation) for 60 minutes, and then washed in TTBS buffer (20 mM Tris-HCl, pH 7.5; 0.5 M NaCl; 0.05% Tween 20) twice for five minutes. The membrane was incubated with an antibody (1:2000 v/v) to the intein portion of the expressed protein (provided by New England Biolabs) for 60 minutes. After washing the membrane again with TTBS (five minutes), the membrane

was exposed to an anti-rabbit antibody (1:3000 v/v) conjugated to alkaline phosphatase for an additional 30 minutes. The membrane was again washed with TTBS. The antibodies were detected using 0.15 mg/ml 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and 0.30 mg/ml nitrobluetetrazolium (NBT) in 0.1 M $NaCO_3$, 1 mM $MgCl_2$, pH 9.8.

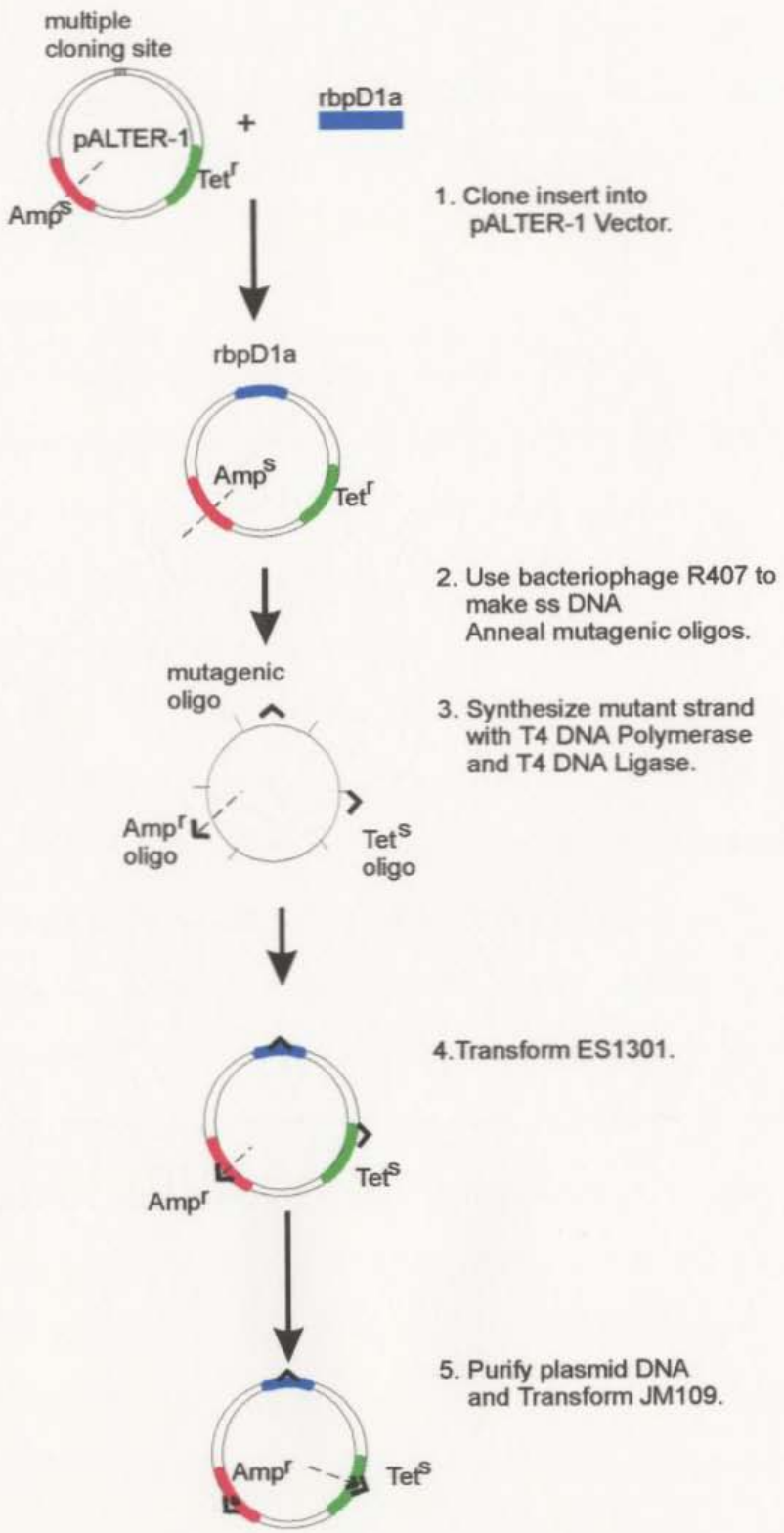
2.6 Expression of RbpD1

2.6.1 Site directed mutagenesis of *rbpD1a*

Cynthia Slade (1998) provided the plasmid pCSS2 which contains the *rbpD1a* gene. The *rbpD1a* gene was digested from pCSS2 using EcoRI and HindIII, cloned into a pALTER-1 plasmid (Promega) and called pRLT2. The plasmid pRLT2 was transformed into *E. coli* strain JM109 and grown on LB plates containing 10 μg/ml tetracycline. Site directed mutagenesis was performed as described in Figure 2.2.

A single colony was used to inoculate a 200 ml LB culture, and the culture grown for 1 hour. The bacteria were then infected with 4×10^8 bacteriophage R407 particles. At an incubation temperature of 37°C, the bacteriophage infect JM109 cells and produce single stranded DNA from both the bacteriophage and host chromosomes. Infection was allowed to occur for seven hours. The JM109 cells were removed from the culture by centrifuging the culture twice for 15 minutes at 12 000 rpm. The resulting supernatant was incubated in a 1:1 ratio with Phage Precipitation Solution (20% polyethylene glycol 8000; 3.75 M ammonium acetate, pH 7.5) at 4°C overnight in order to precipitate the bacteriophage. The bacteriophage was pelletted by centrifugation at 12 000 rpm for 15 minutes at 4°C, and the pellet resuspended in 400 μl TE. The bacteriophage were lysed to

Figure 2.2: Schematic diagram of the Altered Sites II mutagenesis procedure. The gene of interest (in this case *rbpD1a*) was cloned into the multiple cloning site of the pALTER-1 vector. Bacteriophage R407 was used to make ssDNA from the template by infecting the cells with bacteriophage and incubating for seven hours at 37°C with vigorous shaking. The bacteriophage derived ssDNA was isolated from the culture, and mutagenic oligonucleotides were annealed to it. Second strand synthesis was primed from the mutagenic oligonucleotides and completed using T4 DNA polymerase and T4 ligase. The dsDNA plasmids were transformed into the *E. coli* strain ES1301. After growing the transformed ES1301 overnight in a 5 mL culture, the mutated plasmid was isolated and transformed into *E. coli* strain JM109.


multiple
cloning site
pALTER-1
Tet^r
Amp^s
+
rbpD1a
1. Clone insert into
pALTER-1 Vector.
rbpD1a
Amp^s
Tet^r
2. Use bacteriophage R407 to
make ss DNA
Anneal mutagenic oligos.
mutagenic
oligo
3. Synthesize mutant strand
with T4 DNA Polymerase
and T4 DNA Ligase.
Amp^r
oligo
Tet^s
oligo
4.Transform ES1301.
Amp^r
Tet^s
5. Purify plasmid DNA
and Transform JM109.
Amp^r
Tet^s

release single stranded DNA using a solution of 24:1 chloroform:isoamyl alcohol. After centrifugation, the ssDNA remained in the aqueous layer. Bacteriophage debris was removed with successive washings of 25:24:1 phenol:chloroform:isoamyl alcohol; the aqueous layer was retained after each washing. Purified ssDNA was finally isolated by ethanol precipitation.

Mutagenic oligonucleotides were annealed to the ssDNA. The first oligonucleotide annealed to the complement of the *rbpD1a gene* (5' – GAT TCG GAG ACT TGT ATG TCA ATT TAC AT – 3') (Table 2.3), mutating the nucleotides around the start codon into an NcoI site. The second oligonucleotide, provided by Promega, repaired a gene conferring ampicillin resistance; the third knocked out the tetracycline resistance gene. Second strand synthesis was primed from the mutagenic oligonucleotides and extended using T4 DNA polymerase (5 units, 3 to 6 hours, 37°C) and T4 ligase (2 units, overnight, room temperature) to ligate the fragments. The dsDNA plasmids were transformed into the *E. coli* strain ES1301, which is *dut*$^-$ *ung*$^-$ (Kunkel *et al.*, 1991). The *dut*$^-$ mutants lack dUTP pyrophosphatase and therefore contain a high concentration of dUTP which competes with dTTP for incorporation into DNA. Mutants with the *ung*$^-$ genotype lack uracil N-glycosylase, and thus will not remove the incorporated dUTP residues. The template DNA contains many uracil residues. The mutagenic template can then be annealed to this vector, and extended using dTTP rather than dUTP. When the resulting double-stranded plasmid is transformed into a *dut*$^-$ *ung*$^-$ strain, the strand which lacks the mutation will be lost due to the uracils and the mutated strand will be used as the template for making double stranded DNA (Kunkel *et al.*,

1991). The transformed ES1301 were grown overnight in a 5 ml culture, and the plasmid DNA isolated as described in Section 2.2.4. The mutated plasmid was isolated and transformed into *E. coli* strain JM109 for antibiotic selection (Tet^S, Ap^R), and long-term storage.

2.6.2 Expression of RbpD1

After introducing the NcoI site into the *rbpD* gene, *rbpD1* was digested from the plasmid using NcoI and HindIII, purified and then ligated into pTRC99A. The resulting plasmid, designated pRLT3, was transformed into *E. coli* BL21(LysS)DE3 for expression of RbpD1.

Cells containing the plasmid were grown in a 5 ml overnight culture. This was used to inoculate a 500 ml culture, which was grown to an optical density (600 nm) of 0.5-0.7. Because expression from the *trc* promoter in pTRC99A is under the control of the $lacI^Q$ mutation, 1mM IPTG was used to induce expression for 6-8 hours; and, the cells were then centrifuged at 5000 rpm for five minutes. The cells were resuspended in 10 ml of MCAC buffer (50 mM NaH_2PO_4, pH 8.0; 0.5M NaCl; 1% Tween 20; 1 mM PMSF; 10% glycerol), sonicated (200 V; four 1 minute intervals), and then exposed to 1 mM DNaseI for 30 minutes to reduce the viscosity. After removing the cell debris from the extract by further centrifugation (12 000 rpm for 30 minutes), the extract was passed through 0.5 ml of Ni-NTA agarose (Qiagen). Non-specific proteins were washed from the column using 20 mM imidazole in MCAC buffer; RbpD1 was then eluted from the column using buffer containing 100 mM imidazole.

2.7 Storage of Proteins

Various methods of storing the RbpD and RbpD1 were examined. The proteins were initially stored at -70°C in the elution buffer, but lost activity after four or five freeze-thaw cycles. Freeze-drying of RbpD1 caused loss of polymer-binding specificity (Figure 4.1). In order to obtain an acceptable life-span for the protein, RbpD was concentrated to 1 mg/ml; glycerol was then added to a concentration of 50%, and the protein stored at -20°C.

2.8 Binding Experiments

2.8.1 Polymer Binding

RbpD and RbpD1 were bound to Sepharose-4B-bound poly(A), agarose-bound poly(C), or polyacrylhydrazido-agarose-bound poly(G) or poly(U). (For convenience, these will be collectively refered to as "agarose-bound" polymers.) The experiments used the method developed by Sugita and Sugiura (1994). In each binding experiment, 2 μg of protein was used, with 20 μg of agarose-bound polymer. The protein was incubated with each polymer in 1 ml of binding buffer (10 mM Tris-HCl, pH 7.5; 2.5 mM $MgCl_2$; 0.5% Triton-X 100; 1 mM PMSF; 0.2 M NaCl) for 10 minutes at 4°C. The reaction tubes were turned on a rotator to ensure that the protein and polymer continued to stay in contact. The binding reaction mixture was loaded into a 3 ml syringe with a Wizard minicolumn (Promega) attached, so that the unbound protein would pass through the column but the polymer-bound protein would be retained. The polymers were then washed sequentially

with binding buffer plus 2 mM heparin, binding buffer, and then water. The protein was then eluted from the miniprep column using SDS loading buffer. The protein that had bound to the column was visualised on a silver stained SDS polyacrylamide gel (Section 2.4).

2.8.2 Binding to 5'-Untranslated Region DNA

An oligonucleotide was synthesised (5' - TCT CCG AAT TTA ACT CTC TAC ATT TCC TGA ATT CGG AGA – 3') (Table 2.3) which included the conserved sequence of the 5'-untranslated region (5'-UTR) of the *rbpD* transcript. The oligonucleotide was end-labelled with [γ-^{32}P] ATP (3000 Ci/mmol) and subjected to an electrophoretic gel shift mobility assay (EMSA) with nanomolar amounts of RbpD (0, 0.5, 10, 50, 100 and 1000 nM). Binding reactions were incubated at 25°C for 20 minutes in TENT buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA; 250 mM NaCl; 0.5%Triton-X 100) (Laird-Offringa and Belasco, 1996) and electrophoresed on an 8% non-denaturing polyacrylamide gel using a Tris-glycine (25 mM Tris-HCl; 0.19 M glycine; 1 mM EDTA, pH 8.3) buffer (Buratowski and Chodosh, 1999) non-denaturing polyacrylamide gel. Labelled DNA was used in amounts from between 0.012 fmol to 0.12 fmol. The larger amount was generally used because it was easier to visualise on the gel. Gels used in EMSA were 1.5 mm thick; wells were made with a comb that had teeth which were at least 0.5 cm wide to attain the best resolution (Konarska, 1989). Autoradiographic film (Kodak XAR5) was used to detect free or bound DNA.

2.8.3 Binding to 5'-Untranslated Region RNA

Two complementary oligonucleotides were synthesised, corresponding to both halves of the 5'-UTR of the *rbpD* gene (5' – AAT TCT CTC CGA ATT TAA CTC TCT ACA TTT CCT GAA TTC GGA GAC – 3', and 5' –TCG CGT CTC CGA ATT CAG GAA ATG TAG AGA GTT AAA TTC GGA GAG – 3') (Table 2.3). The oligonucleotides also contained sequences for EcoRI and SalI sticky ends. Fifty pmol each of the two oligonucleotides were annealed together by boiling them in a reaction tube and allowing the tube to cool slowly in the water bath until it reached room temperature. The dsDNA fragment was cloned into the plasmid pTZ18R, digested with SalI and EcoRI, to form pRLT7. The Ambion T7 Megashortscript system (Ambion, 1997) was used to transcribe the 5'-UTR of *rbpD*. Transcription was performed after linearisation of pRLT7 with HindIII; this was expected to produce a 50 nucleotide transcript. The transcription reaction generally contained 7.5 mM nucleoside triphosphates, transcription buffer (provided by Ambion), 10 to 20 μCi [α-^{32}P] ATP (3000 Ci/mmol), and 1.5 – 2 μg linearised plasmid (as much as could be isolated and purified from a 5 ml culture of *E. coli* MC1061 cells). The transcription reaction was incubated overnight at 37°C (Ambion, 1997). The transcribed RNA was purified by electrophoresing the RNA on an 8% acrylamide, 8 M urea denaturing gel. Control transcription reactions containing 1 μg of an 80 base-pair cDNA fragment from the 18S rRNA inserted downstream of a T7 promoter (provided by Ambion) were also performed.

The transcribed RNA was isolated by one of two different methods. The first was by the method of Laird-Offringa and Belasco (1996). The gel was exposed to film to locate the bands of radioactivity that indicated the transcribed RNA. The corresponding gel slices were placed in an microfuge tube with 300 µl TE plus 0.2% SDS; slices were ground into a slurry. The tubes containing the slurry were incubated overnight at 4°C with constant rotation and then the slurry was pressed through a 3 ml syringe plugged with aquarium filter floss, and the filtrate eluted. The SDS was removed by extraction with phenol:chloroform (1:1 v/v) and then with chloroform. The RNA was precipitated by the addition of 0.1 volumes of 3 M sodium acetate, pH 5.3, and two volumes of ethanol. After centrifuging the RNA and washing the pellet with 80% ethanol, the RNA was resuspended in an appropriate amount of RNase-free water.

The second method used to isolate the RNA was similar except that the gel slice was first frozen at -70°C, then 200 µl 2 mM EDTA was added and the slice was crushed with a small pestle. One hundred µl of 3 M sodium acetate was added, the slurry was vortexed and pressed through the 3 ml syringe plugged with filter floss as above (Binkley and Tuerk, personal notes). The second method had the advantage of saved time, but did not give the same yield as the method of Laird-Offringa and Belasco (1996).

Electrophoretic mobility shift assays were performed under the same conditions as used in the DNA-5'-UTR binding experiments (Section 2.8.2), except that the RNA was boiled for two minutes prior to the binding experiment in order to remove any secondary structure. In some experiments, TENT buffer was used instead of the buffer that Sugita and Sugiura (1994) successfully used with the polymer binding experiments.

Filter binding experiments were carried out using the procedure as described by Schneider *et al.* (1993). As with the electrophoretic mobility shift assays, the components in the filter binding experiments were incubated at 20°C for twenty minutes, but the amount of RbpD protein varied between 0, 0.5, 1, 5, 10 and 20 μM, and the amount of RNA was increased to about 2500 pmol in order to be able to detect the radioactivity on the filters. Sugita and Sugiura's (1994) buffer was used in the incubation. The incubation mixtures were suctioned through nitrocellulose filters (0.45 μm, Millipore) using a Millipore Vacuum Manifold apparatus.

2.8.4 Binding to Poly(U)

Poly(U) (Sigma) was end-labelled by incubating 30 μCi [γ-^{32}P] ATP (3000 Ci/mmol) with 5 units polynucleotide kinase (PNK), 1x PNK buffer (Pharmacia-Amersham) and 10 pmol poly(U) for 3 hours at 37°C. The labelled poly(U) was separated on an 8 M urea, 8% polyacrylamide gel using Tris-glycine buffer as in 2.8.2. Because the polymer was heterogeneous in length, it appeared as a smear on the autoradiograph. In order to perform a gel shift experiment with the polymer, a slice corresponding to a relatively homogeneous size range of RNA was cut from the upper part of the gel. The labelled RNA was isolated from the gel slice and purified in the same manner as mentioned in Section 2.8.3. For most of the experiments, the size of the purified polymer was approximately 92 nucleotides

Purified poly(U) was subjected to binding by RbpD using the same conditions as mentioned in Section 2.8.2. The binding was done at either 4°C or 25°C for 10 minutes,

using Sugita and Sugiura's (1994) binding buffer, and the reactions run on 6-8% non-denaturing gel. Either 1x TBE (10 mM Tris-HCl, pH 8.3; 10 mM boric acid; 2 mM EDTA) or 1x Tris-Glycine buffer was used in the running buffer and gel in order to determine the optimum ionic conditions to enhance binding. Glycerol (2.5% v/v) was also included in the gels to stabilise protein-RNA interactions during the running of the gel (Buratowski and Chodosh, 1999).

2.9 SELEX

Three oligonucleotides were synthesised for the SELEX protocol (Tuerk and Gold, personal communication). The first, 5LN7.1 (5' - GAA ATT AAT ACG ACT CAC TAT AGG GAG GAC GAT GCG G – 3'), contained a T7 promoter and hybridised to the SELEX template. The second was the SELEX template, designated CH40N (5' – TCC CGC TCG TCG TCT G (N_{40}) CCG CAT CGT CCT C – 3') (Table 2.3). It contained forty random nucleotides in the centre and had two constant ends, making a 69 nucleotide fragment. The third oligonucleotide, 3LN7.1 (5' – TCC CGC TCG TCG TCT G – 3'), had the same constant end as CH40N on the opposite end of the random nucleotides as the complement to the T7 promoter.

5LN7.1 (3 nmol) was annealed to 3 nmol CH40N in 10x annealing buffer (100 mM Tris-HCl, pH 8.0; 100 mM $MgCl_2$) in a 90 µl reaction by heating the mixture to 95°C for five minutes then cooling to 37°C. The annealed oligonucleotides were then extended using Klenow DNA polymerase (45 units), 1mM dNTPs and Klenow buffer (Pharmacia-Amersham) in a total reaction volume of 1 ml. The mixture was incubated at

37°C for 5 hours or overnight. This created a double-stranded template 92 nucleotides in length, which was purified by passing the reaction mixture through a Centricon 30 (Amicon). The double-stranded template remaining in the retentate was ethanol precipitated, and the resulting pellet was resuspended in 100 µl TE.

For each round of SELEX, RNA was transcribed separately in radioactively labelled "hot" and unlabelled "cold" reactions. The "cold" reaction contained RNA polymerase buffer (20% polyethylene glycerol; 200 mM Tris-HCl, pH 8.0; 60 mM $MgCl_2$; 25 mM DTT; 5 mM spermidine; 0.01% Triton-X 100), 2 mM each nucleoside-triphosphate, 2.5 units pyrophosphatase, RNAsin (Promega), 250 pmol template DNA and 600 units T7 RNA polymerase in a 500 µl reaction. The "hot" reaction contained the RNA polymerase buffer, 0.05 mM ATP, 0.5 mM CTP/GTP/UTP, 50 µCi [α-^{32}P] ATP (3000 Ci/mmol), 0.25 units pyrophosphatase, RNasin (10 units), 30 pmol DNA template and 120 units T7 RNA polymerase in a 30 µl reaction. The transcription reactions were allowed to incubated overnight at 37°C.

Filter binding using nitrocellulose filters (0.45 µm, Millipore) and the Millipore Vacuum Manifold apparatus was performed as described in Section 2.8.3. The reactions included 0.1, 0.5, 1, 5, or 10 µM (0, 1.88, 9.38, 18.75, 93.75 or 187.5 pmol) RbpD and 2x Sugita and Sugiura's buffer (1994) in a 75 µl volume. Seventy-five µl of the labelled RNA (diluted to contain about 10^5 cpm/75 µl) was added at timed intervals and the binding reactions were incubated for 20 minutes at either 20°C or at 4°C. The reactions were passed through nitrocellulose filters and the filters washed with 10 ml of Sugita and Sugiura's buffer.

A large scale binding experiment is required for each round so that the number of RNA sequences is limited by binding affinity rather than amount of protein available. For this reason, binding curves were determined, comparing the percent of total RNA bound by the RNA to the amount of protein. In order to achieve the optimum selection of RNA sequences to carry over to the next round of SELEX, the concentration of protein that binds 2-10% of the total RNA in the binding curve should be used in a large scale binding experiment. In these SELEX experiments however, 2-10% of the RNA was not bound, so the experiment was modified to include more protein in order that more RNA would be bound.

Two nmol "cold" RNA was incubated with 2 nmol RbpD in Sugita and Sugiura's buffer for 15 minutes at 4°C or at 25°C in a total reaction volume of 200 µl. The reaction conditions and elution methods were the same as those described above. Bound RNA was isolated from the nitrocellulose filter by elution using 400 µl phenol (equilibrated in 0.1 M sodium acetate, pH 5.2) and 200µl filter elution buffer (8.3 M urea; 20 mM HEPES, pH 7; 50 mM NaCl; 1 mM EDTA). The filter and elution mixture were incubated at room temperature for 30 minutes; they were then microfuged for five minutes. Water (200 µl) was added to help the phases separate, and the mixture was microfuged for five minutes. The RNA was removed with the aqueous upper layer, and ethanol precipitated.

The RNA pellet was resuspended in 27 µl water and heated to 90°C for 5 minutes with 15 µM 3LN7.1 primer, then placed on ice. The reverse transcription reaction was then prepared (1x Transcription Buffer (Gibco BRL); 10mM DTT; 0.5 mM dNTP's;

SuperscriptII (Gibco BRL)) and incubated for 30 minutes at 37°C to reverse transcribe the isolated RNA sequences. Polymerase chain reaction (PCR) was then used to amplify the resulting cDNA. The incubation mixture contained 1x PCR buffer (Promega), 2.5 mM $MgCl_2$, 1.5 μM 5LN7.1, 1 μM 3LN7.1 and 3 units Taq polymerase (Promega). Thirty-five rounds of PCR were performed with an annealing temperature of 55°C for 20 seconds and an elongation step of 72°C for 10 seconds. The amplified fragments were electrophoresed on a 8% non-denaturing polyacrylamide gel and stained with ethidium bromide to confirm the size of the isolated transcripts. The isolated sequences were used in the next round of *in vitro* transcription.

Chapter 3:

Results and Discussion –

Expression and Purification of RbpD and RbpD1

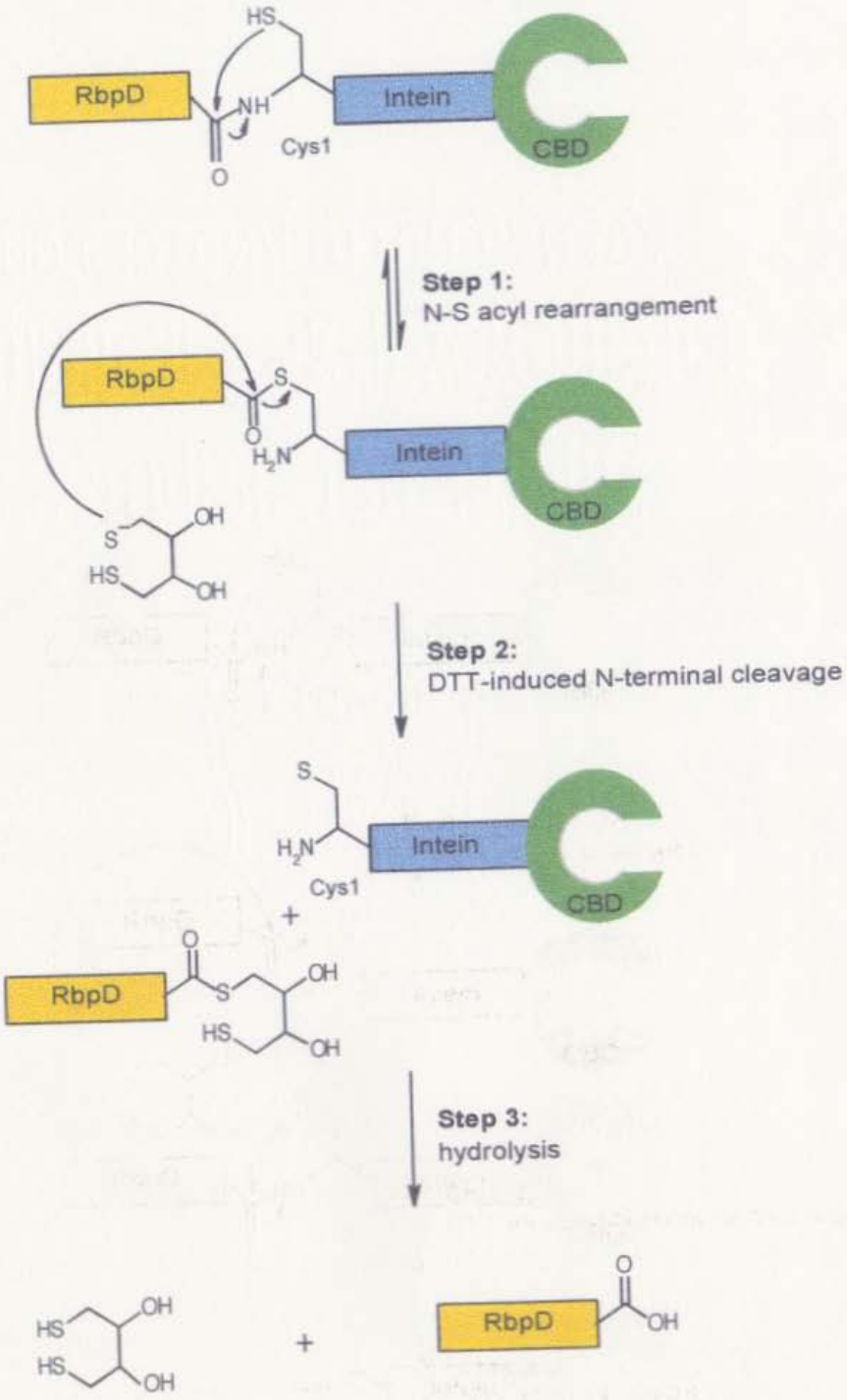
3.1 Expression of RbpD

In order to study the way in which RbpD works in a biochemical manner, we first needed to have access to purified protein. Although a number of methods of protein purification were available to us, we chose to use the IMPACT I system from New England Biolabs (described in this section) and a 6x-histidine tag purification method (described in Section 3.2). The IMPACT I system was chosen to express and purify the RbpD protein because it allowed the wild-type protein to be isolated. Figure 3.1 illustrates how the IMPACT I system works.

3.1.1 Cloning of *rbpD* into pCYB1

Chris Holden (1995) cloned the *rbpD* gene into the plasmid pAnR4.1. The gene was present on a 374 base-pair EcoRI/HindIII fragment. In order to produce the protein, the gene first needed to be cloned into a suitable expression vector. Three issues were considered during cloning. First, because the protein must be expressed as a fusion protein, the gene must be cloned so that it is in frame with the intein and chitin-binding domain coding sequences. Second, the entire construct, consisting of all three coding sequences (*rbpD*, intein gene, and chitin-binding domain gene) must be translated so the stop codon of the *rbpD* gene must be removed. Third, the start codon should not be too far from the ribosome-binding site. While the Shine-Dalgarno sequence (AGGAGGU) is not universal to all genes in bacteria, it does position the ribosome for translation by base-pairing with complementary sequences on the 16S rRNA, and does seem to facilitate expression of foreign genes in *E. coli* (Snyder and Champness, 1997). Because of these

Figure 3.1: Schematic drawing of the splicing mechanism of the IMPACT I system. A fusion protein contains the protein of interest (RbpD) in the N-extein site, the *S. cerevisiae* VMA intein and a chitin-binding domain (CBD) as the C-extein. In the first step, an N-S acyl rearrangement forms a thioester at the first splice junction (Cys1). In the presence of a reducing agent (DTT), the thioester bond will cleave, releasing the protein of interest attached to DTT. The thioester bond between DTT and the protein of interest is then hydrolysed releasing the free protein.


HS
RbpD
NH
Intein
Cys1
CBD
O
Step 1:
N-S acyl rearrangement
RbpD
S
O
H_2N
Intein
CBD
S
OH
HS
OH
Step 2:
DTT-induced N-terminal cleavage
S
H_2N
Intein
Cys1
CBD
+
RbpD
O
S
OH
HS
OH
Step 3:
hydrolysis
HS
OH
HS
OH
+
RbpD
O
OH

considerations, the plasmid pCYB1 (Table 2.1), provided by New England Biolabs (1997), was chosen as the expression vector. The multiple-cloning site (MCS) of pCYB1 contains an NdeI site in the proper orientation and position in relation to a ribosome-binding site (Figure 3.2A). A SapI site exists at the opposite end of the MCS so that the 3' end of the inserted gene is adjacent to the 5' end of the gene for the intein.

Two oligonucleotides (RbpD(NdeI) and RbpD(SapI); Table 2.3) were designed to perform four tasks. The first task was to amplify the *rbpD* gene in the plasmid AnR4.1 (Table 2.1) using the polymerase chain reaction (PCR). The second and third were to introduce mutations into *rbpD* by creating an NdeI site at the 5' end and a SapI site at the 3' end, and the fourth task was to change the stop codon into a cysteine codon (Figure 3.2B). After PCR amplification, the product was digested with the two enzymes, and the gene cloned into the pCYB1 vector, yielding the plasmid pRLT1. The resulting plasmid was confirmed by digesting the plasmid with NdeI and KpnI (the SapI site was lost during the ligation) and observing a 394 base-pair band on an agarose gel (Figure 3.3). As an additional control that the PCR amplification of the plasmid did not change the sequence, the *rbpD* gene within pRLT1 was sequenced on both strands. There was no change in the sequence after the PCR.

3.1.2 Expression of the fusion protein in *E. coli* MC1061

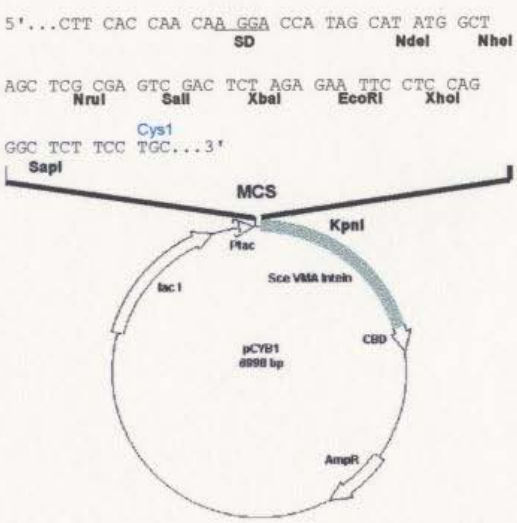
The plasmid pRLT1 was transformed into *E. coli* strain MC1061 in order to express RbpD. The protocol for expression and purification of wild-type RbpD can be found in Figure 2.1 in the Materials and Methods section.

Figure 3.2: Cloning *rbpD* in pCYB1

A) Schematic diagram of the plasmid pCYB1. The 6998 base-pair plasmid contains a multiple cloning site immediately upstream of the *Sce* VMA intein-chitin-binding domain (CBD) fusion gene. Expression of cloned genes is driven by a *tac* promoter and is regulated by a *lacI* gene carrying the I^Q mutation. The *rbpD* gene must be inserted in frame with the genes for the VMA intein and chitin-binding domain, since all three are transcribed from P_{tac}. The *rbpD* gene was inserted between the NdeI and SapI restriction sites. The KpnI site was used to check for the presence of the *rbpD* gene insert, and is indicated on the outside of the plasmid.

B) Mutagenesis of *rbpD*. Oligonucleotides were designed not only to amplify the gene fragment, but to introduce mutations as shown: **i)** The first oligonucleotide introduced an NdeI site at the start codon of the gene. **ii)** The second oligonucleotide was designed for two purposes. It changed the stop codon in the gene to a cysteine codon. This oligonucleotide also introduced a SapI site. **iii)** The PCR amplified gene fragment was digested with NdeI and SapI, and then ligated into a pCYB1 vector. The resulting plasmid was designated pRLT1.

A

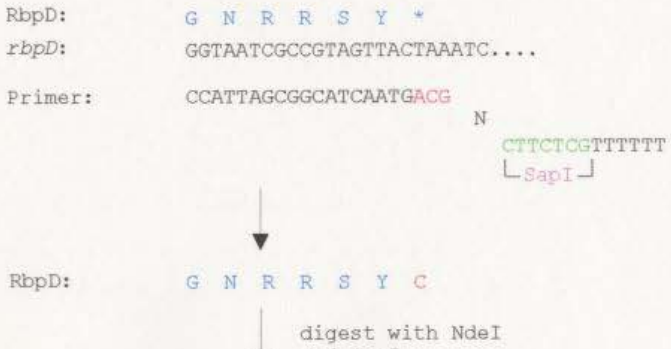

5'...CTT CAC CAA CAA GGA CCA TAG CAT ATG GCT
SD
NdeI
NheI
AGC TCG CGA GTC GAC TCT AGA GAA TTC CTC GAG
NruI
SalI
XbaI
EcoRI
XhoI
Cys1
GGC TCT TCC TGC...3'
SapI
MCS
KpnI
Ptac
lac I
Sce VMA Intein
CBD
pCYB1
6998 bp
AmpR

B

i


NdeI
Primer: GGAGACATCATATGTGTCAA
rbpD: GGAGACATTGTATGTGTCAATTT....
RbpD: M S I Y

ii


RbpD: G N R R S Y *
rbpD: GGTAATCGCCGTAGTTACTAAATC....
Primer: CCATTAGCGGCATCAATGACG
N
CTTCTCGTTTTTT
SapI
RbpD: G N R R S Y C
digest with NdeI
and SapI

iii

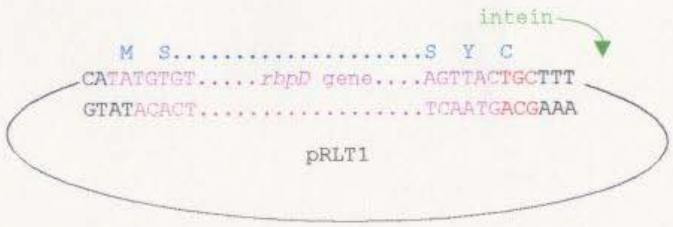
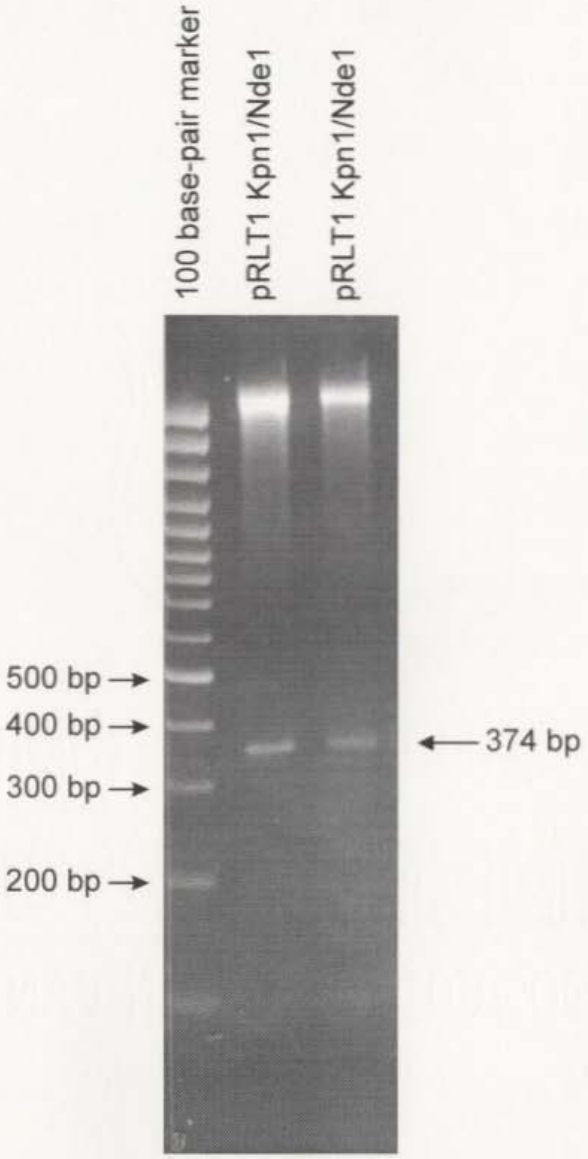

intein
M S.....S Y C
CATATGTGT.....rbpD gene....AGTTACTGCTTT
GTATACACT.....TCAATGACGAAA
pRLT1

Figure 3.3: pRLT1 digested with NdeI and KpnI. Plasmid DNA was digested with the restriction endonucleases NdeI and KpnI to indicate that the *rbpD* gene had been inserted into the multiple cloning site of pCYB1. KpnI was used in place of SapI since the SapI site was lost during the ligation of the insert into the vector. The resulting fragment was 394 base-pairs long.


100 base-pair marker
pRLT1 Kpn1/Nde1
pRLT1 Kpn1/Nde1
500 bp
400 bp
300 bp
200 bp
374 bp

In the first step of attempting to express RbpD, 1mM IPTG was used to induce the cells to express the fusion protein which contained RbpD, the *S. cerevisiae* VMA intein and a chitin binding domain. The results are shown in Figure 3.4, confirming that expression of the fusion protein did occur with this concentration of the IPTG. The chitin-binding domain is 5 kDa, the intein 50 kDa, and the calculated molecular weight of RbpD is 12 kDa, so it was expected that a band about 67 kDa would be observed in the induced cell extract. The induced band on the gel in Figure 3.4 appears larger than expected, slightly larger than 70 kDa. However, cyanobacterial RNA-binding proteins have been found to run higher on SDS gels than their molecular weights would indicate, probably due to the flexible glycine-rich region (Sato, 1994).

When we attempted to purify RbpD, no fusion protein could be seen in the clarified extract lane of the gel (Figure 3.5), nor in the eluted protein lane (data not shown). Because of this, and because we did not seem to be getting much induction, expression was attempted in *E. coli* BL21(DE3)LysS.

3.1.3 Expression of the fusion protein in *E. coli* BL21(DE3)pLysS

The *E. coli* strain BL21(DE3)pLysS is a good strain in which to express proteins because it lacks the *lon* protease and *ompT* outer membrane protease, both of which can degrade proteins during purification (Studier *et al.*, 1990; Grodberg and Dunn, 1988). DE3 is a bacteriophage lysogen and derivative of λ which contains the *lacI* gene, the *lac*UV5 promoter, the beginning of the *lacZ* gene and the T7 RNA polymerase gene. In the presence of IPTG, the DE3 lysogen is induced. This allows expression of T7 RNA

Figure 3.4: Induction of RbpD fusion protein expression in *E.coli* MC1061. The first lane contains the molecular weight marker (MW marker) (Pharmacia-Amersham) which includes the following protein bands: 94 kDa (phosphorylase b); 67 kDa (bovine serum albumin); 43 kDa (ovalbumin); 30 kDa (carbonic anhydrase); 20 kDa (soybean trypsin inhibitor); 14 kDa (α-lactalbumin). The "uninduced" lane contains extracts from cells which were not induced with IPTG. The other lanes in the gel contain extracts from cells induced with IPTG for 1,2,3 and 4 hours at 37 °C and shaking at 225 rpm. The fusion protein (open-headed arrow) appears above the 67 kDa marker band. The amount of protein in the three hour induction extract appears lower due to an error in technique. The gel is an 8-25% acrylamide gradient gel.

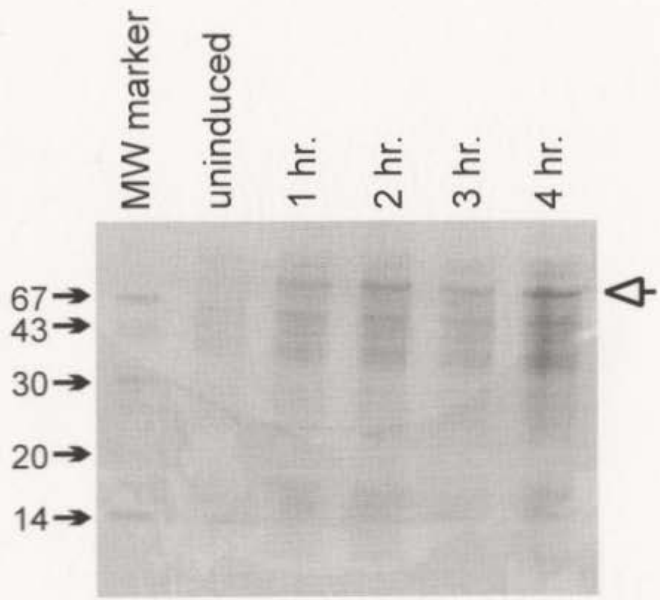
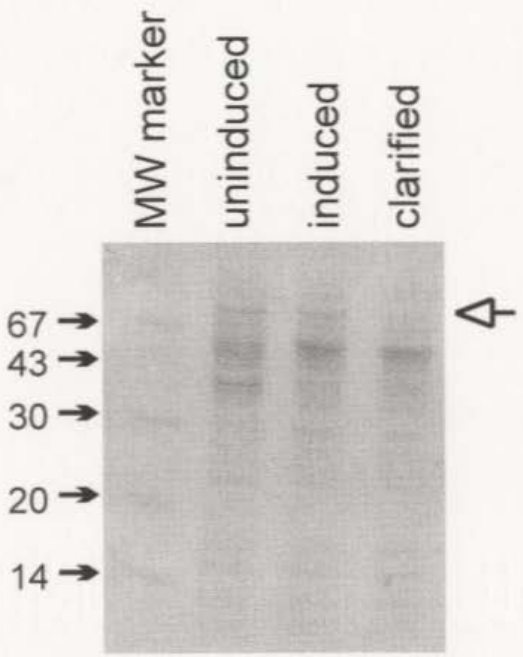

MW marker
uninduced
1 hr.
2 hr.
3 hr.
4 hr.
67
43
30
20
14

Figure 3.5: Expression of the RbpD fusion protein from pRLT1 in *E. coli* strain MC1061. The first lane contains the molecular weight marker. The "uninduced" lane contains cell extract from a culture not exposed to IPTG. The "induced" contains cell extract from a culture exposed to IPTG for four hours. The "clarified lysate" from the same culture as the "induced" lane, but cell debris has been removed by centrifugation. In the initial purification of RbpD, fusion protein (open-headed arrow) could not be detected in the "clarified lysate" lane.


MW marker
uninduced
induced
clarified
67
43
30
20
14

polymerase which, in turn, transcribes the target gene which is under the control of a T7 promoter on a plasmid. Finally, this strain contains the pLysS plasmid which carries a gene for T7 lysozyme. When T7 lysozyme binds to T7 RNA polymerase, it inhibits transcription of genes by the T7 polymerase. This means that there are reduced levels of uninduced transcription in the cell so that the cells will not be killed prematurely if the foreign genes encode proteins toxic to the cells. This also widens the range of proteins that can be expressed in this cell line (Studier *et al.*, 1990; Moffatt and Studier, 1987).

In the first expression experiments, it appeared that a band representing the fusion protein was induced. Due to a misinterpretation of the size of the molecular weight marker, it was thought that the induced protein appeared in the gel above the 67 kDa band (Figure 3.6 A). However, the induced protein remained in the periplasmic fraction rather than in the cytoplasmic fraction when the cells were disrupted using sonication and therefore stayed in the pellet fraction when the cell extract was purified by centrifugation.

Qiagen (1997) recommended the use of Tween-20 in order to solubilise proteins in the periplasmic fraction. This appeared to work, with much of the protein appearing in the Tween-20 soluble fractions (Figure 3.6 B). No matter what was tried though, RbpD could not be eluted from the chitin column during the purification step. Also, no fusion protein appeared in the washing step, indicating that the induced protein was not being eluted from the column.

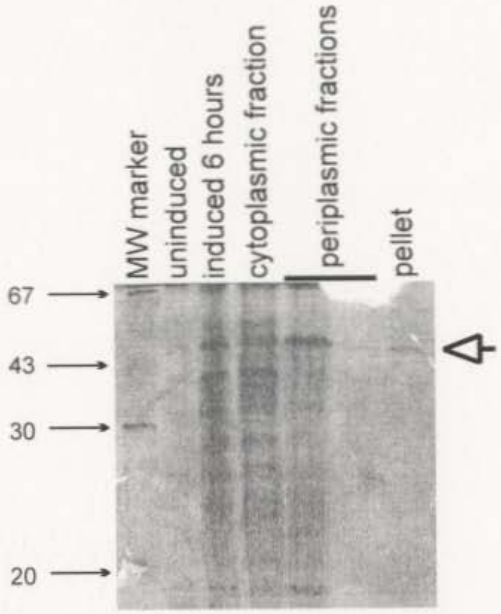
Finally, we examined the molecular weight markers more closely. The protein that was thought to be the 67 kDa marker had a similar (smeared) look to that of ovalbumin (43 kDa). If this marker was ovalbumin, then the bottom marker on the gel

Figure 3.6: Expression of the RbpD fusion protein from pRLT1 in *E. coli* strain BL21(LysS)DE3.

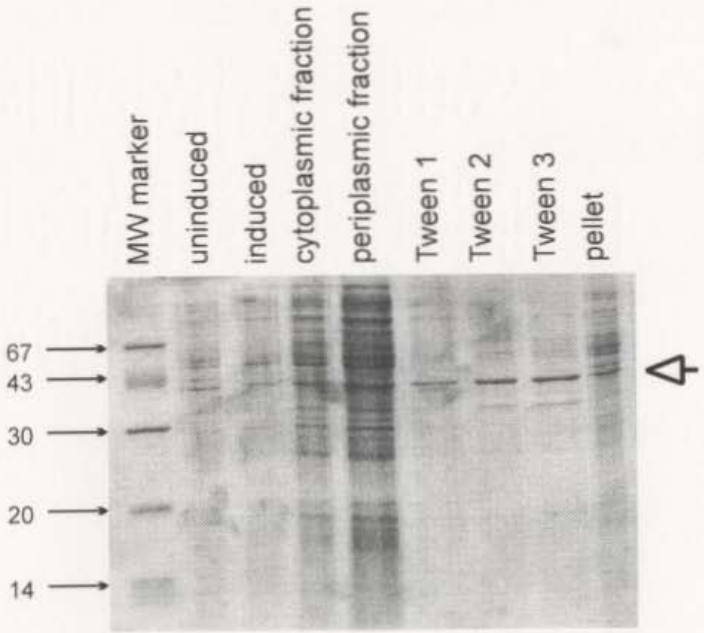
A) The "uninduced" and "induced" lanes contain extracts of *E. coli* BL21 cultures that were not or were exposed to IPTG respectively. The "cytoplasmic fraction" contained only the solubilised proteins from the extract. The "periplasmic fraction" were solubilised in Tween 20 from the pellet left in the clarified extract from the induced culture. The "pellet" contains those proteins that could not be solubilised with the use of detergents.

B) Solubilising the RbpD fusion protein from the periplasmic fraction of BL21. The first five lanes correspond to those in **A)** The Tween 1, 2, and 3 lanes contain successive washed with 0.25% Tween 20. The open-headed arrow represents the band initially thought to represent the fusion protein. The molecular weight marker lane contains the same protein bands as mentioned before except that the 94 kDa band appears to be absent.

A


MW marker
uninduced
induced 6 hours
cytoplasmic fraction
periplasmic fractions
pellet
67
43
30
20

B


MW marker
uninduced
induced
cytoplasmic fraction
periplasmic fraction
Tween 1
Tween 2
Tween 3
pellet
67
43
30
20
14

was the 14 kDa α-lactalbumin and from here with the reassignment of the marker bands, it appeared that the 94 kDa marker was missing from this batch. We ran a number of other gels to confirm this (data not shown). With this new interpretation, it became obvious that the "induced" protein had too low a molecular weight to be the fusion protein. If any induction was occurring, it was at levels too low to clearly observe by Coommassie staining, and we therefore returned to the *E. coli* MC1061 strain.

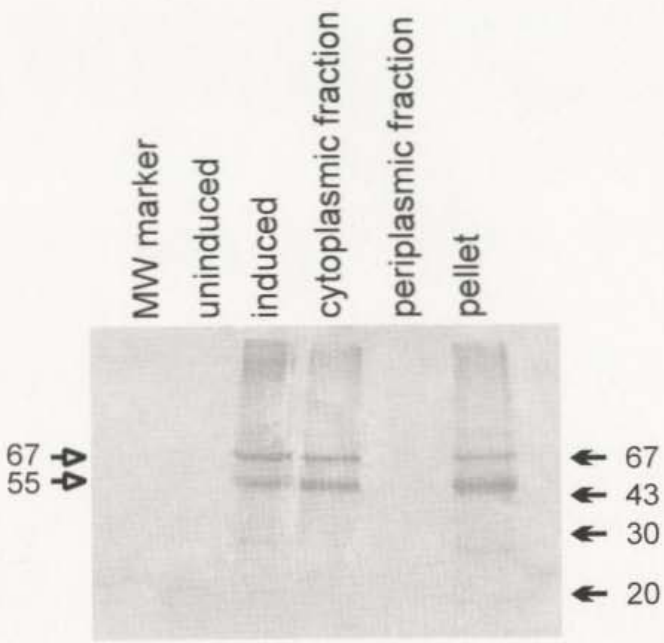
3.1.4 Western Blotting

We confirmed that the protein induced in *E. coli* MC1061 was indeed the RbpD fusion protein by Western blot analysis, using an antibody against the intein portion of the fusion protein (Figure 3.7A). As can be seen in the figure, no fusion protein was produced in the uninduced cells, and no protein was found in the periplasmic fraction. One problem that this experiment brought to our attention was that of *in vivo* cleavage. Two bands were seen: one corresponding to the whole fusion protein (>67 kDa) and one corresponding to the intein-chitin-binding domain (55 kDa). The intein should not cleave except in the presence of a reducing agent, but there are clearly two bands in the Western blot, corresponding to the full fusion protein and the cleaved product containing the intein and chitin-binding domain.

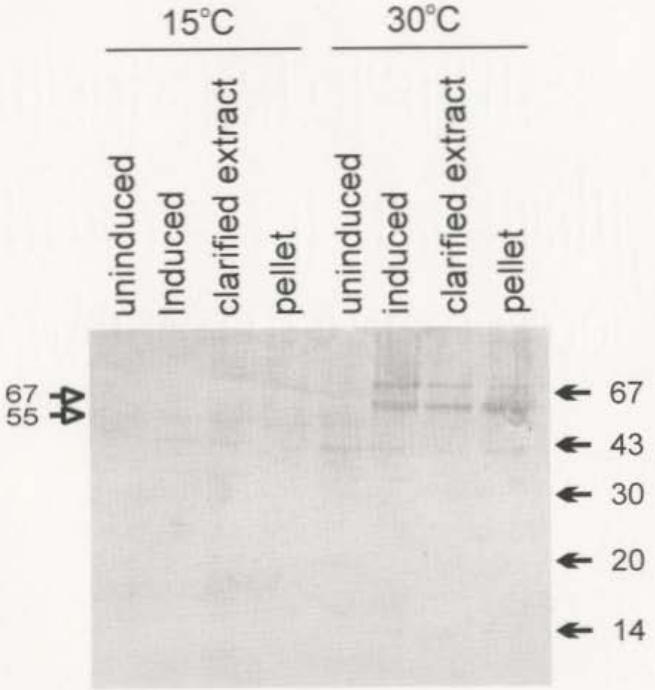
There were four possible reasons for *in vivo* cleavage to have occurred in the fusion protein. The first is simply that the fusion protein may not be stable in *E. coli* cells. The RbpD is from a strain of cyanobacteria, the intein from a eukaryote, and the two may not be stable in the *E. coli* environment. The instability of many proteins in

Figure 3.7: Western blot analysis of the RbpD fusion protein in *E. coli* MC1061.
A) A culture of *E. coli* MC1061 was induced with IPTG at 37°C for four hours. The fractions loaded on each lane correspond to those in figure 3.6. Ten µl of each sample was mixed with an equal volume of Cracking Buffer; the samples were electrophoresed and blotted as described in Section 2.5.3. An antibody against intein portion of the fusion protein was used to detect the protein in each fraction. Positions of the 67 and 55 kDa bands in the molecular weight marker are indicated using open-headed arrows.
B) *E. coli* MC1061 cultures were grown at either 15°C or 30°C. The 15°C culture was induced overnight with IPTG, while the 30°culture was induced for 6 hours. Western blotting was performed as described in **A**. The 55 and 67 kDa bands are indicated with open-headed arrows. Molecular weight markers are indicated with black arrows.

A


MW marker
uninduced
induced
cytoplasmic fraction
periplasmic fraction
pellet
67
55
67
43
30
20

B


15°C
30°C
uninduced
Induced
clarified extract
pellet
uninduced
induced
clarified extract
pellet
67
55
67
43
30
20
14

E. coli is well characterised. Often foreign proteins are recognised as abnormal within the bacteria and are therefore rapidly hydrolysed (Balbas and Bolivar, 1990). However, this is unlikely to be the cause of the problem here since the protein bands on the Western blot do not appear as a smear, but as two distinct bands indicating specific cleavage rather than general proteolysis.

The second possibility was that the temperature of induction might be too high (New England Biolabs, 1997). The induction was carried out at 37°C since that was the temperature at which the cells had been grown, but the fusion protein might not be stable at this temperature. In order to address this possibility, we induced the cells for 6 hours at 30°C and also overnight at 15°C. The amount of *in vivo* cleavage was measured using a second Western blot (Figure 3.7B). In this blot, it appears that the fusion protein lacking the RbpD accounts for about half of the fusion protein present in the 30°C induction experiment. This is the same as the amount of cleavage product formed when the induction temperature was 37°C. At an induction temperature of 15°C, no bands appeared in the Western blot at all, indicating that the fusion protein was not produced at any significant level at this temperature, even if the culture was induced overnight. We decided to keep the induction temperature at 37°C.

The third possible reason to explain the *in vivo* cleavage may have been the presence of the two arginines in the second and third last positions of the RbpD protein. According to New England Biolabs (1997), if arginine is the C-terminal residue of the protein, then the protein will experience 75% *in vivo* cleavage. However, according to the technical support employees of the company, no research has been done to measure

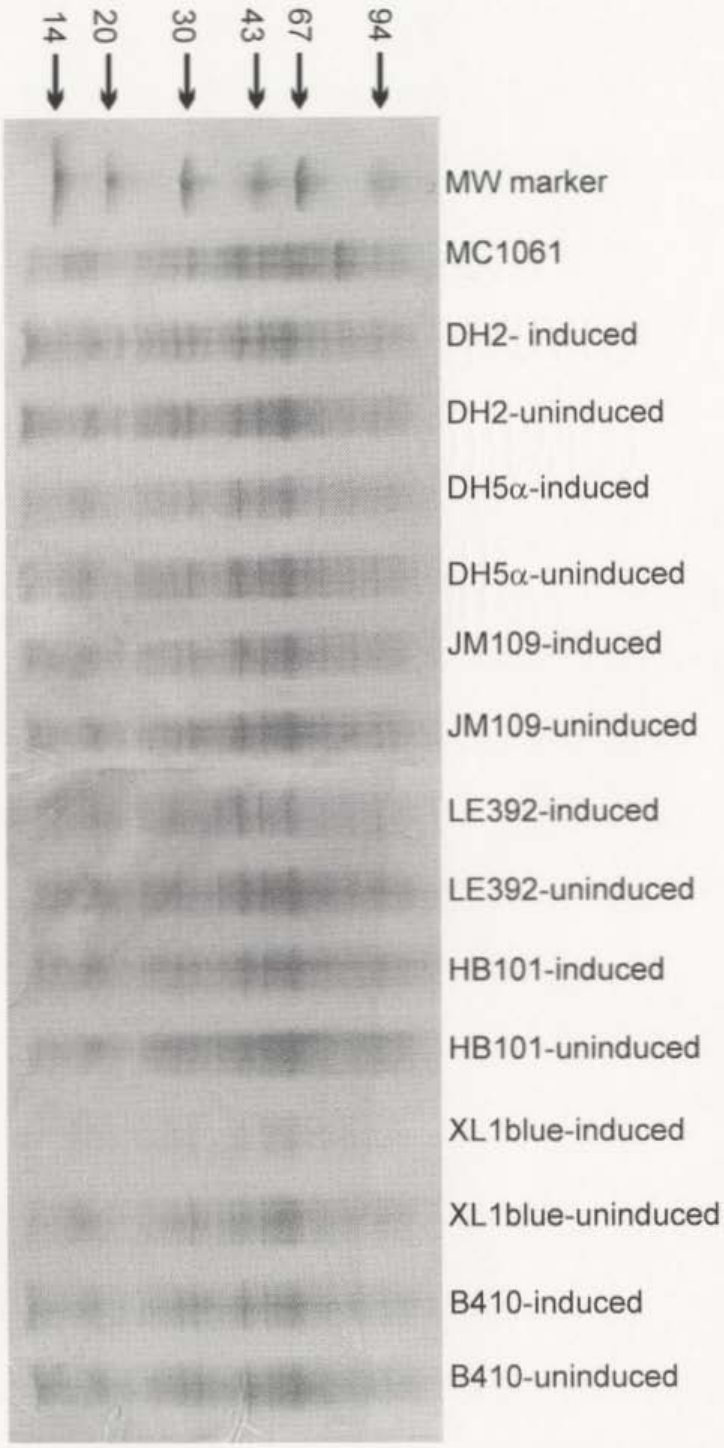
the effects of residues very near to the C-terminal of the protein. It is possible that these residues do have an effect on the cleavage patterns of the fusion protein.

The fourth possible reason for the *in vivo* cleavage is the high reducing environment of the cell. Because *in vitro* fusion protein cleavage depends on the introduction of a reducing agent (New England Biolabs, 1997; Chong *et al.*, 1997), it is possible that, if the reducing environment of the cell is high, then it could cause the cleavage of the RbpD from the intein. In order to test this hypothesis, we decided to express the protein in a number of different cell lines which would have different reducing environments, so that the fusion protein would be more stable.

3.1.5 Expression of RbpD in Other Strains of *E. coli*

We examined the induction of RbpD expression in eight different strains of *E. coli* – MC1061, DH2, DHα5, JM109, LE309, HB101, XL1Blue, B410 (Figure 3.8, Table 2.2). For each strain tested, both a control and an induced sample were grown side by side, so that the same amount of protein would be in both samples. Once the induced samples had been exposed to IPTG for four hours, 1 ml samples were taken from both cultures, pelletted and resuspended in Cracking Buffer. On examination of the gel, only extracts from *E. coli* MC1061 had a band that appeared to be induced by IPTG and which corresponded in size to the induced protein above the 67 kDa band. None of the other strains had any bands that changed in intensity. This result is consistent with the comments from other users of the IMPACT I system on the internet newsgroup

Figure 3.8: Expression of the RbpD fusion protein from pRLT1 in eight strains of *E. coli*. For each of the strains tested, 25 mL cultures were grown to an optical density (600 nm) of 0.5 and then induced for four hours with IPTG. An uninduced control was grown for the same amount of time in parallel with the induced counterpart. The strains tested were MC1061, DH2, DH5α, JM109, LE392, HB101, XL1blue, and B410. Of the strains tested, only MC1061 showed a band that appeared in the induced culture but not in the uninduced culture. The fusion protein migrates above the position of the 67 kDa band shown in the marker lane. The characteristics of all of the strains can be found in Table 2.2.


14
20
30
43
67
94
MW marker
MC1061
DH2- induced
DH2-uninduced
DH5α-induced
DH5α-uninduced
JM109-induced
JM109-uninduced
LE392-induced
LE392-uninduced
HB101-induced
HB101-uninduced
XL1blue-induced
XL1blue-uninduced
B410-induced
B410-uninduced

bionet.molbiol.methds.reagnts who have also had problems with levels of expression of the fusion protein.

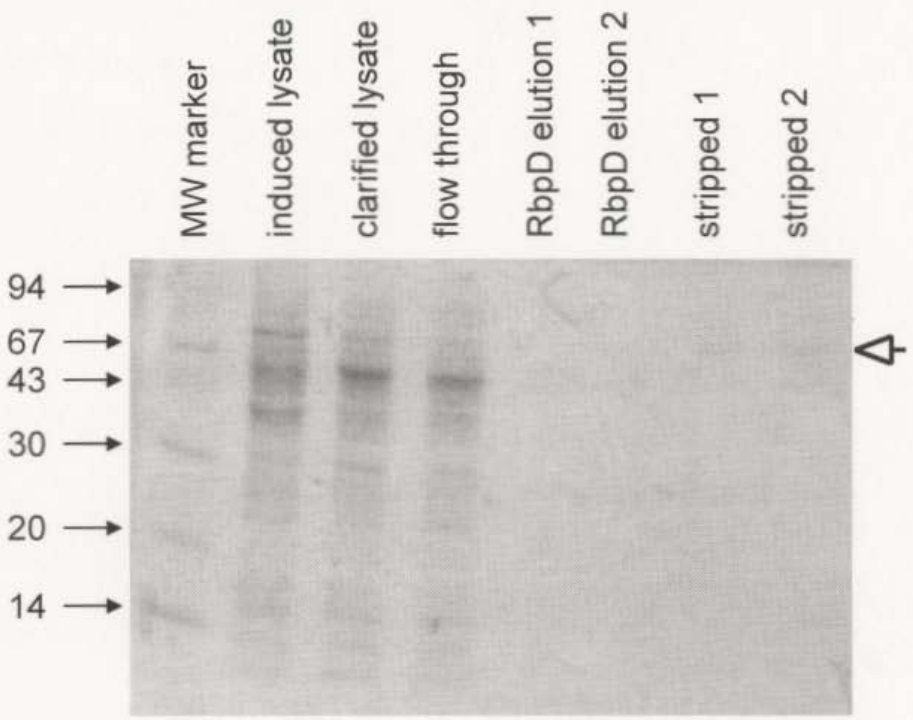
It was decided to use *E. coli* strain MC1061 for purification of RbpD. However, in order to compensate for the *in vivo* cleavage experienced in this strain, larger culture volumes were used to purify the protein.

3.1.6 Purification of RbpD

Once the fusion protein was attached to the chitin column for affinity purification (Section 2.5.2 in Materials and Methods), DTT was introduced as a reducing agent so that cleavage of the RbpD from the fusion protein would occur. While the fusion protein was soluble, the cleaved RbpD appeared to be insoluble in the buffer recommended by New England Biolabs (1997). The company recommended adding up to 0.5% Triton-X 100 if this was a problem. However, when 0.1% Triton-X 100 was added to the Cleavage Buffer, then cleavage did not occur (Figure 3.9). When 0.1% Triton-X 100 was added to the column after the RbpD had been cleaved from the fusion protein then some but not all of the RbpD was solubilised. When the amount of Triton-X 100 was increased to 0.5%, the column clogged up, and nothing could be eluted from it, even if Stripping Buffer was added, although this should have eluted the 55 kDa intein and chitin-binding domain (data not shown). Finally, the Triton-X 100 was removed from the Cleavage Buffer and Elution Buffers.

In a separate attempt to solve the problem of the RbpD insolubility, the salt concentration in the Cleavage Buffer was increased. When 500 mM NaCl was used in

Figure 3.9: RbpD elution with 0.1% Triton-X 100 in Cleavage Buffer. The fusion protein (open headed arrow) is visible in the induced lysate (lane 2) and in the clarified extract (lane 3), but is not present in the flow through (lane 4), which consists of eluate from the chitin column. The cleavage reaction was incubated overnight at 4°C in the presence of 0.1% Triton X-100 and eluted ("RbpD elution" lanes). No RbpD is seen above the 14 kDa molecular weight band. Stripping buffer was used to remove the protein attached to the chitin column through the chitin-binding domain of the fusion protein. The stripped fractions include a band above the 67 kDa molecular weight marker band and corresponds to the total fusion protein.


MW marker
induced lysate
clarified lysate
flow through
RbpD elution 1
RbpD elution 2
stripped 1
stripped 2
94
67
43
30
20
14

the Cleavage Buffer and Elution Buffer rather than the recommended 50 mM (New England Biolabs, 1997), then the free RbpD became soluble, and could be eluted from the column. A higher concentration of NaCl (750 mM) did not allow the fusion protein to bind to the chitin column (data not shown). Therefore, 500 mM NaCl was used in all successive experiments.

Another problem with the protocol for purification of RbpD was that 30 mM DTT (as recommended by New England Biolabs, 1997) was not sufficient to allow the total cleavage of the fusion protein, so that much of the full length fusion protein was eluted only when the Stripping Buffer was passed through the column. This problem was solved by using 50 mM DTT in the Cleavage Solution. The final purification of the protein is demonstrated in Figure 3.10.

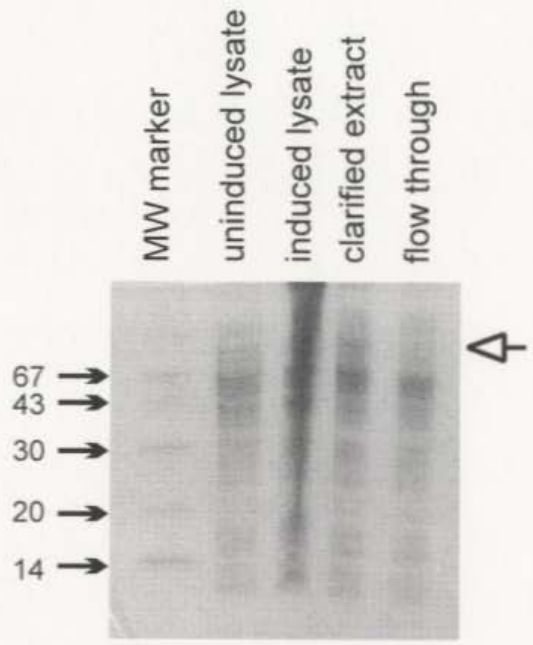
3.2 Expression and Purification of RbpD1

3.2.1 Background

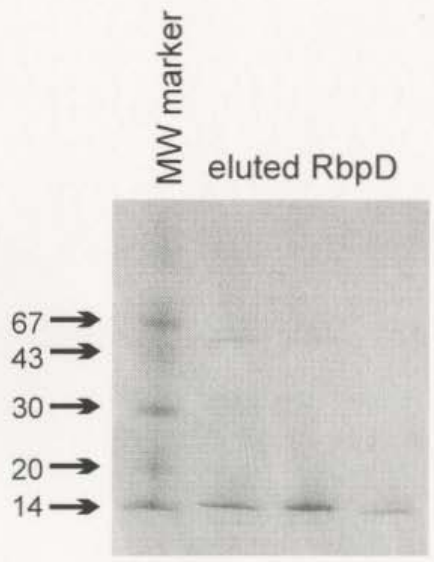
Having a wild-type protein is useful for the initial characterisation of a protein, in that it allows one to be confident that the results represent what the protein might do *in vivo*. There are no extra amino acids attached to the protein that might affect its conformation or function. However, we eventually wanted to be able to do some work with the protein in *Anabaena* 7120. Therefore, it was advantageous to develop an affinity-tagged protein that had all the wild-type properties, but which could be re-isolated from a cellular environment.

Figure 3.10: Expression and elution of RbpD.
(A) RbpD was expressed as a 67 kDa fusion protein in *E. coli* strain MC1061. The fusion protein can be visualised in the clarified extract (open-head arrow), but not in the uninduced lane or in the flow through lane.
(B) Elution of RbpD from the column after overnight cleavage from the fusion protein. Nine 0.5 ml samples were eluted from the column. Four µl of each was mixed with an equal volume of SDS loading buffer, and 1 µl of each sample was electrophoresed on a PhastGel. The concentration of the RbpD in the first and third fractions was estimated to be about 0.5 mg/ml in the samples shown. The concentration of RbpD was estimated by comparing the RbpD band with known concentrations of lysozyme, which were run in parallel on a PhastGel (data not shown). The total yield of RbpD was normally in the range of 0.5 - 1 mg of protein from 3 l of culture. Confirmation that the eluted protein was RbpD was done using the agarose-bound polymer binding experiments.

A


MW marker
uninduced lysate
induced lysate
clarified extract
flow through
67
43
30
20
14

B


MW marker
eluted RbpD
67
43
30
20
14

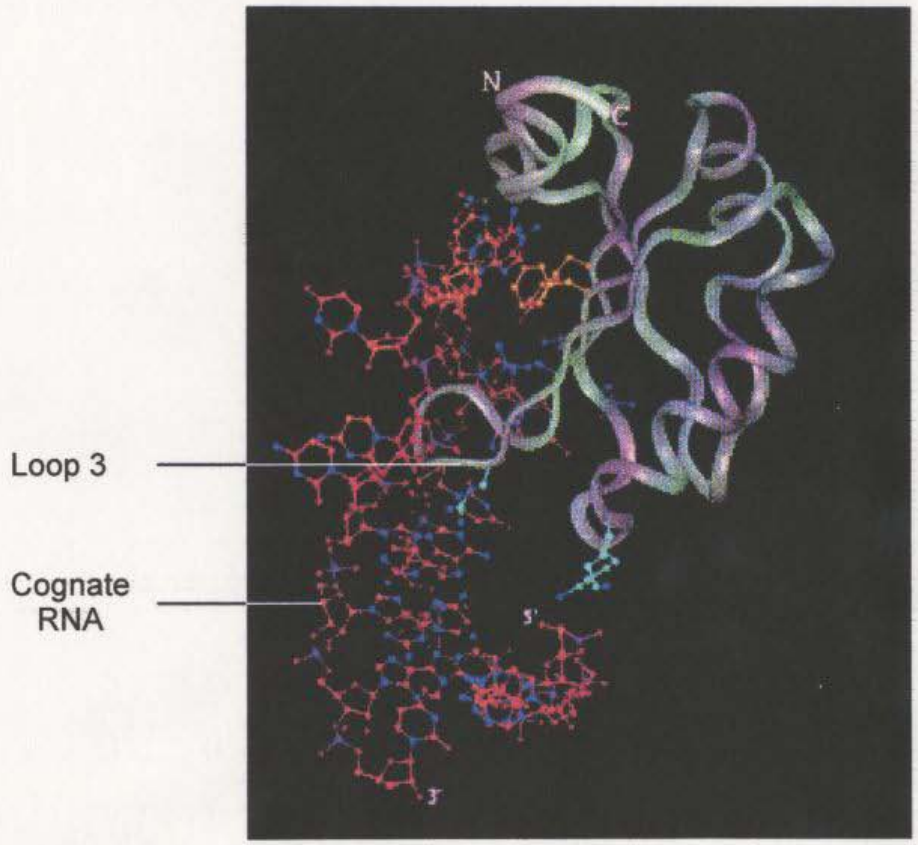
In order to put an affinity-tag into the protein, we used a 6x histidine tag (Qiagen, 1997). Six histidine codons were inserted into the gene for the protein through site directed mutagenesis. Cynthia Slade (1998) used the Promega Altered Sites II system in order to insert a His-coding-tag in the *rbpD* gene. She attempted to insert the His-tag into loop 3 and loop 4 of RbpD, as shown in Figure 3.11. While she was not successful with the insertion of a tag into loop 4, she did insert one into loop 3. The loop 3 and 4 insertions were chosen over N- or C-terminal insertions due to the location of the RNP2 motif at the extreme N-terminal of the protein and also to avoid changing the conserved C-terminal of the protein. Given that we now had the gene with the coding sequence for the histidine-tag, we decided to express this protein in order to be able to perform comparative studies between RbpD1 (RbpD containing a 6x-His-tag in loop 3) and RbpD.

We did have some concerns about the placement of the histidine tag into loop 3 of the protein. The RNA-recognition motifs in the U1A spliceosomal protein are structurally similar to the RRMs in cyanobacterial RNA-binding proteins (Figure 3.12). Both contain a surface formed by four antiparallel β-sheets with two α-helices running behind. The bound RNA is splayed over the β-strands. RNP1 and RNP2 reside in the second and third β-strands, which lie antiparallel to each other, and interact with the RNA from there. Loop three lies between β-strands 2 and 3. In U1A, this loop protrudes through a loop formed by the RNA target (AUUGCAC being the loop) and interacts directly with the RNA (Oubridge *et al.*, 1994; Nagai *et al.*, 1995; Allain *et al.*, 1996). This importance of loop 3 in RNA binding in U1A may hold true for RbpD as well.

Figure 3.11: Illustration of RbpD showing the sites of attempted 6x His-tag insertion. The first His-tag was successfully inserted into loop 3 of the rbpD gene. The loop 4 His-tag insertion was not successful. Also shown in this illustration are the RNP1 (red) and RNP2 (green) sequences, and the glycine-rich region of RbpD (purple).


RNP2
MSIYIGNLSYQVTEEDLK
L
A
F
TERDTPLQVRSVKGYEA
loop 3
His-tag
G
R
P
RNP1
RGFAFVEMETTEAQETAAIEA
L
loop 4
His-tag
D
G
ERPKAKNVKLDRGMWEA
E
R
S
glycine-rich domain
SSPRGGGGSWGNNNRGGGGGG
N
R
R
YS

Figure 3.12: Picture of U1A RRM1 bound to an RNA ligand. U1A interacts with a stem-loop structure in its cognate RNA. The interaction occurs on the face of four anti-parallel β-strands and with loop 3 of the RRM. Loop 3 is shown here protruding through a loop (AUUGCAC) formed by the RNA target This picture was taken from Oubridge *et al.* (1994).


Loop 3
Cognate RNA
N
C
5'
3'

3.2.2 Site-directed mutagenesis of RbpD1a

We chose to express the *rbpD* gene with this His-coding-tag (*rbpD1a*) in a pTRC99A expression vector because it had been successfully used by Tom Belbin (1999). In order to do this, an NcoI site was engineered around the start codon of the gene so that when the gene was cloned into pTRC99A (Pharmacia), it would go into the multiple cloning site in the correct orientation to the promoter and also be the correct distance from the ribosome binding site (Figure 3.13A). This was performed using the Altered Sites II mutagenesis system, explained in Section 2.6.1 (Promega, 1996).

While the procedure sounds fairly straightforward, the mutagenesis took months due to a number of problems. The R407 phage lost its activity after a short time in the freezer even if it was aliquoted out in 200 µl fractions so that it was never freeze-thawed more than once. The ampicillin lost its activity twice. In order to determine whether the drug had lost its activity or whether the antibiotic resistances in the plasmid had been altered, a number of controls needed to be performed. The ampicillin was tested by growing cells lacking any plasmid on ampicillin plates; if the cells were able to grow, then the ampicillin was not working. This problem was eventually solved with the use of ticarcillin. Ticarcillin is also susceptible to β-lactamase, and works in place of ampicillin, but it is not as labile as ampicillin. To test the plasmid for resistances, we were forced to repeatedly digest the plasmid with the restriction enzymes EcoRV or PstI. If the ampicillin resistance gene is in working order, then it contains the PstI site, whereas if the tetracycline resistance gene is in working order, then it contains an EcoRV site (Piechocki and Hines, 1994). Finally, we also had a problem with the *E. coli* ES1301

Figure 3.13: Mutagenesis of *rbpD1a*

A) Oligonucleotide for mutagenesis of *rbpD1a*. The introduction of an NcoI site around the start codon of the gene using the 29 nucleotide oligonucleotide is shown. Introducing the NcoI site also changed the serine in the second amino acid position into an alanine.

B) Diagram of pRLT3. The *rbpD1* gene was cloned into the pALTER1 plasmid and mutated to include an NcoI site at the 5' end of the gene. The *rbpD1* gene was cut from the pALTER-1 plasmid using the introduced NcoI site and the HindIII site already present at the 3' end of the gene fragment, generating a 376 bp fragment. The fragment was then ligated into pTRC99A which had been cut with the same restriction endonucleases, and the ligation carried out using T4 ligase as described in Section 2.2.5. The resulting plasmid was designated pRLT3.

A

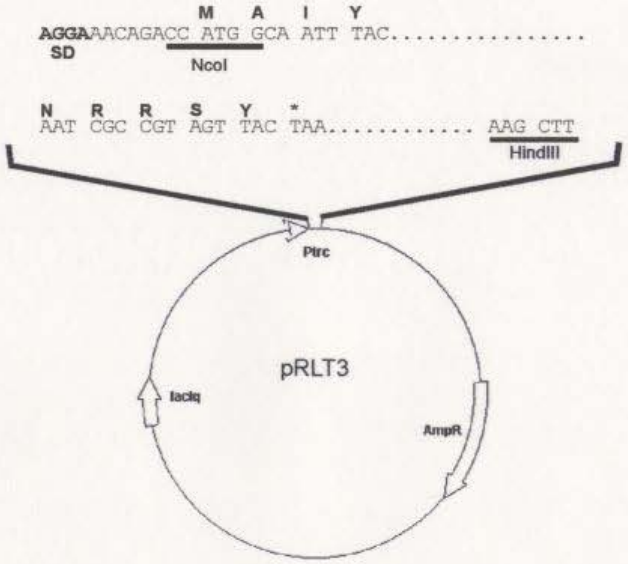
nucleotide sequence for *rbpD1a*

```
GATTCGGAGACTTGTATGTCAATTTACAT...
                M  S  I  Y....
```

oligonucleotide used to introduce mutation

```
                NcoI
GATTCGGAGACTTCCATGGCAATTTACAT
                M  A  I  Y....
```

B



strain losing its competence at various times during these attempts. All of these problems took time both to identify, and to take measures to correct the problem. Eventually though, the NcoI site was incorporated successfully.

3.2.3 Effects of the NcoI site in *rbpD1*

The incorporation of an NcoI site generated an alanine codon in place of the serine codon in the second position (Figure 3.13A). Ideally, we would have preferred to have kept the serine codon, but the pTRC99A expression system required the incorporation of the NcoI site. We felt that these two residues have similar enough properties that the function of the protein should not be affected. Alanine is a small residue and does not affect the overall conformation of the protein to a great extent (Hesselgesser and Horuk, 1997). The alanine replacement should only affect the function of the protein if the amino acid that the alanine is replacing is essential to the function of the protein.

Finally, in the *Synechococcus* sp. strain PCC 6803 RNA-binding proteins, the initial methionine is removed from the protein, leaving serine as the first amino acid residue (Sugita and Sugiura, 1994). Since it is rare that the N-termini of RNA-binding proteins are involved in binding, it is unlikely that this serine is involved in RNA binding. Nevertheless, it was of concern that it is immediately adjacent to the RNP2 motif in RbpD. This motif interacts directly with the RNA ligands of all of the known RNA-binding proteins of the RNP family (Nagai, 1993; Keene, 1996). The *rbpD1a* gene

including the NcoI site (*rbpD1*) was cloned into pTRC99A and called pRLT3 (Figure 3.14). The plasmid was then transformed into the *E. coli* strain BL21(LysS)DE3.

3.2.4 Expression and purification of RbpD1

Cells containing pRLT1 were grown overnight in cultures then inoculated into 25 ml cultures, where they were grown to an optical density (600 nm) of 0.5-0.7. Because expression of the *trc* promoter in pTRC99A is under the control of the $lacI^Q$ mutation, IPTG was used to induce the cells for four hours. In order to obtain the greatest expression of RbpD1, we induced the cells with 1 mM and 2 mM IPTG, to see if higher amounts of the inducer would produce more protein. Figure 3.15 shows that more protein was actually obtained when the cells were induced with the lower concentration of IPTG; possibly the BL21 cells were better able to survive with the lower concentration.

The purification of RbpD1 is explained in Section 2.6.2. Once the histidine-tagged protein had been bound to the Ni-NTA agarose column (Qiagen), non-specific proteins were washed from the column using 20 mM imidazole in MCAC buffer. RbpD1 was then eluted using 100 mM imidazole. The imidazole has a similar structure to the histidine residues, and thus sets up a competition with the 6X-histidine tag. At the 100 mM imidazole concentration, the histidine tagged protein was successfully competed off of the nickel-NTA column. Lower concentrations of the imidazole were tried, but they caused partial elution of the protein. The final purification is shown in Figure 3.16.

Figure 3.14: pRLT3 digested with NcoI and HindIII. In order to indicate that the *rbpD1* gene had been correctly inserted in the pTRC99A plasmid, we digested putative pRLT3 clones with NcoI and HindIII. The digested plasmid DNA was electrophoresed on a 2% agarose gel and visualised with ethidium bromide. The first lane contains λ DNA digested with BstEII as a marker lane. The second lane contains a 100 Base-Pair Plus marker (MBI Fermentas) for the lower molecular weight bands. The pRLT3/NcoI/HindIII lane contains a 376 base-pair band corresponding to *rbpD1*, a 4176 bp pTRC99A band; the third band is from the pLysS plasmid.

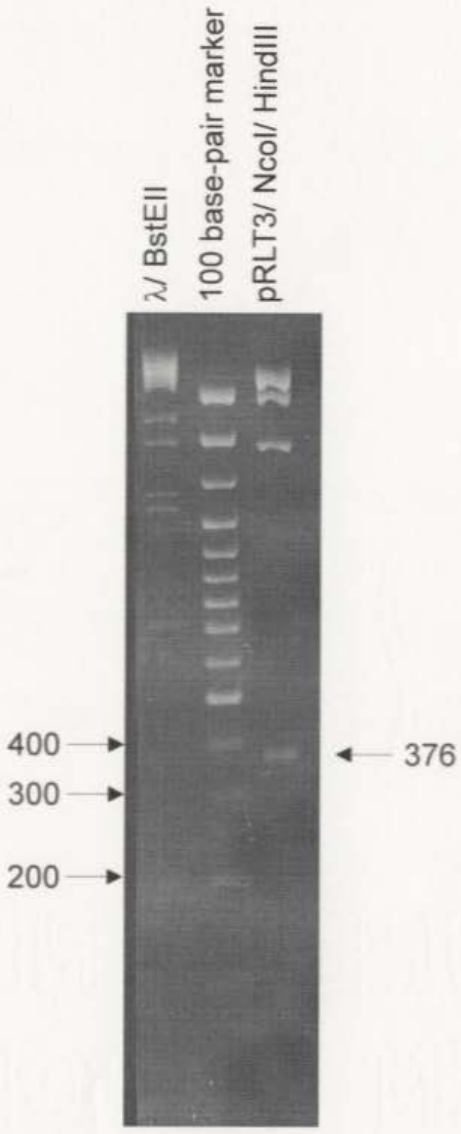

λ/ BstEII
100 base-pair marker
pRLT3/ NcoI/ HindIII
400
300
200
376

Figure 3.15: RbpD1 expression under induction of 1 mM and 2 mM IPTG. Twenty-five ml cultures of *E. coli* BL21 were grown to an optical density of 0.5 and induced for four hours at 37 °C with 1 mM (lane 3) or 2 mM IPTG (lane 4). The cultures were shaken at 225 rpm. An uninduced culture was included for comparison (lane 2). The plasmid was maintained in each of the cultures with the use of 100 μM carbenicillin. Lane 1 is the molecular size marker. The band representing RbpD1 (open head arrow) is stronger in the 1mM IPTG lane than in the 2 mM IPTG lane.

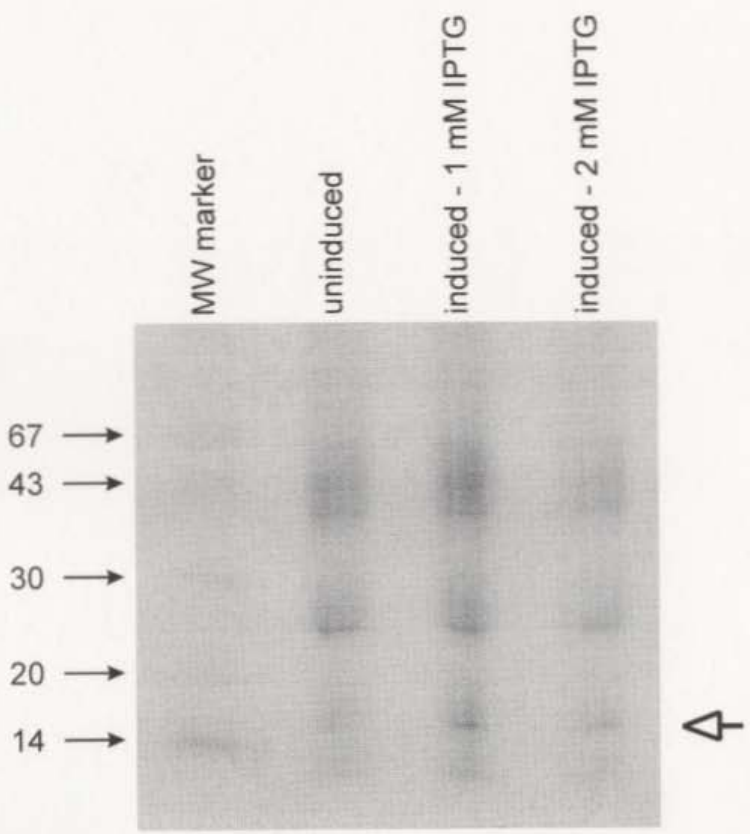
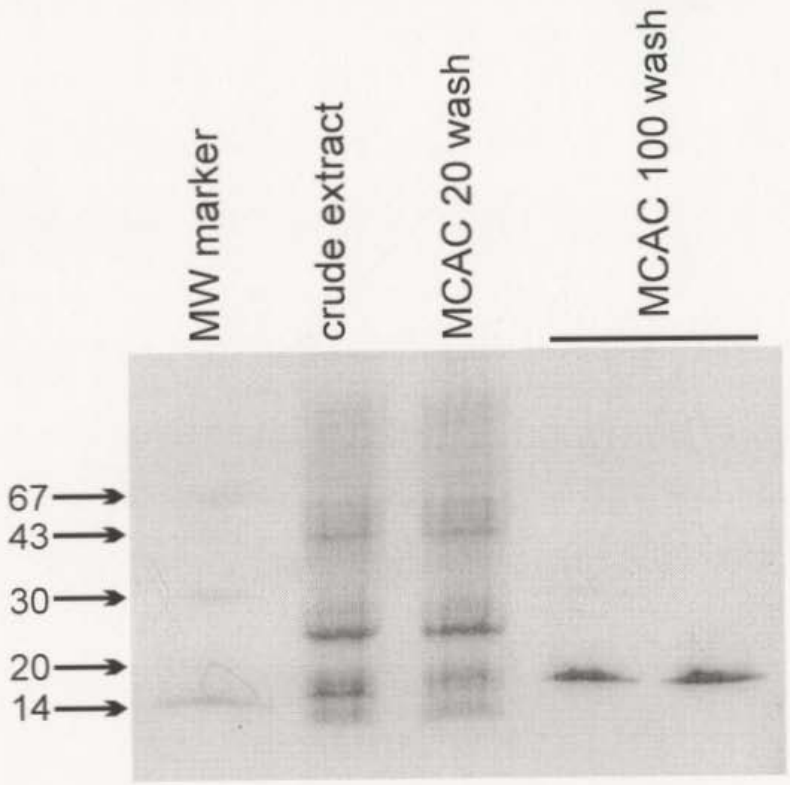

MW marker
uninduced
induced - 1 mM IPTG
induced - 2 mM IPTG
67
43
30
20
14

Figure 3.16: Expression and purification of RbpD1. The protein was expressed in *E. coli* BL21(pLysS)DE3 cells, using 1 mM IPTG for four hours at 37°C. Lane 1 contains the molecular weight marker. Lane 2 contains the crude extract from the induced cells, and lane 3 is the same extract after it has been applied to the Nickel-NTA column and washed with 20 mM imidazole. Lanes 4 and 5 contain the second and fourth RbpD1 fractions eluted from the column with 100 mM imidazole. RbpD1 was eluted from the column in ten 0.5 ml fractions. Four µl of each of the measured fractions was mixed with an equal volume of SDS loading buffer and electrophoresed on a PhastGel. Approximately 4 mg of RbpD1 were recovered from a 500 ml culture. Protein concentration was determined by comparing the intensity of a RbpD1 band to known concentrations of lysozyme electrophoresed on the same gel (data not shown).


MW marker
crude extract
MCAC 20 wash
MCAC 100 wash
67
43
30
20
14

A common problem reported when purifying histidine-tagged proteins is that a number of high-molecular weight bands originating from the bacteria tend to be co-eluted from the Ni-NTA column with the protein of interest (anonymous, 1995). However, there appears to be only one band corresponding to RbpD1 in the elution lane of Figure 3.16, so the extra bands do not seem to have been a problem for this particular purification.

Now that we had successfully expressed and purified both the wild-type and histidine-tagged RbpD proteins, we were in a position to begin to characterise the proteins. This involved determining a binding constant for the protein and determining the specificity of the protein towards specific RNA sequences. These experiments will be discussed in Chapter 4.

Chapter 4:

Results and Discussion -

Characterisation of RbpD

4.1 Binding to agarose-bound polymers

In order to demonstrate the RNA-binding activity of the RbpD and RbpD1 proteins, a binding experiment was done using agarose-bound nucleotide homopolymers, as described by Sugita and Sugiura (1994), and adapted by Tom Belbin (1999). The method is described in Section 2.8.1. In each binding experiment, 2 μg of protein was used with 20 μg of RNA polymer. The binding reactions were incubated for ten minutes at 4°C. The results are shown in Figure 4.1. Both RbpD and RbpD1 bound to poly(G) and poly(U) strongly. Poly(C) was not bound; poly(A) was only slightly bound. The *Synechococcus* sp. strain PCC 7942 RbpA protein also demonstrated slight binding to poly(A) and strong binding to poly(U) and poly(G) (Belbin *et al.*, 1998).

Sugita and Sugiura (1994) reported that two RNP-type RNA-binding proteins from *Synechococcus sp.* strain PCC 6803 bind to poly(G) and poly(U), but not to poly(A) or poly(C). These results match exactly those of the polymer binding studies with plant chloroplast RNA-binding proteins (Hirose *et al.*, 1993) and are similar to our results, with strong poly(G) and poly(U) binding in all cases. Unlike our study, poly(A) was not bound, but binding to poly(A) in our study was very weak and may not have been detected without the modification to the protocol introduced by Tom Belbin (1999). The method of Sugita and Sugiura involved aspirating the supernatant from the agarose-bound polymers after each washing step, which introduced the possibility of aspirating off some of the beads along with the liquid. Tom Belbin introduced the idea of

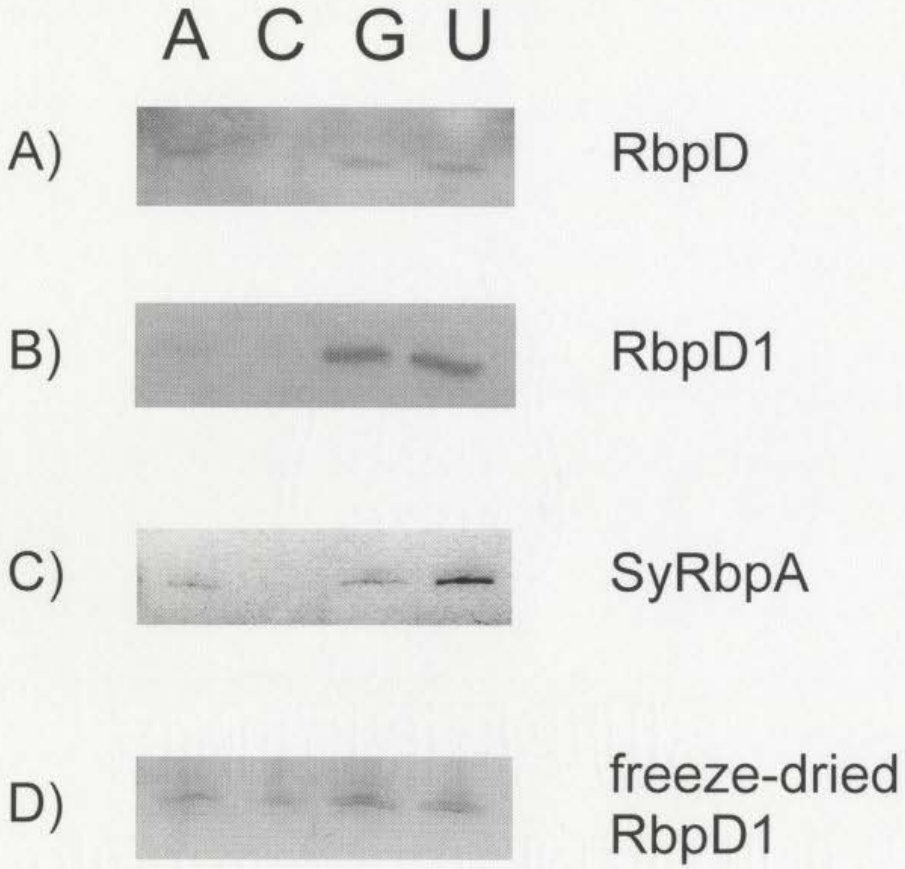
Figure 4.1: Binding of RNA-binding proteins to RNA homopolymers. In each lane 2 μg of RNA-binding protein was bound to 20 μg of RNA homopolymer complexed with Sepharose-4B, polyacrylhydroazido-agarose or agarose beads. Bound protein was separated by SDS polyacrylamide gel electrophoresis using a PhastSystem (Pharmacia-Amersham) SDS gel using a silver stain. Each row represents binding a different protein as follows.

A) RbpD from *Anabaena* 7120

B) RbpD1 from *Anabaena* 7120

C) RbpA from *Synechococcus* 7942 (courtesy of Tom Belbin)

D) Freeze-dried RbpD bound to the polymers. After freeze-drying, RbpD1 lost its binding specificity towards the agarose bound polymers and bound to poly(C) as well poly(A), poly(G) and poly(U).


A C G U
A)
RbpD
B)
RbpD1
C)
SyRbpA
D)
freeze-dried
RbpD1

performing all the washing steps using a Wizard minicolumn. The column was able to retain the beads, minimising the chance of polymer loss, but allowing fluid through. In the final elution step, when the SDS loading buffer was added to the beads, any bound protein passed through the column with the buffer leaving behind the beads.

Sato (1995) found that the RNA-binding proteins RbpA1, RbpA2, RbpB, RbpC Rbp107 and Rbp99 in *Anabaena variabilis* M3 bound strongly to poly(U), less strongly to poly(G), and less strongly still to poly(A). Poly(C) bound the most weakly of all. However, the *Anabaena variabilis* M3 RNA-binding protein RbpD (which corresponds to RbpB of *Anabaena* 7120), lacks a glycine-rich region and appeared to bind to all four of the polymers equally. These results are similar to our results except that we did not observe any binding to Poly(C), possibly because the binding affinity is too low. The difference among Sato's results, our results, and those of Sugita and Sugiura (1994) could be due to the different method used to detect binding. Sato used Western blotting to detect the presence of the proteins rather than silver staining, which may have been a more sensitive technique. We were able to detect poly(A) binding although we used silver staining, but the band was faint.

As mentioned previously, we had some concerns about the positioning of the His-tag in RbpD1. Loop 3 is the most variable and flexible part of the RRM motif in RNA-binding proteins (Shamoo *et al.*, 1996) and is known to interact with the RNA when it is bound to the protein. In fact, NMR (Allain *et al.*, 1996) and X-ray crystallography (Oubridge *et al.*, 1994) demonstrate that, at least in the U1A spliceosomal protein (which contains two RRMs), loop 3 of RRM1 protrudes through the hairpin of the bound RNA

(Figure 3.12). From our polymer binding experiments, it appears that RbpD1 retains the same binding specificity as RbpD, so the location of the His-tag in this case is not an issue.

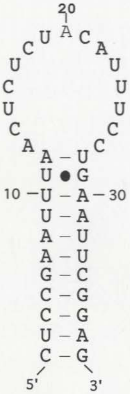
4.2 Binding to 5' – untranslated region of *rbpD* - DNA

The 5' end of the *rbpA1* gene was originally mapped by isolating the expressed RNA and using S1 nuclease mapping of the 5' end of the gene (Sato *et al.*, 1997). The deduction of other 5' untranslated regions was done using sequencing of gene fragments. Most of the known 5'- untranslated regions (5'-UTR) from cyanobacterial RNA-binding proteins are highly conserved (Belbin, 1999, Maruyama *et al.*, 1999). The most conserved portion of this region forms a stem-loop structure when the base-sequence is allowed to fold. The *rbpD* gene of *Anabaena variabilis* M3 (Sato and Nakamura, 1998) is the only gene in the species that does not contain this conserved sequence; it is also the only constitutively expressed cyanobacterial RNA-binding gene.

We obtained the sequence for the 5' – untranslated region of the *rbpD* gene from Dr. Peter Lammers (New Mexico State University) and folded the sequence using the program Loop-D-Loop (D.G. Gilbert, Indiana University) for Macintosh computers. The conserved stem-loop from *rbpD* is shown in Figure 4.2.

The conserved structures in the 5'–UTR of the cyanobacterial RNA-binding protein gene suggested that RbpD might regulate its own expression by binding to the 5'-UTR, either in the ssDNA or RNA form. To test whether RbpD would bind to the 5'-UTR in the DNA form, an oligonucleotide was synthesised (Cortec) that contained the

Figure 4.2: Conserved portion of the 5' – untranslated region from *rbpD*. The diagram shows the portion of the *rbpD* 5'-UTR which is conserved in almost all of the known genes encoding cyanobacterial RNA-binding proteins. The stem loop shown in the figure is also conserved in all these 5'-UTRs.


20
10
30
5'
3'

conserved stem-loop portion of the sequence in the *rbpD* 5'-UTR. We tested the hypothesis that RbpD might bind the sequence in the DNA using an electrophoretic gel-shift mobility assay (EMSA) (Ludwig *et al.*, 1995) as described in Section 2.8.2. There was a possibility that the promoter region of the gene contained in the 5'-untranslated region might be involved in autoregulation of RbpD expression or in the regulation of *rbpD* gene expression. Two things must occur for the EMSA to be successful (providing that binding occurred). The first is that the lifetimes of the complexes must exceed the time it takes for the free DNA to completely leave the well of the gel. Otherwise, the equilibrium between the complex and free probe will change, promoting the dissociation of the complex. The second requirement is that the DNA-protein complex must experience a shift on binding to the protein (Rundlett and Armstrong, 1997; Fried, 1989).

Some of the conditions that were considered while designing the EMSA experiment were the amount of Mg^{2+}, the salt concentration of the running buffer and the temperature of the gel. The amount of Mg^{2+} is critical since this can stabilise the interaction between protein and DNA. The salt concentration of the running buffer is important as well. Too high a salt concentration will disrupt the binding interaction and can disrupt the movement of the complex into the gel matrix. Too low a salt concentration can also disrupt the complex. The EMSA gels were run at 4°C in order to stabilise any interactions between the DNA and RNA (Kerr, 1995; Buratowski and Chodosh, 1999; Klaff *et al.*, 1997).

The oligonucleotide was end-labelled with [γ-^{32}P] ATP (3000 Ci/mmol), and we attempted EMSA experiments with RbpD under the conditions described by

Laird-Offringa and Belasco (1995) - a protocol which called for 0.012 fmol labelled DNA. We used between 0.5 and 1000 nM RbpD for the experiments. The labelled DNA was added after the protein in the reaction tubes to decrease any non-specific binding which might occur (Demczuk, *et al.*, 1990). When no bands were observed on the gel, the amount of labelled DNA used was increased to 0.12 fmol, at which point bands were seen but no shift in mobility (Figure 4.3), indicating that the protein was not binding to the DNA. The reason for the smeared appearance of the bands in the gel is not known, but they are also present in the blank lane to which no protein was added.

4.3 Binding of 5' – untranslated region of *rbpD* - RNA

The failure to observe binding of RbpD to the conserved portion of the 5'-untranslated region suggested that regulation of RbpD expression did not occur through binding of RbpD to DNA. We decided to test whether RbpD could autoregulate its synthesis through binding the conserved stem-loop after it had been transcribed into RNA.

In this set of experiments, two complementary oligonucleotides, which corresponded to the stem-loop structure of the 5'-UTR, were synthesised. The oligonucleotides contained recognition sequences for the EcoRI and SalI 5'-ssDNA overhangs. The two oligonucleotides were annealed together and then ligated into pTZ18R (Table 2.1) to form pRLT7. This allowed the transcription of the 5'-UTR using the Ambion T7 Megashortscript system (Ambion, 1997) (See Figure 4.4).

Figure 4.3: Binding of RbpD to the 5' - untranslated region of *rbpD* in the DNA form. An oligonucleotide containing the 5'-UTR of *rbpD* was end-labelled with [γ-^{32}P] ATP (3000 Ci/mmol) and subjected to an electrophoretic mobility shift assay with nanomolar amounts of RbpD (0, 0.5, 10, 50, 100 and 1000 nM). Although a number of bands appear in the gel, there was no obvious shift in mobility. None of the bands changed in intensity as the amount of protein increased.

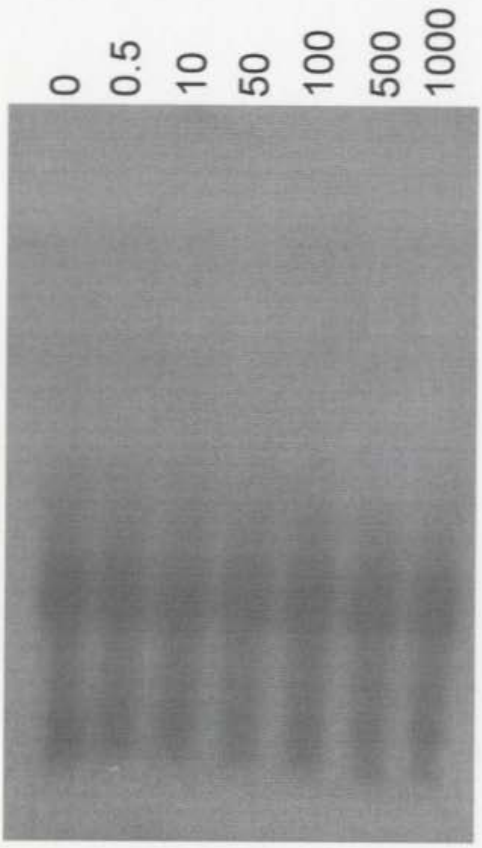
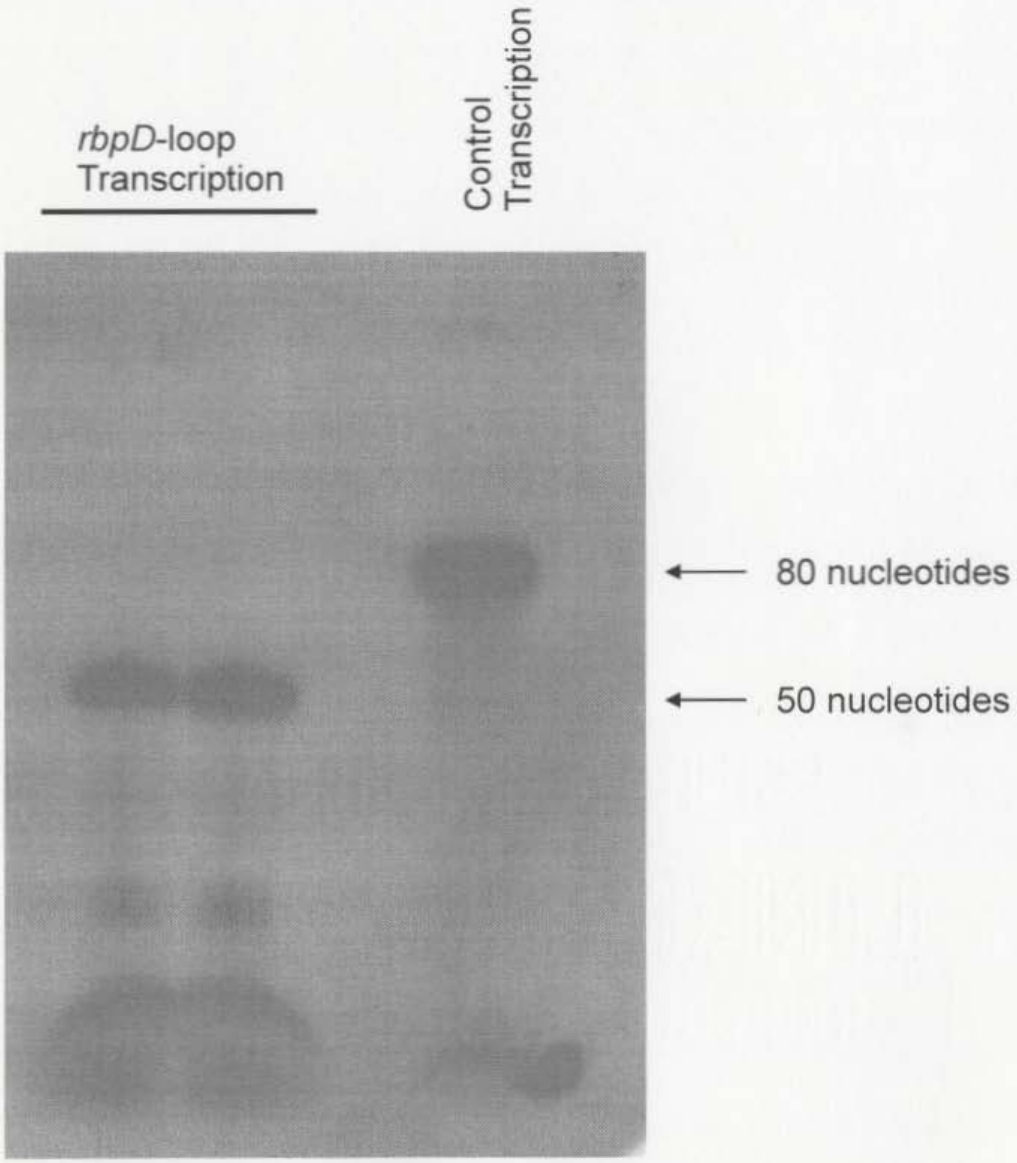

0
0.5
10
50
100
500
1000

Figure 4.4: Transcription of the 5' – untranslated region of *rbpD*. Transcription of the 5'-UTR using the T7 Megashortscript system (Ambion). The transcription product was loaded on an 8% acrylamide, 8 M urea gel, and electrophoresed at 4°C at 150V. The transcript is about fifty nucleotides long, and internally labelled with [α-^{32}P] ATP (3000 Ci/mmol). An 80 nucleotide cDNA from 18S rRNA inserted downstream of a T7 promoter was transcribed as a control to ensure that all the components of the system worked.

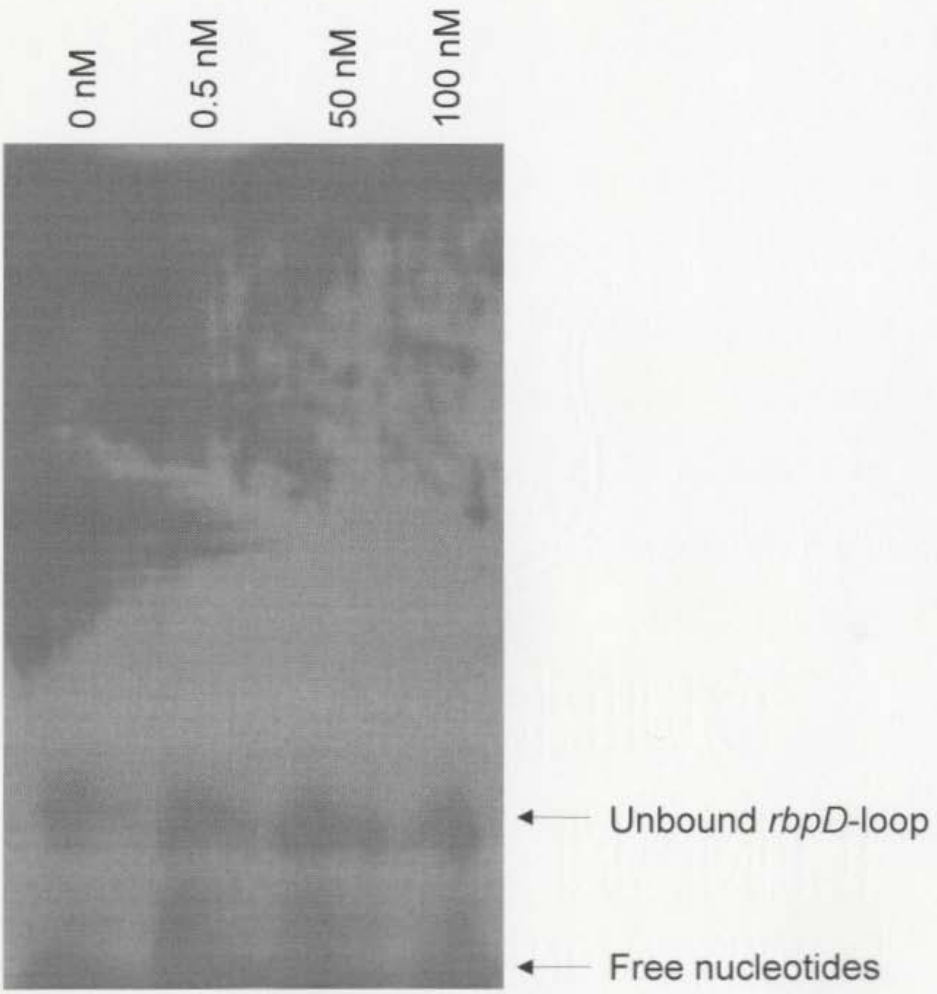

rbpD-loop
Transcription
Control
Transcription
80 nucleotides
50 nucleotides

The transcript was electrophoresed on an 8% acrylamide, 8 M urea gel and purified by the method of Binkley and Tuerk (personal communication) as described in Section 2.8.3. The 8 M urea denaturant prevented RNA degradation during the electrophoretic process. The purified RNA transcript was stored in nanopure water in order to avoid affecting the salt concentrations of buffers used in further experiments.

An electrophoretic mobility shift assay was attempted using 0.12 fmol transcribed RNA, and 0 – 100 nM RbpD. In the first EMSA experiment, TENT buffer was used in the RNA-protein binding reaction. This buffer had been used successfully in the binding of U1A to its cognate RNA by Laird-Offringa and Belasco (1995). However, no binding was observed; the only bands observed in the gel were those corresponding to the RNA band as seen in the lane containing no protein (Figure 4.5). We then switched to the buffer of Sugita and Sugiura (1994) which we had used sucessfully in polymer binding experiments (Section 4.2). The use of this buffer gave the same result – no bands in the gel except those corresponding to the lane containing RNA only (data not shown).

Buratowski and Chodosh (1999) described the effects of different running buffers on the stability of different complexes. The kind and concentration of salts in the buffers can strongly affect whether a complex is observed in the gel or not. In most of the binding experiments, we used a Tris-Glycine buffer (Section 2.8.2) which had been used by Laird-Offringa and Belasco (1995; 1996), and was also one of the buffers suggested by Buratowski and Chodosh (1999). When no binding was observed, we attempted to repeat some of the EMSA experiments on 8% acrylamide gels using TBE buffer. Once again, no binding was observed. Finally, Buratowski and Chodosh suggested that adding

Figure 4.5: Binding of RbpD to the 5' – untranslated region in the RNA form. 0, 0.5, 50 and 100 nM RbpD were incubated with 0.12 fmol transcribed, purified RNA at 25°C for 20 minutes in TENT buffer. Free nucleotides and unbound RNA ("unbound *rbpD*-loop") are present in each lane. None of the lanes appear to have bands above the unbound RNA.


0 nM
0.5 nM
50 nM
100 nM
Unbound *rbpD*-loop
Free nucleotides

2.5% glycerol to the gels would help to stabilise any protein-DNA (or in this case, protein-RNA) complexes formed. Even when this was tried with the EMSA experiments, no binding was observed.

We also tried filter binding experiments (Schneider *et al.*, 1993; Hall and Kranz, 1995; Carey *et al.*, 1983) since this method uses more RNA and protein, which should have been easier to detect. Still no binding was seen. However, Sato and Nakamura (1998) also found that none of the RNA-binding proteins that they tested could bind to the 5'-UTR of *Anabaena variabilis* M3 *rbpA1* gene (which corresponds to RbpC in *Anabaena* 7120). On further examination, Sato found that a kanamycin cassette could be inserted into the *rbpA1* gene *in vivo*, and transcription would still be induced under cold shock conditions, demonstrating that RbpA1 is not involved in regulating its own expression. This suggests that RbpD may not in fact be involved in its own regulation.

4.4 Binding of RbpD to poly(U)

From the polymer-binding experiments in Section 4.1, we knew that RbpD is capable of binding to poly(U), or at least to agarose-bound poly(U). We decided to perform an experiment to determine a basic binding constant of RbpD for poly(U). If we had a binding constant for RbpD, then we could alter the conditions of the binding experiments in order to optimise the binding, and we might be able to detect an interaction between RbpD and RNA for which it had even a weak affinity.

In order to determine the binding constant of RbpD for poly(U) we first needed to have a way to detect the polymer. Secondly, we needed a defined size for the polymer in

order to quantify the amount of binding that was occurring. In order to do this, poly(U) was end labelled with [γ-^{32}P] ATP, and separated by gel electrophoresis. Poly(U) is heterogeneous in size, so all that appeared on the gel was a smear (Figure 4.6A). A slice was cut from the gel in order to purify a small defined size range of molecules. Degraded RNA is often found in the lower part of the gel (because it is smaller and runs faster than the full length fragment). To avoid this degraded RNA, we wanted a fragment that ran fairly high on the gel. However, we did not want too large a fragment or the shift we should see from the binding of the RNA to RbpD might not be visible. The transcribed RNA from the SELEX procedure (see Section 4.5) was 92 base-pairs long, which seems to fill the functional requirements of the fragment to be bound. Using the SELEX transcript as a molecular marker, a slice of gel containing labelled poly(U) lane was cut from the gel and purified as described in Section 2.8.3.

Binding experiments were done using the same conditions as described in Section 4.3 (although only the Tris-Glycine running buffer and Sugita and Sugiura's binding buffers were used), but no binding was seen. All bands on the gel corresponded to the band seen in the lane containing RNA only (Figure 4.6B).

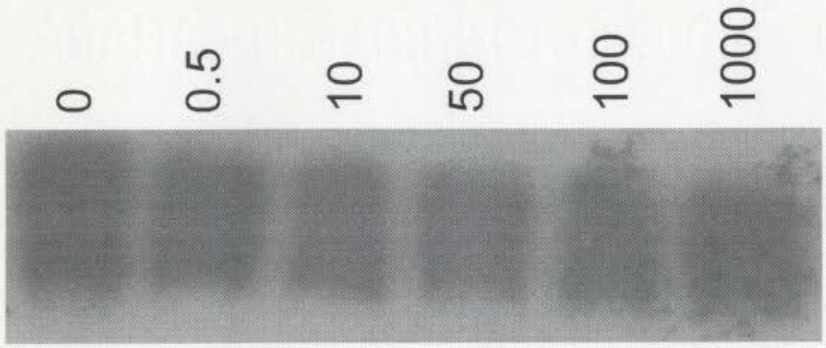
4.5 SELEX

In order to determine the sequence of RNA which will be bound by RbpD, SELEX (Systematic Evolution of Ligands by EXponential enrichment) was performed. The procedure is explained in Figure 1.5. A 69 nucleotide oligonucleotide containing a central region of forty random nucleotides was synthesised (Operon Technologies,

Figure 4.6: PolyU binding experiment.
A) PolyU (Sigma) was end-labelled with [γ-^{32}P] ATP and separated on a polyacrylamide gel. Because the polymer is heterogeneous in length, it appears as a smear on the autoradiograph.
B) Purified polyU from a gel slice taken from a gel corresponding to "A" was subjected to binding with nanomolar amounts of RbpD (0, 0.5, 10, 50, 100 and 1000 nM). The binding was done at 25°C for 10 minutes, using Sugita and Suguira's (1994) binding buffer , and the reactions run on a 8% non-denaturing polyacrylamide gel. No gel shift was observed, indicating that no binding was taking place.

A

B



California). From this, RNA molecules containing all the possible combinations of the forty nucleotide region were transcribed using T7 RNA polymerase. In theory, this RNA is incubated with the protein of interest, which will bind to those RNA molecules with affinity for it. The bound RNA is isolated using nitrocellulose filter binding and reverse-transcription PCR (RT-PCR) is then used to amplify the products to prepare a new pool of templates for the next round of transcription and binding. As the number of rounds of binding increases, RNA sequences with higher affinity should be selected, so that eventually only those with the highest affinity for the protein remain. This method is very powerful because it removes much of the bias in what sequences will be selected.

During the amplification portion of the SELEX method, there is always a possibility that artefacts will arise due to the formation of molecules with an enhanced replication efficiency. This is generally the result of deletions occurring during the amplification; the shorter molecules replicate faster than the full-length molecules. We gel-purified the desired fragment after each round in order to compensate for this possibility (Szostak and Ellington, 1993).

Had SELEX experiment worked perfectly, then we would have seen an increasing affinity for the RNA ligands by the protein as the number of rounds of SELEX increased. This would appear as a increasing slope on a graph of percent total cpm (counts per minute) radioactivity bound (from the labelled RNA) versus concentration protein added. The slope of the graph in each round of SELEX should always be positive.

We performed three rounds of SELEX in our first attempt to discover the sequence of the cognate RNA for RbpD. In each of these rounds we did not see any

increase in binding over the background level. Not only was the slope of each of the lines zero (Fig 4.7A), but the slope did not increase with the number of rounds of SELEX. We felt that this might be due to one of two factors. The first possibility was that the RNA pool was still too large to see the binding. The second possibility was that the high background levels on the filter due to non-specific RNA binding might be obscuring visualisation of the binding (See Figure 4.7A). After each round however, a band of the correct size (92 base-pairs) was isolated after the bound RNA had been eluted from the filter, reverse transcribed and PCR amplified (Figure 4.7B). We performed a control experiment by passing RNA through the nitrocellulose filter in the absence of protein. RNA that was non-specifically bound to the nitrocellulose was then extracted from the filter and subjected to reverse transcription PCR (RT-PCR). A 92 base-pair band was still observed in the control lane (Figure 4.7B lanes 2 and 3) demonstrating that non-specific binding was occurring. Since the intensity of the bands in the control and experimental lanes was equal, we concluded that the level of background binding to the filters was too high to permit detection of any specific protein interactions.

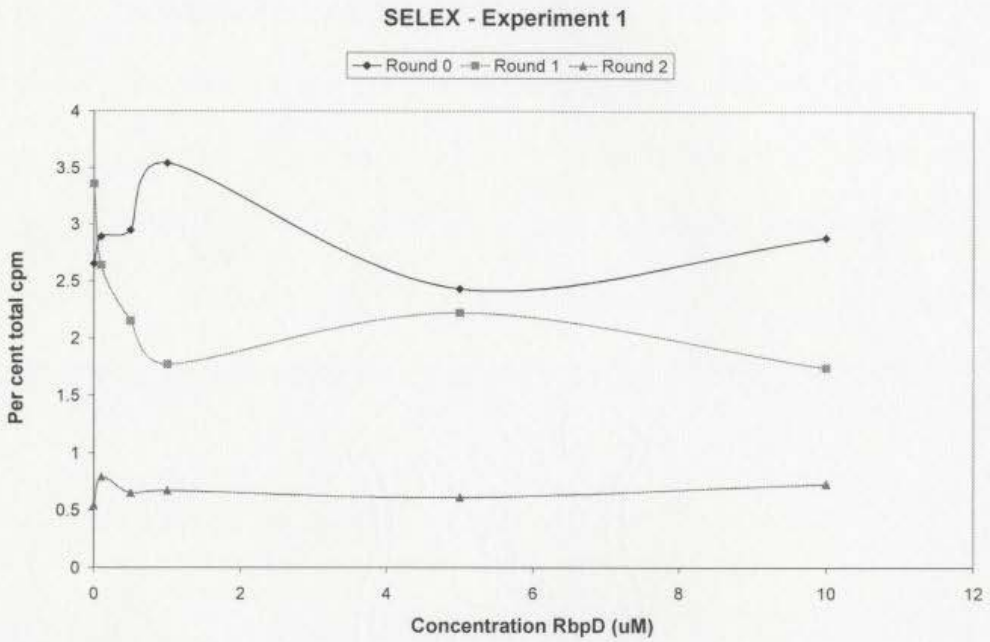
We attempted a second SELEX experiment, with the modification that one or two phenol:chloroform extractions were carried out on the samples after the RNA was purified from the acrylamide gel. It was hoped that this would decrease the background levels of RNA binding to the nitrocellulose filters. In addition, in these new rounds of SELEX, an incubation temperature of 4°C was used rather than 25°C because this was closer to the conditions used in the polymer-binding experiments.

Figure 4.7: SELEX – attempt #1

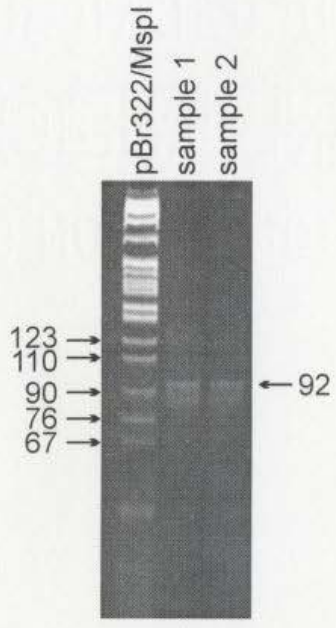
A) Binding curves from the first three rounds on the first attempt at SELEX. The graph compares the total protein in each binding experiment with the total percentage of RNA bound. A curve was calculated for each of the three rounds of SELEX performed. The background level of RNA was between 0.5 and 3.3% in each round of SELEX.

B) After each round of a selection, RNA was eluted from the filter and purified. It was then amplified using RT-PCR, and was electrophoresed on an 8% polyacrylamide gel and viewed by staining with ethidium bromide. This photo shows the 92 base-pair fragment recovered after Round 1. Similar results were obtained from Rounds 2 and 3. Sample 1 and Sample 2 represent samples from each of the duplicate PCR reactions incubated.

A


SELEX - Experiment 1
Round 0
Round 1
Round 2
Per cent total cpm
Concentration RbpD (uM)
0
0.5
1
1.5
2
2.5
3
3.5
4
0
2
4
6
8
10
12

B


pBr322/MspI
sample 1
sample 2
123
110
90
76
67
92

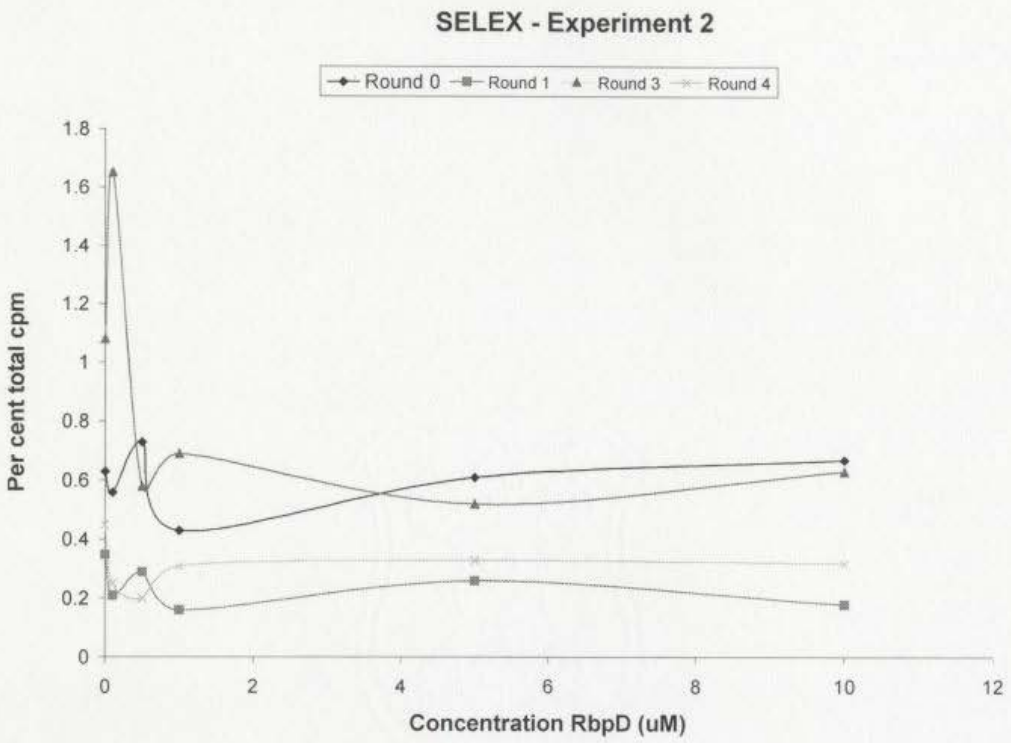
Four rounds of SELEX were completed using the new conditions. While much less background binding was observed in the binding curves, no significant binding above the background levels was seen in any of the rounds (See Figure 4.8A). We completed four rounds of selection as it often takes this number to observe the effects of the amplification on bound RNA sequences. In these four rounds, there was still enough background binding of RNA to the nitrocellulose membrane in the absence of protein to see a band after purification and reverse-transcription PCR (Figure 4.8B, lanes 3 and 4).

Figure 4.8: SELEX – attempt #2

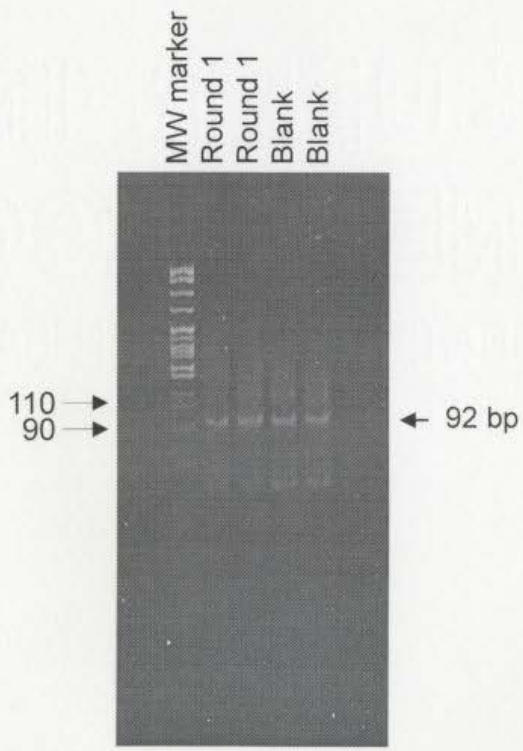
A) Binding curves from the four rounds on the second attempt at SELEX. A binding curve was performed for each round of SELEX, comparing the proportion of total RNA bound by increasing amounts of RbpD. The binding curve for Rounds 0, 1, 3 and 4 are presented in the graph. Each binding reaction was incubated in Sugita and Sugiura's buffer at 4°C for 15 minutes before the bound RNA was isolated on a nitrocellulose filter for scintillation counting.

B) As with the previous figure, the Reverse-transcribed PCR product from the RNA retained on the filter was sized on a 8% non-denaturing polyacrylamide gel. Although little binding was detected in the radioactively labelled reactions, a 92 base-pair fragment was still observed after RT-PCR amplification.

A


SELEX - Experiment 2
Round 0
Round 1
Round 3
Round 4
Per cent total cpm
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
Concentration RbpD (uM)

B


MW marker
Round 1
Round 1
Blank
Blank
110
90
92 bp

Chapter5:

General Discussion

5.1 General Discussion

Given our knowledge of the likely structure of RbpD, it seems reasonable to hypothesise that the RRM motif in RbpD will be involved in the binding of RNA. Specifically, the basic and aromatic amino acids in RNP1 and RNP2 are probably most involved in the alignment and interaction with RNA molecules (See Figure 1.2). The RNP1 motif includes two phenylalanines at residues 44 and 46 in RbpD, which correspond to the phenylalanine residues at residues 54 and 56 in U1A and are the most conserved residues in the RRM domain (Oubridge *et al.*, 1994). These residues are probably most involved in the interaction with RNA ligands. Since the insertion of the hexa-histidine tag in Loop 3 of RbpD did not affect its specificity of binding to the agarose-bound nucleotide polymers, this area of the protein is probably not involved in the interaction of RbpD with RNA. U1A does interact with the RNA through its third loop, but it may not be used by all the RNP-type RNA-binding proteins in their interactions with RNA. In hnRNP C for instance, it is mostly the C-terminal end of the RRM that confers specificity of binding (Dreyfuss, 1993). The C-terminal of the RbpD RRM contains a number of basic residues (Arg70, 80 and 83 and Lys73, 76 and 78) and an aromatic residue (Trp67) which may interact with and determine the sequence of bound RNA.

Given the results of Shahied-Milam *et al.* (1998) that the hnRNP C protein RRM is more involved in protein-protein interactions than in RNA binding and that another domain of the protein is responsible for the RNA binding, the possibility cannot be ignored that the RRM in RbpD may be of secondary importance in RNA binding. This

possibility is lent credence by the claim of Alba and Pagès (1998) that RGG boxes in many plant glycine rich domains are involved in RNA binding. Certainly the glycine rich domain of RbpD contains a number of arginines (Arg88, 99, 107 and 108), and no acidic amino acids to repel RNA binding (Figure 1.2). Two of the arginine residues lie just previous to a row of glycine residues, causing the sequences to resemble RGG boxes.

Although RbpD bound to agarose-bound poly(U) and poly(G), it failed to bind to any of the other RNA molecules presented to it, including free poly(U). This is somewhat disturbing since it indicates that RbpD might be binding to something in the agarose/ Sepharose-4B/ polyacrylhydrazido-agarose rather than the polymers. However, the specificity of the binding of RbpD to poly(U) and poly(G) over poly(A) and poly(C) does indicate that the binding probably was specific to the polymers themselves. It also indicates that RbpD has the conserved binding specificity of most plant and cyanobacterial RNA-binding proteins (Ye and Sugiura, 1992; Ludevid *et al.*, 1992; Maruyama *et al.*, 1999).

The SELEX experiments should have been the most informative set of experiments performed if we could have detected binding. Although binding was not detected, it may have been present in too low an amount to be detected over the background binding of RNA to the nitrocellulose filter. It is possible that we simply out competed any RbpD-bound RNA amplification with nitrocellulose-specific RNA amplification. Another possibility is that the preferred RNA sequence of RbpD may not be well suited to one of the reverse transcriptase steps, or PCR amplification steps (Klug and Famulok, 1994). If something in the sequence or folding of the preferred RNA

ligand is not as suited to these steps as the nitrocellulose specific sequences, then we may not see an amplified fragment corresponding to protein-bound RNA although it was present in the original pool. Given more time, it would be interesting to vary the conditions of the reverse transcription and PCR, both in salt balance and, for the PCR, in temperatures and times, to see if it alters the results of the SELEX.

It would also be interesting to try both the poly(U) binding experiment and SELEX procedures using only the RRM portion of RbpD rather than the full protein. It may change the results of the SELEX experiments especially. For instance, in hnRNP C, the RRM alone binds to a sequence of 5 to 6 Us most strongly (Gorlach *et al.*, 1994; Swanson and Dreyfuss, 1988), but when the whole protein was examined for RNA-binding specificity, hnRNP C preferred blocks of RNA about 700 nucleotides long, a 116 nucleotide ribosomal RNA transcript, as well as the U1, U2, and U6 snRNAs (Shahied-Milam *et al.*, 1998) rather than the poly(U) tract. The full hnRNP C protein did not bind so strongly to the SELEX oligonucleotides, possibly because of the short lengths of the RNA targets (Shahied-Milam *et al.*, 1998). It may be that the RbpD RRM will also bind better to a SELEX target than the whole RbpD protein.

There may be a few things contributing to the lack of binding of RbpD to the RNA in the other assays. The first and most obvious is that perhaps the protein is folding incorrectly after translation. This is a known problem when using an *E. coli* expression system (Balbas and Bolivar, 1990). However, many of the post-translational modifications that would normally cause misfolding of the protein do not occur in RbpD because the features that would cause the misfolding are not present. For instance, RbpD

does not contain a leader peptide, nor does it contain any cysteine residues which often cause misfolding through inappropriate disulphide bond formation. The fact that in the agarose-bound RNA polymer experiments, RbpD bound to poly(G) and poly(U) preferentially over poly(A) and poly(C) indicates that RbpD is acting like the RNA-binding proteins isolated by Sugita and Sugiura (1994) and Sato (1995). Because the structure of the RRM should be similar in RbpD and other RNP-type proteins, circular dichroism studies could be undertaken in the future in order to confirm that RbpD is maintaining its native conformation even when expressed in bacteria. These studies would determine the overall folding and secondary structure of RbpD.

On the same line of reasoning, the lack of binding might be due to the lack of a posttranslation modification event such as a phosphorylation. A 28 kDa RNA-binding protein in spinach chloroplast is dramatically affected in its binding affinity to its cognate RNA by a phosphorylation event. Many hnRNP core proteins are also modified by post-translational phosphorylation (Lisitski and Schuster, 1995). In addition, some proteins in *Anabaena* 7120 require post-translational modification in order to be active. For example, Zhou *et al.* (1998a) propose that HetR, a protein required for heterocyst differentiation may need to be modified by phosphorylation in order to set off the cascade which will start the differentiation process. Because there are no other examples of other prokaryotic RNP-type RNA-binding proteins that have been found, it is too early in this area of research to dismiss the possibility. Very few examples of RNA-binding proteins have been found outside of the cyanobacteria, and even within the cyanobacteria, very little characterisation has been done on the proteins to determine whether

phosphorylation is present or not. Perhaps RbpD requires the modification in order to interact with other proteins to start the process.

It should be fairly simple to determine whether phosphorylation is present in RbpD. Given that the RbpD protein has been isolated, a similar protocol to Lisitski and Schuster (1995) could be followed. These authors incubated 28RNP protein from spinach chloroplasts with at least 100 ng kinase preparation (either maize casein kinase or extracted protein kinase from spinach chloroplasts) in a buffer containing 20 μM [γ-^{32}P] ATP (500 cpm/pmol). The phosphorylated proteins were fractionated by SDS-PAGE and autoradiography. A similar procedure could be carried out to determine whether RbpD is also phosphorylated and to what level the protein might be phosphorylated.

Most likely however, is that RbpD may simply require the presence of a second protein to order to bind specifically to its cognate RNA. This would be very similar to the case of the U2 snRNP B (U2B") RNA-binding protein which is not able to bind to its cognate RNA, hairpin IV of U2 snRNA, unless it is complexed with the U2A protein. Within this complex, both proteins share an AUUGCA hexanucleotide recognition sequence, but each binds specifically to its own hairpin (Oubrige *et al.*, 1994). Another precedent is found with the hnRNP C protein. Shahied-Milan *et al.* (1998) have proposed that the RRM of one protein interacts with other C proteins to mediate tetramer formation rather than participate in the binding of RNA. It may be that RbpD requires binding to a second protein to have any affinity for a cognate RNA. This has been suggested to be a problem in many of the EMSA experiments performed with recombinant protein or protein made by *in vitro* transcription and *in vitro* translation because the protein has no

opportunity to form a heterodimer that would allow it to bind to DNA or RNA (Kerr, 1995). Because the protein was purified from *E. coli* rather than *Anabaena* 7120, RbpD would have no chance to form a complex with a secondary protein.

There are other cases in which no RNA-binding has been demonstrated for an RNA-binding protein. Even in the U1A protein, only the N-terminal RNA-binding domain is able to bind to RNA sequences. The C-terminal domain, for reasons discussed in Section 1.3.4, does not bind to U1, U2, U5 snRNAs, rA_{16}, rC_{16}, rU_{16}, $rA_3U_3GUA_4$, or random RNAs, although it is structurally and sequentially very similar to the N-terminal domain (Lu and Hall, 1995). Perhaps RbpD simply does not bind to RNA.

5.2 Future work

A major priority for future work on RbpD is to determine a binding constant for RbpD and RbpD1. This can be done by the method of Laird-Offringa and Belasco (1995) using poly(U). We know that the cyanobacterial RNA-binding proteins bind to polyU, so a binding curve should be fairly easy to do once the methodology is ironed out. Finding a binding constant for the protein will also allow the optimisation of the buffer conditions. For instance, in some RNA binding experiments, magnesium ions (Mg^{2+}) are eliminated altogether to reduce problems with RNases and the binding works well. (Laird-Offringa and Belasco, 1995; 1996). The RNA-binding buffer of Sugita and Sugiura (1994), which we used contains 2.5 mM $MgCl_2$. Sato has used 5 mM $MgCl_2$ (Sato and Nakamura, 1998). Since the Mg^{2+} concentration will affect how the RNA folds, this will be important to optimise.

Another experiment might include performing site-directed mutagenesis experiments, using again the Altered Sites II protocol, to put a histidine tag in loop 4 of RbpD and to remove the glycine-rich domain of the protein, both of which were started by Cynthia Slade in her honours work (Slade, 1998). Adding a histidine tag into loop 4 of the protein would allow comparisons to be made into the effects of the position of the tag on protein function. It would also provide an alternative if RbpD1 could not be reisolated from *in vitro* experiments, or if the loop 3 histidine tag affected interactions such as protein-protein interactions. Removal of the glycine rich region would allow direct studies on the affect of the glycine rich region on the specificity of RNA binding. It would also allow testing to see whether the region is active in protein-protein interactions *in vivo*.

Probably the most important area of further experiments is finding out whether RbpD does interact with another protein within the cyanobacterial cell, allowing it to bind to a cognate RNA. This could possibly be done using nickel pull-down experiments in which a protein is attached to a nickel column, and cell extract passed through the column. Those proteins with affinity for the attached protein (in this case RbpD) will bind to the column and can then be eluted with the attached protein. If the RbpD required a second protein in complex, and we could find it, then all of the binding experiments might work.

References

Alba, M.M., Pagès, M. (1998) Plant proteins containing the RNA-recognition motif. *Trends Plant Sci.* **3**(1):15-21.

Allain, F.H., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D., Varani, G. (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**:646-650.

Allen, P., Collins, B., Brown, D., Hostomski, Z., Gold, L. (1996) A specific RNA structural motif mediates high affinity binding by the HIV-1 nucleocapsid protein (NCp7). *Virology* **225**:306-315.

Ambion (1997) T7-MEGAshortscript: *In vitro* transcription kit for the large scale synthesis of short transcript RNAs. Ambion Inc. Austin, TX

Anonymous (1995) Purification of His-tag fusion proteins from *Escherichia coli*. *Trends Biochem. Sci.* **20**:285.

Avis, J.M., Allain, F.H-T., Howe, P.W.A., Varani, G., Nagai, K., Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.* **257**:398-411.

Balbas, P., Bolivar, F. (1990) Design and construction of expression plasmid vectors in *Escherichia coli*. *Methods Enzymol.* **185**:14-37.

Bancroft, I. Wolk, C.P., Oren, E.V. (1989) Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **171**(11):5940-5948.

Beesley, C.E., Smith, R.J., Temple, S.J., Lea, P.J. (1994) Cloning and nucleotide sequence of the gene encoding dinitrogenase reductase (*nifH*) from the cyanobacterium *Nostoc* 6720, *Biochim. Biophys. Acta*, **1219**:548-550.

Belbin, T.J. (1999) Characterization of a gene encoding an RNA-binding protein (*rbpA*) in the cyanobacterium *Synechococcus* sp. PCC7942. Ph.D. thesis, Memorial University of Newfoundland.

Belbin, T.J., Tremblay, R.L., Mulligan, M.E. (1998) RNA-binding proteins: expression, purification and RNA-binding studies. Poster presented at the VI Cyanobacterial Molecular Biology Workshop, Asilomar Conference Centre, pacific Grove, California, 24-27 July 1998.

Ben-Porath, J., Zehr, J.P. (1994) Detection and characterization of cyanobacterial *nifH* genes, *Appl. Envir. Mol. Biol.* **60**:880-887.

Bentley, R.C., Keene, J. (1991) Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B" protein and through interactions with U2 snRNP-A' protein. *Mol. Cell Biol.* **11**(4):1829-1839.

Binkley, J., Tuerk, C. One day purification of oligonucleotides. Protocol for SELEX from the lab of Larry Gold.

Birney, E., Kumar, S., Krainer, A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**(25):5803-5816.

Black, T.A., Wolk, C.P. (1994) Analysis of a Het$^-$ mutation in *Anabaena* sp. strain PCC 7120 implicates a secondary metabolite in the regulation of heterocyst spacing. *J. Bacteriol.* **176**(8):2282-2292.

Black, T.A., Cai, Y., Wolk, C.P. (1993) Spatial expression and autoregulation of *hetR*, a gene involved in the control of heterocyst development in *Anabaena*. *Mol. Microbiol.* **9**(1):77-84.

Bolivar, F., Backman, K. (1979) Plasmids of *Escherichia coli* as cloning vectors. *Methods Enzymol.* **68**:245-267.

Bradley, S., Carr, N.G., (1977) Heterocyst development in *Anabaena cylindrica*: the necessity for light as an initial trigger and sequential stages of commitment. *J. Gen. Microbiol.* **101**:291-297.

Buikema, W.J., Haselkorn, R. (1991a) Isolation and complementation of nitrogen fixation mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **173**(6):1879-1885.

Buikema, W.J., Haselkorn, R. (1991b) Characterization of a gene controlling heterocyst differentiation in the cyanobacterium *Anabaena* 7120. *Genes Develop.* **5**:321-330.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987) *BioTechniques* **5**:376-379.

Buratowski, S., Chodosh, L.A. (1999 revision) Mobility shift DNA-binding assay using gel electrophoresis. In "Current Protocols in Molecular Biology". ed. F.A. Ausubel *et al.* John Wiley and Sons Inc. U.S.A. pp. 12.2.1.-12.2.11.

Burd, C.G., Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**(15):1197-1204.

Buvoli, M., Cobianchi, F., Bestagno, M.G., Mangiarotti, A., Bassi, M.T., Biamonti, G., Riva, S. (1990) Alternative splicing in the human gene for the core protein A1 generated another hnRNP protein. *EMBO J.* **9**(4):1229-1235.

Cai, Y., Wolk, P. (1997) *Anabaena* sp. strain PCC 7120 responds to nitrogen deprivation with a cascade-like sequence of transcriptional activations. *J. Bacteriol.* **179**(1):267-271.

Carey, J., Cameron, V., de Haseth, P.L., Uhlenbeck, O.C. (1983) Sequence-specific interaction of R17 coat protein with its ribonucleic acid binding site. *Biochemistry* **22**:2601-1610.

Carpenter, C.D., Kreps, J.A., Simon, A.E. (1994) Genes encoding glycine-rich *Arabidopsis thaliana* proteins with RNA-binding motifs are influenced by cold treatment and an endogenous circadian rhythm. *Plant Physiol.* **104**:1015-1025.

Chamot, D., Magee, W.C., Yu, Esther, Owttrim, G.W. (1999) A cold shock-induced cyanobacterial RNA helicase. *J. Bacteriol.* **181**(6):1728-1732.

Chong, S., Shao, Y., Paulus, H., Benner, J., Perler F.B., Xu, M. (1996) Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. *J. Biol. Chem.* **217**(6): 22159-22168.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Brenner, J., Kucera, R.B., Hervonen, C.A., Pelletier, J.J., Paulus, H., Xu, M.-Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**: 277-281.

Clarke, N.D. (1994) A proposed mechanism for the self-splicing of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **91**:11084-11088.

Davis, J.P., Janjik, N., Javornik. B.E., Zichi, D.A. (1996) Identifying consensus patterns and secondary structure in SELEX sequence sets. *Methods Enzymol.* **267**:302-314.

Demczuk, S., Donovan, M., Franklin, G., Ohlsson, R. (1990) Order of probe and nuclear protein extract addition can determine specificity of protein-DNA complexes in tested mobility shift assays. *Nucleic Acids Res.* **19**(3):677-678.

Derbyshire, V., Belfort, M. (1998) Lightning strikes twice: intron-intein coincidence. *Proc. Natl. Acad. Sci.* U.S.A. **95**:1356-1357.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**:289-321.

Dreyfuss, G., Swanson, M.S., Pinol-Roma, S. (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Soc.* **13**:86-91.

Dunn, M.A., Brown, K., Lightowlers, R., Hughes, M.A. (1996) A low temperature-responsive gene from barley encodes a protein with single-stranded nucleic acid-binding activity which is phosphorylated *in vitro*. *Plant Mol. Biol.* **30**:947-659.

Elhai, J. (1999) Control over heterocyst differentiation. *CyanoNews* **14**(1):ISPP http://chemweb.urich.edu/~biology/cyanonews/cn14-1/ispp-het.htm

Evans, P.R., Oubridge, C., Jessen, T.H., Li, J., Teo, C.H., Pritchard, C., Ito, N., Nagain, K. (1993) Model of the interaction between the U1 A small nuclear ribonucleoprotein and U1 small nuclear RNA stem/loop II. *Biochem. Soc. Trans.* **21**:605-609.

Fogg, G.E. (1949) Growth and heterocyst production in *Anabaena cylinderica* Lemm. II. In relation to carbon and nitrogen metabolism. *Ann. Bot.N.S.* **15**:23-35.

Fraser, C.M. *et al.* (1998) Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**:375-388.

Fried, M.G. (1989) Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* **10**:366-376.

Fukami-Kobayashi, K., Tomoda, S., Go, M. (1993) Evolutionary clustering and functional similarity of RNA-binding proteins. *FEBS Lett.* **335**(2):289-293.

Giovannoni, SJ, Turner, S., Olsen, G.J., Barns, S., Lane, D.J., Pace, N.R. (1988) Evolutionary relationships among cyanobacteria and green chloroplasts. *J. Bacteriol.* **170**:3584-3592.

Gold, L., Brown, D., He, Y-Y., Shtatland, T., Singer, B.S., Wu, Y. (1997) From oligonucleotide shapes to genomic SELEX: Novel biological regulatory loops. *Proc. Natl. Acad. Sci. U.S.A.* **94**:59-64.

Gold, L., Polinski, B., Uhlenbeck, O., Yarus, M. (1995) Diversity of oligonucleotide function. *Annu. Rev. Biochem.* **64**:763-797.

Gold, L., Tuerk, C., Allen, P., Binkley, J., Brown, D., Green, L., MacDougal, S., Schneider, D., Tasset, D., Eddy, S.R. (1993) RNA: the shape of things to come. In R.F. Gesteland, J.F. Atkins (Eds.) The RNA World, Cold Spring Harbor Laboratory Press, pp. 497-510.

Gold, L. (1990) Expression of heterologous proteins in *Escherichia coli*. *Methods Enzymol.* **185**:11-14.

Gorlach, M., Burd, C.G., Dreyfuss, G. (1994) The determinants of RNA-binding specificity if the heterogeneous nuclear ribonucleoprotein C proteins. *J. Biol. Chem.* **267**(37):23074-23078.

Gorlach, M., Wittekind, M., Beckman, R.A., Mueller, L., Dreyfuss, G. (1992) Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.* **11**(9):1289-3295.

Graumann, P., Marahiel, M.A. (1996a) Some like it cold: response of microorganisms to cold shock. *Arch. Microbiol.* **166**:293-300.

Graumann, P., Marahiel, M.A. (1996b) A case of convergent evolution of nucleic acid binding modules. *BioEssays* **18**(4):309-315.

Grodberg, J., Dunn, J.J. (1988) *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* **170**: 1245.

Gubser, C.C., Varani, G. (1996) Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochem.* **35**:2253-2267.

Hall, K.B., Kranz, J.K. (1995) Thermodynamics and mutations in RNA-protein interactions. *Methods Enzymol.* **259**:261-281.

Hall, K.B. (1994) Interaction of RNA hairpins with the human U1A N-terminal RNA binding domain. *Biochem.* **33**:10076-10088.

Hanano, S., Sugita, M., Sugiura, M. (1996) Isolation of a novel RNA-binding protein and its association with a large ribonucleoprotein particle present in the nucleoplasm of tobacco cells. *Plant Mol. Biol.* **31**:57-68.

Haynes, S.R. (1992) The RNP motif protein family. *New Biologist* **4**(5):421-429.

Haselkorn, R. (1998) How cyanobacteria count to 10. *Science* **282**: 891-892.

Haselkorn, R., Buikema, W.J. (1992) Nitrogen fixation in cyanobacteria, In: G. Stacey, R.H. Burris, H.J. Evans, (Eds.), Biological Nitrogen Fixation, Chapman and Hall, pp. 166-190.

Heery, F. Gannon, F. and Powell, R. (1990) A simple method for subcloning DNA fragments from gel slices. *Trends Genet.* **6**:173.

Herdman, M. Janvier, M., Rippka, R., Stanier, R. (1979) Genome size of cyanobacteria. *J. Gen. Microbiol.* **111**:73-85.

Hesselgesser, J., Horuk, R. (1997) Alanine scan mutagenesis of chemokines. *Methods Enzymol.* **287**: 59-69

Hirose, T., Sugita, M., Sugiura, M. (1993) cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue specific alternative splicing. *Nucleic Acids Res.* **21**(17):3981-3987.

Hoffman, D.W., Query, C.C., Golden, B.L., White, S.L., Keene, J.D. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analysed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl. Acad. Sci. U.S.A.* **88**:2495-2499.

Holden, D.C. (1995) Sequence determination of genes encoding RNA-binding proteins in the cyanobacterial species *Anabaena* sp. 7120. B.Sc. (Honours) thesis, Memorial University of Newfoundland.

Iyer, R. (1994) The original Merlin Mini/Maxi-Prep. Computer Newsgroup: bionet.molbio.methds-reagnts.

Jenison, R.D., Gill, S.C., Pardi, A., Poliski, B. (1994) High-resolution molecular discrimination by RNA. *Science* **263**(5152):1425-1429.

Jessen, T-H., Oubridge, C., Teo, C.H., Protchard, ,C., Nagai, K. (1991) Identification of molecualr contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**(11):1447-1456.

Keene, J.D. (1996) Randomization and selection of RNA to identify targetsfor RRM RNA-binding proteins and antibodies. *Methods Enzymol.* **267**:367-383.

Kenan, D.J., Query, C.C., Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Soc.* **16**:214-220.

Kerr, L.D. (1995) Electrophoretic mobility shift assay. *Methods Enzymol.* **254**:619-641.

Kim, Y-J., Baker, B.S. (1993) Isolation of RRM-type RNA-binding proteins genes and the analysis of their relatedness by using a numerical approach. *Mol. Cell Biol.* **13**(1):174-183.

Kirhtein, J.D., Paerl, H.W., Zehr, J.P. (1991) Amplification and sequencing of a *nifH* segment from aquatic micro-organisms and natural communities, *Appl. Envir. Mol. Biol.* **57**:2645-2650.

Klaff, P., Mundt, S.M., Steger, G. (1997) Complex formation of the spinach chloroplast *psbA* mRNA 5' untranslated region with proteins is dependent on the RNA structure. *RNA* **3**:1468-1479.

Klug, S., Famulok, M., (1994) All you wanted to know about SELEX. *Mol. Biol. Rep.* **20**:97-107.

Komarek, J., Anagnostidis, K. (1989) Modern approach to the classification system of cyanophytes. 4-Nostocales. Arch. Hydrobiol, Suppl. 823 *Algological Studies* **56**:247-345.

Konarska, M.M. (1989) Analysis of splicing complexes and small nuclear ribonucleoprotein particles by native gel electrophoresis. *Methods Enzymol.* **180**:442-453.

Kraft, R., Tardiff, J., Krauter, K.S., Leinwand, L.A. (1988) Using mini-prep plasmid DNA for sequencing double stranded templates with Sequenase. *BioTechniques*, **6**(6):544-547.

Krecic, A., Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* **11**:363-371.

Kunkel, T.A., Bebenek, K., McClary, J. (1991) Efficient site-directed mutagenesis using uracil containing DNA. *Methods Enzymol.* **204**:125-139.

Laird-Offringa, I.A., Belasco, J.G. (1995) Analysis of RNA-binding proteins by *in vitro* genetic selection: identification of an amino acid residue important for locking U1A onto its RNA target. *Proc. Natl. Acad. Sci. U.S.A.* **92**:11859-11863.

Laird-Offringa, I.A., Belasco, J.G. (1996) *In vitro* genetic analysis of RNA-binding proteins using phage display libraries. *Methods Enzymol.* **267**:149-168.

Landsman, D. (1992) RNP-1, an RNA-binding motif is conserved in the DNA-binding cold-shock domain. *Nucleic Acids. Res.* **20**(11):2861-2864.

Lazarevik, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C., Karamata, D. (1998) Introns and intein coding sequence in the ribonucleotide reductase genes of *Bacillus subtilis* temperate bacteriophage SPβ. *Proc. Natl. Acad. Sci. U.S.A.* **95**:1692-1697.

Levine, R.A. (1994) Aquarium filter floss: an alternative to silanized glass wool as a porous support matrix. *BioTechniques* **17**(1):67.

Li, Y., Sugiura, M. (1991) Nucleic acid-binding specificities of tobacco chloroplast ribonucleoproteins *Nucleic Acids Res.* **19**(11):2893-2896.

Liang, J., Scappino, L., Haselkorn, R. (1992) The *patA* gene product, which contains a region similar to CheY of *Escherichia coli*, controls heterocyst pattern formation in the cyanobacterium *Anabaena* 7120. *Proc. Natl. Acad. Sci. U.S.A.* **89**:5655-5659.

Lisitski, I., Schuster, G. (1995) Phosphorylation of a chloroplast RNA-binding protein changes its affinity to RNA. *Nucleic Acids Res.* **23**(13):2506-2511.

Lorsch, J.R., Szostak, J.W. (1996) Chance and necessity in the selection of nucleic acid catalysts. *Acc. Chem. Res.* **29**:103-110.

Lorsch, J.R., Szostak, J.W. (1994) In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**(6492):31-36.

Lu, J., Hall, K.B. (1995) An RBP that does not bind RNA: NMR secondary structure determination of the C-terminal RNA binding domain from human U1A protein. *J. Mol. Biol.* **247**:739-752.

Ludevid, M.D., Freire, M.A., Gomez, J., Burd, C.G., Albericio, F., Girolt, E., Dreyfuss, G., Pages, M. (1992) RNA binding characteristics of a 16 kDa glycine-rich protein from maize. *Plant J.* **2**(6):999-1003.

Ludwig, L.B., Hughes, B.J., Schwartz, S.A. (1995) Biotinylated probes in the electrophoretic mobility shift assay to examine specific dsDNA, ssDNA, or RNA-protein interactions. *Nucleic Acids Res.* **23**(18):3792-3793.

Martinez-Romero, E., Oswald-Spring, U., Miranda, M., Garcia, J.L., Fuentes-Ramirez, L.E., Lopez-Reyes, L., Estrada, P., Caballero-Mellado, K. (1997) Towards the application of nitrogen fixation research to forestry and agriculture. In: A. Legocki *et al.* (Eds) Biological Fixation of Nitrogen for Ecology and Sustainable Agriculture. Springer-Verlag, Berlin, Germany. pp. 177-178.

Maruyama, K., Sato, N., Ohta, N. (1999) Conservation of structure and cold-regulation of RNA-binding proteins in cyanobacteria: probable convergent evolution with eukaryotic glycine-rich RNA-binding proteins. *Nucleic Acids Res.* **27**(9):2029-2036.

Mead, D.A., Szczesna-Skorupa, E., and Kemper, B. (1986) Single-stranded DNA "blue" T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Prot. Eng.* **1**(1):67-74.

Meissner, P.S., Sisk, W.P., and Bermen, M.L. (1987) Bacteriophage λ cloning system for the construction of directional cDNA libraries. *Proc. Natl. Acad. Sci. U.S.A.* **84**:4171-4175.

Merrill, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H., Williams, K.R. (1988) Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. *J. Biol. Chem.* **263**(7):3307-3313.

Moffatt, B.A., Studier, F.W. (1987) T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**:221-227.

Mulligan, M.E., Jackman, D.M., Murphy, S.T. (1994) Heterocyst-forming filamentous cyanobacteria encode proteins that resemble eukaryotic RNA-binding proteins of the RNP family. *J. Mol. Biol.* **235**:1162-1170.

Nagai, K. (1996) RNA-protein complexes. *Curr. Biol.* **6**:53-61.

Nagai, K., Oubridge, C., Ito, N., Avis, J., Evans, P. (1995) the RNP domain: a sequence-specific RNA-binding domain involved in processing and transport of RNA. *Trends Biochem. Sci.* **20**:235-240.

Nagai, K. (1993) Recent advances in RNA-protein interaction studies. *Mol. Biol. Rep.* **18**:105-112.

New England Biolabs (1997) NEB Instruction Manual: IMPACT I: One-Step Protein Purification System. New England Biolabs. Beverly, MA.

New England Biolabs (1999) NEB Instruction Manual: IMPACT-CN: One-step Purification of recombinant proteins using a self-cleavable affinity tag. New England Biolabs. Beverly, MA.

Nishiyama, H., Itoh, K., Kaneko, Y., Kishishita, M., Yoshida, O., Fujita, J. (1997) A glycine rich RNA-binding protein mediating cold-inducible suppression of mammalian growth. *J. Cell Biol.* **137**(4):899-908.

Oubridge, C, Ito, N., Evans, P.R., Teo, C., Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*. **372**:432-438.

Ouellette, M.M., Wright, W.E. (1995) Use of reiterative selection for defining protein-nucleic acid interactions. *Curr. Biol.* **6**:65-72.

Qiagen (1997) Instruction Manual: QIAexpressionist: a handbook for high-level expression and purification of 6x His-tagged proteins. Qiagen. Mississauga, ON

Pearson, W.R., Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **85**:2444-2448.

Pharmacia LKB Biotechnology (1986a) SDS PAGE. Separation Techniques File No. 120.

Pharmacia LKB Biotechnology (1986b) SDS PAGE in homogeneous media. Separation Techniques File No. 121.

Pharmacia LKB Biotechnology (1986c) Fast Coomassie staining. Development Techniques File No. 200.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J.T., Belfort, M. (1994) Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucleic Acid. Res.* **22**(7):1125-1127.

Piechocki, M.P., Hines, R.N. (1994) Oligonucleotide design and optimized protocol for site-directed mutagenesis. *BioTechniques* **16**(4):702-707.

Pietrokovski, S. (1996) A new intein in cyanobacteria and its significance for the spread of inteins. *Trends Genet.* **12**(8):287-288.

Pope, B., Kent, H.M. (1996). High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Res.* **24**(3):536-537.

Promega Corp. (1996). Promega Technical Manual: Altered Sites II in vitro mutagenesis systems. Promega Corp. Madison WI.

Rice, D., Mazur, B.J., Haselkorn, R. (1982) Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* 7120. *J. Biol. Chem.* **257**:13157-13163.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., Stanier, R.Y. (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**: 1-61.

Rundlett, K.L., Armstrong, D.W. (1997) Methods for the estimation of binding constants by capillary electrophoresis. *Electrophoresis* **18**: 2194-2202.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning 2nd Ed. Cold Spring Harbor Laboratory Press, NY.

Sanger, F., Niclen, S., Coulsen, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA,* **74**(12):5463-5467.

Sato, N,. Nakamura, A. (1998) Involvement of the 5'-untranslated region in cold-regulated expression of the *rbpA1* gene in the cyanobacterium *Anabaena variabilis* M3. *Nucleic Acids Res.* **26**(9):2192-2199.

Sato, N., Tachikawa, T., Wada, A., Tanaka, A. (1997) Temperature-dependent regulation of the ribosomal small-subunit protein S21 in the cyanobacterium *Anabaena variabilis* M3. *J. Bacteriol.* **179**(22):7063-7071.

Sato, N. (1994) A cold-regulated cyanobacterial gene cluster encodes RNA-binding protein and ribosomal protein S21. *Plant Mol. Biol.* **24**:819-823.

Sato, N. (1995) A family of cold-regulated RNA-binding protein genes in the cyanobacterium *Anabaena variabilis* M3. *Nucleic Acids Res.* **23**(12):2161-2167.

Sato, N., Wada, A. (1996) Disruption analysis of the gene for a cold-regulated RNA-binding protein, *rbpA1*, in *Anabaena*: cold-induced initiation of the heterocyst differentiation pathway. *Plant Cell Phys.* **37**(8):1150-1160.

Scherly, D., Boelens, W., Dathan, N.A., van Venrooj, W.J., Mattaj, I.W. (1990) Major determinants of interaction between small nuclear ribonucleoproteins U1A and U2B" and their cognate RNAs. *Nature* **345**:502-506.

Scherly, D., Boelens, W., van Venrooj, W.J., Dathan, N.A., Hamm, J., Mattaj, I.J. (1989) Identification of the RNA binding segment of human U1 A protein and definition of its binding site in U1 snRNA. *EMBO J.* **8**(13):4163-4170.

Schindelin, H., Marahiel, M.A., Heinemann, U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* **364**:164-170.

Schneider, D., Gold, L., Platt, T. (1993) Selective enrichment of RNA species for tight binding to *Escherichia coli* rho factor. *FASEB J.* **7**:201-107.

Schnuchel, A., Wiltscheck, R., Czslch, M., Herrier, M., Willmski, G., Graumann, P., Marahiel, M.A., Holak, T.A. (1993) Structure in solution of the major cold-shock protein from *Bacillis subtilis*. *Nature* **364**:196-171.

Severinov, K., Muir, T.W. (1998) Expressed protein ligation, a novel method for studying protein-protein interactions in transcription. *J. Biol. Chem.* **273**(26):16205-16209.

Shahied-Milam, L., Soltaninassah, S.R., Iyer, G.V., LeStourgeon, W.M. (1998) The heterogeneous nuclear ribonucleoprotein C Protein tetramer binds U1, U2, and U6 snRNAs through its high affinity RNA binding domain (the bZIP-like motif). *J. Biol. Chem.* **273**(33):21359-21367.

Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R., Steitz, T.A. (1997) Crystal structure of the two RNA-binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature. Struct. Biol.* **4**(3):215-222.

Shao, Y., Xu, M-Q., Paulus, H. (1996) Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process. *Biochemistry* **35**:3810-3815.

Siegal, E.C., Wain, S.L., Meltzer, S.F., Binion, M.L., and Steinburg, J.L. (1982) Mutator mutations in *Escherichia coli* induced by the insertion of Phage Mu and the transposable resistance elements Tn5 and Tn10. *Mut. Res.* **93**:25-33.

Singer, B.S., Shtatland, T., Brown, D., Gold, L. (1997) Libraries for genomic SELEX. *Nucleic Acids Res.* **25**(4):781-786.

Slade, C.S. (1998) Site-directed mutagenesis of the *rbpD* gene of *Anabaena* sp. PCC7120. B.Sc. (Honours) thesis, Memorial University of Newfoundland.

Smith, C.A., Suzuki, J.Y., Bauer, C. (1996) Cloning and characterisation of the chlorophyll biosynthesis gene *chlM* from *Synechocystis* PCC 6803 by complementation of a bacteriochlorophyll biosynthesis mutant of *Rhodobacter capsulatus*. *Plant Mol. Biol.* **30**:1307-1314.

Snyder, L., Champness, W. (1997) Molecular Genetics of Bacteria. ASM Press, Washington, D.C., U.S.A. p. 59.

Steinert, P.M., Mack, J.W., Korge, B.P., Gan, S-Q., Haynes, S., Steven, A.C. (1991) Glycine loops in proteins: their occurrence in certain intermediate filament chains, locicrins and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**:130-139.

Studier, F.W., Rosenburg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* **185**:60-89.

Sugita, M., Sugishita, H., Fujishiro, T., Tsuboi, M., Sugita, C., Endo, T., Sugiura, M. (1997) Organization of a large gene cluster ecoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC 6803: comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. *Gene* **195**:73-79.

Sugita, M., Sugiura, M. (1994) The existence of eukaryotic ribonucleoprotein consensus sequence-type RNA-binding proteins in a prokaryote, Synechococcus 6301. *Nucleic Acids res.* **22**(1):25-31.

Swanson, M.S., Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell Biol.* **8**(5):2237-2241.

Szostak, J.W., Ellington, A.D. (1993) In vitro selection of functional RNA sequences. In R.F. Gesteland, J.F. Atkins (Eds.) The RNA World, Cold Spring Harbor Laboratory Press, pp. 511-534.

Tuerk, C., Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**:505-510.

Thiel, T., Lyons, E.M., Erker, J.C., Ernst, A. (1995) A second nitrogenase in vegetative cells of a heterocyst-forming cyanobacterium. *Proc. Natl. Acad. Sci. U.S.A.* **92**:9358-9362.

Thiel, T. (1993) Characterization of an alternative nitrogenase in the cyanobacterium *Anabaena variabilis*. *J. Bacteriol.* **175**(19):6276-6286.

Thomas, J., Meeks, J.C., Wolk, C.P., Shaffer, P.W., Austin, S.M., Chien, W.-S. (1977) Formation of glutamine from [^{13}N] ammonia, [^{13}N] dintrogen and [^{14}C] glutamate by heterocysts isolated from *Anabaena cylindrica*. *J. Bacteriol.* **129**(8):1545-1555.

Tomb, J.F. *et al.* (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539-547.

van Kammen, A. (1997) Biological nitrogen fixation for sustainable agriculture. In: A. Legocki *et al.* (Eds) Biological Fixation of Nitrogen for Ecology and Sustainable Agriculture. Springer-Verlag, Berlin, Germany. pp. 177-178.

Vance, C.P. (1997) Enhanced agricultural sustainability through biological nitrogen fixation. In: A. Legocki *et al.* (Eds) Biological Fixation of Nitrogen for Ecology and Sustainable Agriculture. Springer-Verlag, Berlin, Germany. pp. 177-178.

Wang, Y. (1988) Double-stranded DNA sequencing with T7 polymerase. *BioTechniques*, **6**(9):843-845.

Weiner, M.P., Anderson, C., Jerpseth, B., Wells, S., Johnson-Browne, B., and Vaillancourt, P. (1994) *Strategies* **7**:41-43.**13**:637-649.

Wilcox, M., Mitchenson, G.J., Smith, R.J., (1973a) Pattern formation in the blue-green alga, *Anabaena* I: Basic mechanisms. *J. Cell Sci.* **12**:707-723.

Wilcox, M., Mitchenson, G.J., Smith, R.J., (1973b) Pattern formation in the blue-green alga, *Anabaena* II: controlled preheterocyst regression. *J. Cell Sci.*

Wilmot, A (1994) Molecular evolution and taxonomy of the cyanobacteria. In: D.A. Bryant (ed.) The Molecular Biology of Cyanobacteria. Kluwer Academic Publishing, Boston, Mass. pp. 1-25.

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G., Mueller, L. (1992) ^{1}H, ^{13}C, and ^{15}N NMR alignments and global folding pattern of the RNA-binding domain of the human hnRNP C proteins. *Biochem.* **31**:3254-6265.

Wolk, C.P. (1996) Heterocyst formation. *Annu. Rev. Genet.* **30**:59-78.

Wolk, C.P., Ernst, A., Elhai, J. (1994) Heterocyst metabolism and development. In: D.A. Bryant (ed.) The Molecular Biology of Cyanobacteria. Kluwer Academic Publishing, Boston, Mass. pp. 770-823.

Wolk, C.P., Elhai, J., Kuritz, T., Holland, D. (1993) Amplified expression of a transcriptional pattern formed during the development of *Anabaena*. *Mol. Microbiol.* **7**(3):441-445.

Wolk, C.P. (1989) Alternative models for the development of the pattern of spaced heterocysts in *Anabaena* (*Cyanophyta*) *Plant Syst. Evol.* **164**:27-31.

Wolk, C.P., Cai, Y., Cardemil, L., Flores, E., Hohn, B., Murray, M., Schmetterer, G., Schrautemeier, B., Wilson, R. (1988) Isolation and complementation of mutants of *Anabaena* sp. PCC 7120 unable to grow aerobically on dinitrogen. *J. Bacteriol.* **170**(3):1239-1244.

Wolk, C.P., Thomas, J., Shaffer, P.W. Austin, S.M, Galonski, A. (1976) Pathway of nitrogen metabolism after fixation of ^{13}N-labelled nitrogen gas by the cyanobacterium, *Anabaena cylindrica*. *J. Biol. Chem.* **131**(16):5027-5034.

Wolk, C.P., Austin, S.M., Bortnis, J., Galonski, A. (1974) Autoradiographic localization of ^{13}N after fixation of ^{13}N-labelled nitrogen gas by a heterocyst-forming blue-green alga. *J. Cell Biol.* **61**:440-453.

Wu, H., Hu, Z., Liu, X. (1998) Protein *trans*-splicing be a split intein encoded in a split DnaE gene of *Synechocystis* sp.PCC6803. *Proc. Natl. Acad. Sci. U.S.A.* **95**:9226-9231.

Xu, M-Q., Comb, D.G., Paulus, H., Noren, C.J., Shao, Y., Perler, F. (1994) Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J.* **13**(23):5517-5522.

Yannisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**(1):103-135.

Ye, L., Sugiura, M. (1992) Domains required for nucleic acid binding actiities in chloroplast ribonucleoproteins. *Nucleic Acids Res.* **20**(23):6275-6279.

Yoon, H., Golden, J.W. (1998) Heterocyst pattern formation controlled by a diffusible peptide. *Science* **282**:935-938.

Zehr, J.P., Mellon, M.T., Hiorns, W.D. (1997) Phylogeny of cyanobacterial *nifH* genes: evolutionary implications and potential applications to natural assemblages, *Microbiol.* (UK) **143**:1443-1450.

Zhou, R., Cao, Z., Zhao, J. (1998a) Characterization of HetR protein turnover in *Anabaena* sp. PCC 7120. *Arch. Microbiol.* **169**:417-423.

Zhou, R., Wei, X., Jiang, N., Li, H., Dong, Y., His, K-L., Zhao, J. (1998b) Evidence that HetR protein is an unusual serine-type protease. *Proc. Natl. Acad. Sci. U.S.A.* **95**:4959-4963.

Appendix 1:

Cloning and Sequencing of *nifH**

from *Chlorogloeopsis sp.* PCC 6192

The ability to fix nitrogen is important in that it allows the nitrogen-fixing cyanobacteria to dominate nitrogen-deficient environments (Zehr *et al.*, 1997) and makes these microorganisms responsible for a large part of the world's fixed nitrogen (Rice *et al.*, 1982). Cyanobacteria are found in such diverse places as marine and coastal areas (Ben-Porath and Zehr, 1994; Kirhtein *et al.*, 1991) and the inside of plant leaves such as the water fern *Azolla*. They are also used extensively in subtropical agriculture as a fixed nitrogen source (Vance, 1997). Most nitrogen-fixing cyanobacteria are filamentous and their ability to fix nitrogen is due to the presence of heterocysts. Under nitrogen deficient conditions these organisms differentiate vegetative cells into anaerobic heterocysts which are the sites of nitrogen fixation.

In all heterocyst-forming cyanobacteria, nitrogenase, which cannot work in the presence of oxygen, is expressed only in the heterocysts. Nitrogenase is composed of two proteins, the MoFe protein encoded by the *nifD* and *nifK* genes, and the Fe protein encoded by *nifH*. The NifH polypeptides of heterocyst-forming cyanobacteria are highly conserved, sharing greater than 94% amino acid sequence identity with each other (Ben-Porath and Zehr, 1994). Most of the sequence differences occur at the C- terminal end of the protein (Beesley *et al.*, 1994). The nucleotide sequences are only slightly less conserved showing 84% - 92% similarity (Ben-Porath and Zehr, 1994).

Rice *et al.* (1982) found a second copy of the *nifH* gene in *Anabaena* 7120 which they designated *nifH**. The nucleotide sequences flanking the *nifH** gene both upstream and downstream show no sequence similarity either with *nifD* or with *nifK*. It has been

suggested that this copy of *nifH* may function as an alternative, molybdenum independent, nitrogenase (Haselkorn and Buikema, 1992).

The Cg30 HindIII fragment was digested with EcoRV and the resulting 1.8 kilobase fragment, Cg9 (Mulligan *et al.*, 1994), was subcloned into pUC18 for sequencing. The resulting plasmid was designated pRLT5 (Table 2.1). Further subclones of Cg9 were generated using the enzymes ApoI, BstUI, HincII and Sau3A. Two oligonucleotides, Oligo.R (5' – GCC AAA AGG AGA AAG CCC G – 3') and Oligo2.R (5' – GGT AAT GTA AAC TCT GTA TAG G – 3') were synthesized to complete the sequence. Sequencing reactions were performed as described in section 2.3.

The Cg30 HindIII fragment was digested with EcoRV and the resulting 1.8 kilobase fragment was subcloned into the SmaI site of the pUC18 vector for sequencing. The resulting plasmid was designated pRLT5 (Table 2.1). Sequencing was performed as described in Section 2.3, using further subclones and oligonucleotide primers. The inferred amino acid sequence was determined using the DNAlysis software program of Dr. Bill Buikema (University of Chicago, IL, USA). Comparison of identities between different sequences was done using the FASTA program (Pearson and Lipman, 1988).

The nucleotide sequence of the *Chlorogloeopsis* 6912 *nifH** gene including the upstream and downstream flanking regions is shown in Figure A.1, and the inferred amino acid sequence is compared with the *Anabaena* 7120 NifH and NifH* genes in Figure A.2. The *Chlorogloeopsis* 6912 NifH* sequence is 97% (289 of 298 residues) identical to the NifH* sequence of *Anabaena* 7120 (An-NifH*). In fact, the

Chlorogloeopsis 6912 and *Anabaena* 7120 NifH* polypeptides are identical in their first 234 residues. The nucleotide sequences of the *Chlorogloeopsis* 6912 and *Anabaena* 7120 *nifH** genes are only slightly less conserved showing 89% identity (787 of 883 nucleotides).

The N-terminal sequences of NifH polypeptides are highly conserved; there is >94% identity among NifH sequences within the heterocyst-forming cyanobacteria (Ben-Porath and Zehr, 1994; most of the differences lie in the last ten C-terminal amino acids. This general pattern is seen in the comparison of the two NifH* sequences as well as in the comparison of all three sequences. However there seems to be a more extensive region of dissimilarity between the NifH sequence and the two NifH* sequences.

Figure A.1: Nucleotide sequence of the *Chlorogloeopsis* 6192 *nifH** gene. The complete sequence of a 1100 base-pair region containing the *nifH** gene and its flanking sequences is shown. The coding sequence of the gene is located between positions 115 and 1008; the inferred amino acid sequence is shown underneath. Subclones used for sequencing were generated with the restriction endonucleases ApoI, BstUI, HincII, and Sau3A. Two oligonucleotides were also used for sequencing. Oligo.R is highlighted in blue and Oligo2.R is highlighted in red. Nucleotides shown in lower case represent nucleotides which could be confirmed by sequencing in only one direction on the DNA.

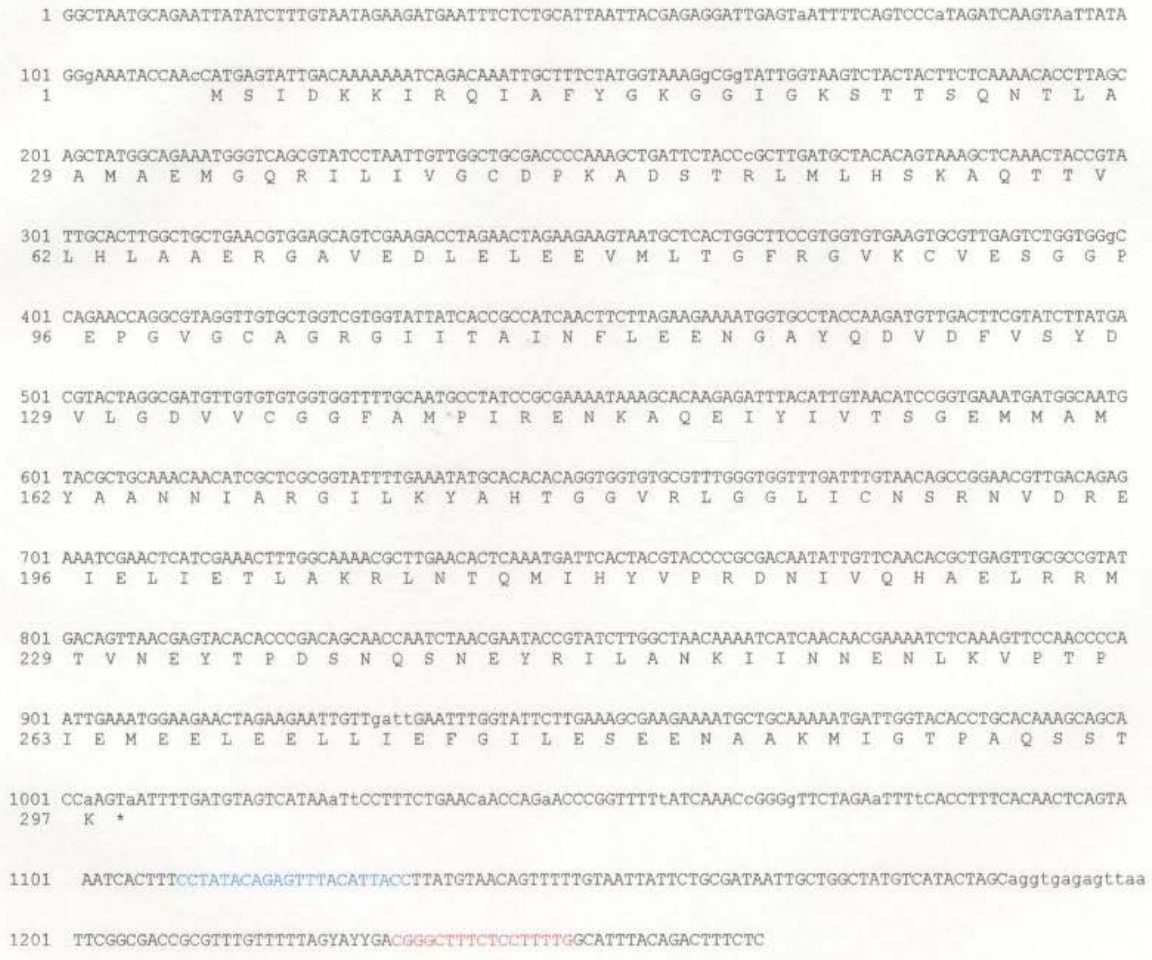

1 GGCTAATGCAGAATTATATCTTTGTAATAGAAGATGAATTTCTCTGCATTAATTACGAGAGGATTGAGTaATTTTCAGTCCCaTAGATCAAGTAaTTATA
101 GGgAAATACCAAcCATGAGTATTGACAAAAAAATCAGACAAATTGCTTTCTATGGTAAAGgCGgTATTGGTAAGTCTACTACTTCTCAAAACACCTTAGC
1 M S I D K K I R Q I A F Y G K G G I G K S T T S Q N T L A
201 AGCTATGGCAGAAATGGGTCAGCGTATCCTAATTGTTGGCTGCGACCCCAAAGCTGATTCTACCcGCTTGATGCTACACAGTAAAGCTCAAACTACCGTA
29 A M A E M G Q R I L I V G C D P K A D S T R L M L H S K A Q T T V
301 TTGCACTTGGCTGCTGAACGTGGAGCAGTCGAAGACCTAGAACTAGAAGAAGTAATGCTCACTGGCTTCCGTGGTGTGAAGTGCGTTGAGTCTGGTGGgC
62 L H L A A E R G A V E D L E L E E V M L T G F R G V K C V E S G G P
401 CAGAACCAGGCGTAGGTTGTGCTGGTCGTGGTATTATCACCGCCATCAACTTCTTAGAAGAAAATGGTGCCTACCAAGATGTTGACTTCGTATCTTATGA
96 E P G V G C A G R G I I T A I N F L E E N G A Y Q D V D F V S Y D
501 CGTACTAGGCGATGTTGTGTGTGGTGGTTTTGCAATGCCTATCCGCGAAAATAAAGCACAAGAGATTTACATTGTAACATCCGGTGAAATGATGGCAATG
129 V L G D V V C G G F A M P I R E N K A Q E I Y I V T S G E M M A M
601 TACGCTGCAAACAACATCGCTCGCGGTATTTTGAAATATGCACACACAGGTGGTGTGCGTTTGGGTGGTTTGATTTGTAACAGCCGGAACGTTGACAGAG
162 Y A A N N I A R G I L K Y A H T G G V R L G G L I C N S R N V D R E
701 AAATCGAACTCATCGAAACTTTGGCAAAACGCTTGAACACTCAAATGATTCACTACGTACCCCGCGACAATATTGTTCAACACGCTGAGTTGCGCCGTAT
196 I E L I E T L A K R L N T Q M I H Y V P R D N I V Q H A E L R R M
801 GACAGTTAACGAGTACACACCCGACAGCAACCAATCTAACGAATACCGTATCTTGGCTAACAAAATCATCAACAACGAAAATCTCAAAGTTCCAACCCCA
229 T V N E Y T P D S N Q S N E Y R I L A N K I I N N E N L K V P T P
901 ATTGAAATGGAAGAACTAGAAGAATTGTTgattGAATTTGGTATTCTTGAAAGCGAAGAAAATGCTGCAAAAATGATTGGTACACCTGCACAAAGCAGCA
263 I E M E E L E E L L I E F G I L E S E E N A A K M I G T P A Q S S T
1001 CCaAGTaATTTTGATGTAGTCATAAaTtCCTTTCTGAACaACCAGaACCCGGTTTTtATCAAACcGGGgTTCTAGAaTTTtCACCTTTCACAACTCAGTA
297 K *
1101 AATCACTTTCCTATACAGAGTTTACATTACCTTATGTAACAGTTTTTGTAATTATTCTGCGATAATTGCTGGCTATGTCATACTAGCaggtgagagttaa
1201 TTCGGCGACCGCGTTTGTTTTTAGYAYYGACGGGCTTTCTCCTTTTGGCATTTACAGACTTTCTC

Figure A.2: Comparison between NifH* sequences of *Chlorogloeopsis* 6912 and *Anabaena* 7120 and the NifH sequence of *Anabaena* 7120. Black lettering represents sequence identities between the proteins. Blue lettering indicates amino acid residues which lack identity with the other two proteins. Dashes indicate places where one sequences lacks an amino acid present in one of the other proteins. ^{Cg}NifH*, ^{An}NifH* = NifH sequences from *Chlorogloeopsis* 6192 and *Anabaena* 7120 respectively. ^{An}NifH = NifH sequence from *Anabaena* 7120. ^{Cg}NifH* is 84% identical (250 of 299 amino acids) to ^{An}NifH; ^{An}NifH* has a comparable identity to ^{An}NifH (84% or 251 of 299 amino acids). ^{An}NifH* and ^{Cg}NifH* share the highest identity among the sequences with 97% identity (289 of 298 amino acids).

```
CgNifH*  1   MSIDKKIRQIAFYGKGGIGKSTTSQNTLAAMAEMGQRILI  40
AnNifH*  1   MSIDKKIRQIAFYGKGGIGKSTTDQNTLAAMAEMGQRILI  40
AnNifH   1   MT-DENIRQIAFYGKGGIGKSTTDQNTLAAMAEMGQRIMI  39

CgNifH*  41  VGCDPKADSTRLMLHSKAQTTVLHLAAERGAVEDLELEEV  80
AnNifH*  41  VGCDPKADSTRLMLHSKAQTTVLHLAAERGAVEDLELEEV  80
AnNifH   40  VGCDPKADSTRLMLHSKAQTTVLHLAAERGAVEDLELHEV  79

CgNifH*  81  MLTGFRGVKCVESGGPEPGVGCAGRGIITAINFLEENGAY  120
AnNifH*  81  MLTGFRGVKCVESGGPEPGVGCAGRGIITAINFLEENGAY  120
AnNifH   80  MLTGFRGVKCVESGGPEPGVGCAGRGIITAINFLEENGAY  119

CgNifH*  121 QDVDFVSYDVLGDVVCGGFAMPIRENKAQEIYIVTSGEMM  160
AnNifH*  121 QDVDFVSYDVLGDVVCGGFAMPIRENKAQEIYIVTSGEMM  160
AnNifH   120 QDLDFVSYDVLGDVVCGGFAMPIREGKAQEIYIVTSGEMM  159

CgNifH*  161 AMYAANNIARGILKYAHTGGVRLGGLICNSRNVDREIELI  200
AnNifH*  161 AMYAANNIARGILKYAHTGGVRLGGLICNSRNVDREIELI  200
AnNifH   160 AMYAANNIARGILKYAHSGGVRLGGLICNSRKVDREDELI  199

CgNifH*  201 ETLAKRLNTQMIHYVPRDNIVQHAELRRMTVNEYTPDSNQ  240
AnNifH*  201 ETLAKRLNTQMIHYVPRDNIVQHAELRRMTVNEYAPDSNQ  240
AnNifH   200 EMNAERLNTQMIHFVPRDNIVQHAELRRMTVNEYAPDSNQ  239

CgNifH*  241 SNEYRILANKIINNENLKVPTPIEMEELEELLIEFGILES  280
AnNifH*  241 GNEYRILANKIINNENLKIPTPIEMEELEELLIEFGILES  280
AnNifH   240 GQEYRALAKKI-NNDKLTIPTPMEMDELEALKIEYGLLDD  278

CgNifH*  281 EENAAKMIGTPAQS---STK                     297
AnNifH*  281 EENAAKMIATTSESK--SSK                     298
AnNifH   279 DTKHSEIIGKPAEATNRSCRN                    299
```