CHARACTERIZATION OF A GENE ENCODING AN RNA-BINDING PROTEIN (rbpA) IN THE CYANOBACTERIUM SYNECHOCOCCUS SP. PCC 7942

THOMAS JAMES BELBIN


CENTRE FOR NFLD. STUDIES
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND

CHARACTERIZATION OF A GENE ENCODING AN RNA-BINDING PROTEIN (*rbpA*) IN THE CYANOBACTERIUM *SYNECHOCOCCUS SP.* PCC 7942

by

Thomas James Belbin

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

Department of Biochemistry
Memorial University of Newfoundland

May 1999

St. John's Newfoundland

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

ABSTRACT

Many species of cyanobacteria possess genes whose products are highly similar to the RNP family of RNA-binding proteins found in eukaryotes. This work describes the characterization of *rbpA*, one of two RNA-binding protein (*rbp*) genes now known to exist in the unicellular cyanobacterium *Synechococcus sp.* PCC 7942. This gene codes for a protein of 107 amino acids. It contains a single RNA Recognition Motif (RRM) as well as an auxiliary domain rich in glycine residues.

Mutation of the *rbpA* gene by insertional inactivation using the spectinomycin resistance omega cassette resulted in a temperature-sensitive phenotype with an altered pigment composition when compared with the wild type organism. This phenotype was not observed in a "control mutant", in which the omega cassette was inserted outside of the *rbpA* gene. Complementation experiments demonstrated that it was possible to rescue the phenotype of the "knock-out" mutant by insertion of a wild type copy of the *rbpA* gene into a neutral site in the cyanobacterial genome.

The function of cyanobacterial RNA-binding proteins is not known. A histidine-tagged form of RbpA (H_6RbpA) was purified using metal chelate affinity chromatography. RNA binding experiments demonstrated that this protein showed a preference for poly(A), poly(G) and poly(U) RNA but not poly(C). This specificity did not appear to be significantly affected by removal of the auxiliary domain. Overall, work presented here suggests that the RbpA protein may affect content of the phycobilisome components in the

photosynthetic apparatus. It also appears to be a protein which is required for growth at lower temperatures.

ACKNOWLEDGEMENTS

I would like to thank Dr. Martin Mulligan, my thesis supervisor, for his support throughout my graduate work. Members of the Mulligan lab, both past and present, have been very helpful. I would especially like to thank Donna Jackman and Shinn-Jia Hwang for their technical assistance. Members of my supervisory committee, Dr. John Robinson and Dr. John T. Brosnan, were always helpful and gave generously of their time. Special thanks to Dr. Larry Gold and Dr. Britta Singer (University of Colorado, Boulder) for their assistance and generous hospitality.

Plasmid vectors pAM990 and pSKS101 were a gift from Dave Cogdell and Susan Golden (Texas A&M University). Plasmid pDW9 was kindly provided by Jim Golden (Texas A&M University).

This work was supported by a grant from the Natural Sciences and Engineering Council of Canada (NSERC). I am also grateful to the School of Graduate Studies of Memorial University for financial support throughout my graduate studies.

Special appreciation to my wife Elizabeth Belbin and to my parents Tom and Ellen Belbin for their enduring support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

bp, base pair

DTT, 1,4-dithiothreitol

E. coli, *Escherichia coli*

EDTA, ethylenediaminetetraacetic acid

Fox$^-$, phenotype incapable of aerobic nitrogen fixation

hr(s), hour(s)

LB, Luria Bertani medium

MCAC, metal chelate affinity chromatography

min, minute

NBT, nitroblue tetrazolium salt

nt, nucleotide

PMSF, phenylmethylsulfonyl fluoride

Rbp, RNA-binding protein

RNP, ribonucleoprotein

rpm, revolutions per minute

RRM, RNA recognition motif

SDS, sodium dodecyl sulfate

sec, seconds

snRNP, small nuclear ribonucleoprotein

SSC, sodium chloride/sodium citrate

Tris, tris(hydroxymethyl)aminomethane

X-Phos, 5-bromo-4-chloro-3-indolyl phosphate toluidinium salt

CHAPTER 1. INTRODUCTION

1.1. RECOGNITION OF RNA BY PROTEINS

1.1.1. The Role of RNA in Cellular Processes

Over the past decade, RNA has emerged as a molecular entity which is involved in a great variety of processes within the cell. The most well known function involves its central role in the transfer and expression of the genetic information found in the nucleotide sequences of genes to the sequence of amino acids in the polypeptide chain of a protein. Central to that process are some of the other key RNA molecules within the cell. Adaptor tRNA molecules provide the crucial link between the sequence of codons in the mRNA and the primary structure of the corresponding polypeptide. In both prokaryotic and eukaryotic systems, the ribosome itself contains ribosomal RNA (rRNA) components which provide the framework necessary to assemble the ribosomal subunits while providing the sequence-specific contacts with both the aminoacyl-tRNAs and the mRNA transcript which are necessary for proper translation of the message (Noller, 1991). The ribonucleoprotein (RNP) complexes responsible for the processing of pre-mRNA transcripts in eukaryotes (hnRNPs and snRNPs) also have several RNA components which form integral parts of the larger spliceosomal machinery (Dreyfuss, 1993; Moore *et al.*, 1993).

In order to appreciate and better understand the diverse mechanisms used for RNA recognition by proteins, we must begin with an outline of the factors influencing RNA secondary and tertiary structure, and how these present unique structural motifs which are capable of specific interaction with proteins.

1.1.2. The Diversity of RNA Structures

RNA is chemically very similar to DNA. The main differences in RNA are the change from a 2'-deoxyribose to a ribose sugar moiety, and the lack of the methyl group of thymine which, in RNA, is replaced by a hydrogen atom in uracil. However, the conformations adopted by DNA and RNA differ substantially. The most recognized RNA conformation is the stem-loop or hairpin loop. This region consists of two segments of complementary RNA which form a double-stranded helical stem by Watson-Crick base pairs. The complementary segments are bridged by a region of single-stranded RNA known as the hairpin. The RNA helices of a stem-loop structure are usually less than 10 bp in length. These helices are interrupted by symmetrical or asymmetrical regions of base pair mismatch known as internal loops; one-sided internal loops are known as internal bulges (Klaff *et al.*, 1996). Internal loops and bulges alter the secondary structure of RNA by introducing a bend in the region separating two RNA helices.

Even greater complexity of RNA structures comes from tertiary interactions in RNA. Base pairs between single-stranded regions of RNA located within a secondary structure or located in separate secondary structures folds RNA into even more complex three-dimensional shapes. One example is the pseudoknot, which forms when the single-stranded loop region bridging a stem-loop base pairs with a single-stranded region outside of this secondary structure. Tertiary interactions in RNA also give rise to structures such as bifurcations and triple helices.

1.1.3. Factors Affecting the Structure of RNA

Similar to proteins, many RNAs must form complex secondary and tertiary structures in order to function in RNA-protein interaction or catalysis. One of the driving forces of protein folding is the thermodynamic advantage of burying hydrophobic residues within the interior of the protein's three-dimensional structure. This thermodynamic effect also promotes the formation of secondary structure in RNA (Doudna and Doherty, 1997). However, other types of molecular interactions have emerged as key players in the proper folding and function of the more complex teriary structures of RNA. One strategy employed in RNA folding is the stacking of one RNA helix upon another. This structure was first observed in the L-shaped 3-dimensional structure of tRNAs, but has since been observed in many other RNA species (Kim *et al.*, 1973; Robertus *et al.*, 1974; Strobel and Doudna, 1997). The main contributor to helical stacking is the 2'-hydroxyl group which distinguishes DNA from RNA. These line the outer edge of the minor groove of RNA and act as hydrogen bond donors and acceptors in the stacking of RNA helices (Strobel and Doudna, 1997).

Another important factor affecting RNA structure and folding is the presence of non-canonical base pairs (those other than G•C and A•U). Their unusual patterns of hydrogen bonding present functional groups in both the major and minor grooves of duplex RNA. These functional groups are therefore available for tertiary interactions with other RNA sequences. As well, the major groove of an RNA helix immediately adjacent to a non-canonical base pair is widened, thereby increasing its accessibility to interaction. The importance of non-canonical base pairs in secondary and

tertiary interactions is underscored by the fact that they are often phylogenetically conserved in nature (Weeks and Crothers, 1993). Non-canonical base pairs have been shown to be directly involved in the action of the self-splicing group I intron of *Tetrahymena* (Pyle, 1994; Strobel *et al.*, 1998).

Another area of recent interest has been the role of divalent cations in RNA folding. Many structured RNAs require divalent cations, such as magnesium, for proper folding (Doudna and Cate, 1997). For example, more than 100 magnesium ions are needed for proper folding of ribonuclease P which is only 375 nucleotides in length (Doudna, 1997). Many RNA ribozymes require magnesium ions for their catalytic activity (Pyle, 1993). It is believed that one of the functions of magnesium ions is to neutralize the negatively charged phosphate groups on the RNA backbone and thereby facilitate the close packing of RNA helices in a tertiary structure (Doudna and Doherty, 1997). In certain instances, RNA structures can also be stabilized by the specific coordination of magnesium ions to sites of high charge density (Laing *et al.*, 1994).

1.1.4. Recognition of RNA by Proteins

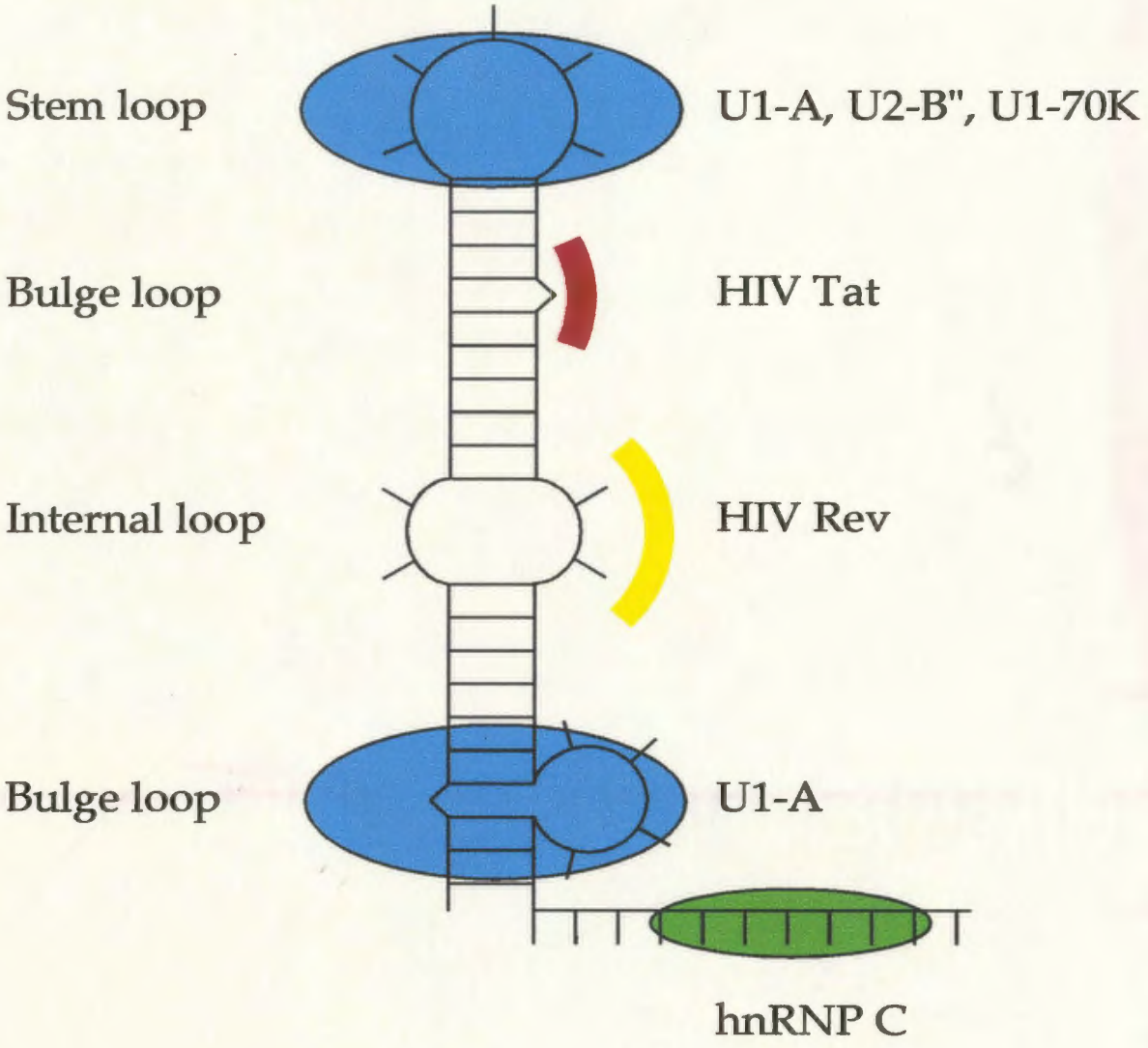
An RNA double helix is generally not amenable to sequence-specific interactions with proteins or other RNA species. Unlike the B-form double helix adopted by DNA, double helices formed by association of complementary RNA sequences are found in the A-form (Varani and Pardi, 1994). The minor groove of an A-form RNA helix is wide (10-11 Å) and shallow (3 Å) and therefore accessible to amino acid side chains of a protein and to nucleotide functional groups of RNA (Steitz, 1990). However, the

minor groove presents only a regular array of hydrogen bond donors and acceptors regardless of the nucleotide sequence present. These functional groups include the 2'-hydroxyl groups of ribose which line the outer edge of the minor groove (Mattaj, 1993). It is easier to distinguish nucleotide sequences by functional groups in the major groove. However, in an A-form RNA helix, the major groove is too narrow (3 Å wide) to allow easy access by functional groups (Steitz, 1990).

Proteins which interact with RNA generally target the single-stranded regions (hairpin loops, internal loops, internal bulges) within a secondary structure (Nagai and Mattaj, 1994). The unpaired nucleotides present functional groups outside of the minor groove which may be targets for sequence-specific recognition by proteins and other RNAs (Varani and Pardi, 1994). As well, the major groove of an A-form RNA helix is widened in the regions adjacent to perturbations such as bulges and internal loops, thereby making other functional groups more accessible to interaction (Weeks and Crothers, 1993). A schematic diagram of some RNA-binding proteins and their sites of interaction with RNA is shown in **Figure 1.1.**

Overall, most of the RNA in living cells exists in some form of complex with proteins. There exists a tremendous variety of processes requiring this RNA-protein interaction, including mRNA biogenesis, pre-mRNA splicing, mRNA polyadenylation, mRNA transport, mRNA stability and development (Nagai and Mattaj, 1994). To date, nine distinct families of RNA-binding proteins have been identified whose members share some common sequence motif (Mattaj, 1993). By far, the largest family of RNA-binding proteins (>250 different proteins) are those which contain the RNA

Figure 1.1. RNA structural elements often recognized by RNA-binding proteins. RNA-protein contacts can appear at hairpin loops as in the case of U1 snRNP protein A (U1-A) and U1 snRNP 70K (U1-70K) (Scherly *et al.*, 1989; Surowy *et al.*, 1989); HIV regulator of virion expression (Rev) protein recognizes an internal loop structure in RNA ; HIV *trans*-activator protein (Tat) recognizes a bulge loop known as the *trans*-activation responsive region (TAR) (Dingwall *et al.*, 1989); U1-A also recognizes a bulge loop structure in RNA as part of its autoregulation of expression (van Gelder *et al.*, 1993); human hnRNP protein C recognizes single-stranded RNA rich in uridine residues (Gorlach *et al.*, 1994).


Stem loop
U1-A, U2-B", U1-70K
Bulge loop
HIV Tat
Internal loop
HIV Rev
Bulge loop
U1-A
hnRNP C

recognition motif (RRM) or ribonucleoprotein (RNP) motif (Birney *et al.*, 1993). The detailed study of the interaction between several members of this family and their RNA targets have revealed a great deal of information about the nature of RNA-protein interactions.

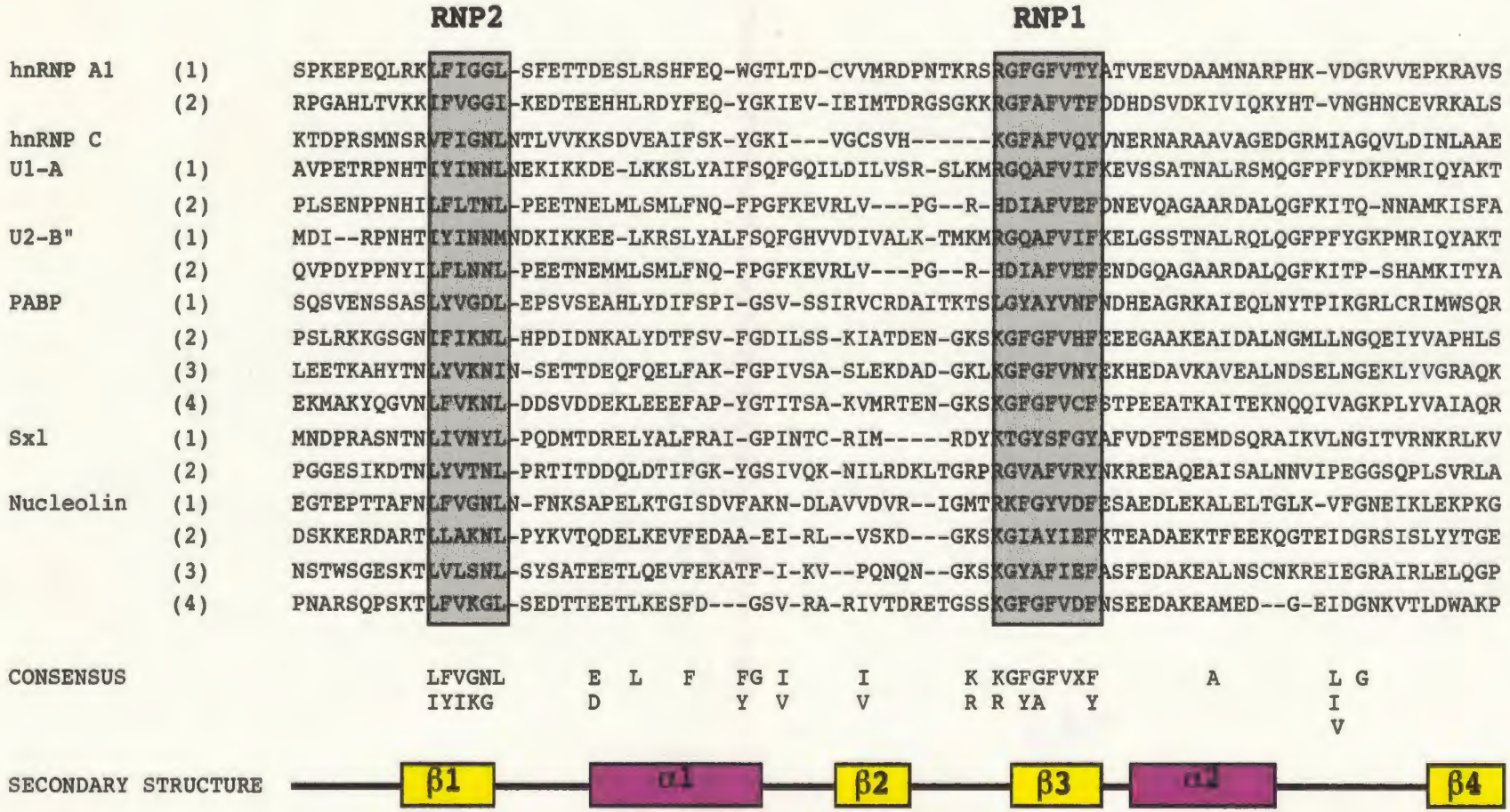
1.2. THE RNP FAMILY OF RNA-BINDING PROTEINS

1.2.1. The RRM Motif

The RNP family of RNA-binding proteins are characterized by the presence of the RRM motif which consists of a conserved sequence of approximately 80 amino acids. This motif has been identified in over 250 proteins with a wide variety of functions **(Figure 1.2)** (Nagai *et al.*, 1995). RNP-type RNA-binding proteins typically have a modular structure. The number of RRM motifs present can vary from as few as one or as many as four in a single polypeptide. They may also be accompanied by other forms of protein domains, known as auxiliary domains, whose function is still unclear. Within the RRM motif are two highly conserved sequences: the RNP1 sequence consists of eight amino acids, of which residues three and five are almost always aromatic amino acids; the RNP2 sequence consists of six amino acids, of which residue two is always an aromatic amino acid (Bandziulis *et al.*, 1989). These two sequences are separated by a region of 30-40 amino acids within the RRM motif.

There is abundant evidence that the RRM motif is responsible for the RNA-binding properties of this family of proteins. Most obvious is the fact that this region is the most highly conserved of all RNP-type RNA-binding proteins sequenced to date. Deletion analysis of a large number of RNP-type RNA-binding proteins, including hnRNP A1 (Riva *et al.*, 1986), hnRNP C (Gorlach *et al.*, 1994), nucleolin (Bugler *et al.*, 1987), polyA-binding protein (Sachs *et al.*, 1987), U1 snRNP protein A (U1-A) (Scherly *et al.*, 1989) and U1 snRNP protein 70K (U1-70K) (Query *et al.*, 1989) have demonstrated that the

Figure 1.2. Alignment of RRM motifs from a diverse group of RNA-binding proteins. Alignments were as described previously (Nagai and Mattaj, 1994); some additional alignments were carried out manually. Proteins containing each sequence are indicated on the left. For those with several RRMs, individual RRMs are numbered. Sequences are taken from human hnRNP A1 (Biamonti *et al.*, 1989), human hnRNP C (Swanson *et al.*, 1987), human U1 snRNP protein A (U1-A) (Sillikens *et al.*, 1987), human U2 snRNP protein B" (U2-B") (Habets *et al.*, 1987), *S. cerevisiae* poly (A) binding protein (PABP) (Adam *et al.*, 1986), *Drosophila* sex-lethal (Sxl) (Bell *et al.*, 1988) and human nucleolin (Srivastava *et al.*, 1990). Conserved positions corresponding to the RNP1 and RNP2 sequences are grey shaded; a consensus sequence is given at the bottom of the figure. The approximate positions of secondary structural elements, based on structural data outlined in Sections 1.2.2 and 1.2.3, are shown as boxed areas below the corresponding sequence. α-helical regions (α1 and α2) are shown in purple; β-strands (β1-β4) are shown in yellow.


RNP2
RNP1
hnRNP A1 (1) SPKEPEQLRKLFIGGL-SFETTDESLRSHFEQ-WGTLTD-CVVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHK-VDGRVVEPKRAVS
(2) RPGAHLTVKKIFVGGI-KEDTEEHHLRDYFEQ-YGKIEV-IEIMTDRGSGKKRGFAFVTFDDHDSVDKIVIQKYHT-VNGHNCEVRKALS
hnRNP C KTDPRSMNSRVFIGNLNTLVVKKSDVEAIFSK-YGKI---VGCSVH------KGFAFVQYVNERNARAAVAGEDGRMIAGQVLDINLAAE
U1-A (1) AVPETRPNHTIYINNLNEKIKKDE-LKKSLYAIFSQFGQILDILVSR-SLKMRGQAFVIFKEVSSATNALRSMQGFPFYDKPMRIQYAKT
(2) PLSENPPNHILFLTNL-PEETNELMLSMLFNQ-FPGFKEVRLV---PG--R-HDIAFVEFDNEVQAGAARDALQGFKITQ-NNAMKISFA
U2-B" (1) MDI--RPNHTIYINNMNDKIKKEE-LKRSLYALFSQFGHVVDIVALK-TMKMRGQAFVIFKELGSSTNALRQLQGFPFYGKPMRIQYAKT
(2) QVPDYPPNYILFLNNL-PEETNEMMLSMLFNQ-FPGFKEVRLV---PG--R-HDIAFVEFENDGQAGAARDALQGFKITP-SHAMKITYA
PABP (1) SQSVENSSASLYVGDL-EPSVSEAHLYDIFSPI-GSV-SSIRVCRDAITKTSLGYAYVNFNDHEAGRKAIEQLNYTPIKGRLCRIMWSQR
(2) PSLRKKGSGNIFIKNL-HPDIDNKALYDTFSV-FGDILSS-KIATDEN-GKSKGFGFVHFEEEGAAKEAIDALNGMLLNGQEIYVAPHLS
(3) LEETKAHYTNLYVKNIN-SETTDEQFQELFAK-FGPIVSA-SLEKDAD-GKLKGFGFVNYEKHEDAVKAVEALNDSELNGEKLYVGRAQK
(4) EKMAKYQGVNLFVKNL-DDSVDDEKLEEEFAP-YGTITSA-KVMRTEN-GKSKGFGFVCFSTPEEATKAITEKNQQIVAGKPLYVAIAQR
Sxl (1) MNDPRASNTNLIVNYL-PQDMTDRELYALFRAI-GPINTC-RIM-----RDYKTGYSFGYAFVDFTSEMDSQRAIKVLNGITVRNKRLKV
(2) PGGESIKDTNLYVTNL-PRTITDDQLDTIFGK-YGSIVQK-NILRDKLTGRPRGVAFVRYNKREEAQEAISALNNVIPEGGSQPLSVRLA
Nucleolin (1) EGTEPTTAFNLFVGNLN-FNKSAPELKTGISDVFAKN-DLAVVDVR--IGMTRKFGYVDFESAEDLEKALELTGLK-VFGNEIKLEKPKG
(2) DSKKERDARTLLAKNL-PYKVTQDELKEVFEDAA-EI-RL--VSKD---GKSKGIAYIEFKTEADAEKTFEEKQGTEIDGRSISLYYTGE
(3) NSTWSGESKTLVLSNL-SYSATEETLQEVFEKATF-I-KV--PQNQN--GKSKGYAFIEFASFEDAKEALNSCNKREIEGRAIRLELQGP
(4) PNARSQPSKTLFVKGL-SEDTTEETLKESFD---GSV-RA-RIVTDRETGSSKGFGFVDFNSEEDAKEAMED--G-EIDGNKVTLDWAKP
CONSENSUS LFVGNL E L F FG I I K KGFGFVXF A L G
IYIKG D Y V V R R YA Y I
V
SECONDARY STRUCTURE β1 α1 β2 β3 α2 β4

RRM motif is required for the RNA binding activity of these proteins. Peptide binding studies have demonstrated that an oligopeptide, containing the conserved RNP1 octamer, can bind to ^{32}P-labelled mRNA (Schwemmle *et al.*, 1989). Taken together, the evidence demonstrates that the presence of an RRM motif is a strong indicator of RNA-binding activity within a protein.

The variety of RNA substrates bound by members of the RNP family of RNA-binding proteins suggests a paradox: how does such a highly-conserved motif recognize and interact with such a wide variety of RNA sequences and structures? It has been suggested that the RNP1 and RNP2 sequences may have a general (non-specific) affinity for RNA, while non-conserved surface residues within the RRM domain, flanking sequence residues outside of the RRM domain, or even auxiliary domain residues in proteins may be responsible for recognition of a specific RNA sequence or tertiary structure (Kenan *et al.*, 1991). Results of some binding studies with members of the RNP family, including U1 snRNP protein A (U1A), hnRNP A1 and hnRNP C partially support this view and will be discussed in more detail later.

In order to better understand the mechanism of RNA recognition by the RNP family of proteins, we must ask the following questions: (1) Which amino acids within the RRM motif are involved in RNA-protein interactions? (2) Which amino acids are responsible for conferring specificity to a given subset of the RNP family? (3) What is the role of any associated auxiliary domains in the overall function of the protein? (4) What other proteins may be involved in conferring RNA binding specificity?

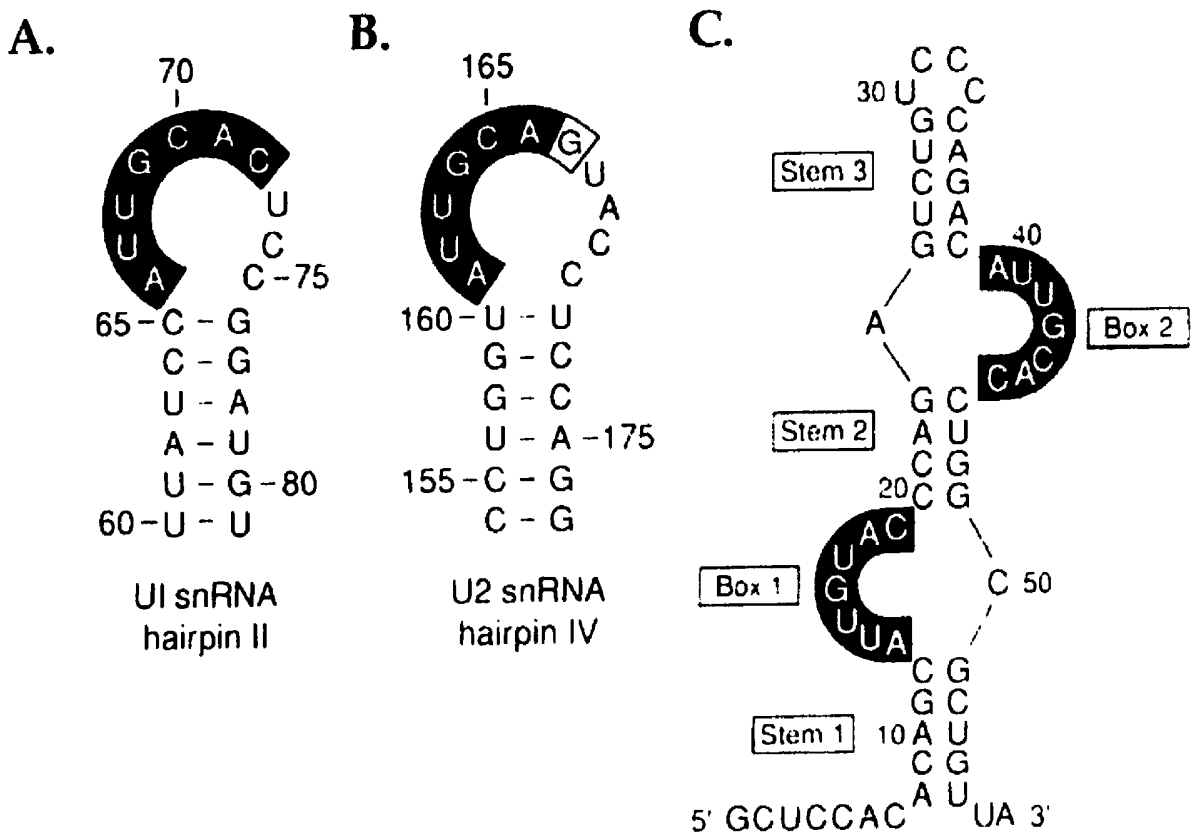
1.2.2. U1 snRNP Protein A as a Model for Studying RNA-Protein Interaction

Most eukaryotic genes contain intervening sequences known as introns which must be removed from the initial RNA transcript (pre-mRNA) to form mature mRNA. Central to that process is the assembly of a macromolecular ribonucleoprotein complex (protein and RNA) known as the spliceosome. The spliceosome is composed of a series of small nuclear ribonucleoproteins (snRNPs), each of which contains an snRNA and several protein components. Assembly of the spliceosome at the excision site of an intron is a complex, but ordered, assembly of individual snRNP components facilitated by protein-protein interactions, RNA-protein interactions and RNA-RNA interactions.

The best characterized spliceosomal snRNP is the human U1 snRNP. It consists of a 165 nt snRNA, a complex of "common" proteins found in all snRNPs, and three U1-specific proteins (U1-70K, U1-C and U1-A) (Baserga and Steitz, 1993). The U1 snRNA folds into a structure which consists of four stem-loops (loops I-IV) (Nagai and Mattaj, 1994b). The U1-70K protein binds specifically to the loop region of stem-loop I and this association is required for proper splice-site selection by U1 snRNP (Surowy *et al.*, 1989). The protein U1-C also associates with U1 snRNA and is required for splicing, but its presence requires the presence of U1-70K (Hamm *et al.*, 1990).

The most studied U1 snRNP protein is U1A, which binds to the second stem loop (loop II) of U1 snRNA (Patton *et al.*, 1989). The U1-A protein actually has several RNA targets to which it binds *in vivo*. The 10 nt loop sequence of stem-loop II of U1 snRNA was shown to be essential for recognition by U1-A (**Figure 1.3A**) (Scherly *et al.*, 1989). This interaction,

Figure 1.3. RNA structures recognized by snRNP proteins U1-A and U2-B". **(A)** Stem loop II of U1 snRNA recognized by U1 snRNP protein A (U1-A), **(B)** Stem loop IV of U2 snRNA recognized by U2 snRNP protein B" (U2-B"), **(C)** 3'-untranslated region (UTR) of U1-A pre-mRNA recognized by two molecules of U1-A. Sequences delineated as recognition elements are shown in reverse type. Reprinted from (Nagai *et al.*, 1995).


A.
70
C A C
G
U
U A
U
C
C –75
65 – C – G
C – G
U – A
A – U
U – G – 80
60 – U – U
UI snRNA
hairpin II
B.
165
C A G
G
U
U A
U
A
C
C
160 – U – U
G – C
G – C
U – A –175
155 – C – G
C – G
U2 snRNA
hairpin IV
C.
C C
30 U C
G C
U A
C G
U A
G C
40
A U U
G
A
C A C
Box 2
Stem 3
G C
Stem 2
A U
C G
20 C G
A C
U
G
U U A
Box 1
C 50
C G
G C
Stem 1
10 A U
C G
A U
5' GCUCCAC UA 3'

along with protein-protein interactions, are part of the incorporation of U1-A into the U1 snRNP complex (Hamm *et al.*, 1988). Two U1-A molecules can also bind specifically to two internal loop sequences found in the 3'-untranslated region (UTR) of U1-A pre-mRNA (**Figure 1.3C**) (van Gelder *et al.*, 1993). This interaction inhibits cleavage and polyadenylation of U1-A mRNA, thereby serving as an autoregulatory mechanism for U1-A expression (Boelens *et al.*, 1993).

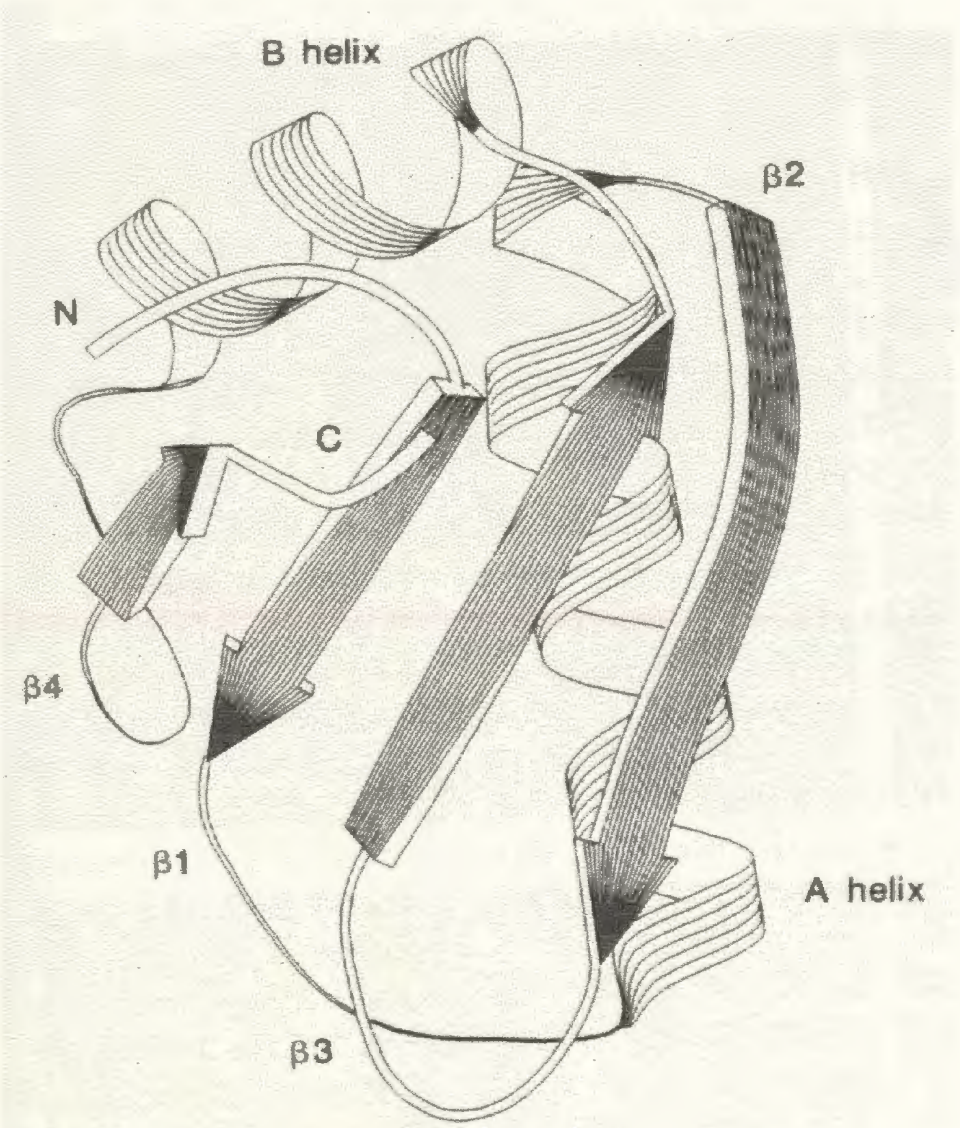
The human U1-A protein is 283 amino acids in length. It contains two copies of the RRM motif which are located at the N-terminal (amino acid residues 1 to 102) and C-terminal (amino acid residues 195 to 283) ends of the protein, respectively (Sillikens *et al.*, 1987). The two RRM motifs are separated by a protease-sensitive region of approximately 100 amino acids which is rich in both proline and lysine residues. Analysis of the binding affinities of polypeptide fragments of the human U1-A protein has revealed that the N-terminal RNA recognition motif (RRM1) was sufficient for binding to stem loop II of U1 snRNA as well as the 3'-UTR of U1-A pre-mRNA (Lutz-Freyermuth *et al.*, 1990; van Gelder *et al.*, 1993). The function of the C-terminal RRM (RRM2) of U1-A remains unknown. Binding experiments with a variety of RNA species have failed to identify a target for this RRM, suggesting that it either does not bind to RNA or perhaps requires the presence of RRM1 for its function (Lu and Hall, 1995).

It was from the N-terminal RRM fragment of U1-A that the first three-dimensional structure of an RRM was obtained by x-ray crystallography (Nagai *et al.*, 1990). The RRM has a β1-α1-β2-β3-α2-β4 structure, with the four β-strands forming an antiparallel β-sheet which is packed against two α-

helices (**Figure 1.4**). The alignment of each structural element is critical to the overall packing and stability of the RRM (Kranz *et al.*, 1996). The two α-helices (α1 and α2) are positioned almost perpendicular to one another, an arrangement most likely controlled by association of conserved hydrophobic amino acids within α1 and α2, as well as the C-terminal amino acid of RNP1 (Wittekind *et al.*, 1992). This association positions itself within the hydrophobic core of the tertiary structure and is likely a major contributor to the proper folding of the RRM. The RNP1 and RNP2 sequences are located on two adjacent β-strands (β3 and β1, respectively). Two of the aromatic amino acids known to be conserved in RNP2 and RNP1 (Tyr13 and Phe56, respectively) are located on the surface of the β-sheet. The overall conformation of the RRM results in a general RNA-binding "platform" where conserved, solvent-exposed, aromatic amino acids in RNP1 and RNP2 are believed to interact with nucleotides in RNA via a mechanism of ring stacking (Kenan *et al.*, 1991). This theory is supported by binding studies which implicate these residues as critical to RNA binding (Jessen *et al.*, 1991). This "open-platform" model for RNA interaction by an RRM therefore exposes the RNA ligand to possible interaction with other proteins or RNAs (Gorlach *et al.*, 1992).

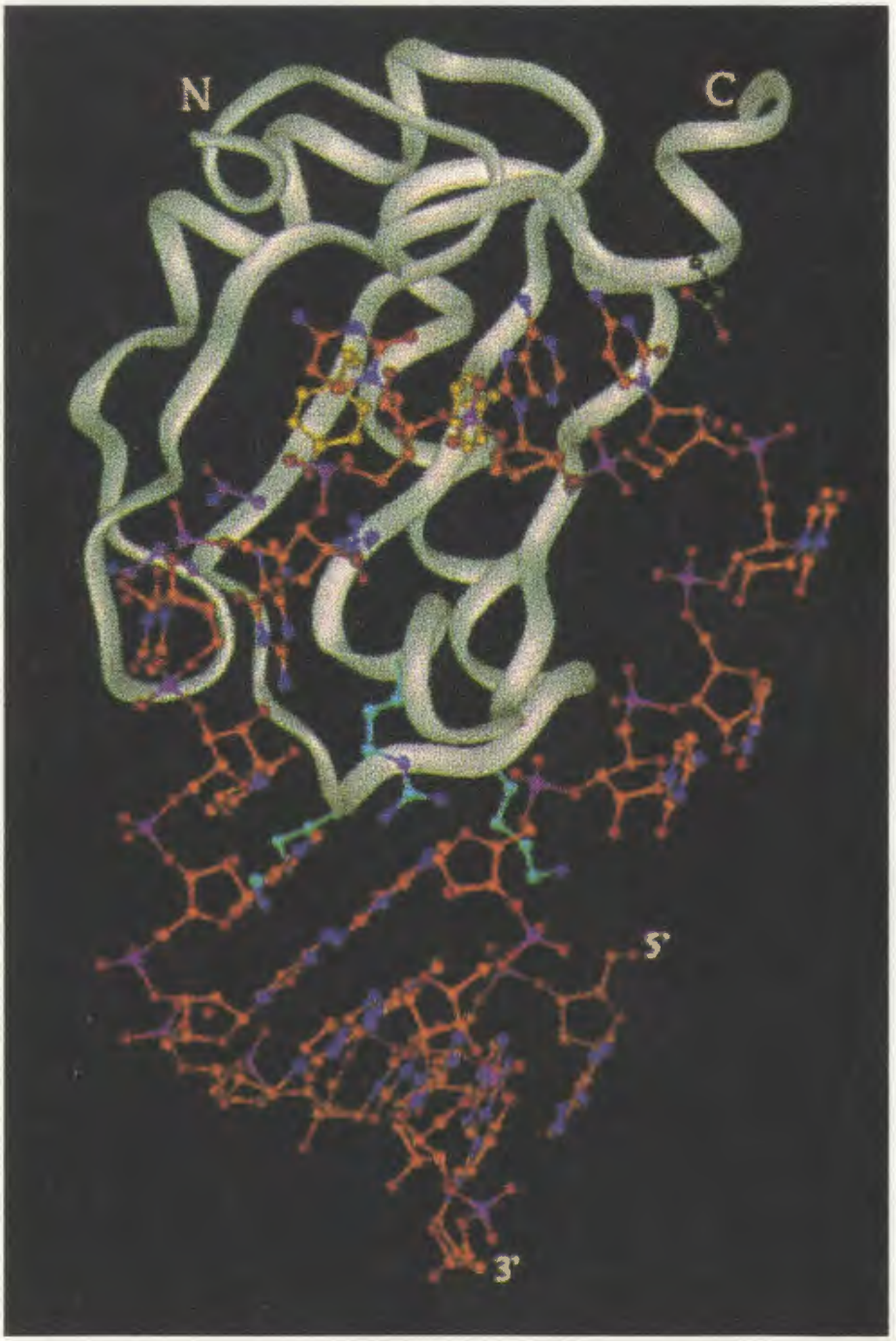
Oubridge *et al.* (1994) provided even more information about RRM-RNA interaction by solving the crystal structure of RRM1 of U1-A complexed with its RNA target, stem-loop II of U1 snRNA (**Figure 1.5**). Within this complex, the 10 nt single-stranded RNA loop lies across the β-sheet platform. The loop sequence specifically recognized by U1-A forms an extensive series of intermolecular interactions with the conserved RNP1 and RNP2 motifs, as

Figure 1.4. Ribbon diagram representation of the RRM motif. Locations of the antiparallel β-sheet (β1-β4), the two α-helices (A and B) and loop 3 are indicated. The two RNP sequences are located on the two inner β-strands (β1 and β3). Reprinted from (Nagai *et al.*, 1995).



loop 3

Figure 1.5. Ribbon representation of the crystal structure of human U1-A RRM1 bound to an RNA sequence representing the loop II region of U1 snRNA. Sidechains important for RNA binding are indicated as follows: nitrogen and oxygen atoms of amino acid side chains are coloured blue and red, respectively; carbon atoms are coloured yellow for Tyr13 (RNP2) and Phe56 (RNP1), pink for Asn15 (RNP2) and Asn15 (RNP2), purple for Gln54 (RNP1), light blue for Lys20 (loop 1) and Lys 22 (loop 1), and grey for Asp92. Reprinted from (Oubridge *et al.*, 1994) and (Nagai *et al.*, 1995).


N
C
5'
3'

well as the C-terminal extension of the RRM. Such interactions include base stacking with aromatic side chains of the RNA-binding "platform" as well as many hydrogen bonds with amide and carbonyl groups in the C-terminal region of the RRM. The polypeptide loop which joins the β2 and β3 strands of the β-sheet (loop 3) actually protrudes through the RNA loop such that the first seven nucleotides (AUUGCAC) fit into a groove formed between loop 3 and the C-terminal region of the RRM. The resulting conformation makes nucleotide bases in the RNA loop available for stacking interactions with conserved aromatic amino acids.

Solution structure NMR of a larger peptide fragment of U1-A (U1-A117) confirmed the RRM conformation observed in both crystal structures and added additional information about the region at the C-terminal of RRM1 (Avis *et al.*, 1996; Howe *et al.*, 1994). Amino acid residues Asp90 to Lys98 form an α-helix (helix C) that lies across the β-sheet structure of the RNA-binding platform. Interactions between conserved residues in helix C (Ile93, Ile94 and Met97) and the β-sheet (Leu44, Phe56 and Ile58) act as a "lid" to stabilize the protein structure by preventing exposure of hydrophobic residues on the surface of the RNA-binding platform (Avis *et al.*, 1996). Helix C moves away from this position upon binding of U1-A to its RNA target, thereby making the hydrophobic bases of the RNA-binding platform (Tyr13, Phe56 and Gln 54) available for base stacking interactions with the RNA target.

The results of structural studies identify the single-stranded RNA loop as the primary recognition site for U1-A. This would therefore imply that recognition of these unpaired bases in the RNA loop are the primary

determinant of U1-A specificity. However, this sequence interacts mainly with the highly conserved residues located on the β-sheet which are the very basis for identifying an RNP-type RNA-binding protein. This would seem to present a paradox; how can the highly conserved β-sheet structure form the basis of interaction with such a diverse set of RNA-binding proteins, each with very different, yet specific, RNA targets?

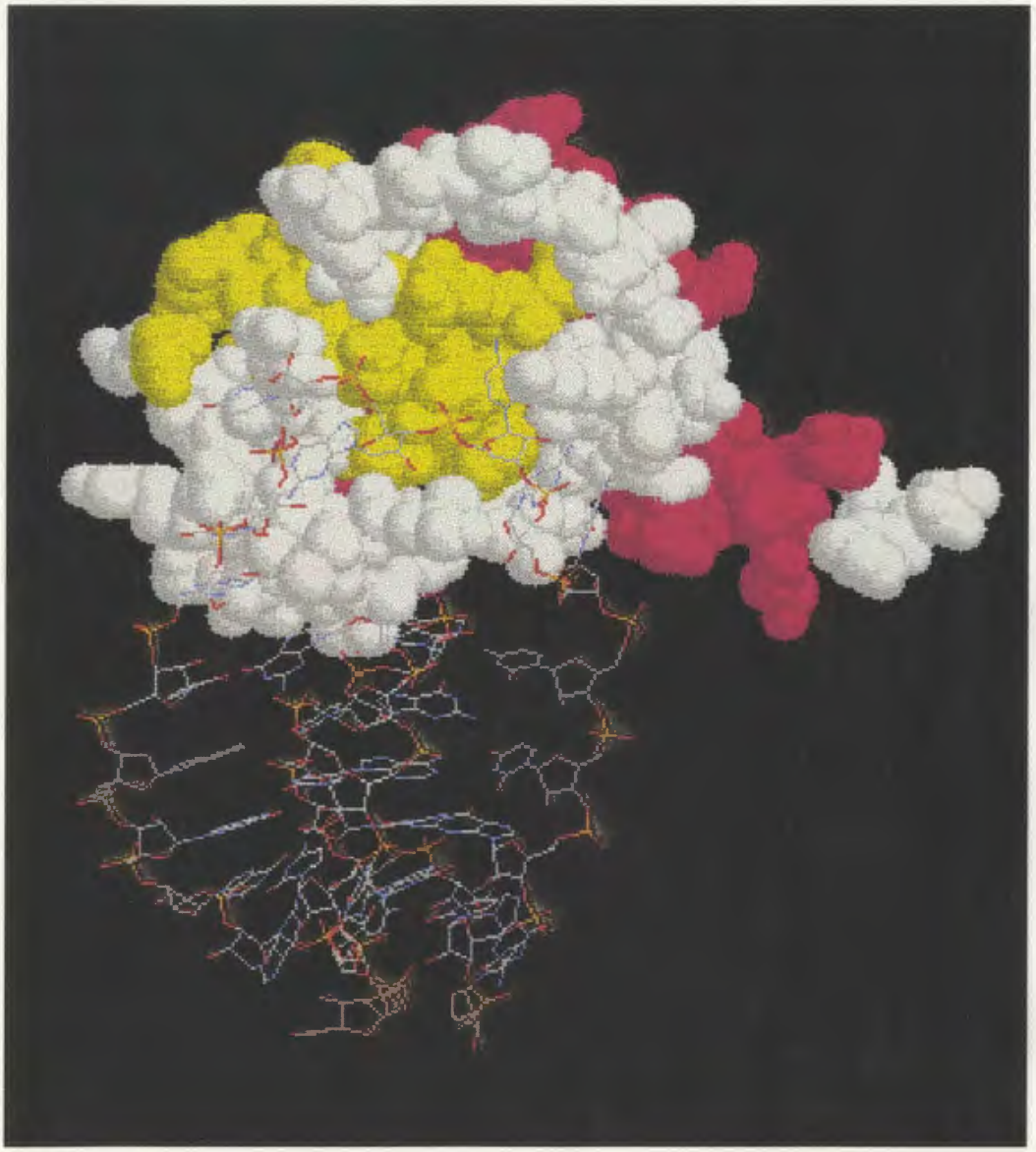
What exactly are the "determinants of specificity" for a given RNP-type RNA-binding protein such as U1-A? Much of the role of the β-sheet is to provide a more general RNA-binding platform while the specificity of RNA interaction is determined mostly by amino acids in the variable loops of the RRM (Kenan *et al.*, 1991). This basis of this specificity was elegantly demonstrated by experiments which compared the RNA binding specificity of U1-A with the closely related U2 snRNP-specific protein U2-B". The protein U2-B", which also contains two RRM motifs, binds specifically to stem loop IV of U2 snRNA (**Figure 1.3B**) but this interaction requires the presence of the U2 snRNP-specific protein U2-A' (Nagai and Mattaj, 1994b). Within the RNA sequence specifically recognized by each protein, there is a difference of only one base in the loop sequence. Without the accessory factor U2-A', U2-B" will recognize both RNA targets (Bentley and Keene, 1991). Both proteins (U1-A and U2-B") require only the first RRM domain for recognition and binding to their individual cognate RNAs. Their RRM regions show 75% sequence identity; the sequence of greatest difference between U1-A and U2-B" is a region of nine amino acids located at the start of loop 3 in the RRM (Kenan *et al.*, 1991). When amino acid residues 40 to 49 of U1-A were substituted with the corresponding residues (37 to 46) of U2-B", this exchange

resulted in a mutated U1-A protein which specifically bound to U2 snRNA (Scherly *et al.*, 1990). The importance of loop 3 in RNA recognition by U1-A is underscored by the structural data obtained by NMR and x-ray crystallography. Loop 3 was observed to be poorly ordered in the crystal structure of the free U1-A fragment (Nagai *et al.*, 1990). However, loop 3 is more highly ordered in the crystal structure of the RNA-protein complex, protruding through the RNA loop formed by stem-loop II of U1 snRNA and forming an extensive series of hydrogen bonds (Oubridge *et al.*, 1994). It has therefore been identified as a critical factor in specific RNA binding by U1-A.

More recently, NMR structural data for the complex between the N-terminal RRM of U1-A and its other RNA target, the polyadenylation inhibition element (PIE) of the 3'-UTR of U1-A mRNA, have also been obtained (**Figure 1.6**) (Allain *et al.*, 1996). As observed with the complex of U1-A with U1 snRNA, the 3'-UTR forms extensive contacts with the four-stranded β-sheet structure. Loop 3 protrudes through the hole of the RNA loop in such a way that RNA bases are splayed out across the RNA-binding platform, forming a set of molecular stacking interactions similar to those observed in the complex with U1 snRNA (Oubridge *et al.*, 1994). However, in this case, the variable loops of the RRM interact with the helical regions of the RNA. As well, there is a folding of the single-stranded RNA loop and reorganization of regions outside of the RRM such as helix C.

The results of all of these structural studies with the protein U1-A suggest that the RNA-protein interaction is a dynamic event. The RNA binding surface of U1-A takes advantage of the flexibility of RNA structure and reorganizes both the RRM and regions outside of the RRM upon RNA

Figure 1.6. Surface representation of the complex between human U1-A and the 3'-untranslated region (UTR) of U1-A pre-mRNA (Allain *et al.*, 1997). The coordinates were obtained from the Image Library of Biological Macromolecules (PDB code: *1aud*) and visualized using the RasMol 2.5 software package (Sayle and Milner-White, 1995; Suhnel, 1996). In this image, the protein is shown in a space-filling model: α-helices are shown in purple, the four-stranded β-sheet is shown in yellow. RNA is shown as a wire-frame structure. Only Box I of U1-A pre-mRNA 3'-UTR is shown.



interaction in order to maximize surface complementarity and functional group recognition (Allain *et al.*, 1997). U1-A appears to bind to RNA targets by a mutual "induced-fit" mechanism where conformational changes in both molecules aid in tight binding. In that regard, the picture of RNA-RRM interaction is much closer to the mechanisms of protein-protein interactions than those observed in DNA-protein recognition (Allain *et al.*, 1997).

1.2.3. Comparison of U1-A With the RNP-Type hnRNP Proteins A1 and C

In eukaryotes, RNA species resulting from transcription by RNA polymerase II must undergo a complex series of processing steps to form mature messenger RNA (mRNA). In the nucleus, the immature forms of RNA transcripts are known as heterogeneous nuclear RNA (hnRNA) due to the diversity of transcript sizes encompassed by this group. Throughout their lifetime in the nucleus, hnRNA are associated with a large group of RNA-binding proteins; the collective name for these complexes are the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes.

A number of techniques have been used to identify components of the hnRNP complexes (Dreyfuss, 1993). Initial studies separated hnRNP complexes from total nuclear extracts by density gradient centrifugation (Wilk *et al.*, 1985). This technique lead to the identification of the "core" hnRNP proteins (A, B and C groups). Additional proteins were identified by uv-crosslinking of RNA-protein complexes, followed by affinity purification of cross-linked polyadenylated RNA (Dreyfuss *et al.*, 1984). By far, the most sensitive method involved immunoprecipitation of hnRNP complexes using monoclonal antibodies raised against previously identified hnRNP proteins

(Pinol-Roma *et al.*, 1988). The specificity of this technique allowed for the isolation of pure, intact hnRNP complexes.

Most of the proteins characterized to date have been purified from human cells. Immunopurification of hnRNP complexes reveal a group of approximately 20 proteins (Pinol-Roma *et al.*, 1988). Most of these human hnRNP proteins have a modular structure. They usually contain one or more RNA-binding domains which may or may not include an RRM motif. Many also include an auxiliary domain whose function is still unclear. It is believed that these domains may be involved in protein-protein interactions, intracellular localization or even RNA-binding specificity (Nagai and Mattaj, 1994a).

One of the most abundant hnRNP proteins is hnRNP A1. It has several roles in eukaryotic RNA metabolism. First, it plays a key role in the packaging of pre-mRNA into a hnRNP particle through the recruitment of other hnRNP proteins (Dreyfuss, 1993). The A1 protein has been implicated in the selection of the 5' and 3' splice sites through recruitment of other splicing factors (Mayeda and Krainer, 1992). Second, as part of this function, A1 has been shown to promote the formation of base-paired double strands in RNA (Munroe and Dong, 1992). It has been suggested that this activity presents the pre-mRNA in a topology suitable for interaction with splicing snRNPs and therefore functions as part of the spliceosome assembly (Buvoli *et al.*, 1992). Third, it has been identified as a carrier of mature mRNA to the cytoplasm (Pinol-Roma and Dreyfuss, 1992).

The RNA structure recognized by A1 is not a stem-loop structure as seen with U1-A but a single-stranded RNA sequence. A1 shows a preference

for binding to both 5' and 3' splice site sequences and this binding was particularly sensitive to mutations affecting these highly conserved sequences (Buvoli *et al.*, 1990; Swanson and Dreyfuss, 1988). Selection/amplification (SELEX) experiments using random RNA pools have identified a high-affinity binding site (UAGGGA/U) for A1 which is highly similar to RNA sequences observed for 5' and 3' splice sites (Burd and Dreyfuss, 1994).

Sequence analysis of human cDNA clones reveals that the human hnRNP A1 protein is 320 amino acids in length with a monomeric molecular weight of 34 kDa (Buvoli *et al.*, 1988). The protein can be divided into two major domains which can be separated by proteolysis. The N-terminal region, which is 184 amino acids in length, consists of two RRM motifs which are arranged in tandem and separated by a linker peptide of approximately 17 amino acids. The peptide fragment (UP1), obtained by proteolysis of hnRNP A1, contains only this N-terminal region of the protein. It has been demonstrated to bind RNA independently although the C-terminal region has been shown to affect cooperativity of RNA binding (Nadler *et al.*, 1991; Riva *et al.*, 1986).

The RRM motifs of A1 share many of the characteristics outlined previously for U1-A. The precise three-dimensional structure of proteolytic fragment UP1 has recently been solved by x-ray crystallography (**Figure 1.7**) (Shamoo *et al.*, 1997). Both RRMs have the characteristic βαββαβ structure, a result which confirms earlier NMR studies (Garrett, 1994). As with U1-A, the four-stranded β-sheet of each RRM forms a flat, solvent-exposed RNA-binding surface with exposed aromatic amino acids. The most highly conserved feature of each RRM are the aromatic amino acids of RNP1 and

Figure 1.7. X-ray crystal structure of UP1, a proteolytic fragment of human hnRNP A1 (Shamoo *et al.*, 1997). The coordinates were obtained from the Image Library of Biological Macromolecules (PDB code: *1up1*) and visualized using the RasMol 2.5 software package (Sayle and Milner-White, 1995; Suhnel, 1996). The protein is shown as a ribbon structure: α-helices are shown in purple, the four-stranded β-sheets are shown in yellow.

RNP2. Photochemical cross-linking experiments with A1 and single-stranded nucleic acids have previously implicated these conserved phenylalanine residues of each RRM in RNA binding (Merrill *et al.*, 1988).

While many of the features of A1 resemble those discussed for U1-A, this protein also has several very distinguishing characteristics. The variable loop 3 of each RRM is rich in basic amino acids such as arginine and lysine, thereby creating an electropositive side to the RNA-binding platform and suggesting a role in RNA binding. However, the residues Lys20 and Lys22 which are critical to the interaction of U1-A with U1 snRNA are not maintained in A1. This change supports an earlier finding that A1 shows the strongest preferences for single-stranded RNA (Nadler *et al.*, 1991). Another key difference with A1 is the fact that two RRM motifs are required for binding to a single-stranded RNA target. Both RRMs actually face the same side of the overall structure of the protein, an orientation that is held in place by antiparallel helix-helix interactions between helix 2 of each RRM (Shamoo *et al.*, 1997). Although both RRMs face to one side of the overall structure, their binding platforms are in fact oriented in opposite directions. This suggests that the RNA ligands bound by each RRM of A1 may be reversed as well (Shamoo *et al.*, 1997). The structure of the two RRMs indicates that RNA sequences bound by each RRM motif pass within 25 Å of one another in an antiparallel manner. A1 could therefore alter the folding topology of an RNA target by the formation of an RNA loop formed when the RNA target binds to the first RRM and then loops back to bind to the second RRM (Shamoo *et al.*, 1997).

Another difference from the features observed in U1-A RNA-binding protein comes from the remaining 124 amino acids which constitute the C-terminal auxiliary domain of A1. Sequence analysis reveals that this region is rich in glycine residues (>40%). More specifically, the glycine-rich domain consists of a regular spacing of aromatic amino acids interspersed between glycine residues (Biamonti and Riva, 1994). Within this region is a 36 amino acid domain of regularly spaced RGG repeat sequences interspersed with aromatic amino acids. This "RGG box" has also been demonstrated to have RNA-binding activity, and therefore represents a second RNA-binding domain within the protein (Kumar and Wilson, 1990). Located just downstream of the RGG box is a 26 amino acid domain known as the "nuclear localization determinant." It is believed that this domain is involved in mediating the movement of A1 between the nucleus and cytoplasm by way of interaction with a carrier protein (Pinol-Roma, 1997; Wieghardt *et al.*, 1995).

The glycine-rich domain in A1 has several functions. It has been shown to confer cooperative RNA binding and therefore is likely involved in protein-protein interactions (Cobianchi *et al.*, 1988). Recent evidence has confirmed that A1 interacts with itself and other RNA-binding proteins through its glycine-rich domain, and that the aromatic residues of this domain are critical for these interactions (Cartegni *et al.*, 1996). The protein binding capacity of this domain may play a role in recruitment of other proteins during assembly of the hnRNP particle.

The auxiliary domain is also the site of several post-translational modifications. Many hnRNP proteins, including A1, are phosphorylated *in*

vivo (Dreyfuss, 1993). In A1, phosphorylation of serine residues within the auxiliary domain completely abolishes the RNA-annealing activity of this protein without affecting its nucleic-acid binding properties (Cobianchi *et al.*, 1993). The auxiliary domain of A1 also contains four arginine residues which are sites for methylation (Kim, 1997). All methylated residues are located within the domain identified as the RGG box. A role has been suggested for A1 methylation in modulating the interaction of A1 with nucleic acids (Kim, 1997).

Other hnRNP proteins which have been studied in detail are the hnRNP C proteins. This group consists of two protein isoforms (C1 and C2) which are identical except for the insertion of 13 amino acids in C2 (Burd *et al.*, 1989). Several functions have been identified for these proteins. First, this highly conserved protein has been shown to be confined to the nucleus in interphase vertebrate cells (Choi and Dreyfuss, 1984). It is therefore believed that the C proteins may act as part of a mechanism which retains unprocessed pre-mRNA transcripts within the nucleus (Legrain and Robash, 1989). Some intron sequences in pre-mRNA have been shown to contain RNA sequences which act as nuclear retention signals (Chang and Sharp, 1989). These sequences have been identified as one of the high-affinity binding sites for C proteins (Swanson and Dreyfuss, 1988a). C1 and C2 have also been implicated in proper selection of splice sites on nuclear pre-mRNA. The C proteins promote annealing of complementary RNA sequences *in vitro* (Portman and Dreyfuss, 1994). Immunoinhibition experiments with nuclear extracts using antibodies to C proteins show that spliceosome assembly is interrupted (Choi *et al.*, 1986). It has been suggested that by facilitating a particular RNA

conformation, the C proteins may be influencing binding of other splicing factors and therefore affecting particular splice-site selection (Nagai and Mattaj, 1994a).

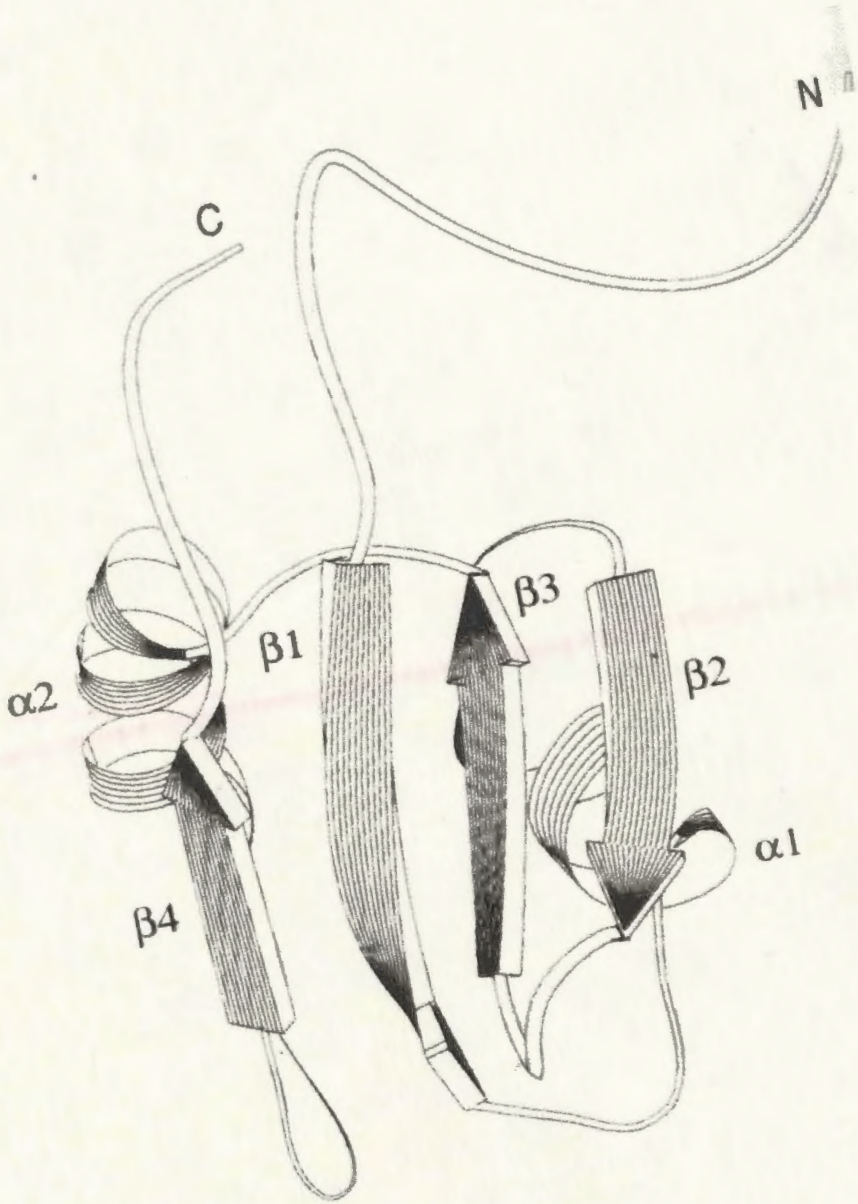
A number of methods have identified uridine (U) rich sequences as the primary RNA targets for hnRNP C. Early experiments with ribohomopolymers showed that hnRNP C bound tightly to poly (U) RNA sequences (Swanson and Dreyfuss, 1988b). Crosslinking and immunoprecipitation experiments showed that hnRNP C was associated with uridine-rich sequences located downstream of the 3' splice site (Swanson and Dreyfuss, 1988a; Wilusz *et al.*, 1988). The location of hnRNP C binding in close proximity to a pre-mRNA splice site supports its possible role in splice site selection. Selection/amplification experiments using a random pool of RNA oligonucleotides have confirmed that hnRNP C1 selects oligo(U) stretches, especially U_6 (Gorlach *et al.*, 1994).

Complementary DNA (cDNA) sequences reveal that the human hnRNP C1 and C2 proteins are 290 amino acids (41 kDa) and 303 amino acids (43 kDa) in length, respectively (Burd *et al.*, 1989). As stated previously, C2 differs only by the presence of 13 additional amino acids in the middle of the protein. The proteins have a single RRM motif at the N-terminal end of the protein as well as a C-terminal auxiliary domain containing several sub-motifs. The N-terminal region of hnRNP C (minimum 94 amino acid element) has been shown to bind poly(U) in the absence of the C-terminal auxiliary region (Gorlach *et al.*, 1994). This single RRM is a typical eukaryotic RNA recognition motif with both of the RNP1 and RNP2 sequences. Deletion analysis showed that binding to poly(U) RNA was abolished by

removal of the first five amino acids of the N-terminal region or by removal of nine amino acids at the C-terminal region of this minimal binding element (Gorlach *et al.*, 1994). It is believed that these C-terminal amino acids are the primary determinant of RNA-binding specificity (Gorlach *et al.*, 1994).

A nuclear magnetic resonance (NMR) structure of the 94 amino acid RNA-binding domain was obtained by Wittekind *et al.* (1992) (**Figure 1.8**). The RRM motif has the typical βαββαβ topology observed in other RRM structures. As expected, NMR structural analysis also revealed that upon binding to U_8 RNA, the residues most affected by RNA interaction were those residing on the RRM platform, as well as some of the N and C-terminal regions of the RRM (Gorlach *et al.*, 1992). However, there were also several key differences from the RRM structures observed previously. In particular, the N and C-terminal regions are more structurally disordered than those observed with other RRMs. This fact combined with the results of deletion analysis suggests that these regions may be the ones responsible for sequence specific recognition (Gorlach *et al.*, 1994). Since the additional 13 amino acids observed in C2 are inserted in this C-terminal region, it has been suggested that these could influence the binding specificity of the two forms of hnRNP C protein (Burd *et al.*, 1989). Another important difference is the complete absence of the variable loop (loop 3) which was shown to be an important determinant of specificity for the RNA-binding platform of U1-A (Scherly *et al.*, 1990). This loop region in U1-A makes critical contacts with the stem portion of U1 snRNA loop II, so it is not suprising that hnRNP C binds a single-stranded RNA target without obvious secondary structure.

Figure 1.8. Ribbon diagram of the RRM (residues 2-94) of hnRNP C (Wittekind *et al.*, 1992). Locations of the antiparallel β-sheet (β1-β4) and the two α-helices (α1 and α2) are indicated. The two RNP sub-motifs are located on the two inner β-strands (β1 and β3). Reprinted from (Dreyfuss, 1993).


N
C
β3
β1
β2
α2
α1
β4

The C-terminal portion of the hnRNP C proteins constitute an auxiliary domain which is rich in acidic amino acids such as aspartate and glutamate. Auxiliary domains of this type are highly similar to transcription activation domains observed in eukaryotic transcription factors such as GCN4 (Hope and Struhl, 1986). While the function of this region is unclear, it is believed to be involved in protein-protein interactions of hnRNP C with other hnRNPs and splicing factors. This domain also consists of two clusters of basic amino acids identified as the nuclear localization signal which confines hnRNP C to the nucleus (Siomi and Dreyfuss, 1995). Also present is a potential NTP binding site whose function is unclear. It is known that like hnRNP A1, the hnRNP C proteins are phosphorylated *in vivo*. These phosphorylations occur via several kinase activities, one of which is RNA-dependent, and are believed to modulate binding of hnRNP C to pre-mRNA (Holcomb and Friedman, 1984; Fung *et al.*, 1997).

1.2.4. An RNP-Type RNA-Binding Protein Involved in Alternative Splicing and Sex-Determination in *Drosophila*

Members of the RNP super-family of RNA-binding proteins are now known to be involved in a great number of RNA processing reactions. One in particular has been studied in great detail and deserves mention here. This protein, known as sex-lethal (Sxl), is involved in the developmental pathway of sex-determination in *Drosophila*. The Sxl protein acts to regulate the alternative splicing of several downstream genes, including the feminizing gene transformer (*tra*), by binding to poly(U) tracts near the proximal splice-site of *tra* pre-mRNA (Inoue *et al.*, 1990). This binding is necessary to produce

a correctly spliced mature *tra* mRNA in females (Nagoshi *et al.*, 1988). Sxl protein also regulates its own sex-specific expression by interaction with its own pre-mRNA. The result is that male *sxl* mRNA transcripts have an extra exon which contains a premature termination codon; only in females is this exon removed and the fully functional Sxl protein produced (Bell *et al.*, 1988).

The Sxl protein has several RNA targets *in vivo*. Selection/amplification experiments from pools of random RNA have identified an AU_nU_nAGU tight-binding RNA ligand, as well one U-rich target flanked by guanine ribonucleotides (Sakashita and Sakamoto, 1994; Horabin and Schedl, 1993). The U-rich sequences bound *in vivo* occur at some distance from the *sxl* splice site, and it is believed that Sxl protein recruits other splicing factors by protein-protein interaction during the splice event (Wang and Bell, 1994). As stated previously, the autoregulation of *sxl* expression is caused by Sxl interaction with its own pre-mRNA by binding cooperatively to adjacent U-rich sequences in its pre-mRNA (Wang and Bell, 1994).

The inferred amino acid sequence of the Sxl protein shows two RRM motifs arranged in tandem in the middle of the protein (Bell *et al.*, 1988). The N-terminal end of the protein is rich in glycine residues. This region has been demonstrated to be required for cooperative RNA binding (Wang and Bell, 1994). As seen with hnRNP A1, a glycine-rich region is believed to function as a protein interaction domain. Protein interaction assays have demonstrated that this is the only protein fragment required for interaction with other Sxl monomers (Wang *et al.*, 1997). The preponderance of glycine residues is believed to fold this region into a series of flexible coils (glycine

loops) with charged and polar residues at the apex of each loop which makes critical inter-molecular contacts (Steinert *et al.*, 1991).

The RRMs of Sxl have been studied, both together and individually, by sequence analysis, nuclear magnetic resonance (NMR) and RNA binding assays. There are conflicting reports as to whether or not the RRMs themselves are sufficient to confer specificity of binding to transformer (*tra*) pre-mRNA polypyrimidine tracts. Sakashita and Sakamoto (1996) demonstrated by UV crosslinking assays that both RRMs were required for proper specificity and strength of RNA binding (Sakashita and Sakamoto, 1996). This somewhat contradicts the results of Wang *et al.* (1997) who used mobility shift experiments with several RNA targets to analyze the RNA-binding activities of each domain of Sxl. They found that an Sxl protein fragment containing the two RRMs lost all specificity for U_8 RNA; specificity was retained only if the protein fragment included either the N-terminal or C-terminal region of the protein. Suprisingly, an RRM1 protein fragment by itself showed only moderately decreased affinity for U_8 RNA. Even more suprising was the fact that RRM2 by itself lost all ability to bind U_8 RNA but instead gained an affinity for U1 snRNA (Wang *et al.*, 1997).

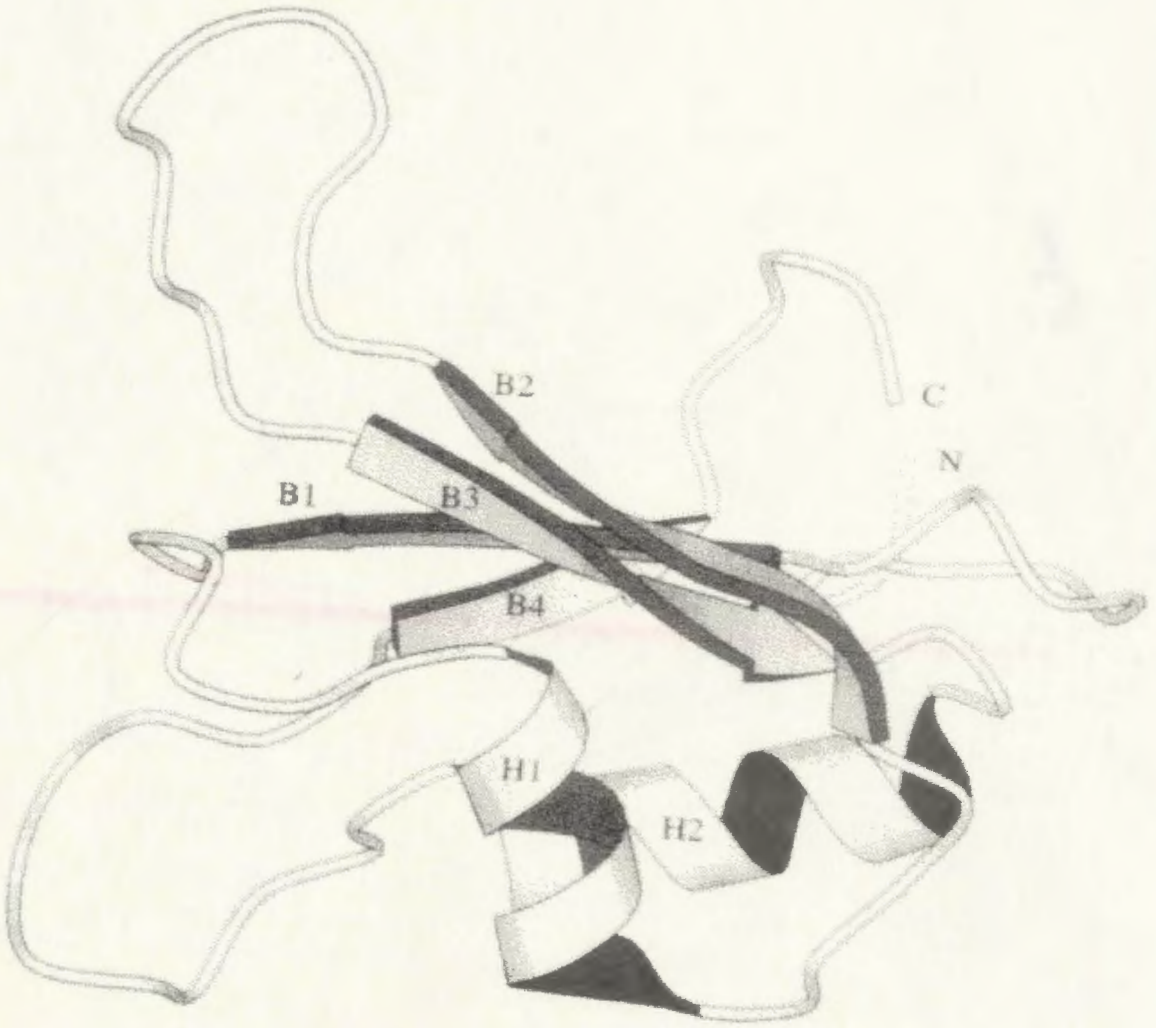
Sequence and structural analysis of the two RRMs of Sxl revealed that they are quite different. The C-terminal RNA recognition motif (RRM2) is quite similar to the RRM motifs observed in other RNA-binding proteins (Haynes, 1992). The sequences of both the RNP1 and RNP2 sequences are highly similar to those of other RRMs with several exceptions. For example, the conserved aromatic amino acid at position 3 of the RNP1 sub-motif is replaced by a valine residue in RRM2 (Bell *et al.*, 1988). The NMR structure of

RRM2 has been determined and is shown in **Figure 1.9** (Lee *et al.*, 1994). As with other RRM-type proteins discussed previously, one of the regions of particular interest is the sequence connecting the β2 and β3 strands (loop 3). In the case of Sxl RRM2, loop 3 consists of a sequence of ten amino acids, the largest loop 3 identified to date within the RNP family of proteins. It has been suggested that the presence of the charged amino acids within this loop play a role in the mechanism of RNA binding by this motif (Lee *et al.*, 1994). As observed with the hnRNP C and U1-A proteins, the two α-helices align themselves in what is essentially a perpendicular arrangement.

The sequence and structural analysis of the N-terminal RRM of Sxl (RRM1) reveals an unusual RRM motif quite different from any of those previously characterized. The three dimensional structure solved by NMR does have the typical βαββαβ pattern seen in other RRM motifs (Inoue *et al.*, 1997). However, amino acids thought to be crucial for RNA binding by the RRM are poorly conserved. First, in the RNP2 sequence, the highly conserved aromatic amino acid at position 2 which has been shown to be involved in base stacking interactions with RNA is replaced with an isoleucine residue (Ile128) (Bell *et al.*, 1988). At position 5 of RNP2, a small conserved amino acid (usually leucine) is replaced by a tyrosine residue (Tyr131).

The differences seen in RRM1 also extend to the RNP1 sequence. The N-terminal end of a typical RNP1 sub-motif almost always consists of a basic amino acid (Arg or Lys) followed by a glycine residue (Bandziulis *et al.*, 1989). Co-crystallography demonstrates that, in U1-A, this RG/KG set of amino acids forms specific contacts with the stem-loop junction of its RNA target

Figure 1.9. Ribbon diagram of the second RRM (RRM2) of *Drosophila* sex-lethal (Sxl) protein (Lee *et al.*, 1994). The four-stranded β-sheet structure (B1-B4) is shown perpendicular to the page; the two α-helices (H1 and H2) are shown below the β-sheet.


B2
C
N
B1
B3
B4
H1
H2

(U1 snRNP loop II) and is therefore crucial to the recognition of an RNA target with that secondary structure (Oubridge *et al.*, 1994). In the RRM1 motif of Sxl, this conserved basic amino acid is replaced by an aromatic amino acid (Phe166); this RRM is the only known example of such a change within the RNP family of proteins. The third major difference lies in the loop connecting the β2 and β3 strands (loop 3). Unlike other RRMs which tend to have charged amino acids at these positions, the RRM1 of Sxl contains three aromatic amino acids (Tyr160, Thr162 and Tyr164) within loop 3. This arrangement is also very rare. It is possible that in the case of RRM1, aromatic amino acids of loop 3 may interact with the nucleotides in an RNA target sequence via a mechanism of ring stacking.

Overall, the two RRMs of Sxl are quite distinct and appear to have very different RNA-binding specificities when examined separately (Sakashita and Sakamoto, 1996; Wang *et al.*, 1997). The actual role of RRM2 is somewhat unclear given that RRM1 appears to show almost normal specificity towards the natural RNA target of Sxl. It is likely that each RRM, along with the auxiliary domain mentioned previously, each have particular functions in the cooperative binding of Sxl to its RNA splice sites.

1.2.5. Evolution of the RNP Family of RNA-Binding Proteins

A comparison of the sequences of RRM motifs from a wide variety of these RNA-binding proteins reveals that, in terms of nucleotide sequences, the RRM is a loosely conserved motif. Within the 80 - 90 amino acid region known as the RRM, only 21 amino acids show a high degree of conservation; 14 of these are accounted for by the RNP1 and RNP2 sequences (Kim and

Baker, 1993). In spite of this low sequence identity, the examples outlined in this chapter demonstrate that the actual three-dimensional structure is very highly conserved. Many of the conserved amino acids of the RRM motif form the hydrophobic core essential to the proper folding of the RNA-binding platform.

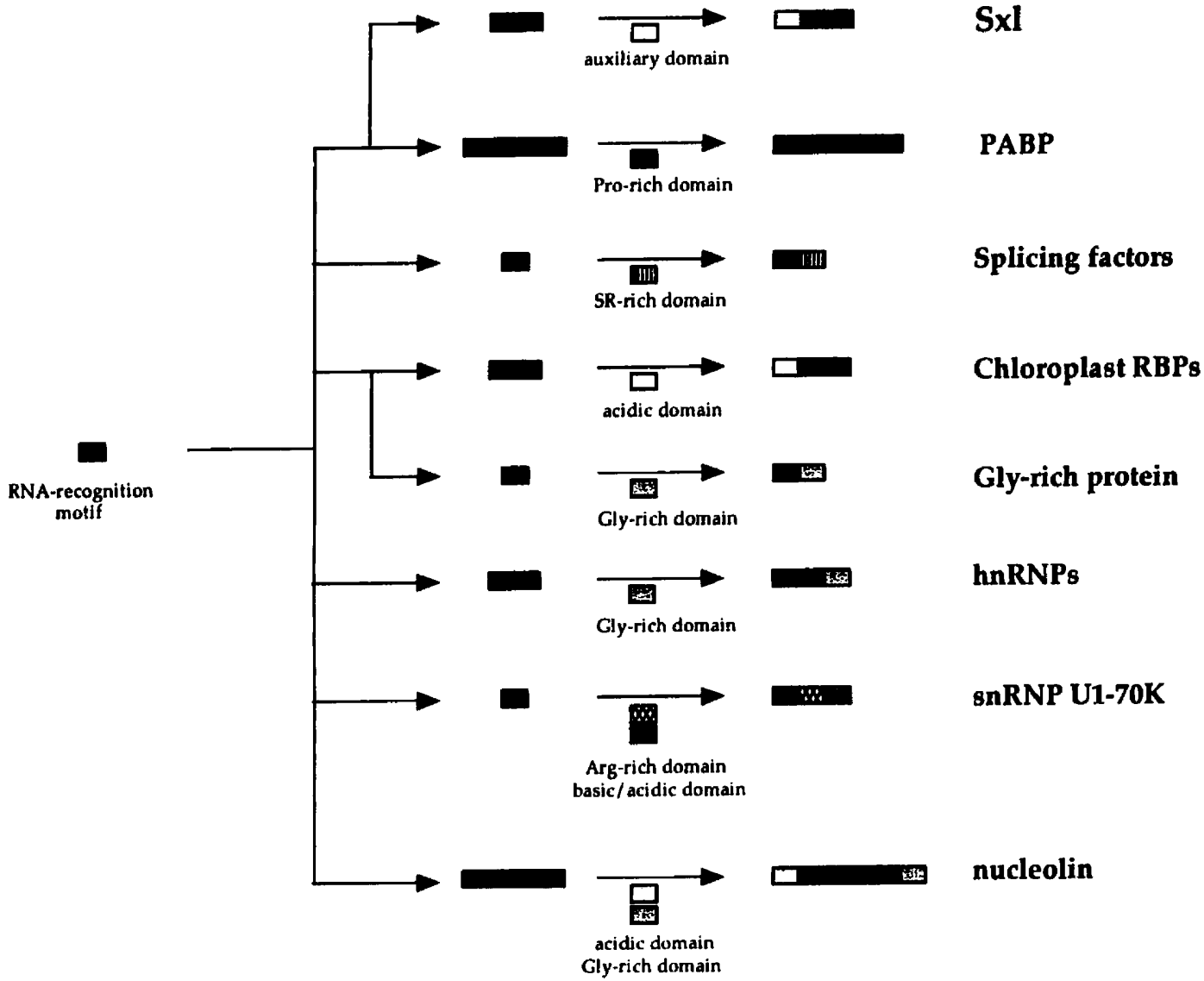
Phylogenetic comparison of a variety of RRM motifs have been used to attempt to answer the following questions: (1) How early in evolutionary history was the RRM established as an RNA-binding motif? (2) How were family members with multiple RRM motifs established and how did they evolve? (3) How did the auxiliary domains of this protein family evolve in comparison to the RRM?

The phylogenetic relationship of large samples of RRM sequences have been rigorously tested (Birney *et al.*, 1993; Fukami-Kobayashi *et al.*, 1993). The results have shown that the RRMs of functionally related sub-groups of this protein family (hnRNPs, PABPs, splicing factors, etc.) cluster on a phylogenetic tree. This correlation between RRM sequence identity and functional relatedness extends to quite diverse species of organisms, indicating that the RRM family of proteins likely evolved from a single ancient ancestral motif (Birney *et al.*, 1993). For example, comparison of the 4 RRMs present in the PABPs of yeast and man indicate that multiple RRMs appeared early in evolution. Results indicate that these multiple RRMs arose by a gene duplication event which predated the divergence of yeast and man. Following duplication, the sequence and function of each RRM is believed to have evolved independently, a theory supported by binding studies with individual RRMs.

A comparison of auxiliary domains from various RNA-binding proteins also indicate a clustering based on function over a diverse group of organisms. This clustering suggests that the fusion of RRMs with these domains occurred before the divergence of the RRM family into these protein sub-groups. A proposed pathway for the evolution of the RRM family of proteins is shown in **Figure 1.10** (Fukami-Kobayashi *et al.*, 1993). In this scheme, duplication of the RRM and fusion with a variety of auxiliary domains subsequently gave rise to a diverse collection of proteins which then evolved independently.

The ancient origin of the RRM motif is underscored by the identification of six chloroplast encoded RNA-binding proteins in tobacco and spinach (Schuster and Gruissem, 1991; Ye *et al.*, 1991). It is now generally believed that eukaryotic cellular organelles such as chloroplasts and mitochondria originated as ancient eubacterial-like endosymbionts (Gray, 1989). In particular, the chloroplast is believed to have evolved from an ancient cyanobacterial endosymbiont (Gray, 1993). The recent discovery of proteins of the RNP family in many species of cyanobacteria therefore has several implications (Mulligan *et al.*, 1994). First, since these proteins form a cluster with chloroplast RNA-binding proteins in a phylogenetic analysis, it provides additional evidence in support of the endosymbiont theory of chloroplast evolution. Second, it positions the emergence of the RRM motif as pre-dating the divergence of prokaryotes and eukaryotes. Third, it implicates an ancient cyanobacterial RRM as the likely progenitor of the modern RNP family of RNA-binding proteins. From this perspective, the study of cyanobacterial RNA-binding proteins would yield information

Figure 1.10. Evolution of the RNP-family of RNA-binding proteins, as proposed by Fukami-Kobayashi *et al.* (1993). Proteins outlined include sex-lethal protein (Sxl), poly (A) binding protein (PABP), splicing factors such as human SF2/ASF (Mayeda and Krainer, 1992), chloroplast RNA-binding proteins from tobacco (Ye *et al.*, 1991) and spinach (Schuster and Gruissem, 1991), plant glycine-rich stress proteins (Cruz-Alvarez and Pellicer, 1987; Dunn *et al.*, 1996; Sturm, 1992; van Nocker and Vierstra, 1993), heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Buvoli *et al.*, 1988), U1 small nuclear ribonucleoprotein 70K (Surowy *et al.*, 1989) and nucleolin (Lapeyre *et al.*, 1987).


RNA-recognition motif
auxiliary domain
Sxl
Pro-rich domain
PABP
SR-rich domain
Splicing factors
acidic domain
Chloroplast RBPs
Gly-rich domain
Gly-rich protein
Gly-rich domain
hnRNPs
Arg-rich domain
basic/acidic domain
snRNP U1-70K
acidic domain
Gly-rich domain
nucleolin

useful in the corresponding studies of the eukaryotic RNP-type RNA-binding proteins.

1.3. RNP-TYPE RNA-BINDING PROTEINS IN CYANOBACTERIA

Cyanobacteria are a very ancient and diverse group of prokaryotes. Early fossil evidence has indicated that the earliest form attributable to cyanobacteria appeared approximately 3.5 billion years ago (Hayes, 1983). The emergence of oxygen-evolving cyanobacteria during the Precambrian Era led to the gradual oxygenation of the Earth's atmosphere and the formation of the stratospheric ozone shield (Garcia-Pichel, 1998). This protection and aerobic conditions were part of the evolutionary transitions which preceded the rise of aerobic microorganisms and ultimately, early plants and animals (Schopf *et al.*, 1983).

Cyanobacteria can be broadly subdivided into 5 distinct groups: Group I (unicellular), Group II (pleurocapsalean), Group III (non-heterocystous, filamentous), Group IV (heterocystous, filamentous, non-branching), Group V (heterocystous, filamentous, branching) (Rippka *et al.*, 1979). Genomic Southern blots from cyanobacterial strains representing most of these groups show that genes encoding RNP-type RNA-binding proteins are widespread in cyanobacteria (**Figure 1.11**) (Mulligan *et al.*, 1994). Under the hybridization conditions used in these blots, copies of *rbp* genes were observed to be relatively abundant (4-10 copies) in the heterocystous species (Group IV and V). The only exception was *Calothrix* which contained only two strongly hybridizing bands; however, two weakly hybridizing signals observed within this organism may also represent *rbp* genes. In contrast to the heterocystous species, non-heterocystous species appeared to contain relatively few copies of *rbp* genes; *Oscillatoria* and *Pseudanabaena* each

Figure 1.11. Southern hybridization of a cyanobacterial *rbp* gene with genomic DNA from other cyanobacteria, as outlined in Mulligan *et al.*, 1994. **Group I:** *Synechococcus sp.* PCC 7942. **Group III:** *Oscillatoria sp.* PCC 7515; *Pseudanabaena sp.* PCC 7403; *LPP* Group B Strain *sp.* PCC 7124; *Phormidium foveolarum* CCAP 1442-1. **Group IV:** *Anabaena sp.* PCC 7120; *Nodularia sp.* PCC 73104; *Cylindrospermum sp.* PCC 7604; *Nostoc MAC sp.* PCC 7911 R2; *Scytonema sp.* PCC 7110; *Calothrix sp.* PCC 7102; *Calothrix sp.* CCAP 1410-1 (representative of *Tolypothrix*). **GroupV:** *Chlorogloeopsis sp.* PCC 6912; *Fischerella sp.* PCC 7414. Also shown are the hybridization analyses of *Escherichia coli* MC1061 and *Rhodobacter capsulatus* SG1003. Bacteriophage λ BstEII DNA fragments used as size markers are indicated.

1.2
2.3
3.6
5.5
14.1
E. coli
R. capsulatus
Synechococcus
Anabaena
I
1.2
2.3
3.6
5.5
14.1
Oscillatoria
Pseudanabaena
LPP
Phormidium
Anabaena
Nodularia
Cylindrospermum
Nostoc
Scytonema
Calothrix
Tolypothrix
Chlorogloeopsis
Fischerella
III
IV
V

contained only two copies whereas *Phormidium* and an *LPP* Group B strain did not contain any *rbp* genes. No copies were observed in the Gram negative bacteria *Escherichia coli* and *Rhodobacter capsulatus*. The significance of these observations is discussed in Chapter 6.

The unicellular species *Synechococcus* 7942 (*Anacystis nidulans* R2) appeared to contain a single hybridizing fragment of 6.4 kb in size. The characterization of that gene and its comparison to other cyanobacterial *rbp* genes formed the basis of my research work. This unicellular strain was chosen for a number of reasons. First, it appeared at the time that this species contained only a single hybridizing fragment representing a single *rbp* gene. Inactivation studies of this relatively simple system would therefore not be complicated by the presence of additional *rbp* genes within the same organism. Second, the unicellular *Synechococcus* species is a very amenable system for mutational analysis of genes. It is a strain which is naturally competent and therefore readily transformable (Shestakov and Khyen, 1970). It also contains a very efficient system of recombination, most likely due to the necessity of this photoautotrophic organism to repair radiation damage to DNA by light (Kuhlemeier and van Arkel, 1987). Vector systems have been established which take advantage of this efficient recombination system to integrate DNA fragments such as antibiotic resistance cassettes into specific locations within the cyanobacterial chromosome (Kuhlemeier *et al.*, 1983).

My work on this *rbp* gene in *Synechococcus* 7942, which I have named *rbpA*, is divided into several sections. Chapter 2 deals with the initial cloning and sequence analysis of the *rbpA* gene, including its comparison with other known cyanobacterial *rbp* genes. Chapter 3 describes the results of

experiments designed to inactivate the *rbpA* gene in *Synechococcus* 7942 and study the resulting phenotype. Chapter 4 summarizes the strategies used to complement the "knock-out" phenotype by re-introduction of an intact copy of the *rbpA* gene. Chapter 5 deals with the isolation of the RbpA gene product and the initial characterization of its RNA-binding properties. Finally, in chapter 6, I bring together and summarize all of these findings while speculating on some of the possible roles for *rbp* genes in *Synechococcus* and in cyanobacteria in general.

CHAPTER 2. CLONING AND SEQUENCE ANALYSIS OF THE *rbpA* GENE

2.1. INTRODUCTION

Proteins belonging to the RNP family of RNA-binding proteins were found exclusively in eukaryotes until Kovacs *et al.* (1993) reported that a number of proteins were precipitated from total protein extracts of *Synechococcus leopoliensis* by anti-RNP or by anti-Sm (core snRNP proteins) sera. Small RNAs, some of which carried the snRNP-unique 5′ $m^{2,2,7}G$ (m^3G) cap structure were also precipitated by the sera and by a monoclonal anti-m^3G antibody.

Our laboratory reported the characterization of the first genes encoding RNP-type RNA-binding protein in two species of filamentous cyanobacteria, *Anabaena sp.* PCC 7120 (*Anabaena* 7120) and *Chlorogloeopsis sp.* PCC 6912 (*Chlorogloeopsis* 6912) (Mulligan *et al.*, 1994). Genomic Southern blots carried out at that time indicated that many species of cyanobacteria possess RNP-type RNA-binding protein (*rbp*) genes (Mulligan *et al.*, 1994). At the same time, two genes were identified in a unicellular species, *Synechococcus sp.* PCC 6301 (*Synechococcus* 6301) (Sugita and Sugiura, 1994). Subsequently, five genes have been characterized in the filamentous strain *Anabaena variabilis* M3 (*Anabaena* M3) (Sato, 1994; Sato, 1995) and our laboratory has characterized two further genes from *Anabaena* 7120 (Holden, 1995; Wu and Mulligan, unpublished). RNA-binding protein genes have also been sequenced in *Agmenellum quadruplicatum* PR-6 (Wagner *et al.*, 1993) and *Synechocystis sp.* PCC 6803 (*Synechocystis* 6803) (Nakamura *et al.*, 1996). All but one of the cyanobacterial genes characterized to date code for a short polypeptide (95-110 amino acids) with a single RNA Recognition Motif

(RRM) and most of them contain an auxiliary domain consisting of a short stretch of glycine residues at the C-terminus (Mulligan and Belbin, 1995).

This chapter outlines the cloning and sequence analysis of an RNA-binding protein gene (*rbpA*) in the unicellular cyanobacterium *Synechococcus* 7942. Comparison of the inferred amino acid sequence with those of the other known cyanobacterial Rbps reveal that the protein has the typical modular structure of a single RRM followed by an auxiliary domain which is rich in glycine residues. It also demonstrates the high degree of structural conservation seen within this family of proteins.

Computational analysis of the nucleotide sequence flanking the *rbpA* gene reveals secondary structure capable of forming in RNA transcripts of *rbpA* which may provide a means for regulating the expression of this gene. These structures have also been seen in the flanking sequence of other cyanobacterial *rbp* genes.

2.2. MATERIALS AND METHODS

2.2.1. Plasmids and Strains

All chemicals used for cloning and sequencing were of reagent grade or better. Restriction endonucleases and other enzymes were purchased from New England Biolabs, Pharmacia, or Bethesda Research Laboratories (BRL), and were used according to the recommendations of the suppliers. A complete list of plasmids used for this work is given in **Table 2.1**. Liquid cultures of all *Escherichia coli* (*E. coli*) strains used were grown in LB medium (1% Bacto®Tryptone (Difco), 0.5% Bacto® Yeast Extract (Difco), 1% NaCl) at 37°C in a Series 25 Incubator Shaker (New Brunswick Scientific Co.) shaking at 200 rpm. LB plates containing 1.5% Bacto® Agar were incubated 12-18 hrs at 37°C in a GCA/Precision Scientific Model 6M Incubator.

Synechococcus 7942 wild type and mutant cultures were grown on BG-11 medium at room temperature with constant illumination

2.2.2. Isolation of Genomic DNA from *Synechococcus* 7942

Genomic DNA from 1 L cultures of *Synechococcus* 7942 was prepared as described by Golden (1987). Cells were collected by centrifugation (5000 rpm, 5 min) and resuspended in 5 mL of TE_{100} (100 mM Tris, 100 mM EDTA, pH 8.0) in a 25 mL Corex tube. Autoclaved glass beads (5 mL), 5 mL of phenol-chloroform-isoamyl alcohol (25:24:1) and 250 µL of 20% SDS were added to the cell suspension and the entire solution vortexed at top speed in 3 min intervals for a total of 9 min, cooling briefly on ice between each interval. This mixture was centrifuged (10,000 rpm, 10 min) and the upper

Table 2.1. Plasmids Used In Sequence Analysis

Plasmid	Description	Reference
pBGS18	Analogue of the cloning vector pUC18 which contains a gene for kanamycin resistance instead of ampicillin resistance	Spratt *et al.*, 1986
pBR322	Amp^r Tet^r	Bolivar *et al.*, 1977
pSyR1	6.3 kb HindIII fragment containing the *Synechococcus* 7942 *rbpA* gene cloned in pBR322	This work
pSyR1.2	1.45 kb NheI/HindIII fragment containing the *Synechococcus* 7942 *rbpA* gene cloned in pUC19	This work
pSyR1.3	4.1 kb NheI fragment containing the *Synechococcus* 7942 *rbpA* gene cloned in pBGS18	This work
pSyR1.4	2.65 kb HindIII/NheI fragment downstream of the *Synechococcus* 7942 *rbpA* gene cloned in pBGS18	This work
pSyR1.5	1.3 kb BamHI/HindIII fragment containing the *Synechococcus* 7942 *rbpA* gene cloned in pUC19	This work
pUC18/19	Amp^r	Yanisch-Perron*et al.*, 1985

aqueous phase transferred to a clean 25 mL Corex tube. This crude preparation was extracted once with phenol, once with phenol-chloroform-isoamyl alcohol (25:24:1), and once with chloroform-isoamyl alcohol (24:1). At each stage, phases were separated by centrifugation (5000 rpm, 5 min). Nucleic acid precipitation was carried out by adding 0.5 volumes of 7.5 M ammonium acetate, pH 7.5 followed by 2.5 volumes of 100% ethanol. The mixture was incubated on ice for 30 min and total nucleic acid pelleted by centrifugation (10,000 rpm, 10 min).

The crude nucleic acid pellet was resuspended in 500 µL of TE (10 mM Tris, 1 mM EDTA, pH 8.0) and transferred to a 1.5 mL microfuge tube. This solution was extracted once with phenol-chloroform (1:1) and once with chloroform. RNA was removed by adding 0.5 volumes of 7.5 M ammonium acetate, pH 7.5, incubating on ice for 30 min, and centrifuging (10,000 rpm, 10 min). The supernatent was transferred to a clean microfuge tube and the genomic DNA precipitated with 100% ethanol as outlined above; intact genomic DNA was spooled from solution with a drawn pasteur pipette, dried in air, and resuspended in TE. The quantity of recovered genomic DNA was determined spectrophotometrically by measuring the absorbance at 260 nm. The quality of the recovered DNA was determined by calculating the ratio of absorbances taken at 260 nm and 280 nm and by agarose gel electrophoresis.

2.2.3. Isolation of Plasmid DNA

Small scale isolation of plasmid DNA from cultures of *E. coli* was carried out using the Merlin mini-prep protocol of Iyer (1993). Cells from 5 mL LB cultures were pelleted by centrifugation and resuspended in 200 µL

of Cell Resuspension Solution (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 100 μg/mL RNase A). Cells were lysed by addition of 200 μL of Cell Lysis Solution (0.2 M NaOH, 1% SDS) and the solution then neutralized by the addition of 200 μL of Neutralization Solution (1.25 M potassium acetate, 1.24 M acetic acid) to precipitate cellular protein and genomic DNA. After a brief centrifugation, the supernatent was transferred to a fresh tube containing 1 mL of DNA-binding resin (Celite resin slurry in 7 M guanidine hydrochloride). After gentle mixing, the slurry was passed through a mini-column (Promega). Plasmid DNA retained in the column was washed with 1-2 mL of Wash Solution (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5 mM EDTA, 50% ethanol) and then eluted into a clean microfuge tube with 60 μL of warm TE. Recovery of plasmid DNA was analyzed by restriction endonuclease digestion, followed by gel electrophoresis.

Large scale isolation of plasmid DNA was carried out as follows. Cells from a 500 mL overnight culture were harvested by centrifugation (10,000 rpm, 10 min), washed with M9 salts (42 mM Na_2HPO_4, 22 mM KH_2PO_4, 9 mM NaCl, 19 mM NH_4Cl) and resuspended in 20 mL of GTE (25 mM Tris, pH 8.0, 10 mM EDTA, 50 mM glucose). Cells were lysed by the addition of 40 mL of Lysis Solution (0.2 M NaOH, 1% SDS) followed by gentle mixing and incubation on ice for 5 min. The solution was neutralized by the addition of 30 mL KAcF (4.79 M potassium acetate, 1.45 M formic acid) and incubated on ice for another 5 min. Precipitated protein and nucleic acid was removed by filtration through two layers of cheesecloth. Plasmid DNA in the filtrate was precipitated by the addition of 0.6 volumes of ice-cold isopropanol;

this solution was centrifuged (10,000 rpm, 10 min) and the supernatent discarded.

The crude plasmid DNA was resuspended in 4 mL of TE, extracted twice with phenol and once with chloroform. To facilitate the removal of contaminating RNA, 4 mL of 7.5 M ammonium acetate, pH 7.5 was added and the solution centrifuged (10,000 rpm, 10 min). The supernatent (approximately 8 mL) was transferred to a fresh tube and the plasmid DNA was precipitated by the addition of 5 mL of isopropanol followed by incubation on ice for 30 min. After centrifugation (10,000 rpm, 10 min), the plasmid DNA pellet was resuspended in 0.5 mL of TE. Its quality and quantity were verified spectrophotometrically and by agarose gel electrophoresis.

2.2.4. Gel Electrophoresis of DNA and Transfer by Southern Blot

Gel electrophoresis of DNA samples was carried out in 0.5X TBE buffer (1X TBE is 100 mM Tris-Cl, pH 8.0, 100 mM boric acid, 2.5 mM EDTA) containing ethidium bromide (50 µg/mL). In circumstances where precise determination of DNA fragment size was necessary, agarose gels were run in 0.5X TBE buffer without ethidium bromide and subsequently stained. DNA fragments were visualized on a Chromato-Vue TM-36 transilluminator (Ultra-Violet Products) and photographed with a Polaroid MP-4 Land camera using Polaroid 665 positive/negative instant Land pack film.

Gels to be used for Southern transfer were first treated with 0.25 M HCl for 15 min, then briefly rinsed with distilled water. DNA was transferred to Hybond N+ nylon membrane (Amersham) using a downward capillary transfer in 0.4 M NaOH (Reed and Mann, 1985; Chomczynski, 1992). The

membrane containing transferred DNA was briefly soaked in 200 mL of a neutralization solution (0.2 M Tris-Cl, pH 7.5, 2X SSC (1X SSC is 150 mM sodium chloride, 15 mM sodium citrate)) and thoroughly dried at 60 °C. Efficient transfer of DNA was confirmed using a Model UVL-56 hand-held transilluminator (UVP Inc.) at a wavelength of 366 nm.

2.2.5. Hybridization and Detection of Digoxygenin (DIG)-Labelled DNA Probes

Detection of DNA hybrids was carried out using an enzyme-linked immunoassay system (Boehringer Mannheim). DNA fragments to be used as probes were labelled by random priming using the Klenow fragment of DNA polymerase. One of the nucleotides incorporated (DIG-dUTP) was linked via a spacer arm to the steroid hapten digoxygenin (DIG). Hybridization of this DIG-labelled probe was detected by an anti-digoxygenin alkaline phosphatase conjugate using a colour reaction with 5-bromo-4-chloro-3-indolyl phosphate (X-phos) and nitroblue tetrazolium salt (NBT).

Approximately 1-2 µg of a linear DNA fragment was used to make a DNA probe. The DNA was completely denatured by boiling for 10 min, quickly cooled on ice, and then incubated overnight at 37°C in the labelling mixture (6.25 A_{260} units/mL random hexanucleotides, 100 µM dATP, 100 µM dCTP, 100 µM dGTP, 65 µM dTTP, 35 µM DIG-dUTP, 2 Units of labelling-grade Klenow enzyme, pH 6.5). This reaction was stopped by the addition of EDTA to a final concentration of 40 mM. Total DNA was ethanol precipitated as outlined above, dried, and resuspended in 50 mL of TE.

Nylon membranes containing transferred DNA were briefly soaked in 2X SSC, then incubated 1-2 hrs at 60°C in Hybridization solution (5X SSC, 0.1%

N-lauroyl sarcosine, 0.8% SDS, 0.5% Blocking Compound (Boehringer Mannheim)). This solution was then replaced with Hybridization solution containing the DIG-labelled DNA fragment, and hybridization was allowed to proceed for 24-48 hrs at 60°C.

Washes and subsequent colour detection of bound probe were carried out using a modified procedure from the manufacturer's recommended protocol. Membranes were washed twice with 2X SSC plus 0.5% SDS (room temperature, 15 min), 3 times with 0.5X SSC plus 0.5% SDS (60°C, 20 min), and twice with 0.5X SSC without SDS (room temperature, 10 min). This was followed by a 30 min room temperature incubation with 1% blocking compound in Buffer I (100 mM Tris-Cl, 150 mM NaCl, pH 7.5) to eliminate nonspecific binding of the antibody. The membrane was then incubated for 30 min in a solution of polyclonal sheep anti-digoxygenin Fab-fragments conjugated to alkaline phosphatase (150 mU/mL) in Buffer I. Unbound antibody was removed by washing the membrane twice with Buffer I (room temperature, 15 min). Hybridized probe was detected using at least 20 mL of colour solution (100 mM Tris-Cl, pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$, 0.338 mg/mL NBT, 0.175 mg/mL X–Phos). For a typical 11 cm by 15 cm nylon membrane, this corresponded to 8.25 cm^2 nylon membrane/mL colour solution. Distinct bands corresponding to the location of hybridized probe typically appeared within 1 hr. The colour reaction was stopped by washing the membrane with TE and allowing it to dry at room temperature overnight.

2.2.6. DNA Cloning and Sequencing

DNA was purified from agarose gel slices using the protocol described by Heery (1990). The gel slice was placed inside a punctured 0.5 mL microfuge tube containing 1-2 mm of aquarium filter floss (Levine, 1994). This was centrifuged at 6000 rpm for 10 min, yielding 100-200 µl of liquid containing the DNA fragment of interest. This DNA was subsequently purified by phenol:chloroform (1:1) extraction and ethanol precipitation as outlined previously. Recovery and quality of the purified fragment were confirmed by agarose gel electrophoresis.

Isolated DNA fragments were cloned into suitable cloning vectors using T4 DNA ligase. Ligation reactions were carried out in ligation buffer (66 mM Tris-Cl, pH 7.5, 6.6 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP) at room temperature for at least 1 hr prior to transformation. T4 DNA ligase was generally inactivated by heating at 65°C for 20 min as this has been shown to increase the efficiency of transformation reactions with *E. coli* (Michelson, 1995). Transformed cells containing plasmid DNA were selected on LB plates containing the appropriate antibiotic. The ligation mixture was used to transform an *E. coli* host using a standard $CaCl_2$ protocol (Sambrook *et al.*, 1989).

Fragments to be sequenced were cloned into plasmid pUC18 or pUC19. Template DNA for sequencing was then prepared by alkali denaturation of plasmid DNA. Approximately 10 µg of plasmid DNA was incubated at 37°C for 5 min in 0.2 M NaOH/0.2 mM EDTA. This solution was quickly cooled on ice, and then neutralized with 3 M sodium acetate, pH 5.2. The denatured

plasmid DNA was ethanol precipitated, resuspended in 20 µl of TE, and stored at -20°C until needed.

Sequencing was carried out on both DNA strands using the dideoxynucleotide sequencing method with Sequenase 2.0® (U.S. Biochemical Corp.) following the recommended protocol. In the primer annealing reaction, 3 - 4 µg of template DNA and 0.5 pmoles of sequencing primer were mixed in Sequenase® reaction buffer (40 mM Tris-Cl pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl). This mixture was heated to 65°C for 5 min, then allowed to cool to 35°C (approximately 1°C/min). The annealed template-primer was then incubated for 5 min at room temperature in 6.67 mM DTT, 0.2 µM dGTP, 0.2 µM dCTP, 0.2 µM dTTP, 5 µCi [α-^{35}S] dATP and 3.25 Units of Sequenase 2.0 DNA polymerase to allow extension from the primer and incorporation of radiolabelled dATP. To terminate the elongation reaction, 3.5 µl was transferred to each of 4 tubes containing 2.5 µl of termination mixture (one for each base), such that the final concentration of each deoxynucleotide was 33 µM and final concentration of dideoxynucleotide was 3.3 µM. This reaction was incubated 5 min at 37°C, then stopped by the addition of 4 µl of stop solution (0.05% Bromophenol Blue, 0.05% Xylene Cyanol FF, 20 mM EDTA, 95% formamide). DNA fragments were separated on a 6% polyacrylamide sequencing gel in 7 M urea, using a Sequi-Gen sequencing apparatus (Biorad) at a constant power of 30 W. The gels were fixed in a 10% methanol/10% acetic acid solution, dried on a DryGel Sr. Slab Gel Dryer Model SE 1160 (Hoefer), and the results visualized by autoradiography using Kodak X-OMAT Scientific Imaging Film.

2.2.7. Computer-Derived Figures and Sequence Analysis

All Southern blots were scanned on a Hewlett Packard ScanJet Plus at 300 dpi using DeskScan II (version 2.1). Analysis of nucleic acid sequences was carried out using the DNAlysis program of Dr. Bill Buikema (University of Chicago). Sequence identity calculations for the nucleic acid and amino acid sequences were carried out using the ALIGN and FASTA software packages on a Macintosh (Pearson & Lipman, 1988). The nucleotide sequence region upstream of *rbpA* was folded using the program Mulfold, displayed with the program loopDloop, and annotated by hand (Jaeger *et al.*, 1989; Zuker, 1989). Multiple alignments of the upstream sequences were carried out using the CLUSTALW Multiple Sequence Alignment Program (Thompson *et al.*, 1994). Alignments of cyanobacterial RBPs and upstream sequences were visualized using DNADraw on a Macintosh (Shapiro, 1995). Nucleotide sequences were retrieved from the GenBank database using the Entrez WWW Server or ExPASy WWW Server (Appel *et al.*, 1994; Benson *et al.*, 1998).

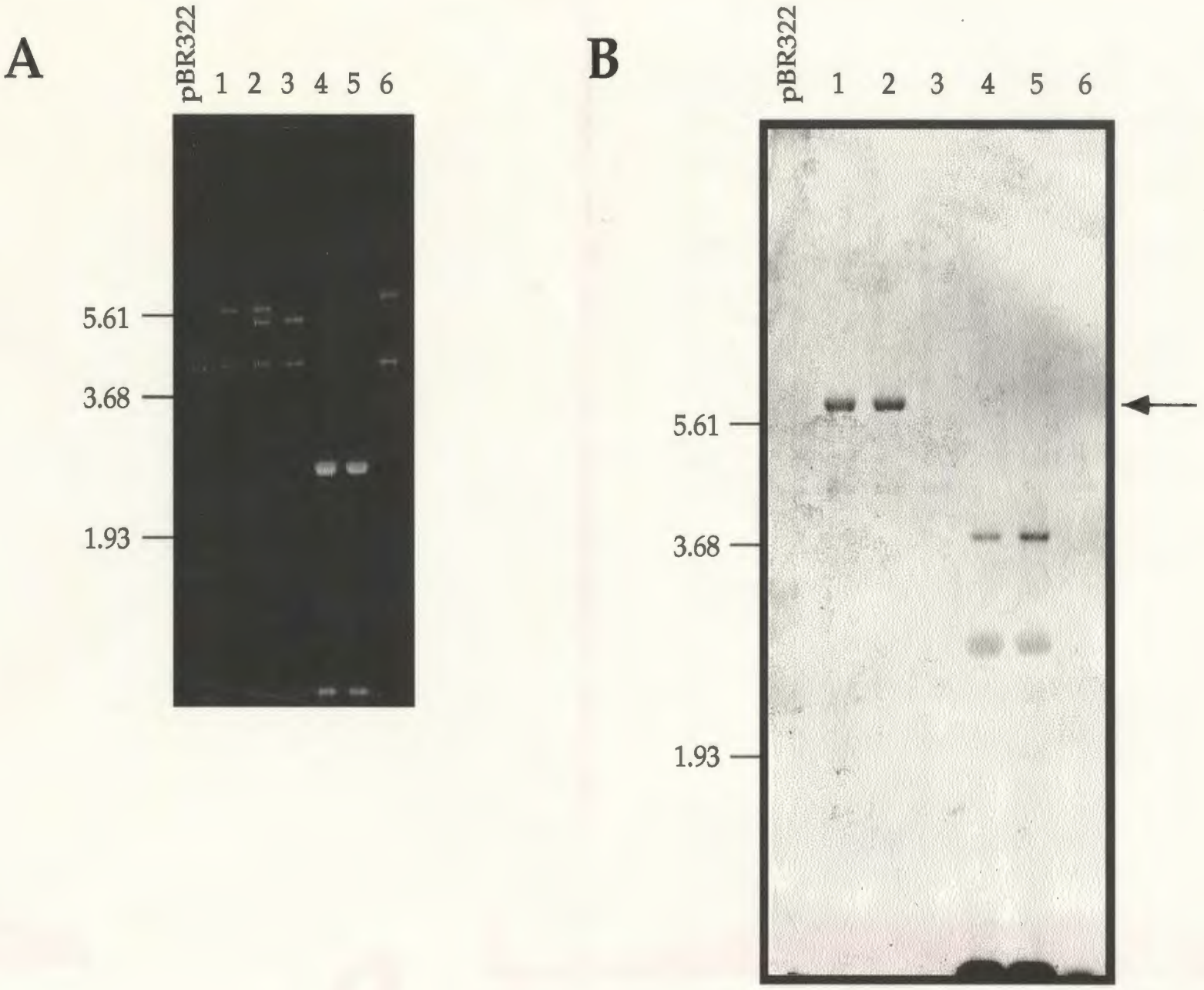
2.3. RESULTS AND DISCUSSION

2.3.1. Cloning and Restriction Mapping of SyR1 and SyR1.2

In order to isolate the 6.3 kb HindIII fragment containing the *rbp* gene, a size-biased (5-10 kb) pBR322 plasmid library was constructed using *Synechococcus* 7942 genomic DNA digested with HindIII. Approximately 800 clones were probed with a digoxygenin-labelled Cg9.2 RBP probe from *Chlorogloeopsis* 6912 (Mulligan *et al.*, 1994). Plasmid DNA from 6 potentially positive clones was isolated and digested with HindIII and analyzed on an agarose gel (**Figure 2.1A**). Four of these plasmids showed a 4.3 kb fragment characteristic of linearized pBR322, and all four contained at least one insert fragment. DNA was then transferred to a nylon membrane and probed with the Cg9.2 RBP probe (**Figure 2.1B**). Of the six clones originally isolated, only plasmid clones #1 and #2 showed a strongly hybridizing signal corresponding to a fragment of length 6.3 kb. Clone #1 was re-streaked to obtain a pure clone; the corresponding plasmid was named pSyR1.

Restriction endonuclease mapping of the 6.3 kb SyR1 fragment was carried out using a variety of enzymes. The fragment SyR1 was digested with combinations of selected restriction enzymes and the results visualized on an agarose gel (**Figure 2.2A**). The results of Southern blot hybridization analysis with a 250 bp MseI DNA fragment containing the *Chlorogloeopsis* 6912 *rbpA* gene (Mulligan *et al.*, 1994) localized the *rbpA* gene to a 1.45 kb NheI/HindIII DNA fragment, which was named SyR1.2 (**Figure 2.2B**). This was cloned into pUC19; the resulting plasmid was named pSyR1.2.

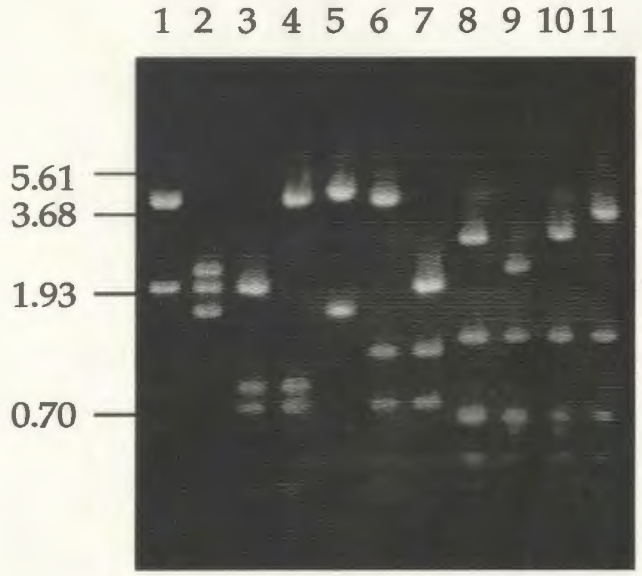
Figure 2.1. Screening of potentially positive clones for presence of the 6.3 kb SyR1 fragment. **(A)** Plasmid DNA from six positive clones (lanes 1-6), along with pBR322 DNA, was digested with HindIII and fractionated on a 0.7% agarose gel. **(B)** Following electrophoresis, DNA was transferred to a nylon membrane and probed with the digoxygenin-labelled Cg9.2 *rbp* fragment from *Chlorogloeopsis* 6912 (Mulligan *et al.*, 1994). The location of the 6.3 kb hybridizing fragment on the blot is indicated by an arrow. The restriction patterns for HindIII digests of clones #4 and #5 were unusual. Clones #4 and #5 contain a 900 bp DNA fragment which gave a strong hybridizing signal. This fragment was not characterized further. In both figures, the locations of bacteriophage λ BstEII fragments used as size markers are indicated.



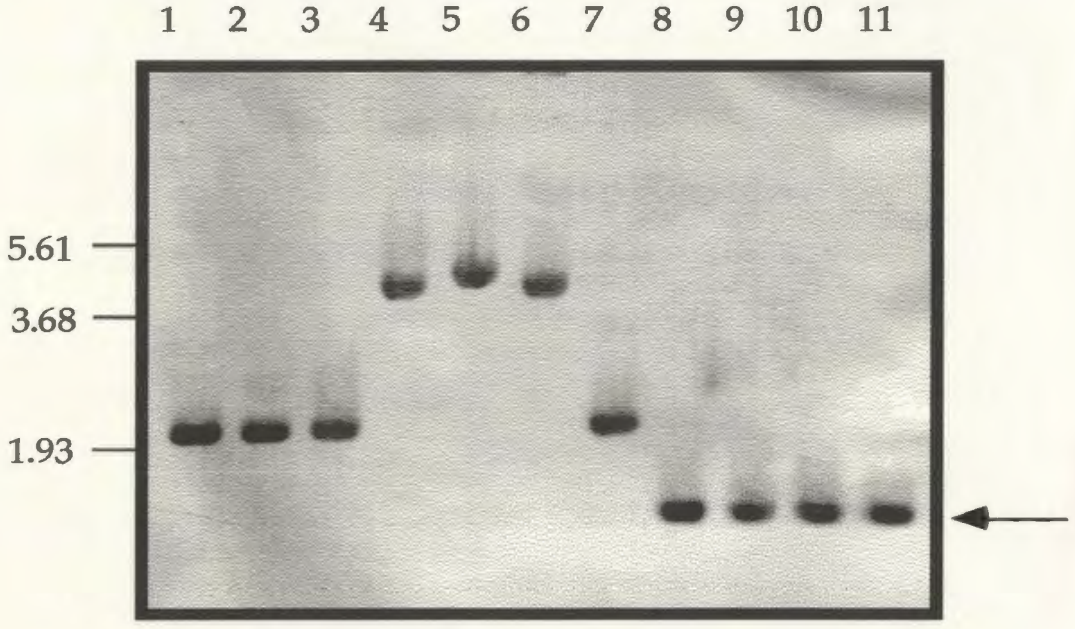
Cg9.2 RBP probe

Figure 2.2. Restriction endonuclease mapping of the SyR1 fragment from *Synechococcus* 7942. **(A)** DNA fragment SyR1 was digested with the indicated enzyme(s) and the products separated on a 0.7% agarose gel. Numbers at the top of each lane correspond to the combinations of enzymes used as follows: (1) EcoRV; (2) EcoRV+ScaI; (3) EcoRV+ScaI+EcoRI; (4) ScaI+EcoRI; (5) ScaI; (6) EcoRI; (7) EcoRV+EcoRI; (8) EcoRV+NheI; (9) EcoRV+ScaI+NheI; (10) ScaI+NheI; (11) NheI. **(B)** Following electrophoresis, DNA was transferred to a nylon membrane and probed with the 250 bp MseI DNA fragment containing the *Chlorogloeopsis* 6912 *rbpA* gene. The 1.45 kb NheI/HindIII hybridizing fragment containing the *Synechococcus* 7942 *rbpA* gene is indicated by the arrow. In both figures, the locations of bacteriophage λ BstEII fragments used as size markers are indicated.

A



B



2.3.2. Cloning and Restriction Mapping of SyR1.3 and SyR1.4

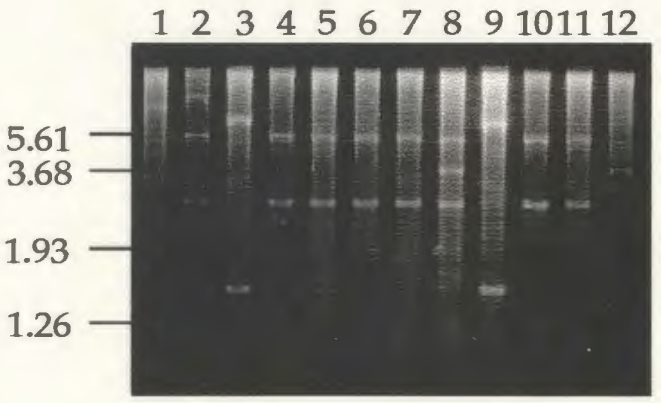
In order to identify and characterize sequence downstream of the *rbpA* gene, it was necessary to clone an overlapping DNA fragment. A Southern blot of NheI-digested *Synechococcus* genomic DNA revealed that the *rbpA* gene was located on a 4.1 kb NheI fragment (data not shown). A size-biased library (3-5.5 kb) of NheI-digested *Synechococcus* 7942 genomic DNA was constructed using the plasmid pBGS18 (Spratt *et al.*, 1986). From the approximately 800 clones probed with the 250 bp MseI *rbp* DNA probe (Mulligan *et al.*, 1994), twelve potential positives were selected. Agarose gel electrophoresis and Southern blot hybridization analysis revealed that ten of these clones contained the 4.1 kb insert and that both orientations were represented (**Figure 2.3**). Clone #3 was designated as clone pSyR1.3; the corresponding 4.1 kb DNA fragment was named SyR1.3. Clone #4 contained the same insert in the reverse orientation and was named pSyR1.3R. The 2.65 kb sequence downstream of *rbpA* was isolated by HindIII digestion of pSyR1.3 and subsequent re-circularization of the plasmid. This new plasmid was named pSyR1.4.

2.3.3. Sequence Analysis of the *rbpA* gene and surrounding sequence

The sequencing strategy for DNA fragments originating from SyR1.2 and SyR1.4 is shown in **Figure 2.4.** More than half of the SyR1.2 DNA fragment has now been sequenced. The sequence of a 707 bp region which included the *rbpA* gene was deposited in the Genbank nucleotide database (Accession No. L48458).

Figure 2.3. Screening of positive clones for presence of the 4.1 kb NheI SyR1.3 fragment. **(A)** Plasmid DNA from twelve positive clones was digested with HindIII and fractionated on a 0.7% agarose gel. **(B)** Following electrophoresis, all DNA was transferred to a nylon membrane and probed with the digoxygenin-labelled 250 bp MseI DNA fragment containing the *rbpA* gene from *Chlorogloeopsis* 6912 (Mulligan *et al.*, 1994). The location of the 1.45 kb SyR1.2 hybridizing fragment on the blot is indicated by an arrow. Clone #3 was designated as pSyR1.3; clone #4 was designated as pSyR1.3R. In both figures, the locations of bacteriophage λ BstEII fragments used as size markers are indicated.

A


1 2 3 4 5 6 7 8 9 10 11 12
5.61
3.68
1.93
1.26

B

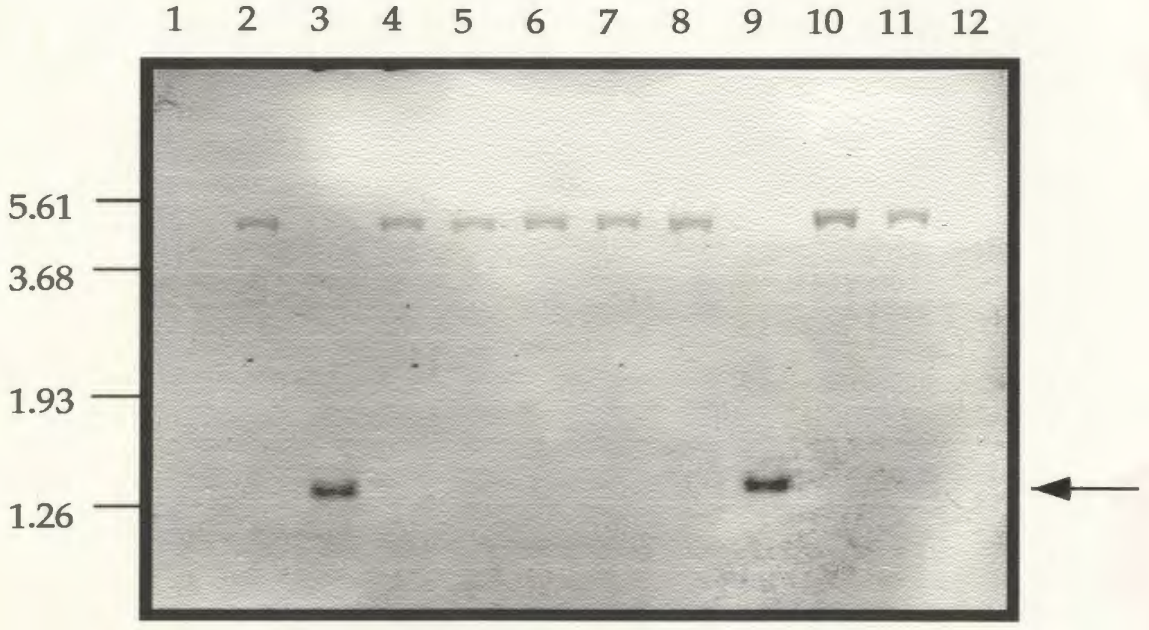
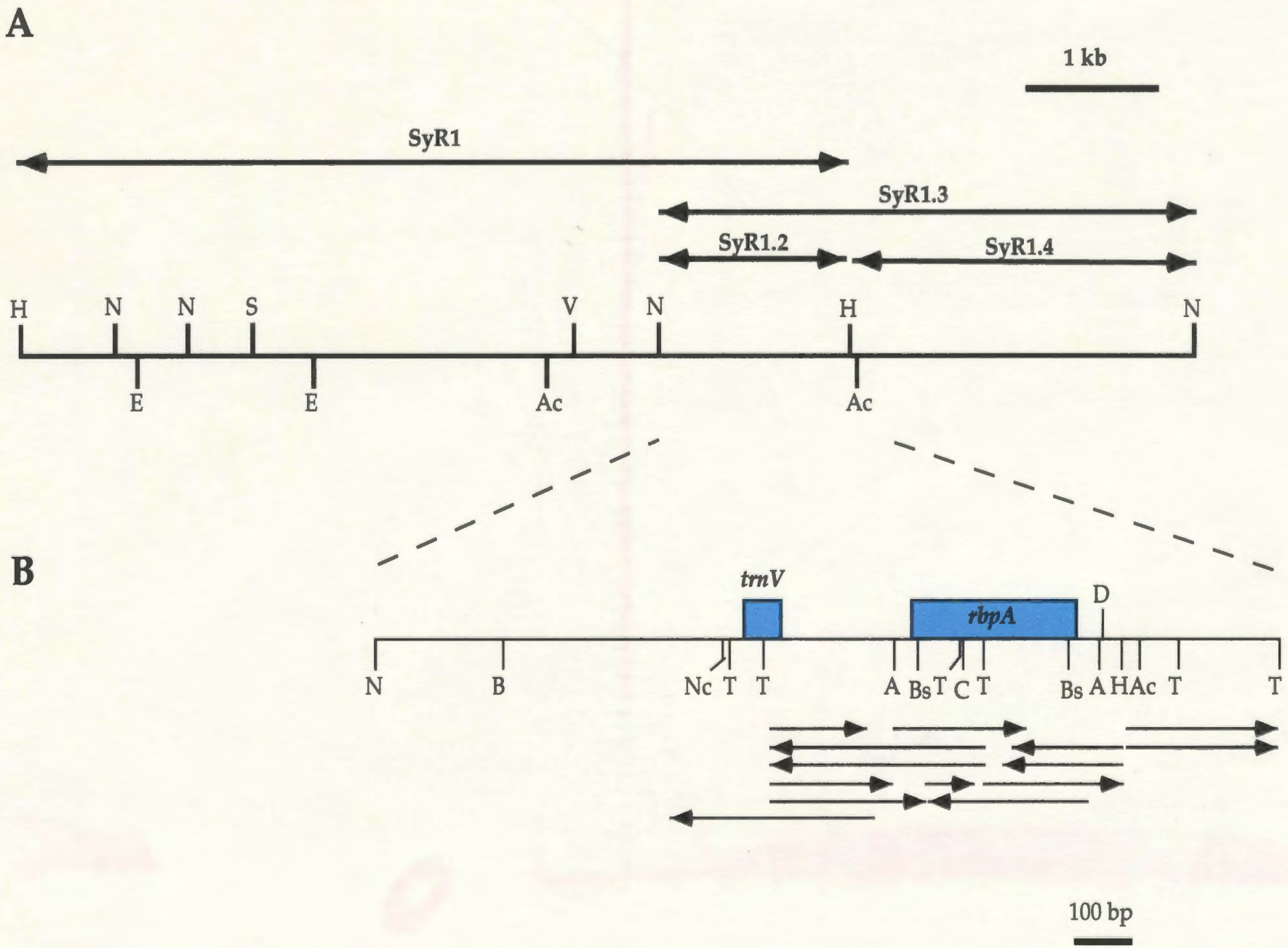

1 2 3 4 5 6 7 8 9 10 11 12
5.61
3.68
1.93
1.26

Figure 2.4. Restriction Map of the *rbpA* region in *Synechococcus* 7942. **(A)** A partial restriction map of the 6.3 kb SyR1 fragment and 4.1 kb SyR1.3 fragment isolated from *Synechococcus* 7942 is shown. The major restriction sites that were used in cloning and sequencing are indicated: Ac, AccI; E, EcoRI; H, HindIII; N, NheI; S, ScaI; V, EcoRV. **(B)** A partial restriction map of the 1.45 kb NheI/HindIII fragment named SyR1.2 is shown in more detail, along with approximately 300 bp of downstream sequence obtained from the 2.65 kb HindIII/NheI fragment named SyR1.4. Restriction sites used for cloning and sequencing are indicated as follows: A, ApoI; Ac, AccI; B, BamHI; Bs, BstEII; C, ClaI; D, DraI; H, HindIII; N, NheI; Nc, NcoI; T, TaqI. The locations of the *rbpA* and $tRNA^{Val}$ (*trnV*) genes are shown as boxes (blue). Arrows indicate the length and direction of various sequencing reactions. Not all TaqI restriction sites are shown in this figure.


A
1 kb
SyR1
SyR1.3
SyR1.2
SyR1.4
H
N
N
S
V
N
H
N
E
E
Ac
Ac
B
trnV
rbpA
D
N
B
Nc
T
T
A
Bs
T
C
T
Bs
A
H
Ac
T
T
100 bp

The nucleotide sequence of a 1207 bp region which included part of the SyR1.4 DNA fragment is shown in **Figure 2.5**. The *rbpA* gene is 324 bp in length. It codes for a protein of 107 amino acids, which contains a single RNA recognition motif (RRM) with both of the highly conserved RNP1 and RNP2 sequences. The RbpA protein also contains an auxiliary glycine-rich domain. This motif is 24 amino acids in length and contains 15 glycine residues (62.5% of the total). This region also contains 3 arginine residues so that RbpA may belong to the RGG box family of RNA-binding proteins as well (Burd and Dreyfuss, 1994). The final tyrosine classifies this gene product with the major subfamily of RNA-binding proteins and distinct from Rbp1 of *Synechococcus* 7942, 12RNP1 of *Synechococcus* 6301, and sequence ssr1480 from *Synechocystis* 6803, the only members of the other subfamilies (Mulligan and Belbin, 1995) (see Section 2.3.4).

This work also confirmed the location of a tRNA-valine (*trnV*) gene upstream of *rbpA* in *Synechococcus* 7942. The sequence of this gene was 100% identical with the corresponding gene identified previously in the closely-related *Synechococcus* 6301 (Sugita and Sugiura, 1994).

2.3.4. Comparison with Other Cyanobacterial RNA-Binding Proteins

A list of inferred amino acid sequences of the known cyanobacterial RNA-binding proteins is shown in **Figure 2.6**. They vary in size from 83 amino acids to 110 amino acids in length. All consist of an 83 amino acid RRM and most also contain an auxiliary domain which is variable in length and composition. The RRM contains both of the RNP1 and RNP2 conserved

Figure 2.5. The complete nucleotide sequence of a 1207 bp region in *Synechococcus* 7942, including the *rbpA* gene and flanking sequence. The sequence of a 707 bp region from positions 194 to 901 has been submitted to Genbank (Accession No. L48548). Lower case letters are used to indicate DNA sequence which has only been sequenced in one direction. Restriction sites used in cloning and sequencing are indicated above the nucleotide sequence. The ClaI and HindIII sites were used for mutagenesis (Sections 3.2.1 and 3.2.2). The two BstEII sites indicate the boundaries of a 290 bp restriction fragment used as a probe in Southern blots. The location of a probable transcription start site at position 374 is indicated (•) (Sugita and Sugiura, 1994). Double underlined sequence indicates a sequence which is potentially a ribosome binding site. Palindromic sequences capable of forming a stem-loop structure in RNA are shown as arrows. The inferred amino acid sequence for the *rbpA* gene is given below the nucleotide sequence. The RNP1 (amino acids 42-49) and RNP2 (amino acids 3-8) sequences are light shaded. The glycine residues of the auxiliary domain are shown in bold type. The tRNA-Val (*trnV*) gene located upstream of *rbpA* is indicated by boxed sequence.

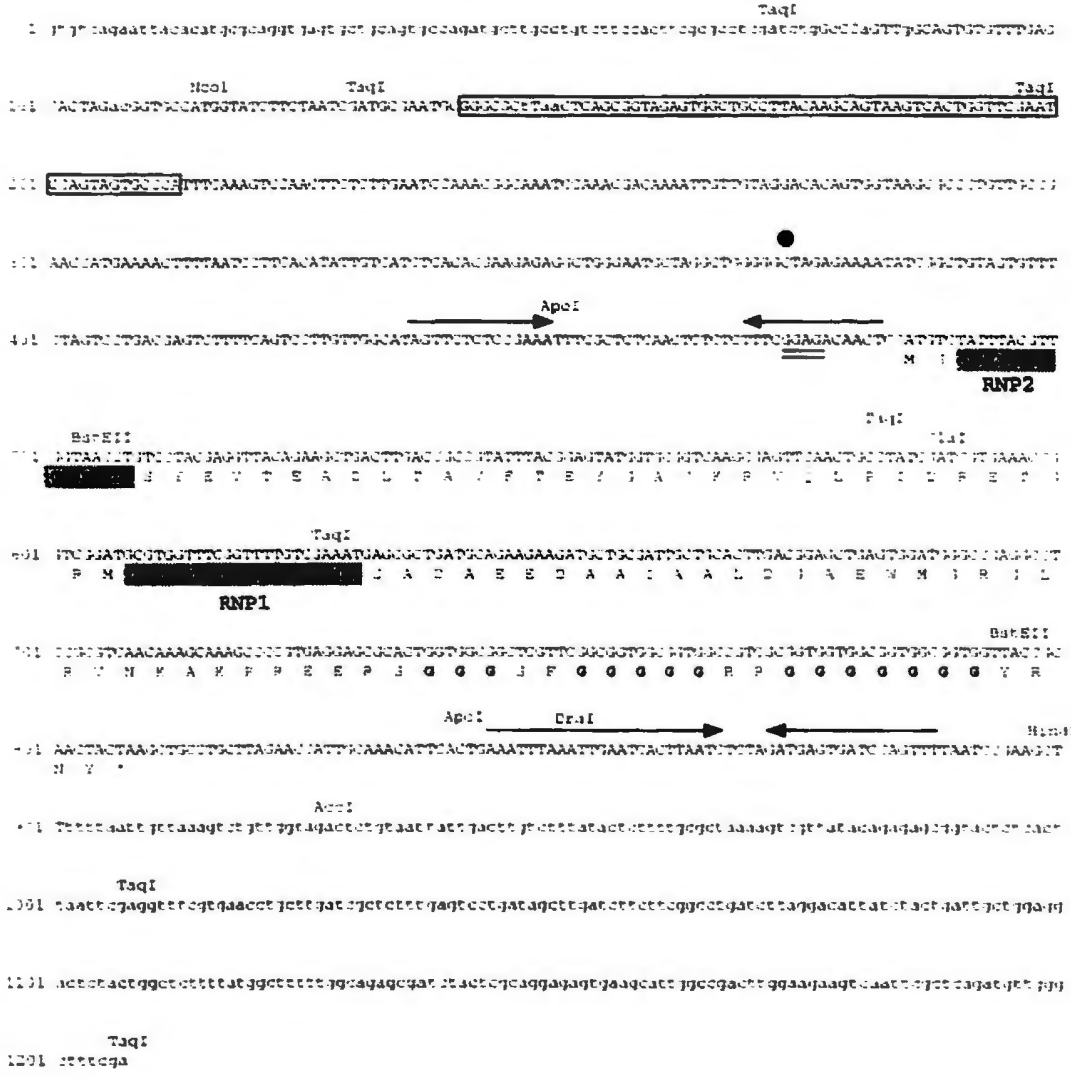
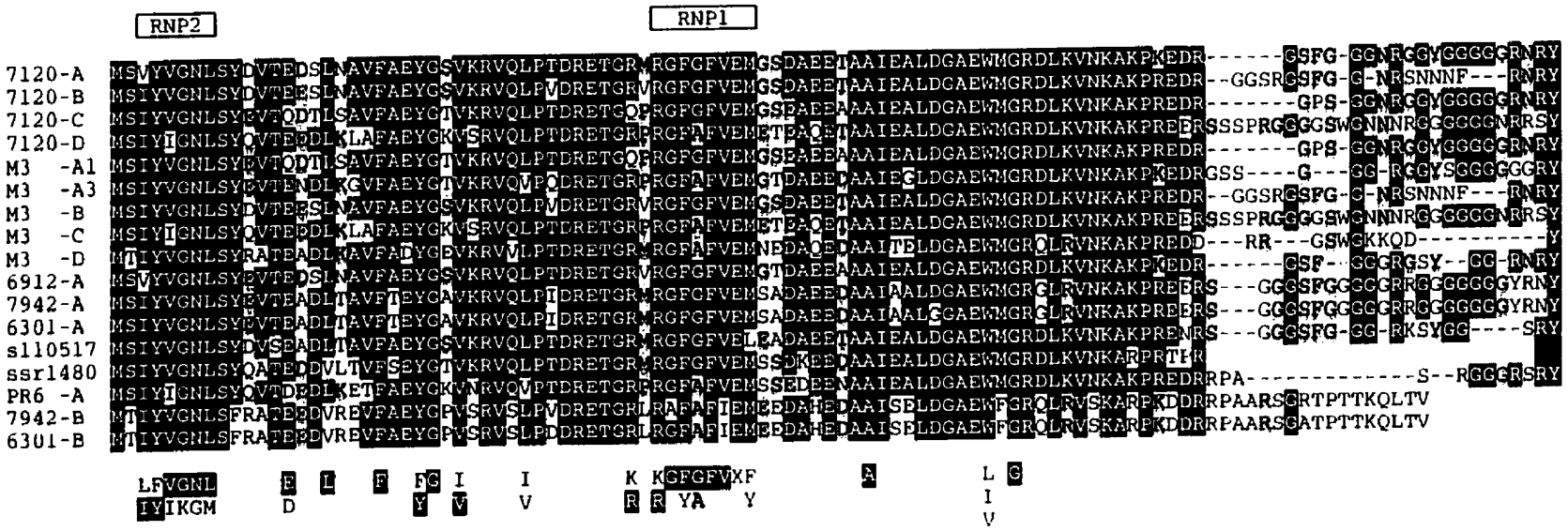

TaqI
NcoI
TaqI
TaqI
ApoI
RNP2
BstEII
TaqI
ClaI
TaqI
RNP1
BstEII
ApoI
DraI
HindIII
AccI
TaqI
TaqI

Figure 2.6. The inferred amino acid sequence of 17 cyanobacterial RNA-binding proteins. Sequences have been aligned manually; residues that occur in at least 13 sequences are shown in reverse type, residues that occur in at least 9 sequences are shaded. The sequences are as follows: 7120-A, 7120-B, 7120-C, and 7120-D are the RbpA, RbpB, RbpC and RbpD proteins, respectively, from *Anabaena* 7120 (Mulligan *et al.*, 1994; Mulligan and Belbin, 1995); M3-A1, M3-A3, M3-B, M3-C and M3-D are the RbpA1, RbpA3, RbpB, RbpC and RbpD proteins, respectively, from *Anabaena variabilis* M3 (Sato, 1995); 6912-A is the RbpA protein from *Chlorogloeopsis* 6912 (Mulligan *et al.*, 1994); 7942-A and 7942-B are the RbpA and Rbp1 proteins, respectively, from *Synechococcus* 7942 (Mulligan and Belbin, 1995; Dolganov and Grossman, 1993); 6301-A and 6301-B are the 12RNP1 and 12RNP2 proteins, respectively, from *Synechococcus* 6301 (Sugita and Sugiura, 1994); PR6-A is the RbpA protein from *Agmenellum quadruplicatum* PR-6 (Wagner *et al.*, 1993); sll0517 and ssr1480 are proteins from *Synechocystis* 6803 (Nakamura *et al.*, 1998). Not shown is the sequence of clone slr0193 from *Synechocystis* 6803 (Nakamura *et al.*, 1998). Conserved residues in eukaryotic RBPs are shown below the sequence alignment (Nagai and Mattaj, 1994). The RNP1 and RNP2 sequences are indicated above the sequence alignment.


RNP2
RNP1
7120-A
7120-B
7120-C
7120-D
M3 -A1
M3 -A3
M3 -B
M3 -C
M3 -D
6912-A
7942-A
6301-A
s110517
ssr1480
PR6 -A
7942-B
6301-B

sequences that are characteristic of the RNP family. A feature common to almost all other cyanobacterial RNA-binding proteins characterized to date has been the presence of a methionine residue at position 8 of the RNP1 motif. In the sll0517 gene product, the methionine is replaced with a leucine residue.

The auxiliary domains are highly variable and have provided the basis for dividing the cyanobacterial Rbps into several distinct classes (Mulligan and Belbin, 1995). In one class are those which contain glycine-rich auxiliary motifs and have tyrosine as the C-terminal amino acid. All of the proteins from the filamentous cyanobacteria *Anabaena* and *Chlorogloeopsis* fall into this class, as do the RbpA proteins from the *Synechococcus* and *Agmenellum* and the sll0517 gene product from *Synechocystis*. In this class, the glycine content of the auxiliary domain ranges from 18% (*Anabaena variabilis* M3 RbpD) to over 60% (*Synechococcus* 7942 RbpA). The second class of cyanobacterial RBPs comprise the RbpB proteins from the *Synechococcus* 7942 and *Synechococcus* 6301. The auxiliary domains of these two proteins do not contain a preponderance of any given amino acid and they contain a single glycine residue. A third class would consist of the ssr1480 gene product from *Synechocystis* 6803, which does not appear to have any form of auxiliary domain. Finally, the slr0193 gene product from *Synechocystis* 6803 contains an auxiliary domain which is significantly longer than any of the cyanobacterial Rbp sequences previously characterized. It contains only 3 glycine residues. Due to its divergent sequence, it is not included in the alignment.

The percentage sequence identities between the RRMs of the known cyanobacterial Rbps are listed in **Tables 2.2, 2.3 and 2.4.** An alignment of all cyanobacterial Rbps revealed that the first 83 amino acids representing the RRM motif were the most highly conserved. The length and sequence variability of the auxiliary domains made sequence identity calculations for this region difficult. Therefore, sequence identity calculations were calculated only for the 83 amino acid RRM segments. Both *Synechococcus* 7942 and *Synechococcus* 6301 are almost identical strains of cyanobacteria (Golden *et al.*, 1989), so the high degree of similarity evident in the sequences of *rbpA* and RbpA is not suprising (see Table 2.2). RbpA from *Synechococcus* 7942 showed 98.8% identity with the 12RNP1 protein from *Synechococcus* 6301, but only 65.1% and 66.3% identity with the second Rbp from the two strains, *Synechococcus* 7942 (Rbp1) and *Synechococcus* 6301 (12RNP2), respectively. RbpA was 73.5% identical to another unicellular cyanobacterium *Agmenellum quadruplicatum* PR-6 and was 79.5% and 85.5% identical to the ssr1480 and sll0517 gene products, respectively, from the unicellular strain *Synechocystis* 6803. RbpA was only 43.4% identical to the slr0193 gene product.

Table 2.3 lists the percentage identities between RRM segments of the filamentous cyanobacterial strains. These genes showed a high degree of conservation; most were greater than 80% identical at the nucleotide level. It should be noted that three sets of gene products were 100% identical to one another at both the nucleotide and amino acid level. Therefore, these represent corresponding genes in the two closely related *Anabaena* strains. The RbpB, RbpC and RbpD gene products from *Anabaena* 7120 (**7120-B, 7120-C**

Table 2.2. Identities of RRM segments[a] of Known Unicellular Cyanobacterial RNA-Binding Proteins

			1.	2.	3.	4.	5.	6.	7.	8.
1.	*Synechococcus* 7942	A[b]	-	65.9	99.6	65.9	67.9	75.1	69.1	46.9
2.		B[c]	65.1	-	65.5	98.8	59.0	59.8	62.7	46.5
3.	*Synechococcus* 6301	12RNP1[d]	98.8	63.9	-	65.5	67.5	74.7	68.7	46.5
4.		12RNP2[d]	66.3	97.6	65.1	-	59.4	60.6	63.1	47.3
5.	*A. quadruplicatum* PR-6	A[e]	73.5	61.4	72.3	62.7	-	71.1	69.1	49.4
6.	*Synechocystis* 6803	sll0517[f]	85.5	62.7	84.3	63.9	75.9	-	67.5	48.1
7.		ssr1480[f]	79.5	66.3	78.3	67.5	75.9	79.5	-	49.2
8.		slr0193[f]	43.4	37.2	43.4	38.4	43.4	41.0	43.4	-

[a] Numbers indicate % identities between 83 amino-acid segments (below the diagonal) or corresponding 249 nucleotide segments (above the diagonal).

[b] This work.

[c] Dolganov and Grossman, 1993.

[d] Sugita and Sugiura, 1994.

[e] Wagner *et al.*, 1993.

[f] Nakamura *et al.*, 1996

Table 2.3. Identities of RRM segments[a] of Known Filamentous Cyanobacterial RNA-Binding Proteins

			1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1.	*Anabaena* 7120	A[b]	-	85.9	80.3	73.1	80.3	79.1	85.9	73.1	67.9	90.0
2.		B[b]	94.0	-	79.5	72.3	79.5	77.1	100	72.3	69.1	87.1
3.		C[c]	86.7	86.7	-	74.3	100	79.1	79.5	74.3	66.3	83.1
4.		D[d]	78.3	80.7	79.5	-	74.3	73.5	72.3	100	68.7	73.5
5.	*Anabaena variabilis* M3	A1[e]	86.7	86.7	100	79.5	-	79.1	79.5	74.3	66.3	83.1
6.		A3[f]	83.1	83.1	83.1	80.7	83.1	-	77.1	73.5	65.5	82.7
7.		B[g]	94.0	100	86.7	80.7	86.7	83.1	-	72.3	69.1	87.1
8.		C[g]	78.3	80.7	79.5	100	79.5	80.7	80.7	-	68.7	73.5
9.		D[g]	74.7	74.7	72.3	73.5	72.3	75.9	74.7	73.5	-	66.3
10.	*Chlorogloeopsis* 6912	A[b]	95.2	91.6	88.0	78.3	88.0	85.5	91.6	78.3	73.5	-

[a] Numbers indicate % identities between 83 amino-acid segments (below the diagonal) or the corresponding 249 nucleotide segments (above the diagonal).
[b] Mulligan and Jackman, 1994
[c] Wu and Mulligan, unpublished
[d] Holden, 1995
[e] Sato, 1994
[f] Sato and Maruyama, 1997
[g] Sato, 1995

Table 2.4. Identities of (A) 249 nucleotide and (B) 83 amino acid RRM segments between Unicellular and Filamentous Cyanobacterial RNA-Binding Proteins

	Filamentous Strains		Unicellular Strains							
			7942-A	7942-B	6301-A	6301-B	PR6-A	6803-sll0517	6803-ssr1480	6803-slr0193
A.										
1.	***Anabaena* 7120**	**A**	73.1	63.1	72.7	63.9	69.1	75.9	70.7	48.3
2.		**B**	75.1	62.2	74.7	62.7	67.9	73.5	67.9	51.0
3.		**C**	71.9	60.2	71.5	61.0	71.9	76.3	68.7	50.0
4.		**D**	67.9	63.5	67.5	63.5	75.9	70.7	67.1	52.2
5.	***Anabaena variabilis* M3**	**A1**	71.9	60.2	71.5	61.0	71.9	76.3	68.7	50.0
6.		**A3**	69.9	65.1	69.5	66.3	71.5	76.7	69.1	48.4
7.		**B**	75.1	62.2	74.7	62.7	67.9	73.5	67.9	51.0
8.		**C**	67.9	63.5	67.5	63.5	75.9	70.7	67.1	52.2
9.		**D**	72.3	67.1	71.9	67.5	67.1	67.1	65.1	48.8
10.	***Chlorogloeopsis* 6912**	**A**	72.3	61.8	71.9	61.8	67.5	77.1	71.5	48.7
B.										
1.	***Anabaena* 7120**	**A**	80.7	63.9	79.5	65.1	77.1	88.0	80.7	39.5
2.		**B**	81.9	66.3	80.7	66.3	79.5	88.0	79.5	43.4
3.		**C**	79.5	60.2	78.3	61.4	80.7	81.9	79.5	41.0
4.		**D**	75.9	65.1	74.7	66.3	86.7	80.7	74.7	42.2
5.	***Anabaena variabilis* M3**	**A1**	79.5	60.2	78.3	61.4	80.7	81.9	79.5	41.0
6.		**A3**	79.5	67.5	78.3	68.7	80.7	79.5	77.1	46.5
7.		**B**	81.9	66.3	80.7	66.3	79.5	88.0	79.5	43.4
8.		**C**	75.9	65.1	74.7	66.3	86.7	80.7	74.7	42.2
9.		**D**	78.3	73.5	77.1	74.7	69.9	75.9	72.3	39.8
10.	***Chlorogloeopsis* 6912**	**A**	80.7	63.9	79.5	65.1	75.9	84.3	78.3	41.9

and 7120-D) correspond to the RbpB, RbpA1 and RbpC (**M3-B, M3-A1 and M3-C**) gene products, respectively, from *Anabaena variabilis* M3. The lowest identities among the filamentous strains were observed for the RbpD protein from *Anabaena variabilis* M3 (Sato, 1995). Its RRM segment showed only 70-75% identity with the other RRMs of the filamentous strains.

Table 2.4 compares RRM segments between the unicellular and filamentous cyanobacterial strains. RbpA from *Synechococcus* 7942 was approximately 80% identical with the inferred amino acid sequences of RBPs from filamentous cyanobacterial strains. This was only marginally higher than average as most of the unicellular RRMs were 60-80% identical to their filamentous counterparts. However, several exceptions should be noted. At the amino acid level, the RbpA sequence from *Agmenellum* was 86.7% identical with RbpD from *Anabaena* 7120 (i.e. RbpC from *Anabaena variabilis* M3). Also, the sll0517 gene product from *Synechocystis* 6803 was 88.0% identical to both RbpA and RbpB from *Anabaena* 7120 (also RbpB from *Anabaena variabilis* M3).

Overall, the comparison of nucleotide and amino acid sequence identities supported the classification for cyanobacterial Rbps. The Rbp1 and 12RNP2 gene products from *Synechococcus* 7942 and *Synechococcus* 6301, respectively, showed a decreased sequence identity when compared to all other known cyanobacterial Rbps but not when compared to each other. Also, the slr0193 gene product from *Synechocystis* 6803 was only 40-50% identical to any of the known cyanobacterial Rbps. The one exception was the ssr1480 gene product from *Synechocystis* 6803. Although it was classified separately based on the fact that it did not contain an auxiliary domain, the

RRM of this protein was greater than 75% identical to the other unicellular Rbps (excluding Rbp1 and 12RNP2, above) and 70-80% identical to all of the Rbps from the filamentous strains at the amino acid sequence level.

2.3.5. Comparison with Eukaryotic RNA-Binding Proteins

The cyanobacterial RRM sequences show a high degree of similarity to the RRM sequences of eukaryotic RNA-binding proteins of the RNP family. Conserved residues observed in eukaryotic RRM sequences are shown in **Figure 2.6** (Nagai and Mattaj, 1994). The cyanobacterial RNP2 sequences are essentially identical to those observed in eukaryotic RNA-binding proteins with only conservative substitutions observed. The RNP1 sequences are also highly similar to those found in eukaryotes, but with several notable exceptions. The RNP1 sequences of almost all cyanobacterial Rbps contain methionine at position 8 in place of an aromatic residue seen in eukaryotic Rbps (Mulligan and Belbin, 1995). This substitution has also been found in the first RRM motif of many plant chloroplast proteins (Breiteneder *et al.*, 1994; Ye *et al.*, 1991). As well, both glycine residues of RNP1 have been replaced by alanine in at least one cyanobacterial sequence. The cyanobacterial sequences also contain leucine at position 32, a conservative substitution from the eukaryotic consensus. All of the cyanobacterial Rbps contain a tryptophan residue at position 67 in place of the aliphatic residue seen in the eukaryotic sequences.

2.3.6. Sequence Analysis of Regions Upstream of Cyanobacterial *rbp* genes

Primer extension experiments have located a putative transcription start site for the 12RNP1 gene of *Synechococcus* 6301 which lies 112 nt upstream of the start codon (Sugita and Sugiura, 1994). An analysis of the corresponding region upstream of *rbpA* in *Synechococcus* 7942 reveals an interesting feature. When transcribed into RNA, this sequence contains complementary sequences capable of folding into three stem-loop structures (**Figure 2.7**). I have designated these structures as stem-loops I, II and III based on their proximity to the start codon for the gene. Stem-loop I was particularly interesting; it contained the complementary consensus sequences UCUCCGAAA and UUUUGGAGA which made up the stem. It was located immediately upstream of the *rbpA* gene in *Synechococcus* 7942 and included the region previously identified as the potential ribosome binding site (see Figure 2.5).

A comparison of the known upstream DNA sequences for the other cyanobacterial *rbp* genes revealed that this stem-loop structure is well conserved among the various strains of cyanobacteria (**Figure 2.8**). All of these are located less than 14 nt upstream of the start codon for each gene with the exception of the *rbpA* gene from *Anabaena* 7120, where the stem-loop is located 56 nt upstream of the start codon for the gene. While there appears to be a high degree of similarity among the stem sequences, the loop region shows a much higher degree of variability. The loops range from 19-24 nt in length. The only region of similarity appears to be a moderately conserved hexameric sequence whose consensus sequence in RNA is UCUCUA.

Figure 2.7. Potential secondary structures formed upstream of the *Synechococcus* 7942 *rbpA* gene. The nucleotide sequence of a 112 nt RNA segment, extending from the reported start site of transcription in *Synechococcus* 6301 (Sugita and Sugiura, 1994) to the start codon for the gene, was folded by using the program Mulfold. The potential stem-loops are numbered (I-III) according to their proximity to the *rbpA* gene. Sequence numbering starts using the first nucleotide preceding the start codon and proceeds in the reverse direction towards the potential transcription start site. Bases in the stem similar to those observed in other sequences upstream of cyanobacterial *rbp* genes are shaded; conserved sequences in the loop are shown in outline type. Bases previously identified as a potential ribosome binding site (see Figure 2.5) are double-underlined.

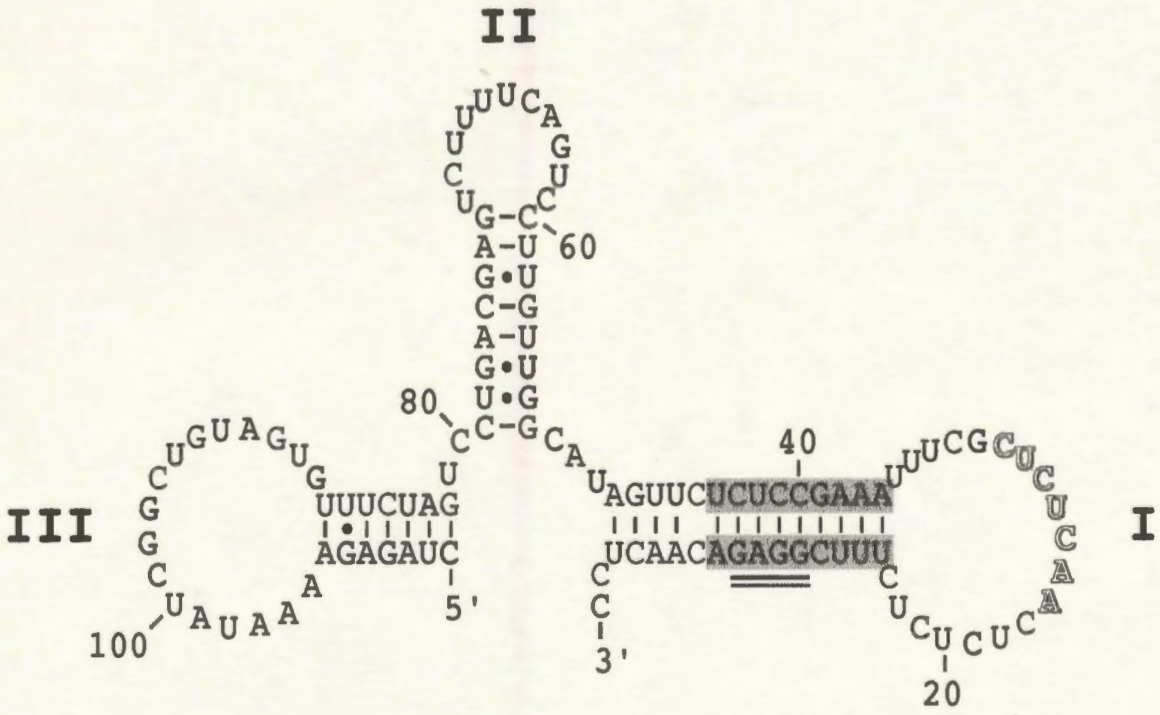
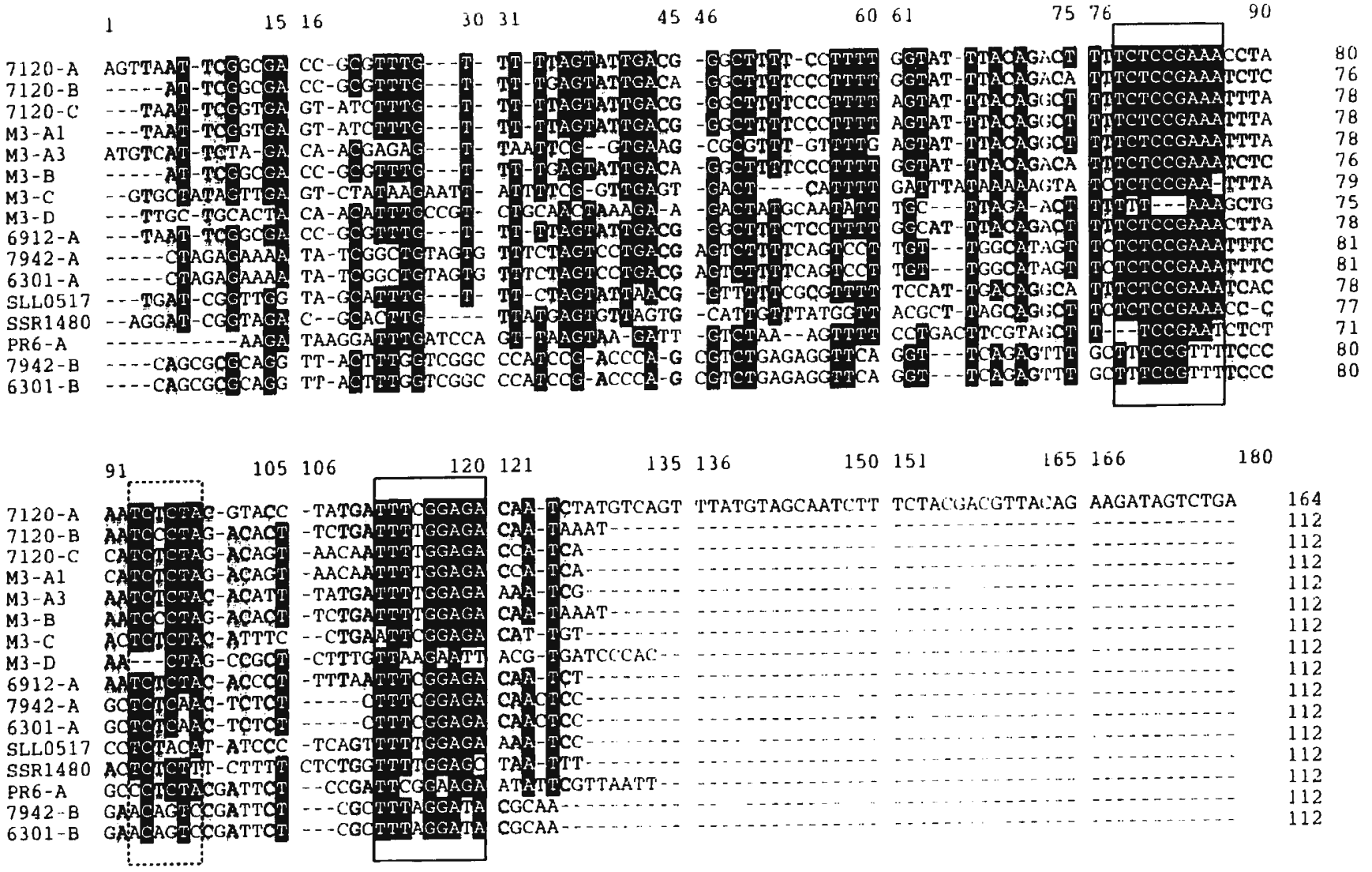

II
III
I
5'
3'
20
40
60
80
100

Figure 2.8. Alignment of the nucleotide sequences upstream of 16 cyanobacterial *rbp* genes. In each case, the sequence of a DNA segment immediately preceding the start codon for the *rbp* gene is shown. Sequence alignments were carried out using the ClustalW Multiple Sequence Alignment Program and the results visualized using DNADraw. Nucleotides that occur in at least 12 sequences (75%) are shown in reverse type while nucleotides that occur in at least 8 sequences (50%) are shaded. Strain and gene abbreviations (see Figure 2.6) indicating the source of each sequence are shown at the left. Nucleotides representing stem-loop I sequences are outlined as follows: highly conserved regions in the stem portion of the stem-loop I are outlined with a solid line, conserved nucleotides included in the loop portion are outlined with a broken line.


1 15 16 30 31 45 46 60 61 75 76 90
7120-A AGTTAAT-TCGGCGA CC-GCGTTTG---T- TT-TTAGTATTGACG -GGCTTTT-CCTTTT GGTAT-TTACAGACT TTTCTCCGAAACCTA 80
7120-B -----AT-TCGGCGA CC-GCGTTTG---T- TT-TGAGTATTGACA -GGCTTTTCCCTTTT GGTAT-TTACAGACA TTTCTCCGAAATCTC 76
7120-C ---TAAT-TCGGTGA GT-ATCTTTG---T- TT-TTAGTATTGACG -GGCTTTTCCCTTTT AGTAT-TTACAGGCT TTTCTCCGAAATTTA 78
M3-A1 ---TAAT-TCGGTGA GT-ATCTTTG---T- TT-TTAGTATTGACG -GGCTTTTCCCTTTT AGTAT-TTACAGGCT TTTCTCCGAAATTTA 78
M3-A3 ATGTCAT-TCTA-GA CA-ACGAGAG---T- TAATTCG--GTGAAG -CGCGTTT-GTTTTG AGTAT-TTACAGGCT TTTCTCCGAAATTTA 78
M3-B -----AT-TCGGCGA CC-GCGTTTG---T- TT-TGAGTATTGACA -GGCTTTTCCCTTTT GGTAT-TTACAGACA TTTCTCCGAAATCTC 76
M3-C --GTGCTATAGTTGA GT-CTATAAGAATT- ATTTTCG-GTTGAGT -GACT----CATTTT GATTTATAAAAAGTA TCTCTCCGAA-TTTA 79
M3-D ---TTGC-TGCACTA CA-ACATTTGCCGT- CTGCAACTAAAGA-A -GACTATGCAATATT TGC---TTAGA-ACT TTTTT---AAAGCTG 75
6912-A ---TAAT-TCGGCGA CC-GCGTTTG---T- TT-TTAGTATTGACG -GGCTTTCTCCTTTT GGCAT-TTACAGACT TTTCTCCGAAACTTA 78
7942-A -----CTAGAGAAAA TA-TCGGCTGTAGTG TTTCTAGTCCTGACG AGTCTTTTCAGTCCT TGT---TGGCATAGT TCTCTCCGAAATTTC 81
6301-A -----CTAGAGAAAA TA-TCGGCTGTAGTG TTTCTAGTCCTGACG AGTCTTTTCAGTCCT TGT---TGGCATAGT TCTCTCCGAAATTTC 81
SLL0517 ---TGAT-CGGTTGG TA-GCATTTG---T- TT-CTAGTATTAACG -GTTTTTCGCGTTTT TCCAT-TGACAGGCA TTTCTCCGAAATCAC 78
SSR1480 --AGGAT-CGGTAGA C--GCACTTG----- TTATGAGTGTTAGTG -CATTGTTTATGGTT ACGCT-TAGCAGGCT TCTCTCCGAAACC-C 77
PR6-A ----------AAGA TAAGGATTTGATCCA GT-TAAGTAA-GATT -GTCTAA--AGTTTT CCTGACTTCGTAGCT T---TCCGAATCTCT 71
7942-B ----CAGCGCGCAGG TT-ACTTTGGTCGGC CCATCCG-ACCCA-G CGTCTGAGAGGTTCA GGT---TCAGAGTTT GCTTTCCGTTTTCCC 80
6301-B ----CAGCGCGCAGG TT-ACTTTGGTCGGC CCATCCG-ACCCA-G CGTCTGAGAGGTTCA GGT---TCAGAGTTT GCTTTCCGTTTTCCC 80
91 105 106 120 121 135 136 150 151 165 166 180
7120-A AATCTCTAC-GTACC -TATGATTTCGGAGA CAA-TCTATGTCAGT TTATGTAGCAATCTT TCTACGACGTTACAG AAGATAGTCTGA 164
7120-B AATCCCTAG-ACACT -TCTGATTTTGGAGA CAA-TAAAT------ --------------- --------------- ------------ 112
7120-C CATCTCTAG-ACAGT -AACAATTTTGGAGA CCA-TCA------- --------------- --------------- ------------ 112
M3-A1 CATCTCTAG-ACAGT -AACAATTTTGGAGA CCA-TCA------- --------------- --------------- ------------ 112
M3-A3 AATCTCTAC-ACATT -TATGATTTTGGAGA AAA-TCG------- --------------- --------------- ------------ 112
M3-B AATCCCTAG-ACACT -TCTGATTTTGGAGA CAA-TAAAT------ --------------- --------------- ------------ 112
M3-C ACTCTCTAC-ATTTC --CTGAATTCGGAGA CAT-TGT------- --------------- --------------- ------------ 112
M3-D AA---CTAG-CCGCT -CTTTGTTAAGAATT ACG-TGATCCCAC-- --------------- --------------- ------------ 112
6912-A AATCTCTAC-ACCCT -TTTAATTTCGGAGA CAA-TCT------- --------------- --------------- ------------ 112
7942-A GCTCTCAAC-TCTCT -----CTTTCGGAGA CAACTCC------- --------------- --------------- ------------ 112
6301-A GCTCTCAAC-TCTCT -----CTTTCGGAGA CAACTCC------- --------------- --------------- ------------ 112
SLL0517 CCTCTACAT-ATCCC -TCAGTTTTTGGAGA AAA-TCC------- --------------- --------------- ------------ 112
SSR1480 ACTCTCTTT-CTTTT CTCTGGTTTTGGAGC TAA-TTT------- --------------- --------------- ------------ 112
PR6-A GCCCTCTACGATTCT --CCGATTCGGAAGA ATATTCGTTAATT-- --------------- --------------- ------------ 112
7942-B GAACAGTCCGATTCT ---CGCTTTAGGATA CGCAA---------- --------------- --------------- ------------ 112
6301-B GAACAGTCCGATTCT ---CGCTTTAGGATA CGCAA---------- --------------- --------------- ------------ 112

The function of this stem-loop structure is unclear. However, its proximity to the start codon of the gene suggests that it may play a role in regulating expression of the gene at the translational level. One possibility is that by binding to nucleotides in the stem-loop, a cyanobacterial Rbp may inhibit translation of its own RNA transcript by interfering with binding by the ribosomal machinery. Autoregulation of expression by an RNP-type RNA-binding protein has been observed previously. The human U1A protein regulates its own expression by binding to a polyadenylation inhibition element (PIE) located in the 3'-untranslated region of U1A pre-mRNA (van Gelder *et al.*, 1993).

A second possibility is that the stem-loop may be the site of regulation of these genes by cold. This theory is supported by the fact that in *Anabaena variabilis* M3, interruption of the stem-loop structure by interposon mutagenesis disrupts regulation of the *rbpA1* gene by a decrease in temperature (Sato and Wada, 1996). This observation has been confirmed by deletion analysis of the 5' untranslated region of *rbpA1* (Sato and Nakamura, 1998).

CHAPTER 3. INACTIVATION OF *rbpA* by INTERPOSON MUTAGENESIS

3.1. INTRODUCTION

3.1.1. Techniques for Mutagenesis in Cyanobacteria

In prokaryotes, one of the most powerful approaches to studying the function of specific cellular systems is through the analysis of mutants which have been somehow altered with respect to that system. In cyanobacteria, the traditional approaches to generating mutants used chemicals and UV light (Golden, 1988). For example, auxotrophic mutants of *Anabaena variabilis* ATCC 29413 have been isolated using the potent mutagen N-methyl-N'-nitro-N-nitrosoguanine (MNNG or NTG) (Currier *et al.*, 1977). Both diethyl sulfate mutagenesis and UV mutagenesis have been used to obtain mutants of *Anabaena* 7120 which are unable to fix nitrogen (Buikema and Haselkorn, 1991b; Wolk *et al.*, 1988). The main problem with these approaches lies in the risk of inducing multiple mutations at closely linked sites.

The use of chemical and UV mutagenesis has gradually been replaced with other techniques made possible by the development of gene cloning and transfer methods. For example, techniques of transposon mutagenesis have been used to isolate mutants with a particular phenotype (Golden, 1988). This technique generally involved transforming a cyanobacterial host with a transposon-containing vector. The transposon is capable of inserting itself into the cyanobacterial genome, and it contains a gene which codes for resistance to a particular antibiotic. The mutated region, carrying the inserted transposon, is subsequently cloned into *E. coli*; the segment flanking the transposon is used to isolate the wild type gene sequence from a library of cyanobacterial genomic DNA. Mutagenesis using transposon *Tn*901 has been

used successfully in *Synechococcus* 7942 to isolate genes involved in methionine biosynthesis as well as genes involved in nitrate assimilation (Madueno *et al.*, 1988; Tandeau de Marsac *et al.*, 1982).

Transposons have also been used in filamentous strains of cyanobacteria. Transposon *Tn*5 has been used in the study of nitrogen fixation in *Anabaena* 7120 through the isolation of mutants unable to grow without a source of combined nitrogen (Borthakur and Haselkorn, 1989). Several transposons created for use in cyanobacteria contain the *luxAB* genes, encoding luciferase, as an *in situ* reporter of transcription (Wolk *et al.*, 1991). When transposons containing these genes are incorporated into a gene that is being expressed, it produces a luminescent cell. This system has been successfully used to study changes in patterns of gene expression in response to environmental changes and heterocyst differentiation (Elhai and Wolk, 1990; Wolk *et al.*, 1991). One gene in particular, expressed in response to a decrease in incubation temperature, was later identified as one of the RNA-binding protein genes (*rbpA*) in *Anabaena* 7120 (Panoff, 1993). Transposons have also been used in the mapping of photosynthesis and nitrogen fixation genes in *Anabaena* 7120 (Kuritz *et al.*, 1993).

The most popular technique for inactivation mutagenesis has been the interruption of specifically targeted genes by insertion of a foreign DNA fragment. Once the gene of interest is cloned into a suitable *E. coli* plasmid, a DNA fragment containing an antibiotic resistance gene is inserted within the coding region for the cloned gene. Upon transfer of this plasmid into a wild type cell, a homologous double recombination event replaces the wild type copy of the gene with the mutated copy. Mutants are then selected by their

antibiotic-resistance phenotype. This approach has been successfully used in *Synechococcus* 7942 to study each of three photosystem II-related (*psbA*) genes by selectively inactivating each gene and characterizing the mutant phenotype (Golden *et al.*, 1986). To date, it has also been used in the study of other genes in *Synechococcus* 7942, including those involved in nitrate transport (Omata *et al.*, 1993) and phycobiliprotein structure and function (Bhalerao *et al.*, 1993; Bhalerao *et al.*, 1994). It has also been used to study genes in other unicellular cyanobacteria, including *Synechocystis* 6803 (Smart *et al.*, 1993) and *Synechococcus* 7002 (Zhou *et al.*, 1992).

Interposon mutagenesis has also been successfully used in filamentous strains of cyanobacteria. In *Anabaena* 7120, interposon mutants have been used to study processes such as nitrogen fixation (Golden and Wiest, 1988), and heterocyst differentiation (Buikema and Haselkorn, 1991a). Examples of interposon mutagenesis in *Anabaena variabilis* 29413 have also been reported in the literature (Thiel, 1993). One of the problems encountered with interposon mutagenesis procedures for filamentous strains of cyanobacteria, including *Anabaena* 7120, is the high frequency of single recombination events due to conjugal transfer of the mutated DNA. This differs from unicellular strains such as *Synechococcus* 7942 where the procedure yields mainly double recombinants (Thiel, 1994). This problem has been overcome in *Anabaena* 7120 by the development of a system using the conditionally lethal *sacB* gene which selects for double recombinants by growth on solid media containing 5% sucrose (Cai and Wolk, 1990).

3.1.2. Mutagenesis of RNA-Binding Protein Genes in Cyanobacteria

Inactivation of cyanobacterial *rbp* genes has been used in other species of cyanobacteria to identify and characterize knock-out phenotypes. For example, in *Anabaena variabilis* M3, the *rbpA1* gene has been mutated by interposon mutagenesis (Sato and Wada, 1996). The resulting knock-out strain K1 grew well at 38°C, but growth at 22°C resulted in a derepression of events related to heterocyst differentiation. At both temperatures, the cells were greenish-yellow in appearance as compared to the blue-green colour of the wild type strain. This bleaching was shown to be due to the loss of the phycobiliprotein phycocyanin. Interposon mutants of the *Anabaena* 7120 *rbpB* gene have also been constructed (Holden, 1995). This mutant was shown to have a Fox^- phenotype (Wolk *et al.*, 1994), although it remained fully capable of forming differentiated heterocysts.

As stated in Chapter 2, the number of *rbp* genes present in cyanobacterial strains varies greatly. It was initially thought that the unicellular cyanobacterium *Synechococcus* 7942 had only a single *rbp* gene and so would make a good model for inactivation mutagenesis studies (see Section 1.3). Although a second *rbp* gene has been identified in *Synechococcus* 7942 (Dolganov and Grossman, 1994), it nevertheless remains a relatively simple system to use for mutagenesis. This strain has shown itself to be particularly amenable to genetic manipulation since it is easily transformed by plasmids as large as 50 kb in size (Kuhlemeier and van Arkel, 1987). It has also been shown to have a highly efficient system of recombination, a system which likely functions in repair of damage caused by solar radiation (Kuhlemeier and van Arkel, 1987). These properties can be

exploited to move foreign DNA into the cyanobacterial chromosome by flanking the foreign sequence with DNA homologous to the site of insertion.

The objectives of this mutagenesis study are therefore two-fold. The first objective was to create two mutants of *Synechococcus* 7942. In the first mutagenesis procedure, the object was to create a knock-out mutant by inserting an antibiotic resistance cassette within the coding sequence for the *rbpA* gene. As a control, a second mutant would be created to contain the same antibiotic resistance cassette inserted outside of the *rbpA* gene. The nucleotide sequence of the *rbpA* gene and its flanking sequence have revealed several restriction sites quite useful in these mutagenesis strategies (see Figure 2.5). The ClaI site was chosen as a convenient site at which to inactivate the gene by interposon mutagenesis. This site is located between the sequences for the RNP1 and RNP2 conserved sequences in the gene; insertion at this site should therefore preclude synthesis of any functional product. The HindIII site located downstream of the *rbpA* gene was used in the second mutagenesis experiment.

Having established the genotypes of these two mutants, the second objective would be to characterize the phenotypes resulting from this mutagenesis. Comparison of the phenotypes of the two mutants to each other and to the wild type strain would identify aspects of *Synechococcus* 7942 growth which are affected by the absence of a functional *rbpA* gene and present clues as to its possible function.

3.2. MATERIALS AND METHODS

3.2.1. Construction of pSyR1.2(Sm^r) and Mutant SY-RBPA1

Plasmids used in mutagenesis experiments are listed in **Table 3.1.** In order to investigate the functional significance of the *rbpA* gene, interposon mutagenesis was used to inactivate the gene. This was carried out using the omega (Ω) cassette originally isolated from plasmid R100.1 and modified by DNA manipulation to give a fragment flanked by both transcription and translation termination signals (Prentki and Krisch, 1984). The cassette used in our mutagenesis experiments was isolated from pDW9, which contained the spectinomycin resistance cassette cloned in the plasmid pUC1819H3 (Golden and Wiest, 1988).

Construction of the inactivated gene is outlined in **Figure 3.1.** The resulting construct contained the omega cassette inserted at the ClaI site within the *rbpA* gene. This construct was used to transform *E. coli* JM109 using the method described previously (Section 2.2.6); transformants were subsequently selected for growth on LB agar containing spectinomycin (100 μg/mL). From six potential positive clones, plasmid DNA was analyzed for the correct restriction pattern. Digestion of these clones with HindIII confirmed that clones #2 and #3 contained the correct plasmid construct (**Figure 3.2**). Clone #2 was re-streaked to obtain a pure clone; this plasmid was named pSyR1.2(Sm^r). The total size of this plasmid is 6.1 kb.

A fragment containing the entire *rbpA* gene plus the inserted cassette was removed by digestion of the plasmid pSyR1.2(Sm^r) with EcoRI and PvuII. This 3.65 kb fragment was separated by electrophoresis on a 0.8% agarose gel

Table 3.1. Plasmids Used In Mutagenesis Experiments

Plasmid	Description	Reference
pUC1819H$_3$	Analogue of the pUC plasmid which contains a duplicated and inverted multiple cloning site (MCS).	Golden and Wiest, 1988
pDW9	2 kb omega cassette fragment with the spectinomycin resistance gene in pUC1819H$_3$.	Golden and Wiest, 1988
pSyR1.2(Sm^r)	pSyR1.2 containing the 2 kb omega cassette inserted at the ClaI site within the *rbpA* gene (see Figure 3.1).	This work
pTJB1	4.1 kb *rbpA* construct in pBGS18 (see Figure 3.3).	This work
pTJB1(Sm^r)	pTJB1 containing the 2 kb omega cassette inserted at the HindIII site located downstream of *rbpA* (see Figure 3.3).	This work

Figure 3.1. Construction of plasmid pSyR1.2(Sm^r). Plasmid pSyR1.2 was linearized by digestion with ClaI, a unique restriction site within the *rbpA* gene (blue), resulting in the formation of a 4.1 kb DNA fragment. At the same time, the 2 kb omega cassette containing a gene for streptomycin and spectinomycin resistance (black) was isolated from the plasmid pDW9 following digestion with AccI. Both DNA fragments (4.1 kb and 2 kb) were purified by electrophoresis on a 0.8% agarose gel followed by gel purification as described previously (Section 2.2.6). These fragments were ligated with T4 DNA ligase using ligation conditions described previously (Section 2.2.6). The new plasmid, named pSyR1.2(Sm^r), now contained an *rbpA* gene interrupted by the omega cassette. Plasmids and DNA fragments are not drawn to scale. Parentheses indicate that a restriction site is no longer completely intact. Abbreviations: Amp, gene for ampicillin resistance; ori, origin of replication.

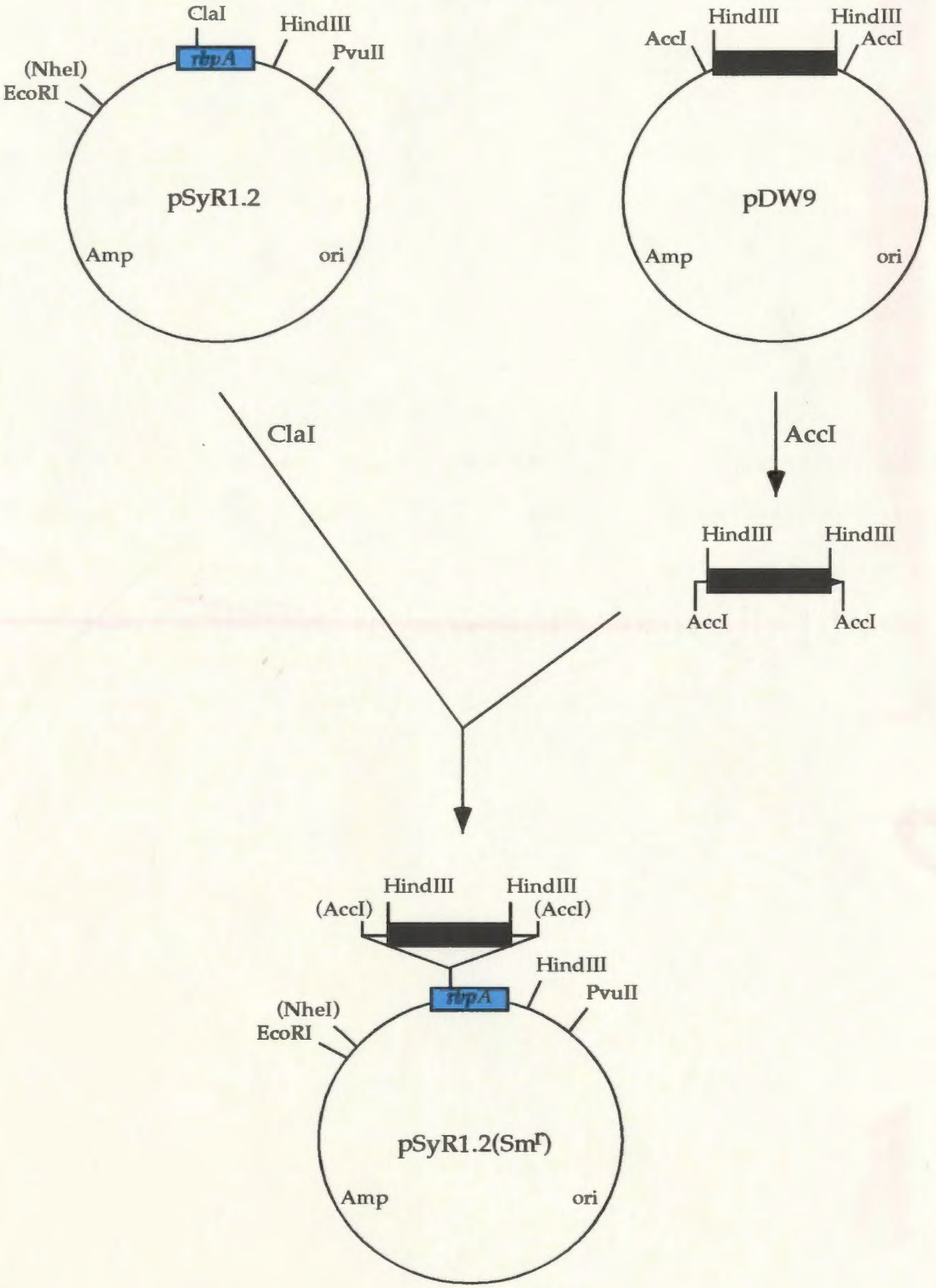
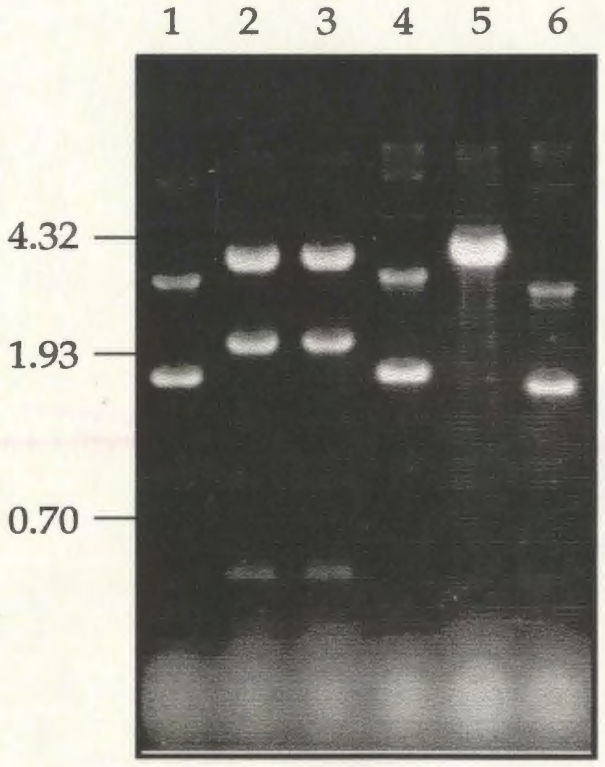

ClaI
HindIII
PvuII
(NheI)
EcoRI
tetA
pSyR1.2
Amp
ori
HindIII
HindIII
AccI
AccI
pDW9
Amp
ori
ClaI
AccI
HindIII
HindIII
AccI
AccI
HindIII
HindIII
(AccI)
(AccI)
HindIII
PvuII
(NheI)
EcoRI
tetA
pSyR1.2(Sm^r)
Amp
ori

Figure 3.2. Analysis of pSyR1.2(Sm^r) clones by restriction digestion. Plasmid DNA from six clones (lanes 1-6) was digested with HindIII and fractionated on a 0.7% agarose gel. Clones #2 and #3 have the correct pattern based on the restriction map outlined in Figure 3.1. This would consist of a 3.8 kb fragment (vector with some insert), a 2 kb fragment (omega cassette) and a 310 bp fragment (part of *rbpA* gene plus some downstream sequence). Clones #1, #4 and #6 have restriction patterns differing greatly from that expected and were not characterized further. The locations of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated.


1
2
3
4
5
6
4.32
1.93
0.70

and gel purified (see Section 2.2.6). This purified DNA fragment was used to create the SY-RBPA1 mutant of *Synechococcus* 7942 by transformation of wild type *Synechococcus* 7942 competent cells (see Section 3.2.3).

3.2.2. Contruction of pTJB1(Sm^r) and Mutant SY-RBPA3

In this plasmid construct, the strategy was to insert the omega cassette into the HindIII site located downstream of the *rbpA* gene (see Figure 2.5). The first step of this strategy involved the construction of pTJB1, a plasmid in which this HindIII restriction site was unique (**Figure 3.3**). Subsequent digestion of pTJB1 with HindIII resulted in the formation of a linear 7.75 kb DNA fragment which was ligated with a HindIII fragment containing the 2 kb omega cassette as before. The resulting ligation mixture was used to transform *E. coli* MC1061; these cells were plated on LB agar containing 40 µg/mL kanamycin.

Plasmid DNA from six potentially positive clones was analyzed for the correct restriction pattern. To do this, plasmid DNA was digested separately with HindIII and EcoRI (**Figure 3.4**). Both clones #4 and #5 had the correct restriction pattern for both enzymes. Clone #4 was re-streaked to obtain a pure clone; this plasmid was named pTJB1(Sm^r). Digestion of this plasmid with EcoRI was used to remove the entire *rbpA* region plus inserted cassette as a 6.1 kb DNA fragment. This fragment was separated on a 0.8% agarose gel and purified from an agarose gel slice as described previously (Section 2.2.6). The 6.1 kb DNA fragment was used to create the SY-RBPA3 mutant of *Synechococcus* 7942 by transformation of wild type *Synechococcus* 7942 competent cells (Section 3.2.3).

Figure 3.3. Construction of plasmid pTJB1(Sm^r). Plasmid pSyR1.2 was digested with HindIII and NdeI to give a 1.75 kb fragment which included the 1.45 kb NheI/HindIII *rbpA* fragment and approximately 300 bp of pUC19 between the NdeI and NheI restriction sites. Digestion of pSyR1.3 with HindIII and NdeI yielded a 6 kb fragment which included the 2.65 kb HindIII/NheI downstream fragment (SyR1.4) and most of the pBGS18 vector. Ligation of the 1.75 kb and 6 kb fragments restored the NheI/HindIII *rbpA* region upstream of the HindIII site. The resulting plasmid was named pTJB1. Plasmid pTJB1 was digested with HindIII resulting in the formation of a linear 7.75 kb linearized vector. At the same time, the 2 kb omega cassette was removed from pDW9 by digestion with HindIII. Both DNA fragments were purified on a 0.8% agarose gel as described previously (Section 2.2.6). Both fragments were ligated with T4 DNA ligase using conditions described previously (Section 2.2.6); this final plasmid construct was named pTJB1(Sm^r). Note that sizes of plasmids and DNA fragments are not drawn to scale. Parentheses indicate that a restriction site is no longer completely intact. Abbreviations: Amp, gene for ampicillin resistance; Kan, gene for kanamycin resistance; Sm/Sp, gene for streptomycin and spectinomycin resistance; ori, origin of replication.

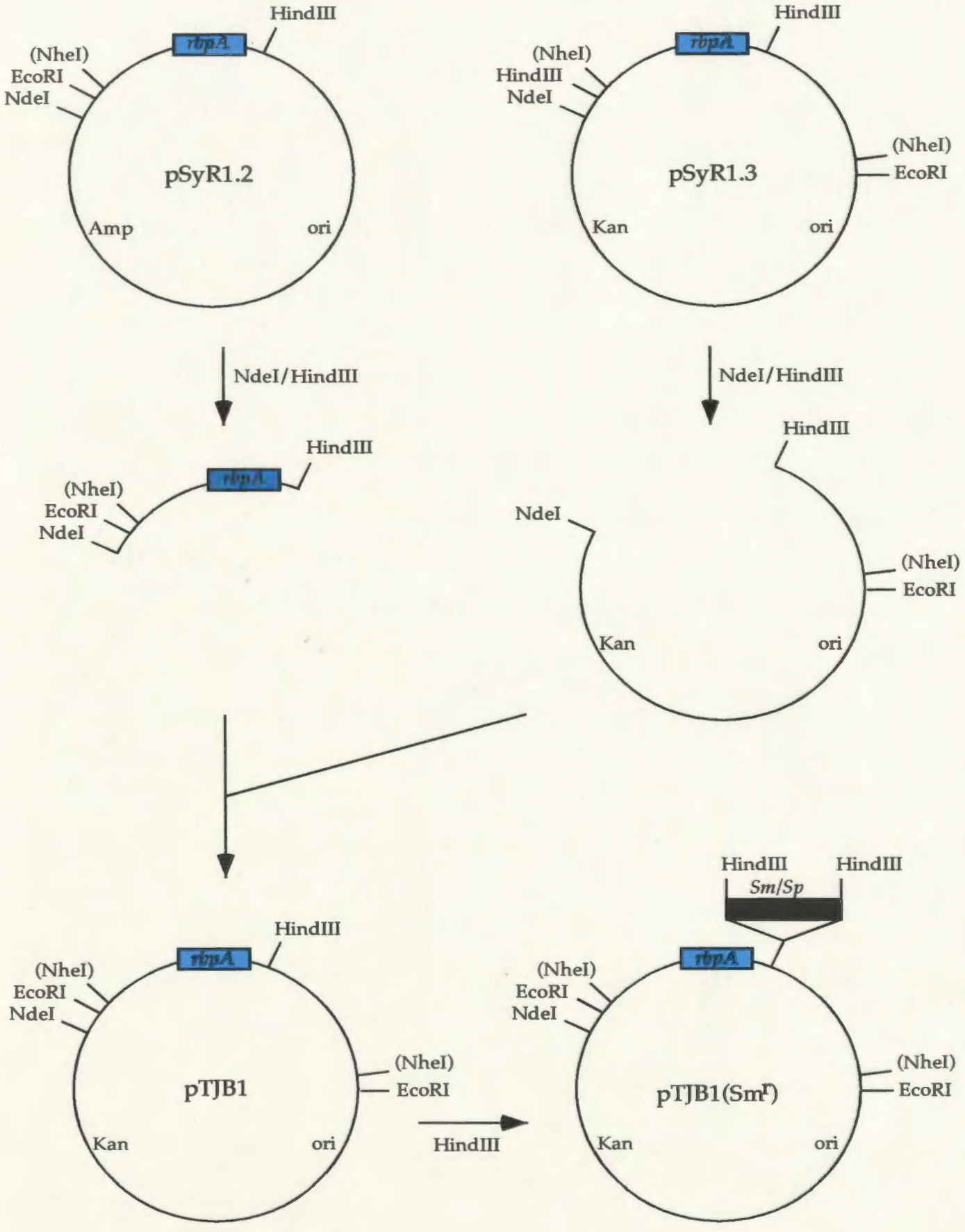
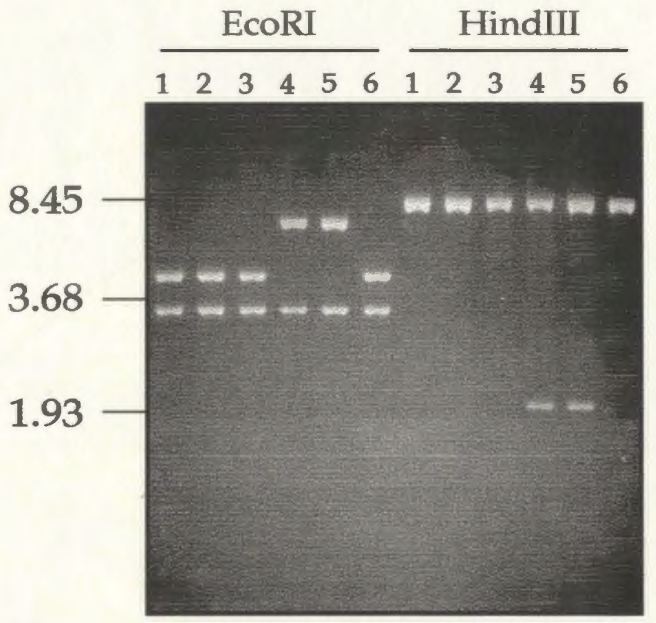

HindIII
rbpA
(NheI)
EcoRI
NdeI
pSyR1.2
Amp
ori
HindIII
rbpA
(NheI)
HindIII
NdeI
pSyR1.3
(NheI)
EcoRI
Kan
ori
NdeI/HindIII
NdeI/HindIII
HindIII
rbpA
(NheI)
EcoRI
NdeI
HindIII
NdeI
(NheI)
EcoRI
Kan
ori
HindIII
rbpA
(NheI)
EcoRI
NdeI
pTJB1
(NheI)
EcoRI
Kan
ori
HindIII
HindIII
HindIII
Sm/Sp
rbpA
(NheI)
EcoRI
NdeI
pTJB1(Sm^r)
(NheI)
EcoRI
Kan
ori

Figure 3.4. Analysis of pTJB1(Sm^r) clones by restriction digestion. Plasmid DNA from six clones was digested with EcoRI (lanes 1-6) or with HindIII (lanes 7-12) and fractionated on a 0.8% agarose gel. Digestion with EcoRI should remove the entire construct as a 6.1 kb DNA fragment, separating it from the 3.4 kb pBGS18 vector fragment. Digestion with HindIII would remove only the 2 kb omega cassette. Clones #4 and #5 have the correct pattern based on the restriction map outlined in Figure 3.3. Clones #1, #2, #3 and #6 appear to be recircularized pTJB1 plasmid. The locations of bacteriophage λ BstEII DNA fragments used as size markers are indicated.


EcoRI
HindIII
1 2 3 4 5 6 1 2 3 4 5 6
8.45
3.68
1.93

3.2.3. Transformation of *Synechococcus* 7942 Strains and Selection of Mutants

Transformation of *Synechococcus* 7942 cells was carried out as described by Golden (1988). Cells from 30 mL of a late-log phase culture were pelleted by centrifugation (5000 rpm, 5 min), washed briefly with 10 mM NaCl, and resuspended in 2 mL of BG-11 media (Allen and Stanier, 1968). DNA fragments were mixed with 0.5 mL of *Synechococcus* 7942 cells, and the resulting mixture inverted by continuous rotation in the dark overnight. The mixture was diluted 10-fold in liquid BG-11 media prior to plating onto solid BG-11 media containing 1.5% agar, 10 mM sodium thiosulphate, 50 µg/mL cycloheximide and 40 µg/mL spectinomycin. Cycloheximide was added to inhibit fungal growth, while sodium thiosulphate was added since it was found to promote cyanobacterial growth on solid media (Golden, 1986). Plates were then incubated at room temperature with constant illumination (approx. 750 ft. candles of light). Transformants capable of growing on the antibiotic typically appeared as single colonies within two weeks. These were re-streaked onto solid BG-11 media supplemented as before, and single clones used to inoculate 100 mL liquid BG-11 cultures.

3.2.4. Analysis of *rbpA* Expression in Wild Type and Mutant *Synechococcus* Strains

For Northern blot analysis of *rbpA* transcripts, total RNA from each *Synechococcus* strain was prepared as described by Kormanec and Farkasovsky (1994). Approximately 5-10 µg of total RNA was denatured in glyoxal/DMSO solution (1 M glyoxal, 50% DMSO, 10 mM sodium phosphate, pH 7.0) at 50°C for 1 hr (McMaster and Carmichael, 1977; Sambrook *et al.*,

1989). Gel-loading buffer (1X is 10% glycerol, 10 mM sodium phosphate pH 7.0, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added to each sample prior to loading onto a 1.5% agarose gel. Gel electrophoresis was carried out at 3-4 V/cm in 10 mM sodium phosphate pH 7.0. RNA was blotted onto a Hybond N+ transfer membrane in 10X SSC using the downward capillary technique (Chomczynski, 1992). Hybridization with DNA probes and subsequent detection were carried out as described previously (Chapter 2).

3.2.5. Spectrophotometric Measurement of Cyanobacterial Growth

Initial characterization of the wild type and mutant *Synechococcus* 7942 phenotypes was carried out on 100 mL and 500 mL laboratory stock cultures of each strain maintained at room temperature. Spectinomycin was omitted from mutant *Synechococcus* 7942 cultures in order to maintain identical growth conditions between mutant and wild type cultures. Whole cell spectra in the range of 400 nm to 800 nm were measured in a Shimadzu UV-260 recording spectrophotometer. Ratios of absorbance peaks were measured by measuring the absorbance of each peak, subtracting the absorbance due to light scattering (800 nm) and taking the ratio of the resulting two numbers. Phycocyanin (PC), allophycocyanin (AP) and chlorophyll a were measured at 72 hr intervals using the method described by Tandeau de Marsac and Houmard (1988). For measurement of phycobiliprotein content, 2 mL of culture solution was centrifuged at 10,000 x g for 10 min at 4°C and the cell pellet resuspended in 2 mL of sodium acetate buffer (20 mM sodium acetate pH 5.5, 10 mM Na_2EDTA, 3 mM sodium azide). Cells were broken by

sonication at 4°C in a Branson 1200 ultrasonic water bath. Membrane fragments were precipitated with 1% streptomycin sulfate (30 min, 4°C) and eliminated by centrifugation at 10,000 x g for 10 min at 4°C. The absorbance of the supernatant was measured at 620 nm and 650 nm. Phycobiliprotein concentrations were calculated using the following equations:

$$PC\ (mg/mL) = \frac{OD_{620nm} - (0.7 * OD_{650nm})}{7.38}$$

$$AP\ (mg/mL) = \frac{OD_{650nm} - (0.19 * OD_{620nm})}{5.65}$$

For determination of chlorophyll a, 1 mL of culture solution was centrifuged at 10,000 x g for 10 min at 4°C. The cell pellet was extracted twice with 90% methanol for 1 hr at 4°C in the absence of light. The absorbance of the combined extract was measured at 665 nm. Chlorophyll a content (chl a) was calculated using the following equation:

$$chl\ a\ (\mu g/mL) = OD_{665nm} * 13.9$$

Intact phycobilisomes were prepared from 500 mL laboratory stock cultures as described by Glazer (1988). Stationary phase cultures were harvested by centrifugation (5000 rpm, 10 min), washed with 0.65 M $NaKPO_4$ buffer (0.65 M Na_2PO_4/0.65 M K_2PO_4, pH 8.0) and resuspended in the same buffer to a final cell density of 1.2 g wet weight cells/mL. The cells were broken by passage through a French Pressure cell at 20,000 psi, and the broken

cell suspension incubated in 1% Triton X-100 for 1 hr to release phycobilisomes. Broken cell debris was removed by centrifugation (31,000 x g, 30 min), leaving an intense blue supernatent which represented a crude preparation of phycobilisomes.

In order to isolate intact phycobilisomes, approximately 1.5 mL of this crude preparation was loaded onto a sucrose density step gradient, which consisted of 1 mL, 2 mL, 2 mL, 2 mL, and 2 mL steps of 2.0 M, 1.0 M, 0.75 M, 0.50 M, and 0.25 M sucrose, respectively, all in 0.75 M $NaKPO_4$ buffer at pH 8.0. These gradients were centrifuged at 24,000 rpm for 13 hrs at 18°C. Intact phycobilisomes appeared as an intense blue layer at the 0.75 M sucrose step, and were subsequently removed from the gradient by syringe and used directly for spectrophotometric analysis. Determination of intact phycobilisome content in each culture was calculated by measuring the absorbance of each phycobilisome preparation at 621 nm according to equations described by Yamanaka (1978).

A second series of growth experiments were carried out on wild type and mutant *Synechococcus* 7942 strains in an incubator (Conviron) which allowed for constant illumination and controlled temperature conditions. Growth rates of 100 mL cultures of wild type and mutant strains were monitored by removing 1 mL aliquots at approximately 24 hr intervals and measuring the absorbance at 750 nm in a Spectronic 601 spectrophotometer. Growth experiments were carried out separately at temperatures of 30°C and 21°C under constant illumination. Due to variation in the number of viable cells used for each inoculation, data points for growth experiments were not

recorded until the density of the culture reached an absorbance (750 nm) of approximately 0.10.

3.3. RESULTS AND DISUCSSION

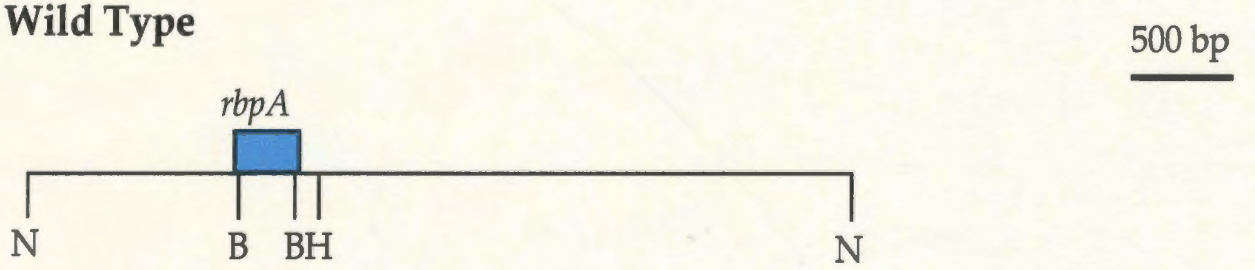
3.3.1. Confirmation of Mutant *Synechococcus* 7942 Genotypes

Inactivation of the *rbpA* gene in *Synechococcus* 7942 utilized the ClaI restriction site. A 2 kb DNA fragment containing the omega cassette, which carries a gene coding for spectinomycin resistance (Prentki & Krisch, 1984), was cloned into the ClaI restriction site and the mutated gene was then recombined into the *Synechococcus* 7942 genome. The resulting mutant, named **SY-RBPA1**, was viable, indicating that *rbpA* was not an essential gene for growth of *Synechococcus* 7942. In order to ensure that any observations regarding this mutant were not due merely to the presence of the omega cassette, a second mutant was constructed in which the cassette was inserted into the HindIII site located just downstream of the *rbpA* gene. This site is downstream of a palindromic sequence capable of forming a stem-loop structure in RNA, and it is outside of any identifiable open reading frame or known control feature (see Figure 2.5). This mutant was named **SY-RBPA3**. The restriction maps for the wild type strain and these two mutant strains in the region of the *rbpA* gene are shown in **Figure 3.5**.

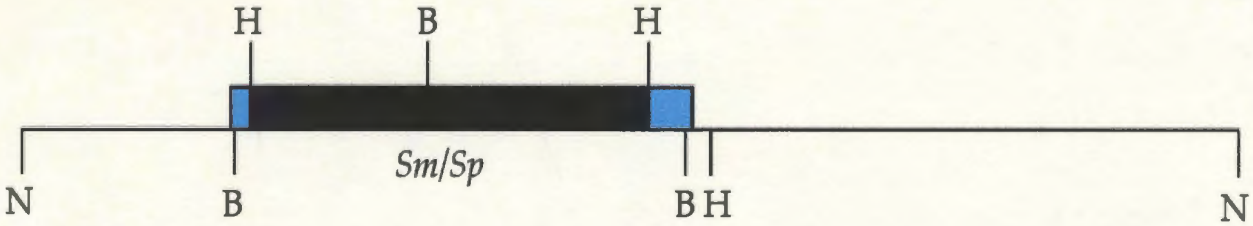
The genotypes of these mutant strains were confirmed and compared to that of wild type *Synechococcus* 7942 by Southern blot hybridization (**Figure 3.6**). Comparison of the HindIII genomic digests revealed that the original 6.3 kb hybridizing fragment observed in the wild type organism was replaced with a 0.3 kb fragment in SY-RBPA1 (**Figure 3.6A**). Mutant strain SY-RBPA3 had the same 6.3 kb hybridizing band as seen with the wild type strain. This is consistent with the maps shown in Figure 3.5. Comparison of the

Figure 3.5. Restriction maps of *rbpA* gene and surrounding sequence in wild type and mutant strains of *Synechococcus* 7942. **(A)** The 4.1 kb NheI DNA fragment in the wild type strain contains a single HindIII site. Within the *rbpA* gene are two BstEII sites. Insertion of the omega cassette introduces two HindIII sites and a BstEII site into this region. **(A)** In mutant SY-RBPA1, the 2 kb omega cassette (black) is inserted within the coding sequence for the *rbpA* gene (blue). **(C)** In mutant SY-RBPA3, the omega cassette in inserted at the HindIII site located downstream of the *rbpA* gene. Abbreviations for restriction sites are as follows: B, Bst EII; H, HindIII; N, NheI.

A. Wild Type



B. SY-RBPA1



C. SY-RBPA3

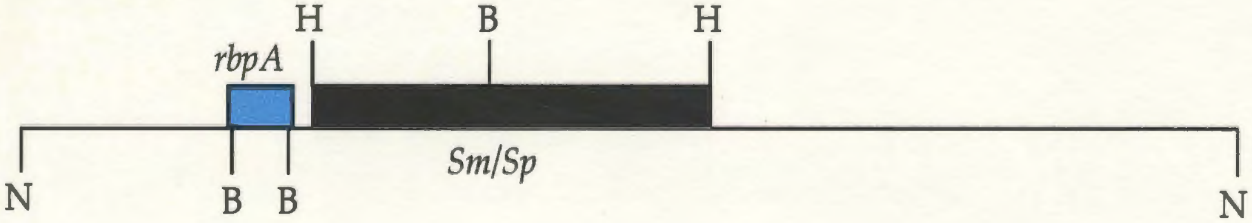
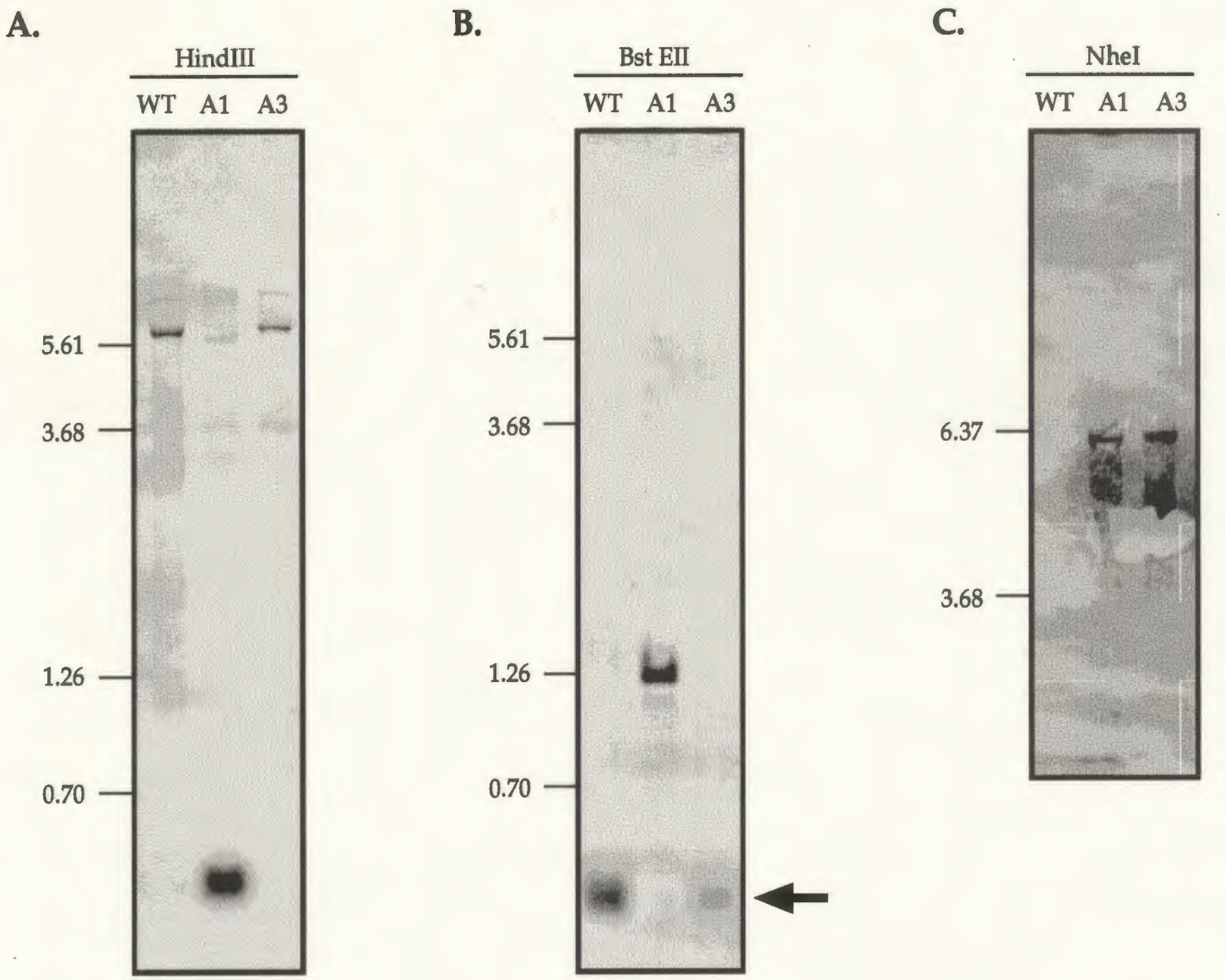


Figure 3.6. Analysis of wild type and mutant *Synechococcus* 7942 genotypes by Southern hybridization. Genomic DNA (3-5 μg) from *Synechococcus* 7942 (WT), mutant SY-RBPA1 (A1) and mutant SY-RBPA3 (A3) was digested with **(A)** HindIII, **(B)** BstEII and **(C)** NheI. Probes used in the hybridization analysis consisted of a 290 bp BstEII DNA fragment containing sequence totally internal to the *rbpA* gene (*rbpA*) and a 2 kb HindIII fragment containing the entire omega cassette (*Sm/Sp*). An arrow is used to indicate the location of the 290 bp BstEII hybridizing fragment which is used as an indicator of an intact *rbpA* gene. The locations of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left of each blot. Sizes of smaller hybridizing bands were determined using a 100 bp DNA ladder (Gibco BRL).

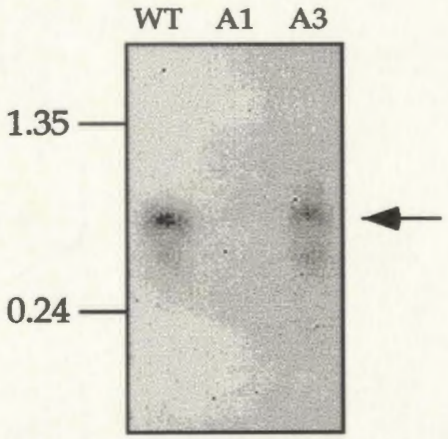

A.
HindIII
WT A1 A3
5.61
3.68
1.26
0.70
B.
Bst EII
WT A1 A3
5.61
3.68
1.26
0.70
C.
NheI
WT A1 A3
6.37
3.68
Probe:
rbpA
rbpA
Sm/Sp

BstEII digests of genomic DNA from all strains demonstrated that the original 290 bp BstEII DNA fragment present in the wild type organism was replaced with 1.3 kb and 1.1 kb hybridizing fragments due to the presence of the inserted omega cassette (**Figure 3.6B**). Mutant strain SY-RBPA3 had a hybridization pattern identical to that of the wild type strain as expected.

The genotypes of these three strains were further confirmed using a DNA probe corresponding to the omega cassette used in the mutagenesis procedure (**Figure 3.6C**). Both mutant strains contained a 6.1 kb fragment which hybridized with the omega cassette probe as expected from the restriction maps in Figure 3.5. No hybridizing sequences were observed in the wild type genomic DNA.

The results of Northern blot hybridization analysis of total RNA isolated from these three strains of *Synechococcus* 7942 revealed that the *rbpA* gene is expressed as a 550 nt RNA transcript in the wild type organism (**Figure 3.7**). This result is consistent with that observed in *Synechococcus* 6301 (Sugita and Sugiura, 1994). A similar sized transcript was also observed in mutant SY-RBPA3. However, no transcript was detected in RNA isolated from mutant SY-RBPA1. These results, combined with those obtained from genomic Southern blot analysis, confirmed that the *rbpA* gene is not expressed in the mutant SY-RBPA1. It also confirmed that insertion of the omega cassette at the downstream HindIII site did not interfere with the expression of the *rbpA* gene.

Figure 3.7. Transcription of *rbpA* in wild type and mutant *Synechococcus* 7942 strains. Total RNA isolated from cultures of *Synechococcus* 7942 was used for Northern hybridization analysis, using the digoxygenin-labelled 290 bp DNA probe. The location of RNA ladder (Gibco BRL) fragments used for estimating sizes of hybridizing bands are indicated at the left.



rbpA

3.3.2. Effects of *rbpA* Inactivation on *Synechococcus* 7942 Pigment Composition and Phycobilisome Content

In *Synechococcus* species, the major pigments responsible for photosynthesis are chlorophyll a (chl a) and the phycobiliproteins phycocyanin (PC) and allophycocyanin (AP). Each phycobiliprotein exists as two dissimilar subunits (α and β) which differ in molecular weight and composition. In the case of phycocyanin, the α subunit consists of a 17.7 kD polypeptide which is attached via a cysteine residue to a single tetrapyrrole chromophore (phycocyanobilin); the β-subunit consists of a 19 kD polypeptide attached to two phycocyanobilin chromophores (Glazer *et al.*, 1973). These proteins assemble into a water soluble, macromolecular protein complex known as the phycobilisome which attaches to the cytoplasmic surface of cyanobacterial thylakoid membranes (reviewed in Glazer (1985)). PC and AP phycobiliproteins assemble into $(\alpha\beta)_6$ hexameric complexes which form the basic subunits of the phycobilisome. The AP hexamers are localized to the core of the phycobilisome along with nonpigmented polypeptides which anchor the phycobilisome to the cytoplasmic surface of the thylakoid membranes. The PC hexamers comprise the rod-like structures which protrude from the central core. The rod-like structures of the phycobilisome are responsible for a unidirectional flow of captured light energy toward the core of the phycobilisome (Glazer, 1989). This overall structure is stabilized by a separate group of nonpigmented linker polypeptides.

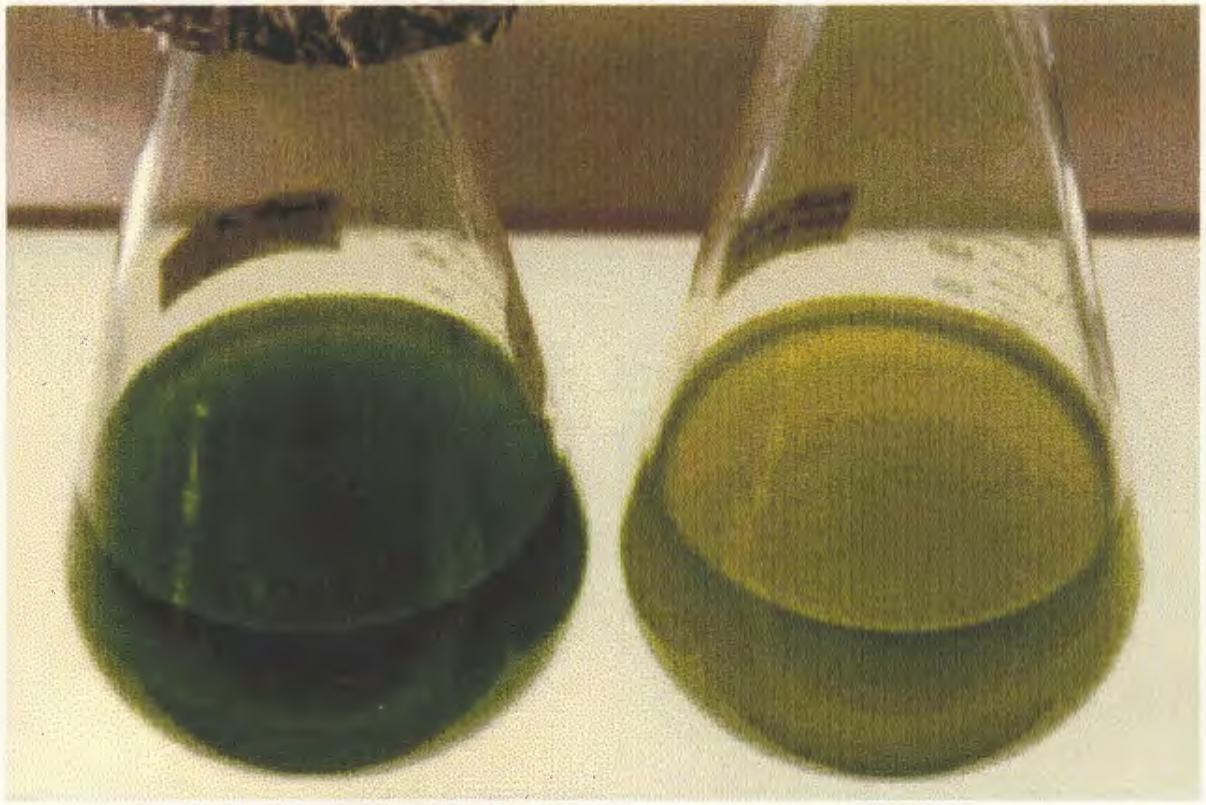
One of the first observations made concerning the mutant *Synechococcus* 7942 phenotypes was that cultures of SY-RBPA1 appeared to have a bleached appearance when compared to the deep blue-green colour

observed in the wild type culture (**Figure 3.8**). This bleached appearance was not observed in the mutant SY-RBPA3. Whole cell absorbance spectra for each strain in the range from 400 nm to 800 nm were recorded in order to identify the component which could be responsible for the bleaching effect. Results of whole cell spectra indicated that the relative heights of absorbance peaks at 630 nm and 680 nm were affected by the inactivation of *rbpA* (**Figure 3.9**). More specifically, the mutant SY-RBPA1 showed a reduction in the ratio of the phycobiliprotein (630 nm) absorbance peak to the chlorophyll a (680 nm) absorbance peak (ratio 0.99) relative to the ratio observed in the wild type strain (ratio 1.13). This effect was not observed in the mutant SY-RBPA3, where the peak ratio (ratio 1.24) was actually higher than that observed for the wild type culture. This result suggested that inactivation of the *rbpA* gene affected the phycobiliprotein content of these cells.

For growing stock cultures of wild type and mutant *Synechococcus* 7942, PC, AP and chlorophyll a were measured using the method described by Tandeau de Marsac and Houmard (1988). A reduction in PC, AP, and chlorophyll a content was observed in the mutant SY-RBPA1 culture relative to the wild type culture (**Figure 3.10**). There also appeared to be an increase in the ratio of PC to AP. These effects were not observed in the mutant SY-RBPA3.

In order to confirm the apparent effect of *rbpA* inactivation on the phycobiliprotein content of SY-RBPA1 cells, total intact phycobilisomes were isolated by centrifugation using a sucrose density gradient, according to the method of Glazer (1988) (**Figure 3.11**). Approximately 7 mg of intact

Figure 3.8. Pigmentation of stock cultures of wild type *Synechococcus* 7942 and mutant SY-RBPA1 grown at room temperature. This photograph, taken 40 days after inoculation, shows the difference in pigmentation between the two strains.



Wild Type SY-RBPA1

Figure 3.9. Representative whole cell spectra of wild type (red), SY-RBPA1 (green) and SY-RBPA3 (blue) *Synechococcus* 7942 strains. Cells were diluted with BG-11 media to an approximate absorbance (800 nm) of 0.2 prior to spectral measurement. The peaks at 680 nm and 440 nm are due to the presence of chlorophyll a. The peak at 630 nm is due to the presence of phycobiliproteins, mainly phycocyanin (PC). Other peaks within the range of 400 to 550 nm are due to carotenoid absorbance.

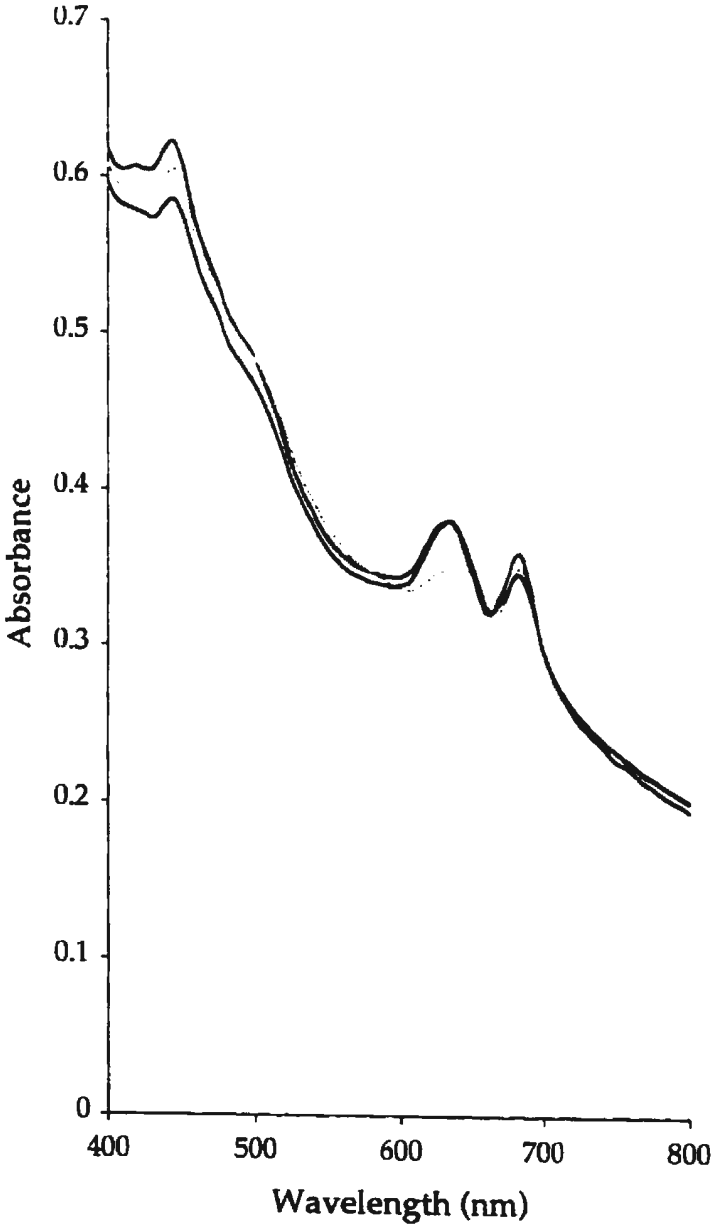

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance
400
500
600
700
800
Wavelength (nm)

Figure 3.10. Measurements of **(A)** phycocyanin (PC), **(B)** allophycocyanin (AP), **(C)** chlorophyll a (chl a), and **(D)** PC/AP ratio in growing stock cultures of wild type (red), SY-RBPA1 (green), and SY-RBPA3 (blue) *Synechococcus* 7942 maintained in liquid BG-11 culture media at 25°C. Phycobiliproteins and chlorophyll a were measured using the method of Tandeau de Marsac and Houmard (1988).

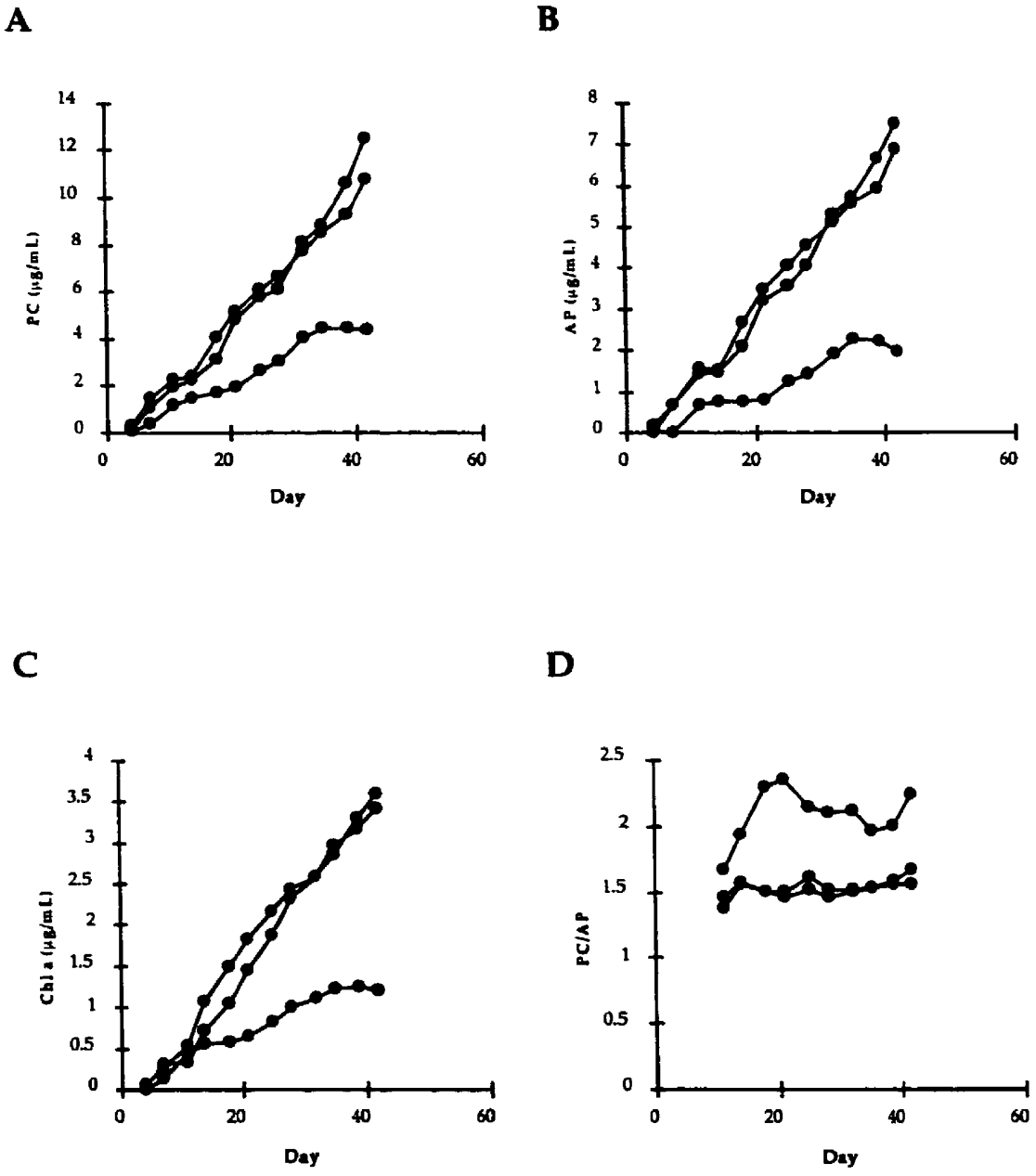
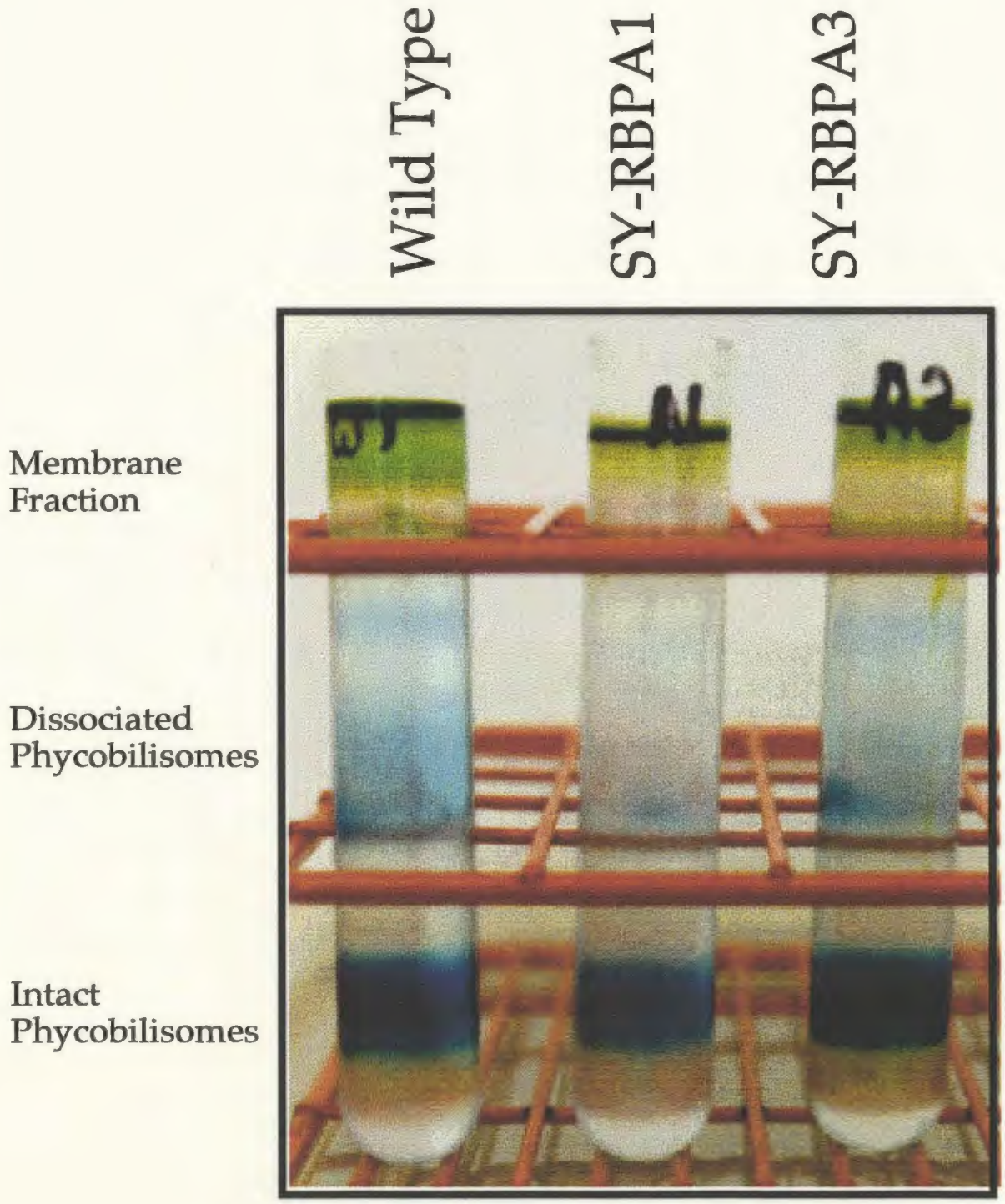

A
B
C
D
PC (µg/mL)
AP (µg/mL)
Chl a (µg/mL)
PC/AP
Day
0
20
40
60

Figure 3.11. Isolation of intact phycobilisomes from wild type and mutant strains of *Synechococcus* 7942 using sucrose density centrifugation. For each strain, 2 mL of a crude phycobilisome preparation was loaded onto a sucrose density gradient and centrifuged at 36,000 rpm for 13 hrs at 18°C. Intact phycobilisomes appear as an intense blue band at the 0.75 M step. Minor bands at lower density steps are due to small quantities of dissociated phycobilisome complexes.


Wild Type
SY-RBPA1
SY-RBPA3
Membrane Fraction
Dissociated Phycobilisomes
Intact Phycobilisomes

phycobilisomes were isolated per gram wet weight of SY-RBPA1 cells. This amount was less than half the amount isolated from wild type cells (15 mg intact phycobilisomes per gram wet weight of cells) and lower than that isolated from the mutant SY-RBPA3 (11 mg intact phycobilisomes per gram wet weight of cells). Fluorescence spectra on samples of intact phycobilisomes indicated that when excited at a wavelength of 580 nm, intact phycobiliproteins derived from wild type *Synechococcus* 7942 showed maximum fluorescence at a wavelength of approximately 660 nm (data not shown). However, the fluorescence spectra for phycobilisomes isolated from mutant strain SY-RBPA1 was blue-shifted, with a maximum fluoresence at a wavelength of 650 nm. This shift is consistent with the dissociation of phycocyanin into trimers and monomers and may represent a decrease in the stability of phycobilisome complexes (Yamanaka *et al.*, 1978). The spectral properties of phycobilisomes isolated from strain SY-RBPA3 were comparable to those of the wild type organism.

Overall, the inactivation experiment indicated that the RbpA protein in *Synechococcus* 7942 affected the regulation of the phycobilisome content in the cell. Whole cell spectra demonstrated that the mutant SY-RBPA1 strain had a reduced ratio of phycobiliprotein to chlorophyll a content relative to the wild type strain. Reduction of phycobiliprotein content was also reported in mutant (K1 and K8) cultures of *A. variabilis* M3 in which another *rbp* gene (*rbpA1*) had been either inactivated or its expression affected by interposon mutagenesis (Sato, 1996). In mutant SY-RBPA1, there was a decrease in the amount of intact phycobilisomes isolated from cultures of this knock-out strain relative to the wild type. Spectral measurements also showed a slight

increase in the ratio of PC to AP in the knock-out strain, an effect not observed in wild type *Synechococcus* 7942 cells. The role of RbpA in regulation of phycobiliprotein content is consistent with expression studies on the corresponding *rbp* gene in *Synechococcus* 6301 (12RNP1) which demonstrated that this gene was regulated in a light-dependent manner (Sugita and Sugiura, 1994). Regulation of expression in response to light intensity is a pattern often observed in genes related to photosynthesis.

There are a number of possible ways in which this RNA-binding protein might contribute to regulating phycobilisome content in this cyanobacterium. First, the genes coding for phycobilisome components are localized in two multicistronic operons in *Synechococcus* 7942. The *cpc* operon contains genes coding for both the α and β forms of PC. The genes encoding these phycocyanins are duplicated and separated by an intergenic region containing the genes for various linker proteins (Kalla *et al.*, 1989). The *apc* operon, containing genes coding for the α and β forms of AP, is structured in a similar way. These phycobiliprotein operons produce polycistronic RNA transcripts of varying length and in varying amounts, depending on the growth conditions at any given time (Kalla *et al.*, 1993). RbpA may play a role in regulating the levels of these transcripts under different lighting conditions, so that the absence of a functional RbpA protein in the SY-RBPA1 mutant may perturb the cells' ability to grow under certain environmental conditions. A second possibility is that RbpA might actually be a regulatory component of phycobilisomes. Immunological experiments using an anti-RNP antibody with subcellular membrane fractions of the closely related strain *Synechococcus leopoliensis* have revealed that a small

snRNP-like protein is associated with the thylakoid/phycobilisome fraction of this cyanobacterium (Vijay and Kovacs, 1993).

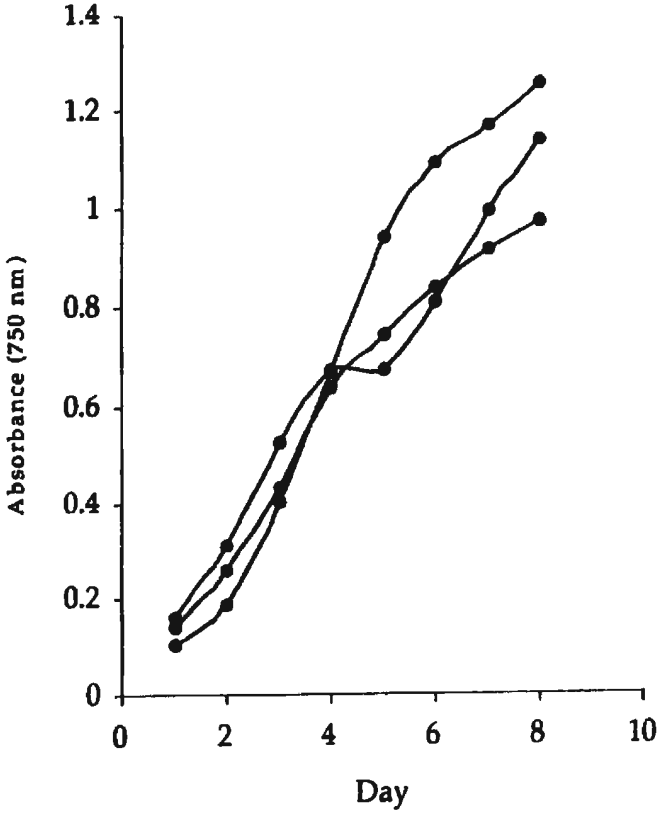
3.3.3. Effect of Temperature on Growth Rates of Wild Type and Mutant *Synechococcus* 7942

The growth rates of wild type and mutant *Synechococcus* 7942 cultures were measured under controlled conditions of temperature and light in order to investigate the effects of *rbpA* inactivation on phenotype. Growth rates were measured by monitoring the apparent absorbance due to light scattering of growing cultures at 750 nm. At normal growth temperature (30°C), all three strains grew at approximately the same rate (**Figure 3.12A**). Doubling times for the three strains varied from approximately 2.5 days for the knock-out phenotype SY-RBPA1 to 1.8 days for the wild type strain. When strains were grown at a lower temperature (21°C), both the wild type and mutant SY-RBPA3 reduced their doubling times to approximately 3 days (**Figure 3.12B**). The mutant SY-RBPA3 appeared to grow at a rate slightly lower than the wild type strain. The most noticeable difference was observed in the mutant SY-RBPA1. Its growth rate was so severely affected by this decrease in temperature that a doubling time could not be accurately determined.

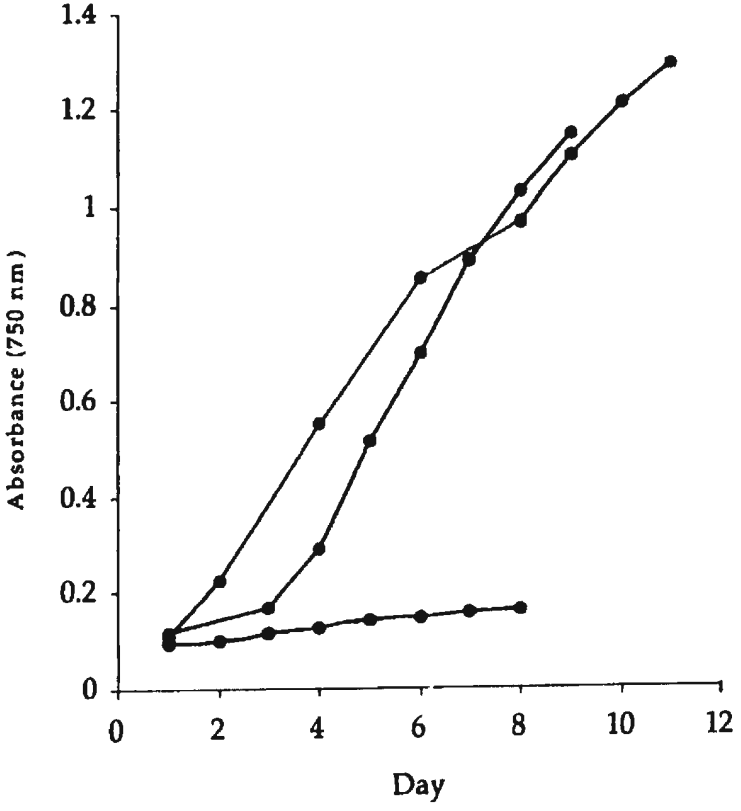
It was concluded from this experiment that the *rbpA* gene was essential for growth at low temperatures. Knock-out strains SY-RBPA1 appeared to grow very slowly at a temperature of 21°C. This effect was not observed in the mutant strain SY-RBPA3 or in the wild type strain. This observation agreed with a finding in *Synechococcus* 6301 that the level of the corresponding RNA-binding protein gene (12RNP1) transcript increased two-fold when

Figure 3.12. Growth of wild type (red), SY-RBPA1 (green), and SY-RBPA3 (blue) *Synechococcus* 7942 strains in BG-11 growth media. Cultures (100 mL) were incubated with shaking (200 rpm) under constant illumination at **(A)** 30°C and **(B)** 21°C. The density of each culture was monitored by measuring light scattering at 750 nm. Each data point represents the average of absorbances taken for two cyanobacterial cultures grown in parallel under identical conditions.

A.


Absorbance (750 nm)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
Day

B.


Absorbance (750 nm)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
Day

cultures grown at 30°C were exposed for 3 hours to a lower temperature of 20°C (Sugita and Sugiura, 1994).

Many of the cyanobacterial Rbps characterized to date accumulate when cells are grown at low temperature (Sato, 1994; Sato, 1995). This observation is consistent with previous results obtained with *Anabaena* 7120 mutant TLT2, obtained by insertion of a *lux*-bearing transposon into the genome (Panoff, 1993). Mutant TLT2 increased in luminescent intensity when transferred to from 30°C to 20°C. DNA sequencing revealed that the *tlt2* gene was in fact the *rbpA* gene of *Anabaena* 7120.

The site of regulation of the *rbp* genes by cold is likely found in the conserved upstream elements associated with all of the cyanobacterial *rbp* genes characterized to date (see Figure 2.8). Both interposon mutagenesis and deletion analysis of the region upstream of *rbpA1* in *A. variabilis* M3 has implicated the region in the regulation of *rbpA1* expression by a decrease in temperature (Sato and Wada, 1996; Sato and Nakamura, 1998). It has been suggested that several polypeptides (75 kDa, 72 kDa and 35 kDa) which form complexes with this upstream element may act through a model of transcriptional repression to regulate expression of *rbpA1* by temperature (Sato and Nakamura, 1998).

If the cyanobacterial Rbps are required for growth at colder temperatures, what is their mechanism of action? They may function as RNA-chaperones which maintain the correct secondary and tertiary structures of cellular RNA during periods of temperature stress. It is also possible that the cyanobacterial Rbps alter the pattern of gene expression in the cyanobacterial cell exposed to cold stress, either by protecting certain

transcripts from degradation or by facilitating their degradation through binding. One candidate for this mode of regulation are the acyl-lipid desaturases in cyanobacteria (reviewed in Murata and Wada, 1995). In *Synechococcus* 7002 and *Synechocystis* 6803, two unicellular strains of cyanobacteria, there are changes in both lipid composition and expression of lipid desaturase genes in response to a shift in temperature (Sakamoto *et al.*, 1997; Los *et al.*, 1997). The increases in mRNAs transcribed from these genes are caused by a combination of enhanced transcription and increased mRNA stability in cells grown at the lower temperature of 22°C relative to those grown at 38°C.

If cyanobacterial RBPs function in cold tolerance, they could be a counterpart of the bacterial cold shock proteins of *E. coli* and *B. subtilis* (Graumann and Marahiel, 1996a). These proteins contain motifs similar to the cyanobacterial Rbps and share a similar single-stranded nucleic acid binding surface (Schnuchel *et al.*, 1993). No bacterial cold shock proteins have yet been identified in cyanobacteria. At the same time, RNP-type RNA-binding proteins have not been identified in any other prokaryotic organism. It has been suggested that these two families of proteins may represent a case of convergent evolution with respect to their cold shock and nucleic-acid binding functions (Graumann and Marahiel, 1996b).

CHAPTER 4. COMPLEMENTATION OF *SYNECHOCOCCUS* 7942 SY-RBPA1 MUTANT PHENOTYPE

4.1. INTRODUCTION

The next step in the establishment of a true knock-out phenotype for any given gene is to determine whether or not it is possible to restore the mutant phenotype to that of the wild type organism by re-introducing an intact copy of the gene of interest. In the case of the mutant strain SY-RBPA1, I attempted to complement its mutant phenotype by re-introducing an intact copy of the *rbpA* gene into the cyanobacterial genome by homologous recombination.

The site chosen for insertion of the *rbpA* gene had been previously designated as a "neutral-site" since foreign DNA fragments could be inserted into this region without any deleterious effects on the cell (Bustos and Golden, 1992). I decided to use a neutral-site integration vector named pAM990 which was specifically designed for use in *Synechococcus* 7492 (Li and Golden, 1993). This plasmid, which was incapable of replicating in *Synechococcus* 7942, contained a 2.8 kb BamHI fragment of chromosomal DNA whose sequence corresponded to a neutral-site in the chromosome. Within this region was a selectable spectinomycin resistance marker (see Figure 4.1). The usefulness of this marker was limited by the fact that spectinomycin selection had already been used in the initial mutagenesis of *rbpA* (Chapter 3). It was therefore necessary to replace the spectinomycin resistance cassette of pAM990 with one which coded for resistance to the antibiotic kanamycin. This newly constructed neutral-site vector was named pMUN3. It contained two restriction sites (SmaI and BglII) which could be

used for cloning of DNA fragments for recombination into the neutral-site of the cyanobacterial chromosome.

A homologous double recombination event with the corresponding region in the cyanobacterial genome resulted in the integration of the cloned DNA fragment into the genome along with a selectable kanamycin resistance marker. The *Synechococcus* 7942 strains resulting from recombination events with this vector were subsequently selected for growth on spectinomycin and kanamycin.

The genotypes of the neutral-site strains were confirmed using Southern blot hybridization analysis of genomic DNA from each strain. The phenotypes of the neutral-site strains were also tested for pigment alteration and low-temperature growth relative to both the wild type and earlier mutant *Synechococcus* 7942 strains.

4.2. MATERIAL AND METHODS

4.2.1. Construction of Neutral-Site Integration Vector pMUN3

Plasmids used in these complementation experiments are listed in **Table 4.1**. The neutral site mutant strains of *Synechococcus* 7942 were created using the neutral-site vector pMUN3, a plasmid derived from the neutral-site vector pAM990 (Li and Golden, 1993). Construction of pMUN3 is outlined in **Figure 4.1**. It was imperative that in the last step of pMUN3 construction the 1.2 kb SalI DNA fragment containing the upstream neutral-site flanking sequence (ns#1) be incorporated in the correct orientation in order to facilitate a double recombination event with the corresponding region in the *Synechococcus* 7942 genome. The correct orientation of this insert was confirmed by digestion of pMUN3 with SalI, BamHI, and HindIII **(Figure 4.2)**. Both plasmid clones which were screened appeared to have the correct restriction pattern based on the map shown. The result was a pAM990-based neutral-site vector which permitted incorporation of a kanamycin resistance cassette (*Kan*) into the neutral-site of the *Synechococcus* 7942 genome. The new vector has a total size of 12.8 kb. The SmaI and BglII restriction sites located between the kanamycin resistance cassette (*Kan*) and the promoterless *lacZ* gene on this plasmid were subsequently used for insertion of DNA fragments containing the *rbpA* gene.

4.2.2. Construction of pMUN3.1 and Mutant SY-RBPA6

Plasmid pMUN3.1 was constructed for the creation of mutant SY-RBPA6 **(Figure 4.3)**. This plasmid contained a 1.1 kb DNA fragment with the

Table 4.1. Plasmids Used In Complementation Experiments

Plasmid	Description	Reference
pAM990	pBR328-based plasmid containing a chromosomal segment of *Synechococcus* 7942 DNA along with the omega cassette and a promoterless *lacZ* gene.	Li and Golden, 1993
pSKS101	pUC7-based plasmid containing the 1.2 kb SmaI/HindIII kanamycin resistance cassette from transposon *Tn*5.	Shapira *et al.*, 1983
pMUN3	pAM990-based plasmid in which the omega cassette has been replaced with a cassette for kanamycin resistance (see Figure 4.1).	This work
pMUN3.1	pMUN3 containing a 1.1 kb SmaI/DraI DNA fragment with an intact *rbpA* gene (see Figure 4.3).	This work
pTJB2	pUC19 plasmid containing a 1.6 kb BamHI DNA fragment with an intact *rbpA* gene (see Figure 4.5).	This work
pMUN3.3	pMUN3 containing a 1.6 kb BamHI DNA fragment with an intact *rbpA* gene (see Figure 4.7).	This work

Figure 4.1. Construction of the neutral-site integration vector pMUN3. The neutral site vector pAM990 was digested with SmaI and SalI and the resulting DNA fragments separated on a 0.8% agarose gel. Both the 9.8 kb SmaI/SalI DNA fragment which included the majority of the plasmid sequence, and the 1.2 kb SalI DNA fragment containing the upstream neutral site flanking sequence (ns#1; dark blue) were purified. At the same time, a 1.8 kb SmaI/SalI DNA fragment containing a kanamycin resistance (*Kan*) cassette (yellow) was purified following digestion of the plasmid pSKS101 with SalI and SmaI. Ligation of this fragment with the 9.8 kb pAM990 fragment introduced a *Kan* cassette into this new vector, named pMUN2. In order then to re-introduce the upstream neutral site flanking sequence (ns#1), pMUN2 was linearized with SalI and ligated with the 1.2 kb SalI DNA fragment isolated from pAM990. The resulting plasmid was named pMUN3. Note that sizes of plasmids and DNA fragments are not drawn to scale. Restriction sites immediately adjacent to one another are listed as a group (top to bottom in clockwise order).

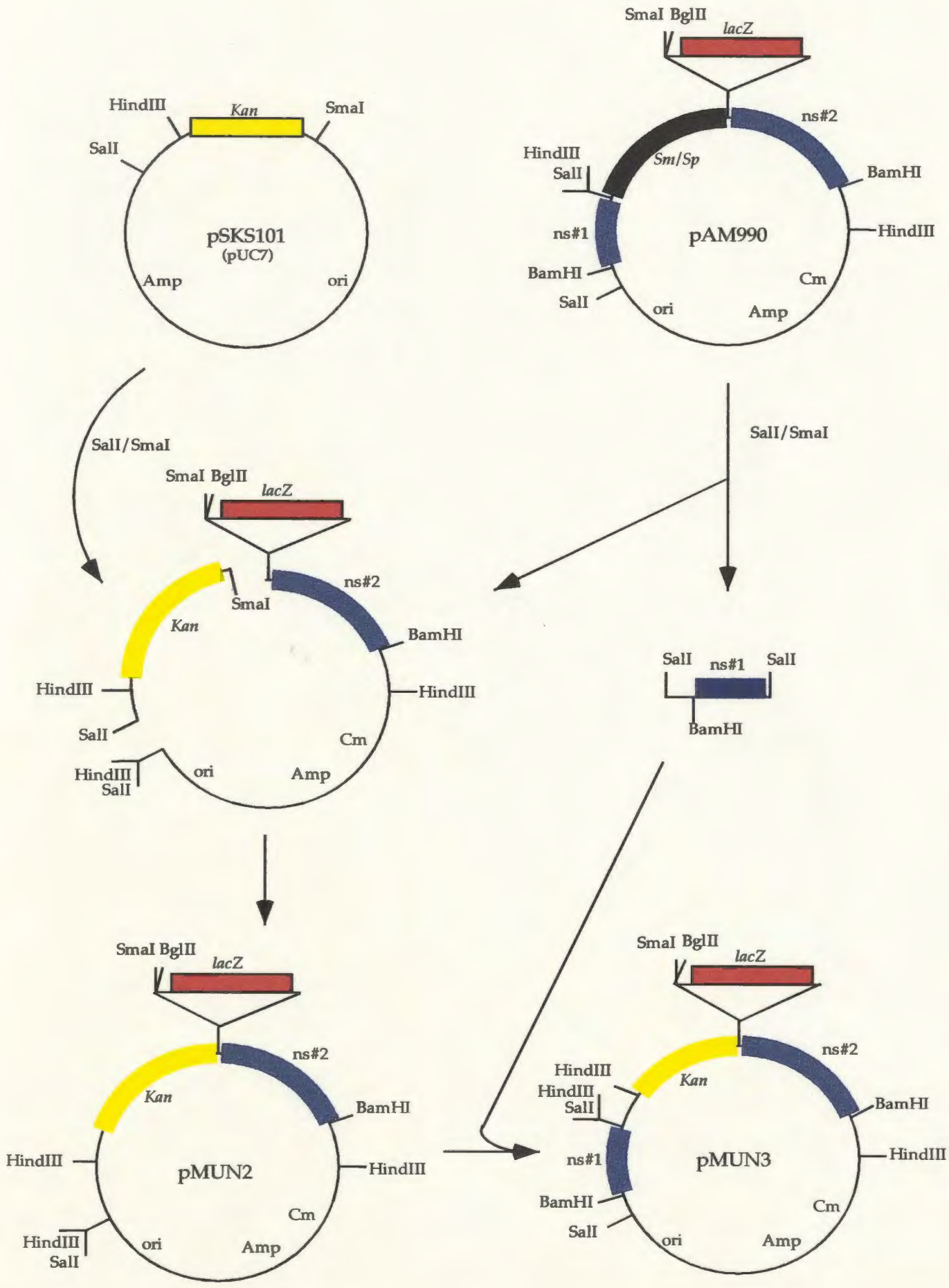

HindIII
Kan
SmaI
SalI
pSKS101
(pUC7)
Amp
ori
SmaI BglII
lacZ
ns#2
HindIII
Sm/Sp
SalI
BamHI
ns#1
pAM990
HindIII
BamHI
Cm
SalI
ori
Amp
SalI/SmaI
SalI/SmaI
SmaI BglII
lacZ
ns#2
SmaI
Kan
BamHI
HindIII
HindIII
SalI
Cm
HindIII
SalI
ori
Amp
SalI
ns#1
SalI
BamHI
SmaI BglII
lacZ
ns#2
Kan
BamHI
HindIII
pMUN2
HindIII
Cm
HindIII
SalI
ori
Amp
SmaI BglII
lacZ
ns#2
HindIII
HindIII
Kan
SalI
BamHI
ns#1
pMUN3
HindIII
BamHI
Cm
SalI
ori
Amp

Figure 4.2. Analysis of pMUN3 clones by restriction digestion. Plasmid DNA from two clones was fractionated on a 0.7% agarose gel following digestion with SalI, BamHI or HindIII. Digestion with SalI removed the upstream neutral site flanking sequence (ns#1) as a 1.2 kb DNA fragment, separating it from the 11.6 kb vector fragment. Digestion with BamHI separated the entire neutral-site construct as a 7.9 kb DNA fragment; the remaining vector fragment was 4.9 kb in length. Digestion with HindIII cut pMUN3 into three segments. A 6.8 kb DNA fragment contained part of the *Kan* cassette, the promoterless *lacZ* gene, and the downstream neutral-site (ns#2) sequence. A 5.5 kb DNA fragment contained the pBR322 sequence that was the parental vector for pAM990. A 0.5 kb DNA fragment contained the sequence separating two HindIII sites within the *Kan* cassette was not observed since it ran off the end of the gel. The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left.

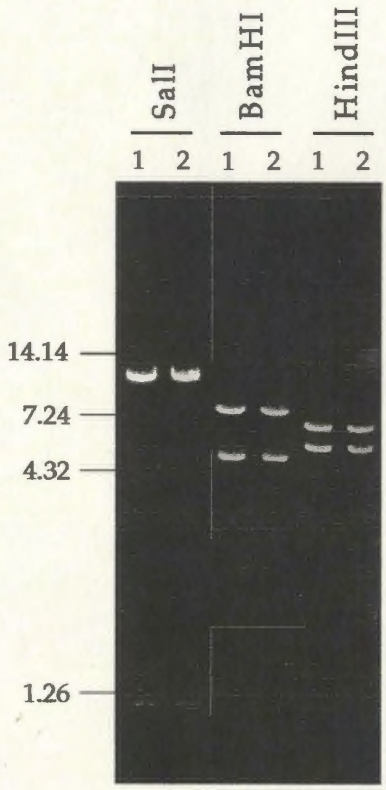
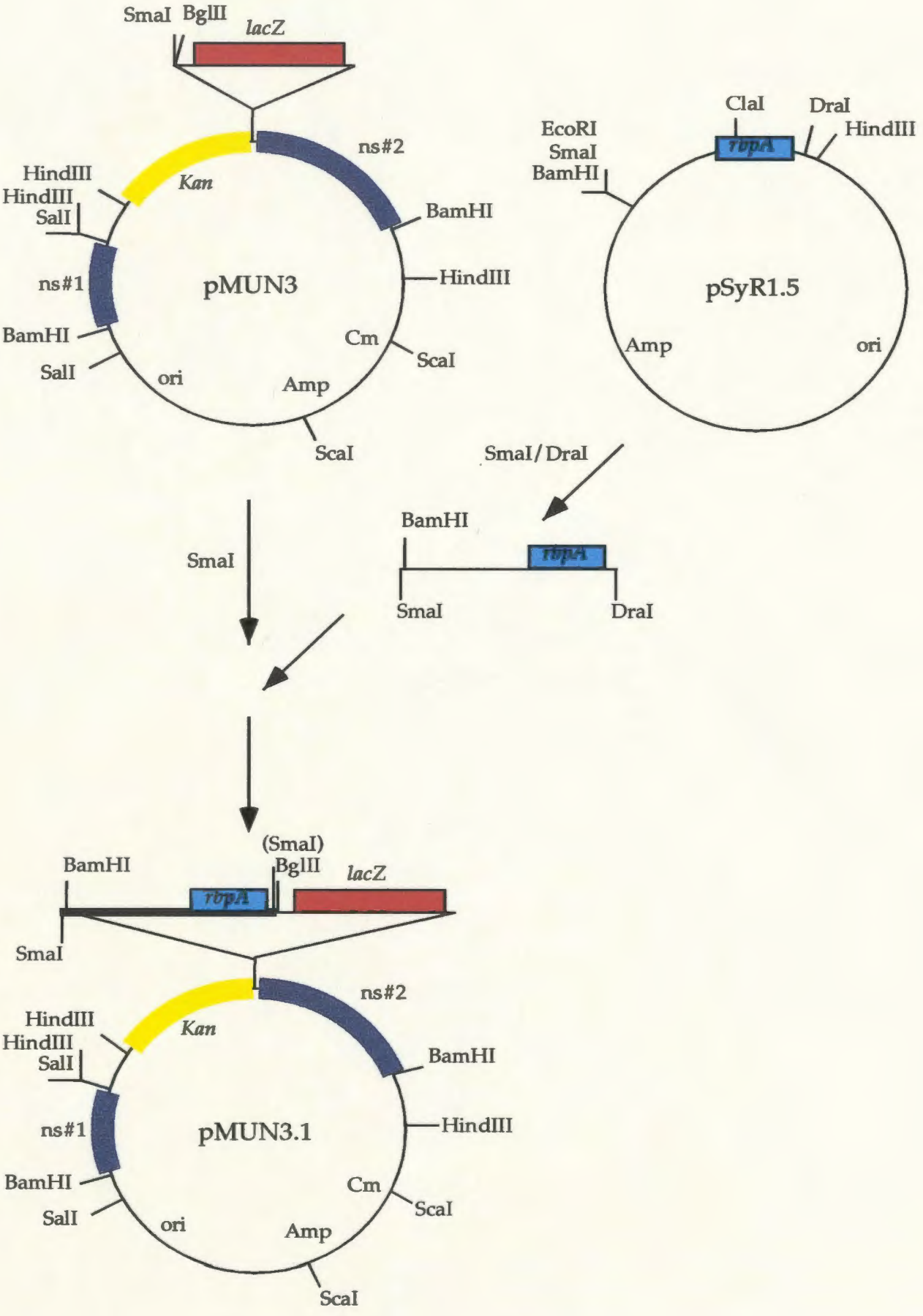

SalI
BamHI
HindIII
1 2 1 2 1 2
14.14
7.24
4.32
1.26

Figure 4.3. Construction of plasmid pMUN3.1. Plasmid pSyR1.5 (see Table 2.1) was digested with SmaI and DraI. The resulting 1.1 kb SmaI/DraI DNA fragment, containing the *rbpA* gene, was purified on a 0.8% agarose gel and ligated into the SmaI site of pMUN3. The resulting plasmid, now 13.9 kb in size, was named pMUN3.1. The location of the 1.1 kb DNA fragment within the plasmid pMUN3.1 is shown as a thick line. Plasmids and DNA fragments are not drawn to scale. Parentheses indicate that a restriction site is no longer completely intact.


SmaI BglII
lacZ
ns#2
HindIII
HindIII
SalI
Kan
BamHI
ns#1
pMUN3
HindIII
BamHI
Cm
ScaI
SalI
ori
Amp
ScaI
ClaI
DraI
EcoRI
SmaI
BamHI
HindIII
tbpA
pSyR1.5
Amp
ori
SmaI/DraI
BamHI
tbpA
SmaI
DraI
SmaI
(SmaI)
BamHI
BglII
lacZ
tbpA
SmaI
ns#2
HindIII
HindIII
SalI
Kan
BamHI
ns#1
pMUN3.1
HindIII
BamHI
Cm
ScaI
SalI
ori
Amp
ScaI

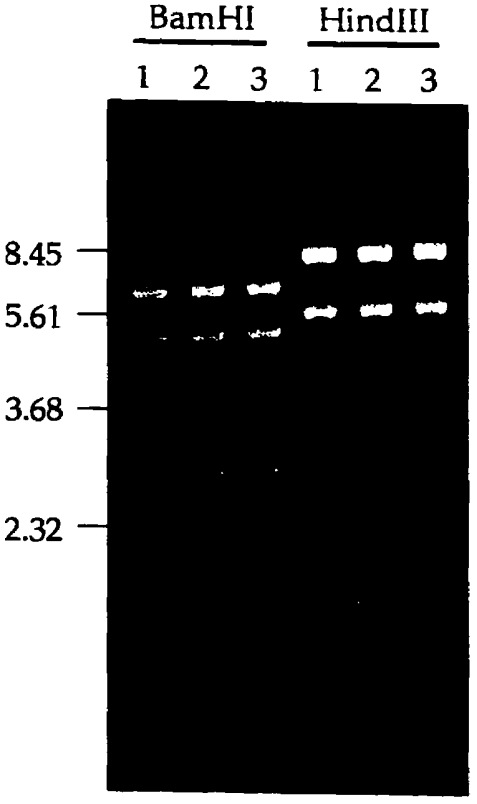
rbpA gene cloned into the SmaI site of pMUN3. The resulting plasmid, now 13.9 kb in size, was named pMUN3.1. Three plasmid clones were screened for the presence of the insertion by digestion with HindIII (**Figure 4.4**). All three clones isolated appeared to have a restriction pattern which indicated insertion of the 1.1 kb DNA fragment. However, it was also desirable to have the *rbpA* gene oriented in the same direction as the promoterless *lacZ* gene in pMUN3.1. This arrangement would be useful in future experiments designed to study *rbpA* expression by measuring β-galactosidase activity in *Synechococcus* 7942 cellular extracts. A BamHI restriction site located adjacent to the SmaI site on the 1.1 kb DNA insert was used for this purpose. Orientation of the inserted 1.1 kb DNA fragment was determined by digestion of the three plasmid clones with BamHI. All three clones appeared to have this "correct" orientation based on the expected restriction map.

Digestion of plasmid pMUN3.1 with ScaI removed the entire construct as a 12.6 kb DNA fragment. This fragment was separated on a 0.7% agarose gel and purified from an agarose gel slice as described previously (Section 2.2.6). The 12.6 kb DNA fragment was used to create the SY-RBPA6 mutant of *Synechococcus* 7942 by transformation of mutant SY-RBPA1 competent cells (Section 4.2.5).

4.2.3. Creation of Control Mutant SY-RBPA7 Using Plasmid pMUN3.

As with the mutagenesis experiments in Chapter 3, it was necessary to examine any effects due to the presence of the inserted *lacZ* gene and *Kan* cassette. A control mutant SY-RBPA7 was therefore created using the original vector pMUN3. Digestion of pMUN3 with ScaI removed a 11.5 kb DNA

Figure 4.4. Analysis of potential pMUN3.1 clones by restriction analysis. Three clones were screened for the presence of the 1.1 kb insert by digestion with HindIII. The presence of the 1.1 kb *rbpA* SmaI/DraI DNA fragment increased the size of the HindIII fragment observed to 7.9 kb from the 6.8 kb fragment observed in HindIII digests of pMUN3 (see Figure 4.2). The 5.5 kb and 0.5 kb (not visible) HindIII digestion products seen with pMUN3 remain unchanged in pMUN3.1. The orientation of the 1.1 kb insert was confirmed by digestion with BamHI. The 1.1 kb blunt-end fragment, containing the *rbpA* gene, contained a single BamHI site adjacent to the SmaI site. Digestion of pMUN3.1 with BamHI divided the construct into three DNA fragments with sizes of 6.2 kb, 4.9 kb and 2.8 kb. The 6.2 kb fragment contained the *rbpA* DNA fragment as well as the promoterless *lacZ* gene and the downstream neutral-site flanking sequence (ns#2). The 4.9 kb fragment contained only vector sequence. The 2.7 kb fragment contained the neutral site flanking sequence (ns#1) as well as the *Kan* cassette. All three clones have the correct restriction pattern based on the expected restriction map. The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left.


BamHI
HindIII
1 2 3 1 2 3
8.45
5.61
3.68
2.32

fragment containing the neutral-site flanking sequence as well as the *lacZ* gene and *Kan* cassette. This fragment was separated on a 0.7% agarose gel and purified from an agarose gel slice as described previously (Section 2.2.6). The 11.5 kb DNA fragment was used to create the SY-RBPA7 mutant of *Synechococcus* 7942 by transformation of mutant SY-RBPA1 competent cells (Section 4.2.5).

4.2.4. Construction of pMUN3.3 for Mutant SY-RBPA8

Plasmid pMUN3.1, used for creation of mutant SY-RBPA6, contained only 40 bp of *Synechococcus* 7942 DNA downstream of *rbpA*. It therefore did not contain the downstream palindromic sequence capable of forming a stem-loop in RNA (see Figure 2.5). Initial observations of the SY-RBPA6 growth phenotype suggested that the inserted 1.1 kb DNA fragment did not rescue the SY-RBPA1 mutant phenotype. It was therefore decided to create another mutant which incorporated a larger DNA fragment into the neutral-site of the chromosome.

Construction of this new neutral-site vector, named pMUN3.3, occured in two stages. The first stage is outlined in **Figure 4.5**. The effective result of this stage was the addition of a BamHI site to the end of the 351 bp downstream DNA fragment.

The next step was the re-introduction of the upstream sequence containing the *rbpA* gene. The 1.4 kb HindIII fragment isolated from pSyR1.3 was cloned into the single HindIII site in pSyR1.6. The orientation of this inserted *rbpA*-containing DNA fragment was determined by digestion with BamHI (**Figure 4.6**). Six clones were screened for the correct orientation of the

Figure 4.5. Construction of plasmid pTJB2. Digestion of plasmid pSyR1.3 with both HindIII and HincII produced three fragments when separated on a 0.8% agarose gel. Two of these DNA fragments were isolated from an agarose gel slice as outlined previously (Section 2.2.6). A 1.4 kb HindIII fragment was isolated which represented the SyR1.2 region containing the *rbpA* gene. A second fragment, 351 bp in length, was isolated which contained the sequence extending from the HindIII site downstream of *rbpA* to the HincII site located 351 bp further downstream of the gene. This latter fragment was first cloned into pUC19 to generate pSyR1.6. The 1.4 kb fragment was cloned into the HindIII site of pSyR1.6; the orientation of the insertion was determined by screening with BamHI and HindIII (see Figure 4.6). Digestion of this new plasmid, named pTJB2, removed a 1.6 kb DNA fragment containing the *rbpA* gene. The location of downstream sequence capable of forming secondary structure in RNA is indicated by a stem-loop structure. Plasmids and DNA fragments are not drawn to scale.

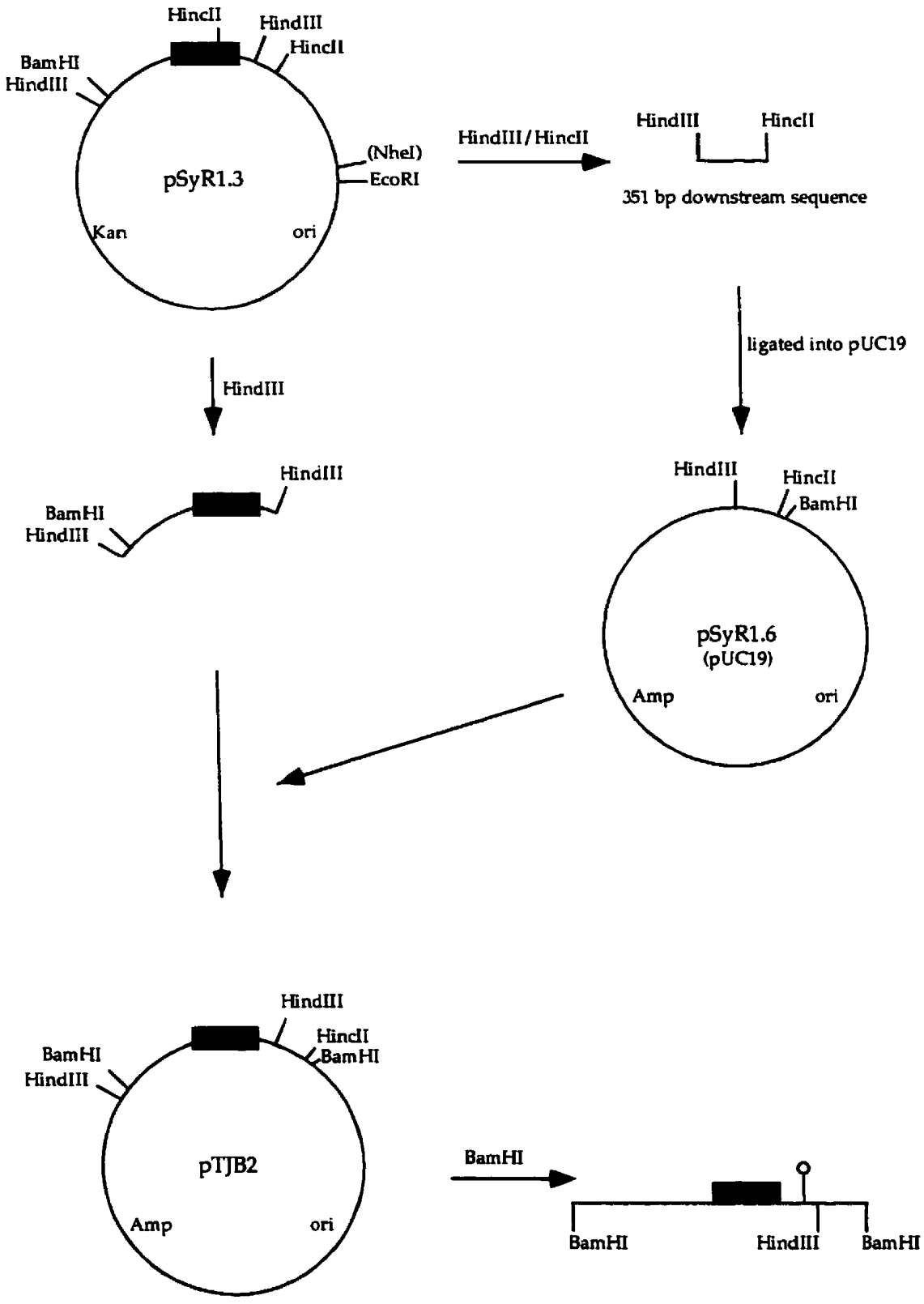
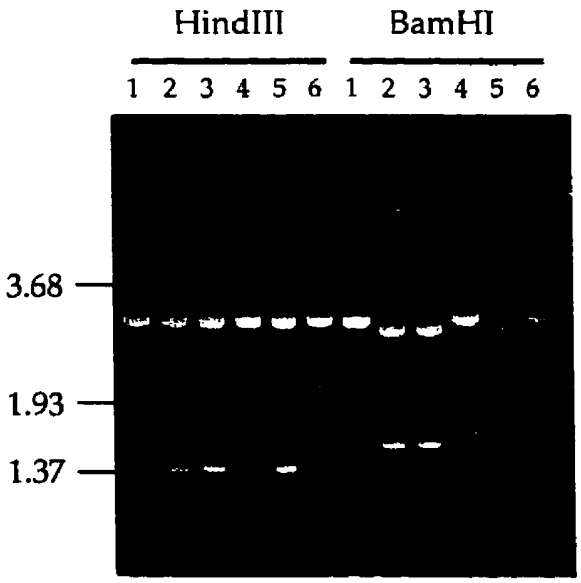

HincII
HindIII
HincII
BamHI
HindIII
(NheI)
EcoRI
pSyR1.3
Kan
ori
HindIII / HincII
HindIII
HincII
351 bp downstream sequence
ligated into pUC19
HindIII
HindIII
BamHI
HindIII
HindIII
HincII
BamHI
pSyR1.6
(pUC19)
Amp
ori
HindIII
HincII
BamHI
BamHI
HindIII
pTJB2
Amp
ori
BamHI
BamHI
HindIII
BamHI

Figure 4.6. Analysis of pTJB2 clones by restriction analysis. Six plasmid clones were screened for the presence of the 1.4 kb SyR1.2 fragment by digestion with HindIII (lanes 1-6) (see Figure 4.5). The orientation of the SyR1.2 fragment was determined by digestion with BamHI (lanes 7-12). Plasmid clones containing the correct insert orientation produced a 1.6 kb DNA fragment which represented the SyR1.5 fragment (see Table 2.1) plus the 351 bp sequence originally cloned in pSyR1.6. Clones #2 and #3 appear to contain the correct restriction pattern based on the restriction map of pTJB2. The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left.


HindIII
BamHI
1 2 3 4 5 6 1 2 3 4 5 6
3.68
1.93
1.37

insert, two of which appeared to have the correct restriction pattern based on the map outlined in Figure 4.5. One of these (clone #2) was carried forward to stage two of the construction. This intermediate construct was named pTJB2. It contained a 1.6 kb DNA fragment with BamHI ends which could subsequently be cloned into the neutral-site vector pMUN3. This 1.6 kb fragment contained 438 bp of *Synechococcus* 7942 genomic DNA downstream of the *rbpA* gene.

Stage two of this construction is outlined in **Figure 4.7**. The direct insertion of the 1.6 kb DNA fragment into the BglII site of pMUN3 was complicated by the discovery of a second BglII site within the *Kan* cassette sequence of pMUN3. A different strategy was therefore implemented utilizing the original neutral site vector pAM990 which contained only a single BglII site. The first step involved cloning the 1.6 kb BamHI DNA fragment containing the *rbpA* gene from stage one into the unique BglII site of pAM990. Four potential positive clones were screened for the presence of the inserted fragment by digestion with HindIII (**Figure 4.8A**). Conveniently, the DNA fragments produced by this restriction digest could also be used to confirm the orientation of the insert since the 1.6 kb BamHI fragment contained a HindIII site located immediately downstream of the *rbpA* gene. Of the four clones screened, all contained the 1.6 kb BamHI fragment; three of these (clones #1, 3 and 4) appeared to be in the correct orientation. Clone #2 contained the 1.6 kb insert in the opposite orientation. Clone #1 was carried forward to the next step. This plasmid was named pTJB3.

The next step of the construction involved replacing the omega cassette with the kanamycin resistance (*Kan*) cassette used in the previous constructs

Figure 4.7. Construction of pMUN3.3 from the intermediate construct pTJB2. The 1.6 kb BamHI fragment from pTJB2 was cloned into the BglII site of pAM990. Digestion of this new plasmid, named pTJB3, with SalI and SmaI released the omega cassette as well as the 1.2 kb upstream neutral-site flanking sequence (ns#1). These fragments were replaced with the 1.8 kb kanamycin resistance cassette (*Kan*) originally isolated from pSKS101 (see Figure 4.1). The 1.2 kb fragment containing upstream neutral-site flanking sequence (ns#1) was then re-inserted into the unique SalI site. The correct orientation of this insert was confirmed by digestion with HindIII (see Figure 4.8B). The new plasmid, named pMUN3.3, was 14.4 kb in size. Plasmids and DNA fragments are not drawn to scale. Brackets indicate that a restriction site is no longer completely intact.

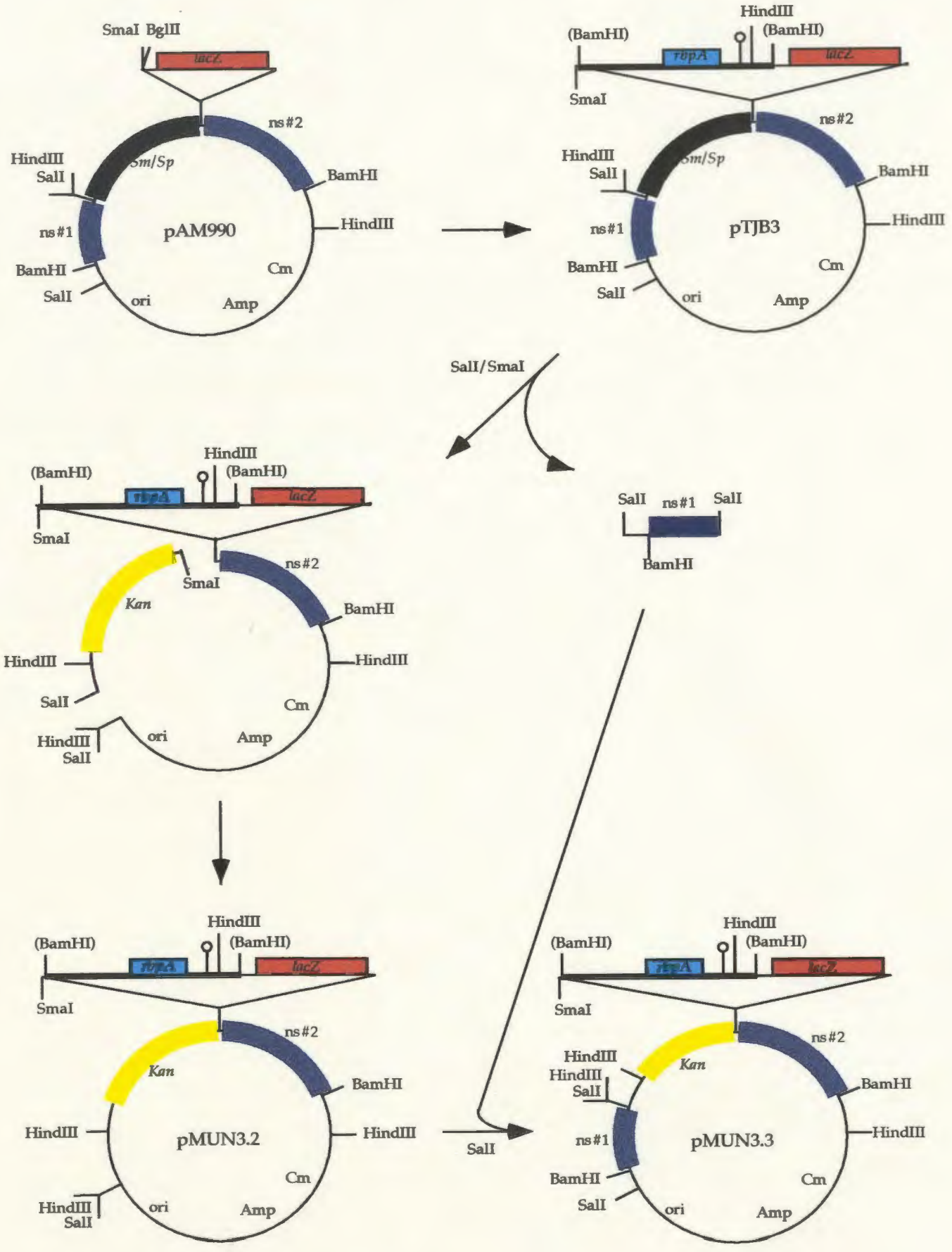
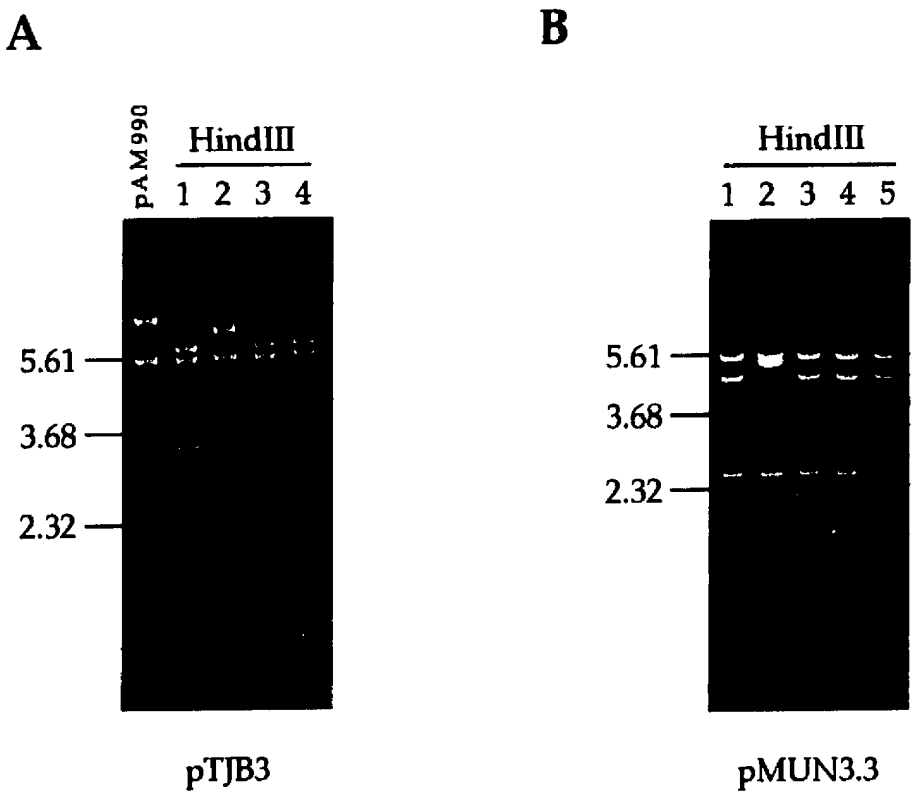

SmaI BglII
lacZ
ns#2
HindIII
SalI
Sm/Sp
BamHI
ns#1
pAM990
HindIII
BamHI
Cm
SalI
ori
Amp
HindIII
(BamHI)
(BamHI)
tetA
lacZ
SmaI
ns#2
HindIII
SalI
Sm/Sp
BamHI
ns#1
pTJB3
HindIII
BamHI
Cm
SalI
ori
Amp
SalI/SmaI
SalI
ns#1
SalI
BamHI
HindIII
(BamHI)
(BamHI)
tetA
lacZ
SmaI
SmaI
ns#2
Kan
BamHI
HindIII
HindIII
SalI
Cm
HindIII
SalI
ori
Amp
HindIII
(BamHI)
(BamHI)
tetA
lacZ
SmaI
ns#2
Kan
BamHI
HindIII
pMUN3.2
HindIII
Cm
HindIII
SalI
ori
Amp
SalI
HindIII
(BamHI)
(BamHI)
tetA
lacZ
SmaI
ns#2
HindIII
HindIII
SalI
Kan
BamHI
ns#1
pMUN3.3
HindIII
BamHI
Cm
SalI
ori
Amp

Figure 4.8. Restriction analysis of pTJB3 and pMUN3.3 plasmid clones. In each case, plasmid DNA was digested with HindIII and the resulting DNA fragments separated on a 0.7% agarose gel. **(A).** Digestion of pTJB3 with HindIII would cut the plasmid into three fragments of sizes 5.8 kb, 5.5 kb and 3.3 kb, based on the restriction map in Figure 4.7. The 5.8 kb DNA fragment included a small portion of the 1.6 kb insert, the promoterless *lacZ* gene and the downstream neutral-site flanking sequence (ns#2). The 5.5 kb DNA fragment contained the pBR322 sequence along with the upstream neutral-site sequence (ns#1). Finally, the 3.3 kb DNA fragment contained the omega cassette plus most of the 1.6 kb insert which contained the *rbpA* gene. Three of the clones (clones #1, #3 and #4) appeared to contain the 1.6 kb BamHI DNA fragment in the correct orientation. Plasmid pAM990 DNA was included for the purpose of comparison. **(B).** Digestion of pMUN3.3 with HindIII cut the plasmid into four fragments of sizes 5.8 kb, 5.5 kb, 2.6 kb and 0.5 kb. The 5.5 kb DNA fragment contained the pBR322 sequence as well as the upstream neutral-site sequence (ns#1). The 0.5 kb DNA fragment contained the sequence within the *Kan* cassette which separated the two HindIII sites. The 2.6 kb fragment contained most of the *Kan* cassette along with most of the 1.6 kb insert containing the *rbpA* gene. Finally, the 5.8 kb fragment contained a small portion of the inserted sequence, all of the promoterless *lacZ* gene and all of the downstream neutral-site flanking sequence (ns#2). Clone #2 had the correct restriction pattern based on the map shown in Figure 4.7. The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left.


A
pAM990
HindIII
1 2 3 4
5.61
3.68
2.32
pTJB3
B
HindIII
1 2 3 4 5
5.61
3.68
2.32
pMUN3.3

(see Figure 4.1). To do this, the plasmid pTJB3 was digested with SalI and SmaI. Three digestion products were separated on an 0.7% agarose gel. The 1.2 kb DNA fragment containing the upstream neutral-site sequence (ns#1) was isolated, as was the 11.4 kb fragment which represented the remainder of the construct. The 2 kb fragment containing the omega cassette was discarded. Ligation of the 11.4 kb vector fragment with the 1.8 kb *Kan* cassette from pSKS101 produced the intermediate pMUN3.2. In the final step, the 1.2 kb fragment which contained the upstream neutral-site flanking sequence (ns#1) was inserted into the unique SalI site of pMUN3.2. A total of twenty clones were screened for the presence of this insertion in the correct orientation. Five of these are shown in **Figure 4.8B**. Only one (clone #2) had the correct restriction pattern. The total size of the new plasmid, named pMUN3.3, was 14.4 kb. Further characterization using the restriction enzymes BamHI, EcoRV and SalI all gave the correct restriction patterns as determined from known sequence information (data not shown). The entire neutral-site region was subsequently isolated from pMUN3.3 by digestion with BamHI. This 9.5 kb DNA fragment was used to create the mutant SY-RBPA8 by transformation of mutant SY-RBPA1 competent cells (Section 4.2.5).

4.2.5. Transformation of Mutant SY-RBPA1 and Selection of Double Resistant Phenotypes

Preparation of *Synechococcus* 7942 mutant SY-RBPA1 competent cells was identical to that described previously (Section 3.2.3). Following transformation, cells were plated onto solid BG-11 media containing both 40 µg/mL spectinomycin and 20 µg/mL kanamycin. All other growth conditions were as described previously (Section 3.2.5). Transformants

capable of growing on both antibiotics appeared as single colonies within two weeks. These were re-streaked onto solid BG-11 media supplemented with both antibiotics and single clones used to inoculate 100 mL liquid BG-11 cultures.

4.3. RESULTS AND DISCUSSION

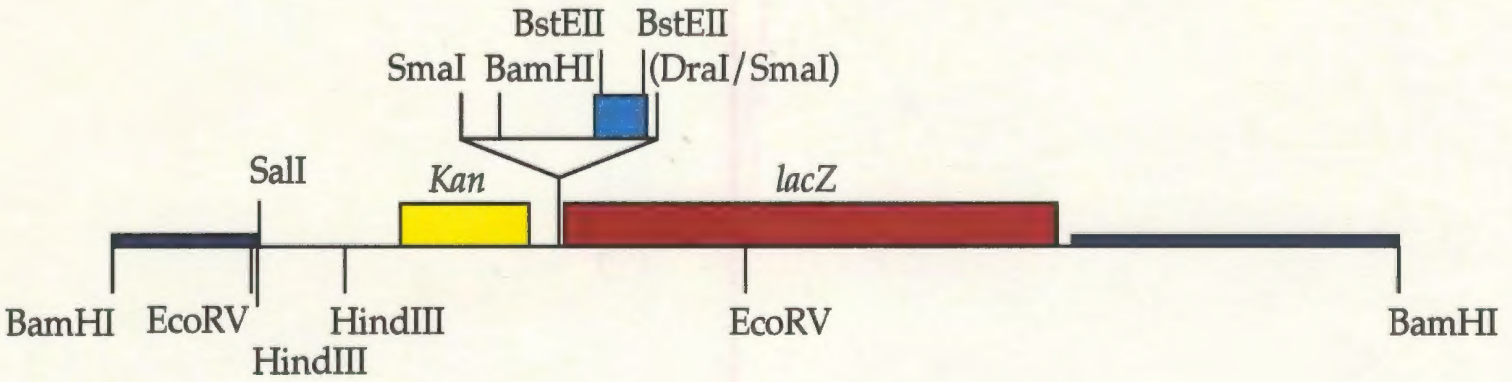
4.3.1. Confirmation of SY-RBPA6, SY-RBPA7 and SY-RBPA8 Mutant Genotypes

The first complementation experiment attempted to rescue the mutant phenotype of SY-RBPA1 by insertion of a 1.1 kb SmaI/DraI DNA fragment, containing the *rbpA* gene, into the neutral site in the cyanobacterial genome. The resulting mutant, now resistant to both spectinomycin and kanamycin, was named SY-RBPA6. As a control, a second mutant, named SY-RBPA7, was constructed. This strain also contained the *Kan* cassette and *lacZ* gene integrated into the neutral site but did not contain an intact *rbpA* gene. The effective result was that mutant SY-RBPA7 was identical to mutant SY-RBPA6 except for the presence of an intact *rbpA* gene. The restriction maps for these two strains in the region of the "neutral-site" are shown in **Figure 4.9**.

The genotypes of both of these strains was confirmed by Southern blot hybridization (**Figure 4.10**). When genomic DNA isolated from SY-RBPA6 and SY-RBPA7 was digested with BstEII was probed with the 290 bp BstEII DNA fragment containing a portion of the *rbpA* gene, the sizes of hybridizing fragments matched what was expected from the restriction maps for the two strains. The 1.3 kb and 1.1 kb hybridizing fragments were observed in both strains, just as they were observed in the knock-out mutant SY-RBPA1 (see Figure 3.6). This indicated that the original *rbpA* gene still contained the omega cassette inserted at the ClaI site. The 290 bp hybridizing fragment which was indicative of an intact *rbpA* gene was only observed in the lane

Figure 4.9. Restriction maps of neutral-site region in two strains of *Synechococcus* 7942. **(A).** In mutant SY-RBPA6, a 1.1 kb DNA fragment containing the *rbpA* gene (light blue) was inserted into the chromosomal neutral site (dark blue), along with a promoterless *lacZ* gene (red) and a gene for kanamycin resistance (yellow). **(B).** In mutant SY-RBPA7, only the *lacZ* and kanamycin resistance genes were inserted into the neutral-site. Only relevant restriction sites are shown.

A. SY-RBPA6



B. SY-RBPA7

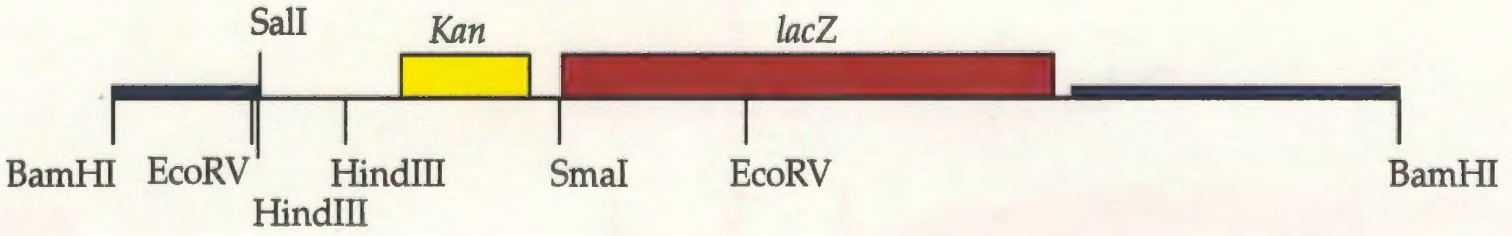
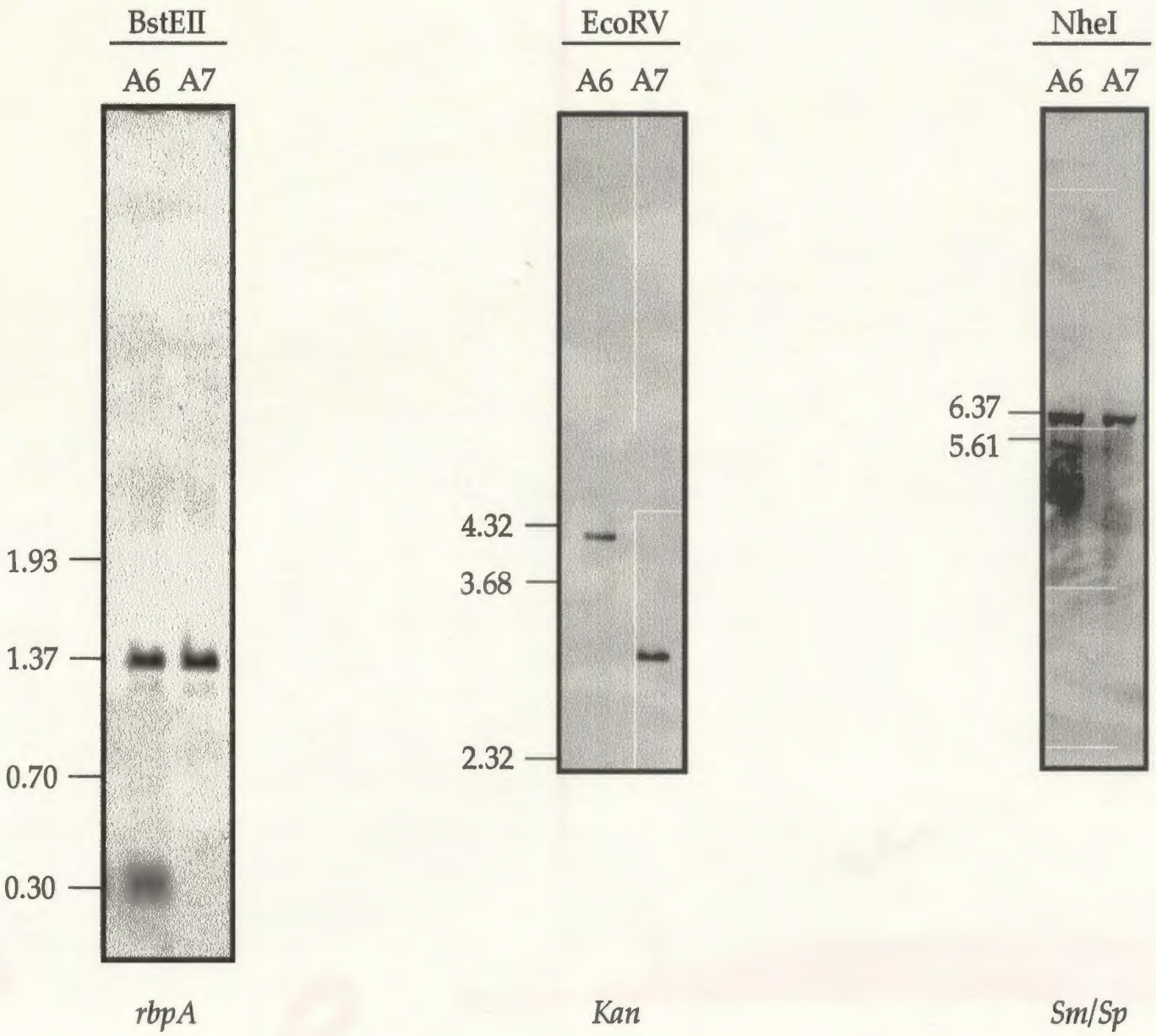


Figure 4.10. Analysis of neutral-site mutant *Synechococcus* 7942 genotypes by Southern blot hybridization. Genomic DNA from strain SY-RBPA6 (A6) and from strain SY-RBPA7 (A7) was digested with BstEII, EcoRV or NheI. Digoxygenin-labelled DNA probes used in each hybridization analysis consisted of a 290 bp BstEII DNA fragment containing sequence totally internal to the *rbpA* gene (*rbpA*), a 1.2 kb SalI/SmaI DNA fragment containing the gene for kanamycin resistance derived from pSKS101 (*Kan*) and a 2 kb HindIII DNA fragment containing the entire omega cassette from pDW9 (*Sm/Sp*). The location of DNA fragments derived from BstEII-digested bacteriophage λ DNA as well as a 100 bp DNA ladder (Gibco BRL) used as size markers are indicated at the left of each blot. Only size markers in the size range of the hybridization signals are shown.


BstEII
A6 A7
1.93
1.37
0.70
0.30
rbpA
EcoRV
A6 A7
4.32
3.68
2.32
Kan
NheI
A6 A7
6.37
5.61
Sm/Sp

containing genomic DNA from SY-RBPA6. The presence of this fragment confirmed the presence of an intact *rbpA* gene at the neutral site of SY-RBPA6.

A second Southern blot confirmed the genotype of these two strains using the *Kan* cassette probe derived from the plasmid pSKS101 (**Figure 4.10**). The restriction maps for these two strains showed two EcoRV restriction sites, one within the upstream neutral-site (ns#1) flanking sequence and one within the coding sequence of the *lacZ* gene (see Figure 4.9). Mutant SY-RBPA7 genomic DNA digested with EcoRV and probed with the *Kan* cassette probe revealed a 3.0 kb hybridizing fragment matching the size expected from the restriction map shown in Figure 4.9. In SY-RBPA6, the insertion of the 1.1 kb DNA fragment containing the *rbpA* gene increased the size of the hybridizing EcoRV fragment to 4.1 kb. No hybridizing fragments were observed in any of the other strains using the *Kan* cassette probe (data not shown).

A third Southern blot (**Figure 4.10**) was used to confirm that the recombination events used in the complementation experiments had no effect on the inactivated *rbpA* gene or the omega cassette. Genomic DNA from both strains was digested with NheI and probed with *Sm/Sp* cassette probe. A 6.1 kb hybridizing fragment was observed in both strains. This size was identical to that observed for the mutant SY-RBPA1 (see Figures 3.5 and 3.6C) and confirmed that the original *rbpA* gene still contained the omega cassette.

In the second complementation experiment, the mutant SY-RBPA8 was created. In this strain, a 1.4 kb DNA fragment which included the *rbpA*

gene as well as the palindromic sequence downstream of *rbpA* was recombined into the neutral site of the SY-RBPA1 genome. The restriction map for this strain in the region of the "neutral site" is shown in **Figure 4.11**. The genotype of this strain was confirmed by Southern blot hybridization as with the previous two strains (**Figure 4.12**). Genomic DNA digested with BstEII and probed with the 290 bp BstEII DNA fragment containing a portion of the *rbpA* gene. As with mutant SY-RBPA6, the 290 bp hybridizing fragment which was indicative of an intact *rbpA* gene was observed in the lane containing genomic DNA from SY-RBPA8. The presence of this fragment confirmed the presence of an intact *rbpA* gene at the neutral site of SY-RBPA8.

4.3.2. Characterization of SY-RBPA6, SY-RBPA7 and SY-RBPA8 Mutant Phenotypes

Stock cultures of strains SY-RBPA6 and SY-RBPA8 possessed the same blue-green colour as observed in the wild type organism (not shown). The strain SY-RBPA7, which did not contain a copy of *rbpA* at the neutral site, possessed the bleached appearance as was observed in the knock-out phenotype SY-RBPA1 from which it was derived. Unfortunately, the SY-RBPA7 strain was lost shortly after these initial observations were carried out and so further analysis of SY-RBPA7 was not possible. Measurements of whole cell absorbance spectra using the previously outlined method (Section 3.2.5) with the neutral-site strains in the range of 400 nm to 800 nm indicated that both neutral site strain SY-RBPA6 and SY-RBPA8 showed an increase in the ratio of phycobiliprotein (630 nm) to chlorophyll a (680 nm)

Figure 4.11. Restriction map of neutral-site region in mutant SY-RBPA8. A 1.4 kb DNA fragment containing the *rbpA* gene (blue) plus the downstream palindromic sequence was inserted into the chromosomal neutral site at the BglII restriction site. Also shown are the location of the kanamycin resistance gene (yellow) and the *lacZ* gene (red). Only the relevant restriction sites are shown.

SY-RBPA8

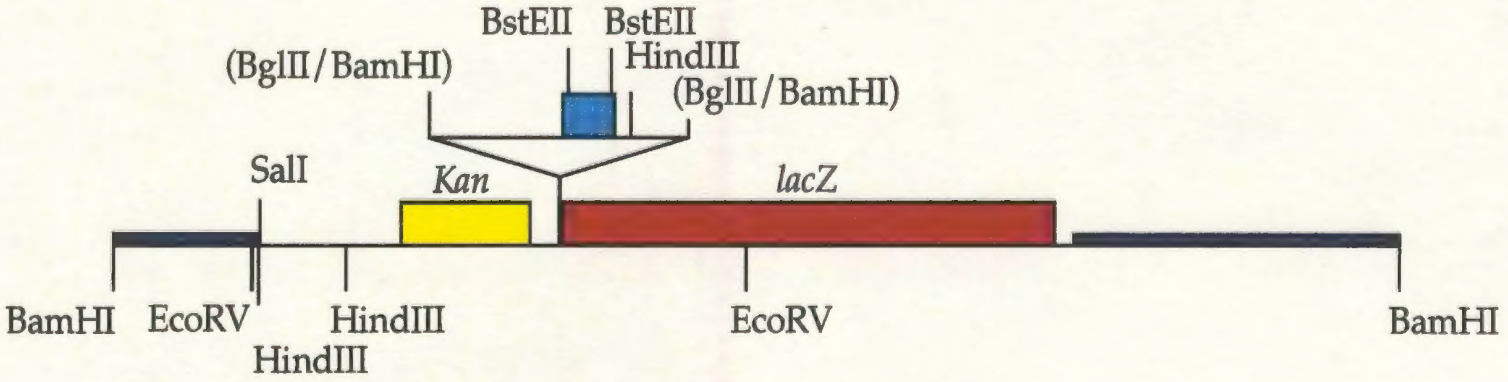
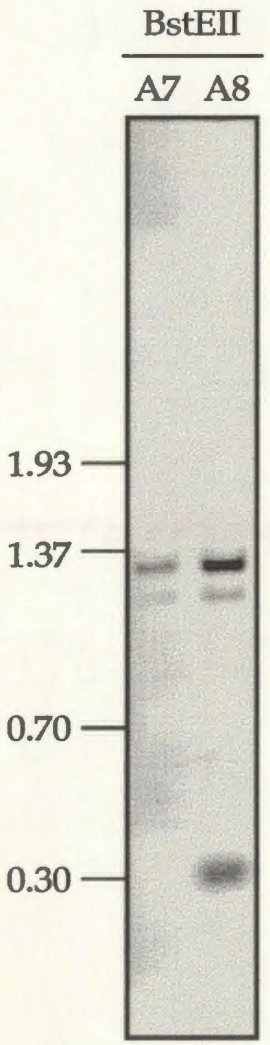

500 bp
BstEII
BstEII
HindIII
(BglII/BamHI)
(BglII/BamHI)
SalI
Kan
lacZ
BamHI
EcoRV
HindIII
HindIII
EcoRV
BamHI

Figure 4.12. Analysis of *Synechococcus* mutant SY-RBPA8 genotype by Southern hybridization. Genomic DNA from strains SY-RBPA8 (A8) and SY-RBPA7 (A7) was digested with BstEII and probed with the *rbpA* probe described previously (see Figure 4.10). The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are indicated at the left.



rbpA

absorbance peaks relative to the knock-strain SY-RBPA1 (**Figure 4.13**). Strain SY-RBPA6 had an absorbance peak ratio of 1.05; strain SY-RBPA8 had an absorbance peak ratio of 1.08. These results indicated that the reduced phycocyanin content observed in the knock-out phenotype SY-RBPA1 was not observed in those strains which contained an intact *rbpA* gene at the neutral site (SY-RBPA6 and SY-RBPA8). The blue-green colour observed in stock cultures of SY-RBPA6 and SY-RBPA8 further supported the idea that the increase in the ratio of these absorbance peaks was due to an increase in the amount of phycobiliproteins present in these strains.

Growth of the two neutral-site strains (SY-RBPA6 and SY-RBPA8) in liquid BG-11 growth media was monitored, as with the previous three strains, under controlled conditions of temperature and illumination (**Figure 4.14**). Both strains appeared to grow at an incubation temperature of 30°C (approximate doubling time 2 days), although apparently not as well as the wild type organism. At an incubation temperature of 21°C, strains SY-RBPA6 and SY-RBPA8 were able to grow at an rate comparable to the wild type organism (approximate doubling time 2.5 days).

Overall, the results suggested that the re-introduction of an intact *rbpA* gene into the knock-out strain SY-RBPA1 was sufficient to restore this strain to a wild type phenotype. Both the cold-sensitivity of the knock-out strain and its bleached appearance were eliminated by the introduction of an intact *rbpA* gene in mutants SY-RBPA6 and SY-RBPA8. Furthermore, the absence of the stem-loop structure downstream of *rbpA* appeared to have little effect on the overall phenotype of the strain. This suggested that while this stem-

Figure 4.13. Representative whole cell spectra of SY-RBPA1 (green), SY-RBPA6 (blue) and SY-RBPA8 (purple) *Synechococcus* 7942 strains. As before, cells were diluted with BG-11 media to an absorbance (800 nm) of 0.2 prior to spectral measurement (see Figure 3.9).

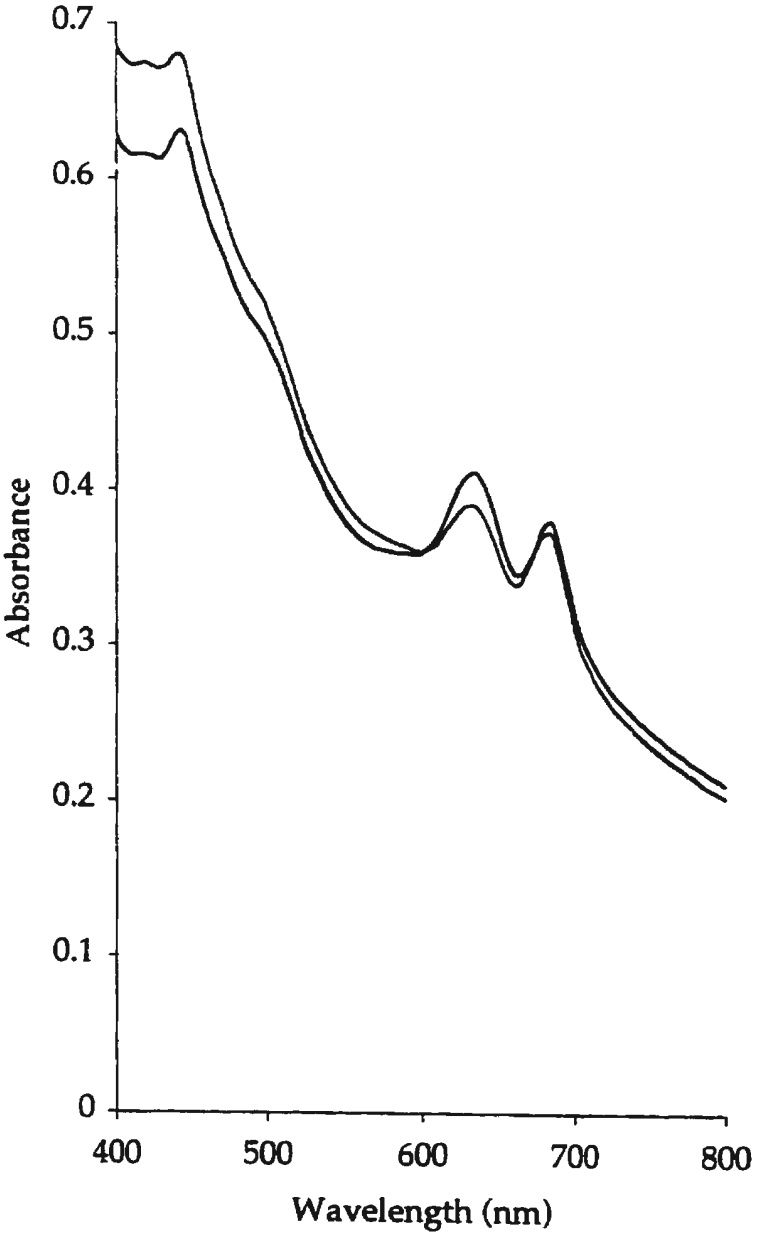
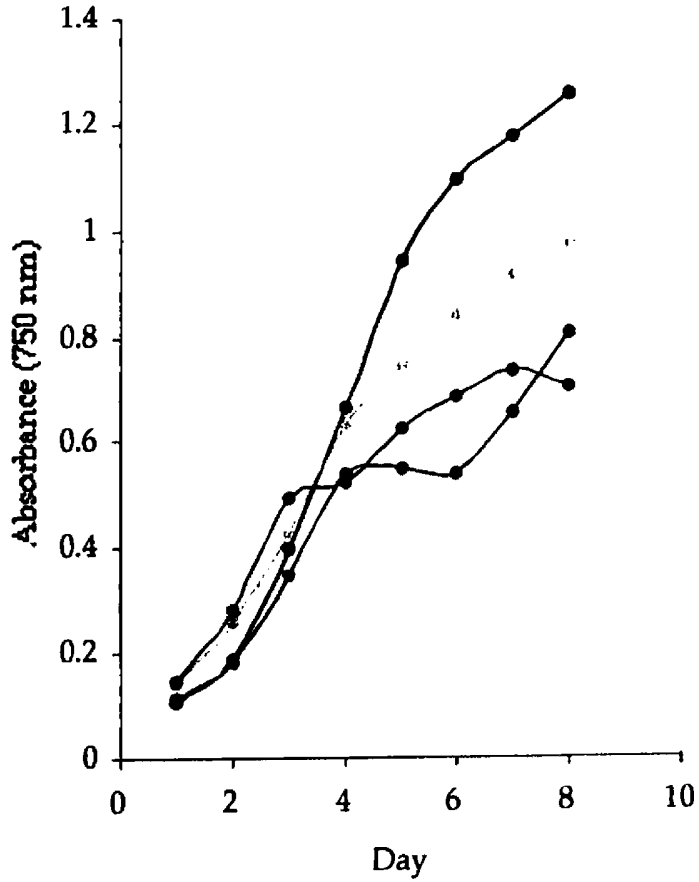
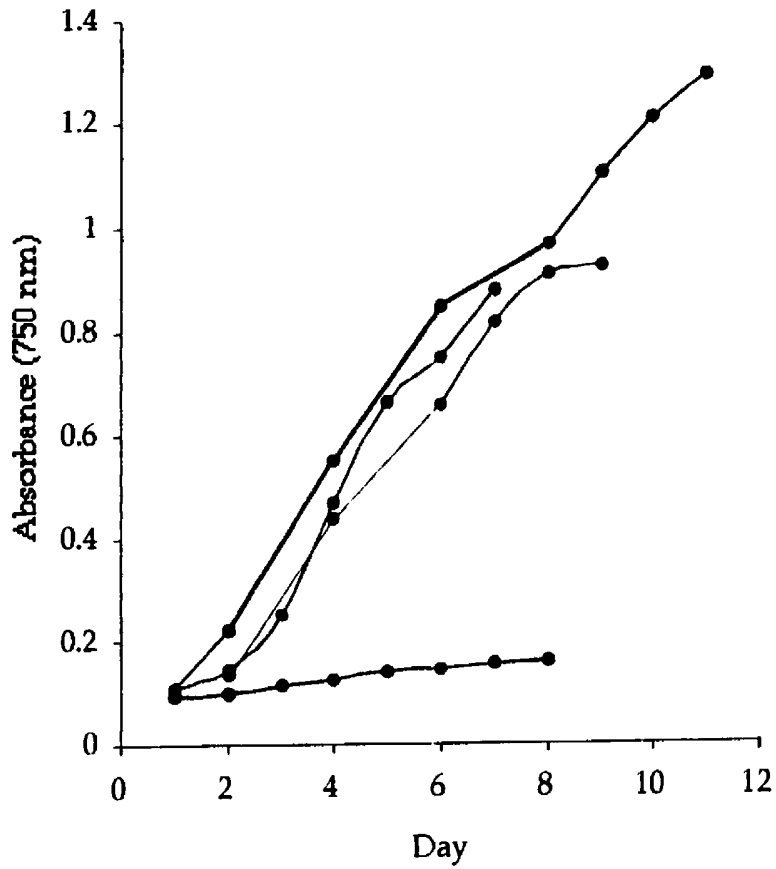

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance
400
500
600
700
800
Wavelength (nm)

Figure 4.14. Growth of wild type (red), SY-RBPA1 (green), SY-RBPA6 (blue) and SY-RBPA8 (purple) *Synechococcus* 7942 strains in BG-11 growth media. Cultures (100 mL) were incubated with shaking (200 rpm) under constant illumination at **(A)** 30°C and **(B)** 21°C. The density of each culture was monitored by measuring light scattering at 750 nm. Each data point represents the average of absorbances taken for two cyanobacterial cultures grown in parallel under identical conditions.

A.


Absorbance (750 nm)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
Day

B.


Absorbance (750 nm)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
10
12
Day

loop structure may function as a transcription terminator for *rbpA*, it does not appear to be a point at which expression of the gene is regulated.

CHAPTER 5. EXPRESSION, PURIFICATION AND RNA-BINDING STUDIES OF H_6RbpA AND DELETION PRODUCT H_6RbpA(ΔGly)

5.1. INTRODUCTION

5.1.1. Affinity Purification of Histidine-Tagged Proteins by Metal Chelate Affinity Chromatography (MCAC)

Affinity chromatography using immobilized metal ions has become an established method for the purification of proteins. First described in 1975, it took advantage of the fact that some surface amino acids, acting as electron donors, allow a protein to bind reversibly to metal ions which were immobilized on a support matrix (Porath *et al.*, 1975). The first experiments used the tridentate chelating ligand iminodiacetic acid (IDA) complexed with zinc (Zn^{2+}) or nickel (Ni^{2+}) ions to purify proteins (reviewed in Sulkowski, 1989). However, the ligand bound to three sites within the coordination sphere of a divalent metal ion (Hochuli *et al.*, 1987). In the case of nickel (coordination number of six), that left three coordination sites available for interaction with the protein of interest. The result was that the matrix did not bind metal ions tightly, allowing them to be easily washed off during protein loading and washing steps.

An improvement on this ligand came with the development of the quadridentate nitrilotriacetic acid (NTA) adsorbant (Hochuli *et al.*, 1987). The NTA ligand had four sites capable of interaction with the metal ion, resulting in an interaction which was stable over a wide variety of buffer conditions. Overall, this chelating resin was demonstrated to bind histidine-tagged proteins 1000X more tightly than IDA allowing for more stringent conditions for purification and less contamination due to non-specific protein interactions (Hochuli, 1989). Proteins containing histidine tags bind to a Ni-

NTA resin with a K_d of 10^{-13} M at a pH of 8.0, an affinity greater than that observed between antibody and antigen (Hoffman and Roeder, 1991; Qiagen, 1995).

Removal of a histidine-tagged protein from a Ni-NTA resin is accomplished in one of two ways. First, a protein may be eluted by using buffers of decreasing pH. The pK_a of the imidazole group of the histidine residue is approximately 5.97. Therefore, dropping the pH of the buffer below this value causes protonation of the histidine residues resulting in dissociation of the protein from the Ni-NTA resin (Hochuli *et al.*, 1988). However, one disadvantage of this method comes from the fact that some proteins are adversely affected by sudden alterations in pH or prolonged exposure to conditions of low pH. An alternative procedure uses increasing concentrations of imidazole to compete with the protein's histidine tag for binding to the Ni-NTA moiety. This technique offers the added advantage of minimizing non-specific protein contaminants by including low concentrations of imidazole during the loading and washing steps (Janknecht *et al.*, 1991).

5.1.2. Site-Directed Mutagenesis Using the Altered Sites II *In Vitro* Mutagenesis System

The ability to specifically modify particular amino acids in a protein sequence has contributed greatly to the understanding of protein structure and function. In recent years, oligonucleotide-directed mutagenesis has become the established technique for studying the role of particular amino acids in the biological activity of a protein (reviewed in Smith, 1985).

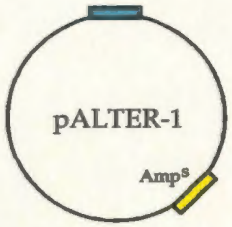
However, limiting factors such as the processivity of the DNA polymerase involved and correction of the oligonucleotide mismatch by the host strain's mismatch repair system have usually resulted in a mutation frequency of only 1-2%.

I chose to carry out site-directed mutagenesis experiments using the Altered Sites II *in vitro* mutagenesis system (Promega). This system was an improvement over previous systems in that it used antibiotic selection as a means to obtain a high frequency of mutants **(Figure 5.1)**. The pALTER-1 vector contains genes for both tetracycline and ampicillin resistance, but the ampicillin resistance gene has been inactivated by introduction of a frameshift mutation. Mutagenesis is carried out by hybridizing recombinant pALTER-1 single stranded DNA (ssDNA) to a mutagenic oligonucleotide which is complementary to the single-stranded template except for a region of mismatch corresponding to the desired mutation. Simultaneously, a second oligonucleotide is annealed to the ssDNA template which repairs the ampicillin resistance gene and therefore restores ampicillin resistance to the mutant strain. Overall, selection of transformants with ampicillin results in a high percentage of mutants (80-90%).

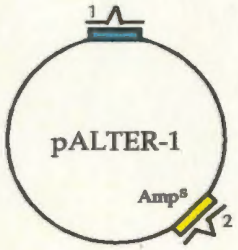
5.1.3. Objectives of This Study

1. To express a cyanobacterial RNA-binding protein (**RbpA**) in *E. coli* and subsequently purify it from total cellular extracts by metal chelate affinity chromatography using Ni^{2+}-NTA (Qiagen). The coding sequence for the *rbpA* gene was altered in order to introduce six histidine residues at the N-

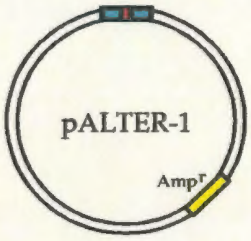
Figure 5.1. Schematic diagram of the Altered Sites II *in vitro* mutagenesis procedure. The system uses antibiotic selection as a means of increasing the frequency of mutants. The plasmid pALTER-1 contains an ampicillin gene (yellow box) which has been inactivated by a frameshift mutation. In the first step, ssDNA corresponding to the minus (-) strand of a recombinant pALTER-1 plasmid is hybridized with two oligonucleotides: **1).** the mutagenic oligonucleotide introduces a sequence change in the cloned DNA fragment (green box), and **2).** the ampicillin repair oligonucleotide restores the correct coding sequence for the ampicillin resistance gene. Both oligonucleotides are subsequently extended with T4 DNA polymerase to obtain a double stranded structure. DNA nicks are sealed with T4 DNA ligase and the resulting duplexes are transformed into *E. coli* ES1301. This strain is used for transformation since its mismatch repair activity has been abolished. Selection using ampicillin greatly increases the frequency of mutants.



Anneal oligonucleotides:
1. mutagenic oligonucleotide
2. Ampicillin repair oligonucleotide



Mutant strand synthesis with
T4 DNA polymerase and ligation
with T4 DNA ligase



Transform into *E. coli* ES1301 and grow
overnight with ampicillin selection

Prepare plasmid DNA

Transform into *E. coli* JM109

Screen clones for presence of mutation

terminal end of the protein. X-ray crystallographic data of other RRM domains showed that this location should be a solvent exposed region in which the histidine tag could be easily accessible (Oubridge *et al.*, 1994). The modified protein was referred to as **H_6RbpA**.

2. To assess the RNA binding capabilities of the expressed protein. We measured the ability of H_6RbpA to bind to an immobilized matrix which was covalently linked to RNA homopolymers.

3. To determine the role of the auxiliary glycine-rich domain in RNA binding. Site-directed mutagenesis was used to insert or modify particular nucleotide bases resulting in the introduction of a restriction enzyme site. Digestion of cloned DNA with this enzyme specifically removed the segment of the gene coding for the auxiliary glycine-rich domain. This derivative of protein H_6RbpA, named **H_6RbpA(ΔGly)**, which did not contain the auxiliary glycine-rich region, was subsequently expressed. The RNA binding ability of this protein was compared to that of the full length sequence.

4. To examine the effects of the histidine tag on the proper functioning of RbpA *in vivo*. The mutated *rbpA* gene, containing the coding sequence for a histidine tag, was re-introduced into a neutral site in the cyanobacterial chromosome and the phenotype of this strain (**SY-RBPA8H**) was characterized.

5.2. MATERIALS AND METHODS

5.2.1. Plasmids and Strains

Plasmids used in site-directed mutagenesis and expression of the *rbpA* gene are outlined in **Table 5.1.** Plasmid pALTER-1 and *E. coli* strain ES1301 were purchased from Promega. Plasmid pTRC99A used in the expression of RbpA protein in *E.coli* BL21(DE3) was purchased from Pharmacia.

5.2.2. Amplification of DNA fragments and Insertion of Restriction Sites by Polymerase Chain Reaction (PCR)

Custom synthesized oligonucleotides used in PCR amplification were purchased from Gibco BRL. Their names, sequences and parameters are listed in **Table 5.2.** PCR reactions were carried out in a 10 µL reaction volume using a PCR reaction buffer (20 mM Tris-HCl pH 8.4, 1.5 mM $MgCl_2$, 50 mM KCl, 1% Triton X-100, 0.2 mM dNTPs) supplied by the manufacturer (Gibco BRL). Both oligonucleotide primers were added to a final concentration of 0.5 µM. *Taq* DNA polymerase (Gibco BRL) was added to a final concentration of 25 mU/µL.

All PCR incubations were carried out in a PTC-200 Peltier Thermal Cycler (MJ Research). PCR reactions were initially incubated at 95°C for 3 min. This was followed by 30 cycles of denaturation (95°C for 45 sec), primer annealling (50°C for 90 sec) and synthesis (72°C for 60 sec). PCR products were electrophoresed on a 0.7 - 1.0% agarose gel and the products isolated from an agarose gel slice as described previously (Heery *et al.*, 1990) (see Chapter 2). In circumstances where restriction sites have been introduced, PCR products

Table 5.1. Plasmids Used In Protein Studies

Plasmid	Description	Reference
pALTER-1	$Amp^{s}Tet^{r}$	Promega
pTRC99A	Amp^{r}	Pharmacia
pTJB4	1.6 kb BamHI fragment containing the *rbpA* gene in pALTER-1	This work
pTJB4H	1.6 kb BamHI fragment containing His-tagged *rbpA* gene in pALTER-1	This work
pUC4H	1.6 kb BamHI fragment containing His-tagged *rbpA* gene in pUC19	This work
pTJB5H	1.2 kb BamHI/HindIII fragment containing His-tagged *rbpA* in pALTER-1	This work
pTJB5H(Bst)	1.2 kb BamHI fragment containing His-tagged *rbpA* gene with introduced BstEII restriction site	This work

Table 5.1. Plasmids Used In Protein Studies (continued)

Plasmid	Description	Reference
pUC4H(ΔGly)	1.6 kb BamHI fragment containing His-tagged *rbpA* gene without the auxiliary Gly-rich domain in pUC19	This work
pH_6RBPA	433 bp NcoI/HindIII fragment containing His-tagged *rbpA* gene in pTRC99A	This work
pH_6RBPA(ΔGly)	379 bp NcoI/HindIII fragment containing His-tagged *rbpA* gene without the auxiliary Gly-rich domain in pTRC99A	This work
pAM990H	1.6 kb BamHI DNA fragment containing His-tagged *rbpA* gene in pAM990 (Chapter 4)	This work
pTJB25	identical to pMUN3.3 (Figure 4.7) except for the presence of a His-tagged *rbpA* gene	This work

Table 5.2. List of Oligonucleotides Used for Site-Directed Mutagenesis and PCR.

Name	Length	Sequence	T_m (°C)*	GC (%)
1. HIS-RBPA	47mer	5'-GAC AAC TCC ATG TCT CAT CAC CAT CAC CAT CAC TCT ATT TAC GTT GG-3'	62.9	44
2. RBPA-REV1	18mer	5'-GGC GGT CAA GTC AGC TTC-3'	58.0	61
3. RBPA-NCOI	25mer	5'-GGA GAC AAC TCC ATG GCT CAT CAC C-3'	72.8	56
4. RBPA-BSTEII	24mer	5'-GCG CAG TGG TTA CCG GCT CGT TCG-3'	77.0	66
5. RBPA-REV2	18mer	5'-CAG AGT CTA CCA ACA GAC-3'	48.6	50
6. AMP-REPAIR	27mer	5'-GTT GCC ATT GCT GCA GGC ATC GTG GTG-3'	77.7	59

* Oligonucleotide melting temperatures (T_m) were calculated using software at location *http://alces.med.umn.edu/rawtm.html* on the World Wide Web according to the method of Breslauer *et al.*, 1986. All calculations were carried out for a salt concentration of 50 mM. Calculations were carried out for an oligonucleotide concentration of 100 nM for oligonucleotides used in site-directed mutagenesis (1 and 4), 20 nM for the AMP-REPAIR oligonucleotide used in site-directed mutagenesis (6), 500 nM for oligonucleotides used in PCR reactions (3 and 5), and 25 nM for the RBPA-REV1 oligonucleotide used in dideoxynucleotide sequencing reactions (2).

were digested with the appropriate enzyme prior to electrophoresis. The resulting cohesive ends were then utilized for cloning into a suitable cloning vector such as pUC19.

5.2.3. Preparation of Single-Stranded Phagemid DNA (ssDNA) for Site-Directed Mutagenesis

The first step in site-directed mutagenesis was the preparation of phagemid single-stranded DNA (ssDNA). A 25 mL LB culture containing 10 μg/mL tetracycline in a 250 mL Erlenmeyer flask was inoculated with 0.5 mL of an overnight culture of *E. coli* JM109 containing one of the recombinant pALTER-1 plasmids. This culture was incubated at 37°C for 1 hr with shaking (200 rpm). The culture was subsequently infected with bacteriophage R408 ($5X10^{10}$ pfu/mL; 200 μL). This volume represents a multiplicity of infection (m.o.i.) of approximately 10-20 phage particles per cell. Following infection, the culture was incubated for a further 6 hrs under the same incubation conditions as above.

Cells were pelleted twice by centrifugation (12,000 rpm, 15 min). The resulting supernatent, containing phage particles, was transferred to a fresh tube containing 6 mL (0.25 volumes) of phage precipitation solution (3.75 M ammonium acetate pH 7.5, 20% PEG-8000). Phage particles were precipitated by incubating at 4°C overnight and then centrifuging at 12,000 rpm for 15 min. Once thoroughly drained, the phage pellet was resuspended in 400 μL of TE buffer. This solution was extracted once with chloroform:isoamyl alcohol (24:1), twice with TE-saturated phenol:chloroform:isoamyl alcohol (25:24:1), and once with chloroform. Nucleic acid was precipitated by adding 200 μL

(0.5 volumes) of 7.5 M ammonium acetate plus 1.2 mL (2 volumes) of 100% ethanol and incubating the resulting mixture for 1 hr at -20°C. The nucleic acid was pelleted by centrifugation (14,000 rpm, 10 min), washed with 70% ethanol, air-dried at room temperature, and resuspended in 20 µL of sterile water. The quality and quantity of recovered phagemid DNA was confirmed by agarose gel electrophoresis.

5.2.4. Annealling of Oligonucleotides and Second Strand Synthesis

Custom synthesized oligonucleotides used in site-directed mutagenesis and DNA sequencing were purchased from Gibco BRL. The names, sequences and parameters of these oligonucleotides are listed in **Table 5.2.** The mutagenesis reaction was carried out in two steps: annealing of oligonucleotides to the template ssDNA, and second strand DNA synthesis (see Figure 5.1). The annealing reaction was carried out in a buffer containing 20 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl. It contained the following :

10 µL (0.05 pmol)	recombinant pALTER-1 ssDNA template
1 µL (0.25 pmol)	AMP-REPAIR oligonucleotide (2.2 ng/µL)
1 µL (1.25 pmol)	Mutagenic oligonucleotide

The total volume of the reaction mixture was 20 µL. This solution was heated to 75°C for 5 minutes then allowed to cool slowly (1-2°C/minute) to room temperature.

The second step of the mutagenesis procedure involved synthesis of the second (mutagenic) DNA strand. To the cooled reaction mixture, the following were added:

3 µL	10X synthesis buffer (100 mM Tris-HCl pH 7.5, 5 mM dNTPs, 10 mM ATP, 20 mM DTT)
1 µL (5-10 U)	T4 DNA Polymerase
1 µL (1-3 U)	T4 DNA Ligase

The total volume of the synthesis reaction mixture was 30 µL. This solution was incubated at 37°C for 1-2 hrs to perform mutant strand DNA synthesis and ligation.

5.2.5. Transformation of *E.coli* ES1301 *mutS* Competent Cells and Selection of Mutant pALTER clones

E. coli ES1301 and JM109 competent cells were prepared using a modified rubidium chloride protocol as recommended by Promega (Hanahan, 1985) and stored at -80°C until needed. In a pre-chilled 17 X 100 mm (Falcon 2059) polypropylene tube, 100 µL of *E. coli* ES1301 competent cells were mixed with 10 µL of the mutagenesis reaction mixture and incubated on ice for 10 min. Cells were then heat shocked for 45 sec at 42°C. LB (900 µL) was added to the transformation mixture which was then incubated at 37°C for 1 hr with shaking (200 rpm). 500 µL of the transformation mixture was used to inoculate 4.5 mL of LB media containing 10 µg/mL tetracycline plus 100 µg/mL carbenicillin. The resulting 5 mL culture was incubated overnight at 37°C with shaking.

Plasmid DNA was prepared from 5 mL *E. coli* ES1301 cultures using the standard Merlin miniprep protocol (Iyer, 1993) (see Chapter 2). Recovered plasmid DNA was then used to transform *E. coli* JM109 competent cells. Single clones were subsequently screened for the presence of the mutation, either by restriction analysis or by DNA sequencing.

5.2.6. Expression of His-tagged RNA-Binding Proteins Using pTRC99A in *E.coli* BL21

Conditions for growth and induction of *E. coli* BL21 were as described previously (Petty, 1987). Cultures of *E.coli* BL21 (100 mL-500 mL) containing recombinant pTRC99A plasmids were grown at 37°C with shaking (200 rpm). When the cultures reached mid-log phase (A_{600} of 0.6-0.8), expression of the recombinant protein was induced by adding IPTG to a final concentration of 1-2 mM. Following induction, cultures were incubated at 37°C for 4 hrs. Cells were pelleted by centrifugation at 6000 rpm for 10 min. Cell pellets not immediately used were stored at -80°C until needed.

Total cellular protein was prepared using a non-denaturing protocol (Petty, 1987). Cell pellets (approximately 2 g wet weight) were resuspended in 10 mL of Resuspension Buffer (50 mM Na_2HPO_4 pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM PMSF). Triton X-100 was added to a final concentration of 0.1% and the cells were lysed by sonication. A 150X protease inhibitor cocktail (1mg/mL leupeptin, 1 mg/mL pepstatin A) was added to reduce protein degradation. DNase I was added to a final concentration of 10 µg/mL to reduce the viscosity of the solution. The solution was incubated on ice for 15 min, then centrifuged at 15,000 rpm for 15 min to pellet cell debris. The

supernatent containing total cellular protein was transferred to a clean tube and stored at 4°C until needed.

5.2.7. Isolation of His-tagged Proteins by Metal Chelate Affinity Chromatography (MCAC)

Columns and NTA resin were prepared according to the recommendations of the supplier (Qiagen). Approximately 2 mL of a 50% slurry of pre-charged Ni^{2+}-NTA resin was added to a 4 mL Econo-column (1 cm. diameter, Biorad) and equilibrated with Resuspension buffer. The bed volume of the column was approximately 1 mL. Total protein extract (10 mL) was loaded onto the column and allowed to flow-through at a rate of 10-15 mL/hr. The column was washed with 25 mL Wash Buffer I (50 mM Na_2HPO_4 pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM PMSF, 20 mM imidazole) and 25 mL of Wash Buffer II (50 mM Na_2HPO_4 pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM PMSF, 40 mM imidazole). His-tagged proteins were eluted from the column with 5 mL of Elution Buffer (50 mM Na_2HPO_4 pH 8.0, 500 mM NaCl, 10% glycerol, 1 mM PMSF, 100 mM imidazole). Samples were collected as 0.5 mL fractions which were analyzed for the presence of the His-tagged protein by SDS-PAGE.

5.2.8. Proteins Electrophoresis Using the PhastGel® Electrophoresis System

Separation of proteins was carried out using a PhastGel® Electrophoresis System (Pharmacia). Protein Sample Buffer (3X is 187.5 mM Tris-HCl pH 6.8, 6% SDS, 30% glycerol, 0.3% bromophenol blue) was added to each protein sample and the tubes boiled for 2 min prior to loading on a 8-

25% gradient polyacrylamide gel containing 0.112 M Tris-HCl pH 6.4, 0.112 M acetate (Pharmacia). Protein samples containing imidazole were not boiled but instead were incubated at 37°C for 10 min prior to loading. Gels were run at 250V (10 mA) for approximately 30 min at 15°C with buffer strips containing 0.20 M tricine, 0.20 M Tris-HCl pH 8.1, 0.55% SDS, 3% agarose. Following electrophoresis, gels were treated for 30 min with staining solution (0.1% PhastGel Blue R, 30% methanol, 10% acetic acid), washed with a destaining solution (30% methanol, 10% acetic acid), and stored in a preserving solution (5% glycerol, 10% acetic acid). Gels were documented by drying at room temperature or by digitizing with a scanner.

5.2.9. Nucleic Acid Binding Assay for Cyanobacterial Rbps

The nucleic acid binding assay was carried out essentially as described by Sugita and Sugiura (1994). RNA resins used for binding experiments were as follows: polyadenylic acid (poly(A) RNA) attached to a matrix of cross-linked 4% beaded agarose (0.5 - 1.0 mg RNA polymer per mL resin), polycytidylic acid (poly(C) RNA) attached to a matrix of cross-linked 4% beaded agarose (0.25 - 1.0 mg RNA polymer per mL resin), polyguanylic acid (poly(G) RNA) attached to a matrix of polyacrylhydrazido-agarose (1.5 - 4.5 mg RNA polymer per mL resin) and polyuridylic acid (poly(U) RNA) attached to a matrix of polyacrylhydrazido-agarose (0.1 - 1.0 mg RNA polymer per mL resin) (Sigma). Purified RNA-binding protein (5 µg) was incubated with 20 - 30 µg of RNA polymer in 1 mL of RNA-binding buffer (10 mM Tris-HCl pH 7.5, 0.2 M NaCl, 2.5 mM $MgCl_2$, 0.5% Triton X-100, 1 mM PMSF). This mixture was incubated for 10 min at 4°C with rotation. The total mixture was

then applied to a mini-column (Promega) which retained the immobilized RNA polymer along with any bound protein. The retained RNA polymers were washed once with RNA-binding buffer containing 1 mg/mL heparin, once with RNA-binding buffer, and once with distilled water. Bound protein was eluted from each column by adding 20 µL of protein sample buffer (see Section 5.2.8). The resulting proteins were separated by SDS-PAGE and visualized by Coomassie staining as described above.

5.2.10. Construction of Plasmid pTJB25 and Mutant SY-RBPA8H

Construction of plasmid pTJB25 is outlined in **Figure 5.2**. The strategy employed was essentially identical to that described for the construction of pMUN3.3 (see Chapter 4). A 1.6 kb BamHI DNA fragment containing the modified *rbpA* gene was ligated into the BglII site of pAM990. This plasmid was named pAM990H. Potential pAM990H plasmid clones were screened for the presence and orientation of the inserted DNA fragment by digestion with HindIII (**Figure 5.3A**). Of the eight clones tested, one (clone#1) had the correct insert. Introduction of the kanamycin resistance cassette (*Kan*) and re-insertion of the upstream neutral-site flanking sequence (n.s.#1) were carried out as before for plasmid pMUN3.3 (**Figure 5.2**). The final plasmid, named pTJB25, was confirmed by digestion with HindIII (**Figure 5.3B**). Its restriction pattern matched exactly with that observed for pMUN3.3 since the additional sequence corresponding to the histidine tag was essentially negligible in the migration of the 2.6 kb DNA fragment (see Figure 4.8). The restriction pattern of plasmid pTJB25 was also confirmed by digestion with BamHI (not shown). As with the plasmid pMUN3.3, the entire construct, including the kanamycin

Figure 5.2. Construction of plasmid pTJB25. The construction of this plasmid was essentially identical to the strategy outlined for pMUN3.3 (see Figure 4.7) except that the 1.6 kb BamHI DNA fragment was derived from pUC4H and therefore contained an *rbpA* gene which included the coding sequence for a histidine tag (orange).

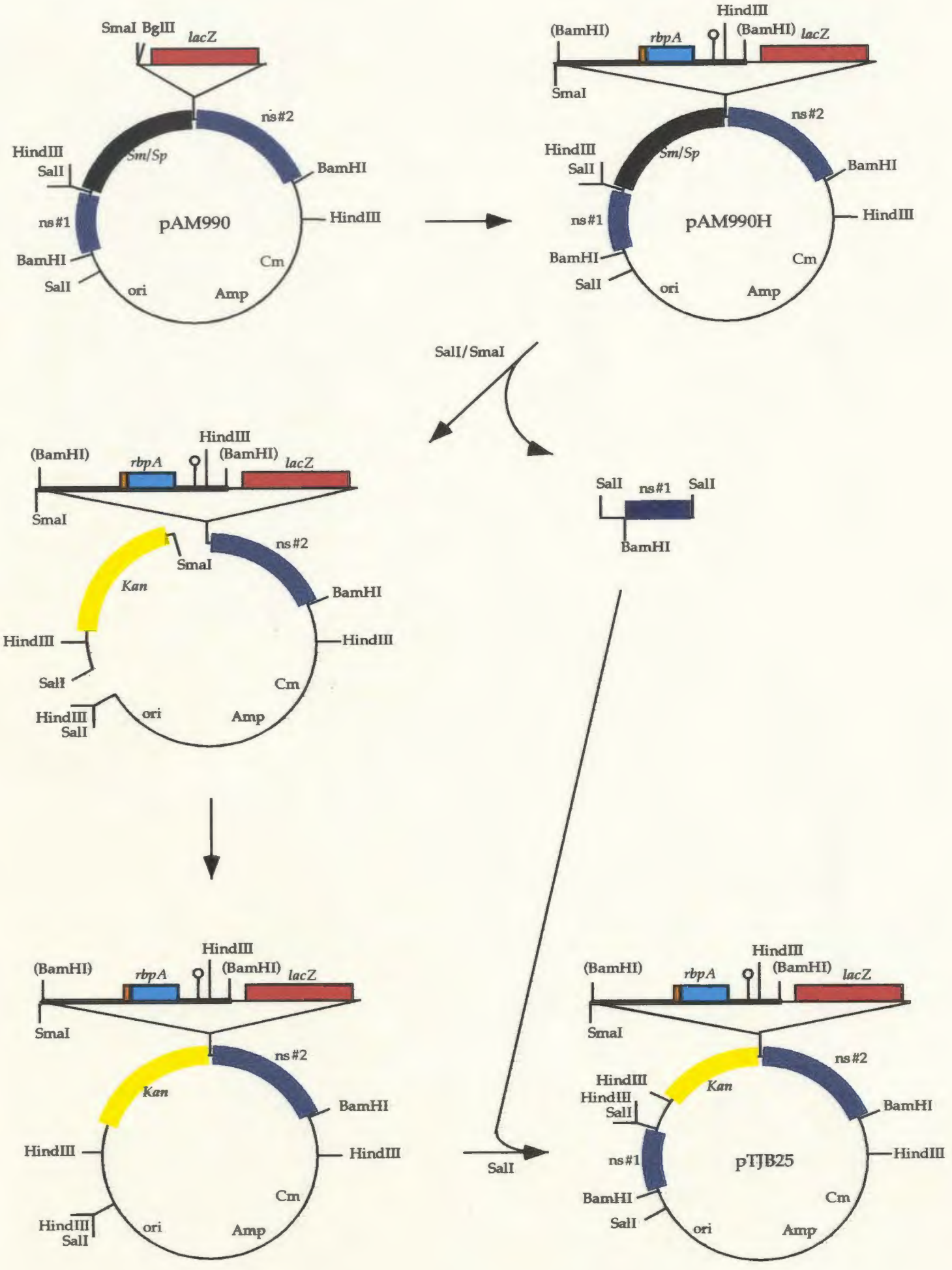
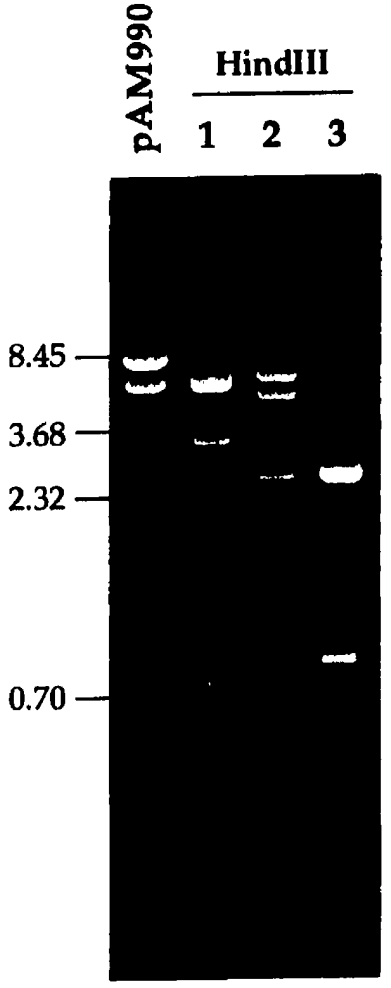
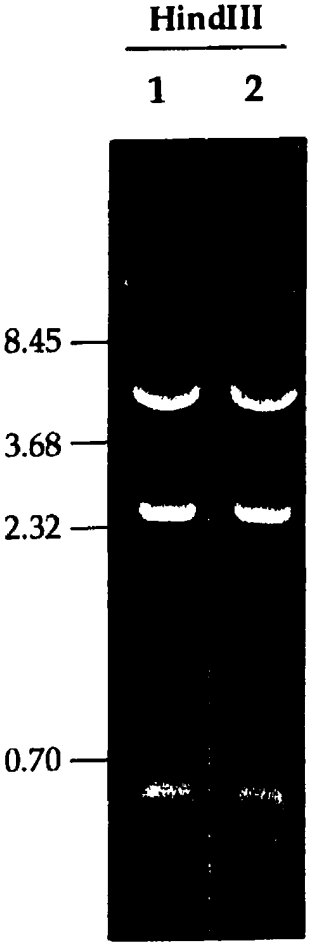

SmaI BglII
lacZ
ns#2
HindIII
SalI
Sm/Sp
BamHI
HindIII
ns#1
pAM990
BamHI
Cm
SalI
ori
Amp
HindIII
(BamHI)
rbpA
(BamHI)
lacZ
SmaI
ns#2
HindIII
SalI
Sm/Sp
BamHI
HindIII
ns#1
pAM990H
BamHI
Cm
SalI
ori
Amp
SalI/SmaI
HindIII
(BamHI)
rbpA
(BamHI)
lacZ
SmaI
SmaI
ns#2
Kan
BamHI
HindIII
HindIII
SalI
Cm
HindIII
SalI
ori
Amp
SalI
ns#1
SalI
BamHI
HindIII
(BamHI)
rbpA
(BamHI)
lacZ
SmaI
ns#2
Kan
BamHI
HindIII
HindIII
SalI
Cm
HindIII
SalI
ori
Amp
SalI
HindIII
(BamHI)
rbpA
(BamHI)
lacZ
SmaI
ns#2
HindIII
HindIII
SalI
Kan
BamHI
ns#1
pTJB25
HindIII
BamHI
Cm
SalI
ori
Amp

Figure 5.3. Analysis of potential **(A)** pAM990H and **(B)** pTJB25 clones by restriction digestion with HindIII. **(A)** Plasmid DNA from three of eight potential pAM990H clones was digested with HindIII and the products resolved on a 0.8% agarose gel. The three predicted DNA fragments included a 5.8 kb fragment which included the *lacZ* gene and downstream neutral-site flanking sequence (n.s.#2), a 5.46 kb fragment which included most of the pBR322 vector sequence plus the upstream neutral-site flanking sequence (n.s.#1), and a 3.2 kb fragment containing the omega cassette and most of the 1.6 kb fragment containing the *rbpA* gene (see Figure 5.3). Of eight clones screened, only one (clone #1) had the expected restriction pattern. Neutral-site vector pAM990 is included as a control. **(B)** Plasmid DNA from two potential pTJB25 clones was digested with HindIII and separated on a 0.8% agarose gel. The location of BstEII-digested bacteriophage λ DNA fragments used as size markers are shown at the left.

A.



B.



resistance cassette (*Kan*), modified *rbpA* gene and promoterless *lacZ* gene, was isolated from pTJB25 by digestion with BamHI. This 9.5 kb DNA fragment was used to create the mutant SY-RBPA8H by transformation of SY-RBPA1 competent cells using the previously described method (see Section 4.2.5).

5.3. RESULTS AND DISCUSSION

5.3.1. Introduction of Hexahistidine Tag by Site-Directed Mutagenesis

The purification of proteins using a histidine tag was chosen as a purification strategy since it offered many advantages over other fusion protein systems. The coding sequence for a histidine tag could easily be inserted into a gene by methods of PCR or site-directed mutagenesis. The presence of six histidine residues added less the 1 kDa to the overall molecular weight of the protein. Finally, with the hundreds of proteins purified to date using this strategy, it is rare to find instances where the presence of the six histidines interfered with the normal biological activity of the protein.

The strategy for introduction of a histidine tag into the coding sequence of the *rbpA* gene is outlined in **Figure 5.4**. In order to modify the coding region of the *rbpA* gene, a 1.6 kb DNA fragment containing BamHI cohesive ends was cloned into the BamHI site of pALTER-1. The fragment contained a complete copy of the *rbpA* gene and was identical to the 1.6 kb DNA fragment used in complementation experiments (see Figure 4.5). The success of the mutagenesis strategy also relied on the DNA fragment being ligated in the proper orientation to allow hybridization of single stranded plasmid DNA with the mutagenic oligonucleotide (**Figure 5.5A**). Four clones screened for the presence of a recombinant pALTER-1 plasmid by digestion with BamHI and HindIII (**Figure 5.5B**). Only one of these (clone #3) contained the 1.6 kb DNA fragment in the proper orientation for mutagenesis. This plasmid was named pTJB4.

Figure 5.4. Strategy for construction of the histidine-tagged variant of RbpA. The inferred amino acid sequence of the RbpA protein in shown in outlined type below the nucleotide sequence. Regions which are the focus of a mutagenesis step are shown in bold type. Mutagenesis with the HIS-RBPA oligonucleotide introduced a 21 bp sequence which coded for six histidine residues as well as a serine residue **(2)**. For purposes of cloning into the expression vector pTRC99A, the RBPA-NCOI oligonucleotide introduced an NcoI restriction site at the start of the *rbpA* gene **(3)**. This modification altered the second amino acid of the inferred RbpA sequence from serine to alanine.

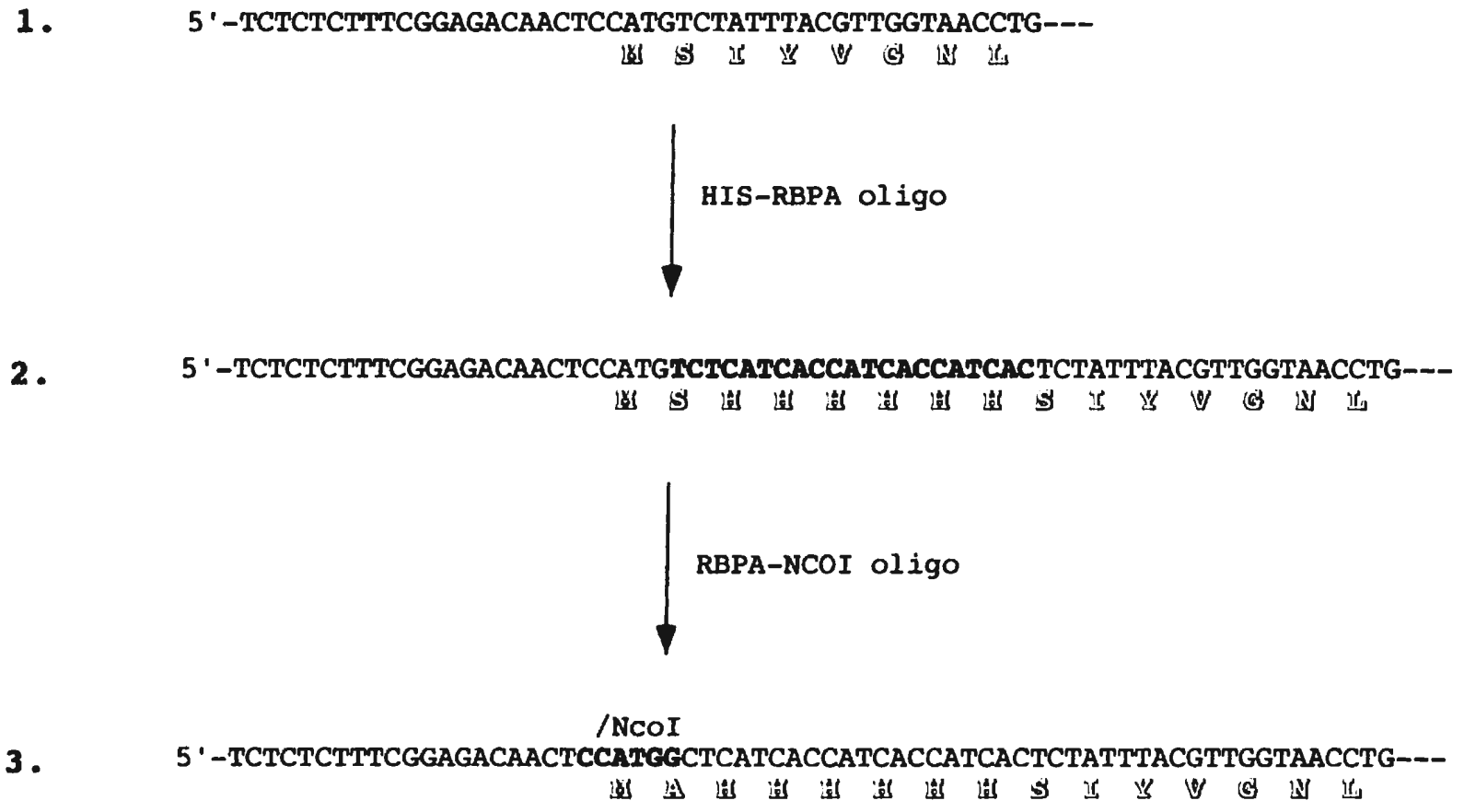
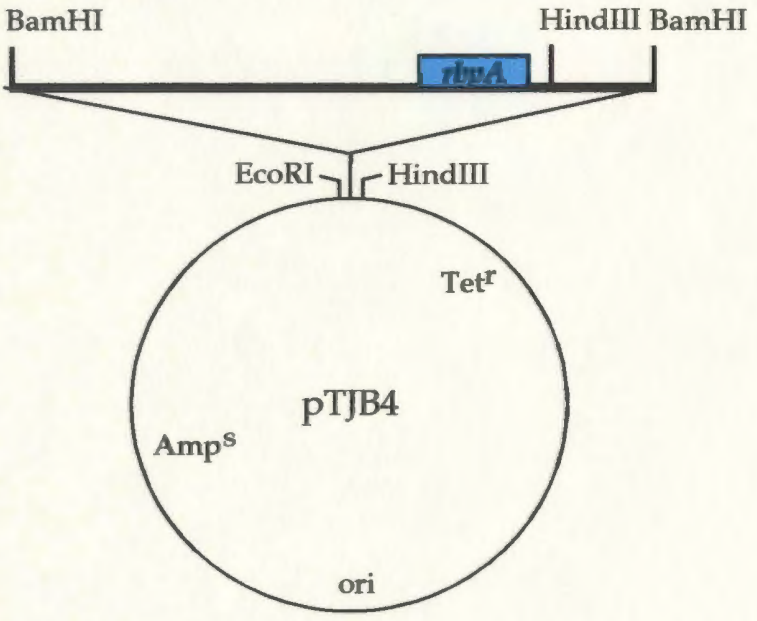
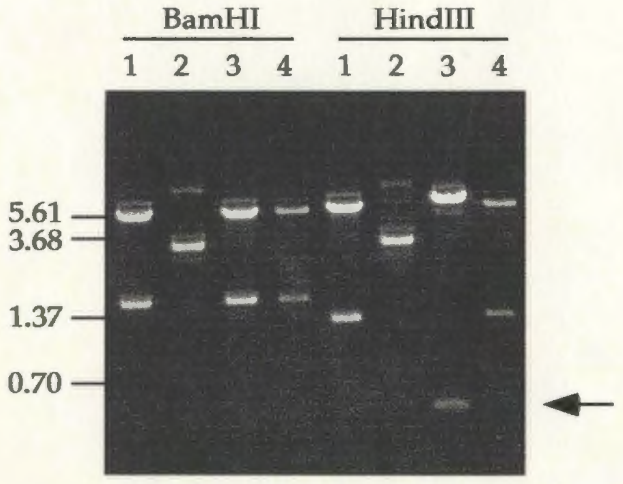

1.
5'-TCTCTCTTTCGGAGACAACTCCATGTCTATTTACGTTGGTAACCTG---
M S I Y V G N L
HIS-RBPA oligo
2.
5'-TCTCTCTTTCGGAGACAACTCCATGTCTCATCACCATCACCATCACTCTATTTACGTTGGTAACCTG---
M S H H H H H H S I Y V G N L
RBPA-NCOI oligo
/NcoI
3.
5'-TCTCTCTTTCGGAGACAACTCCATGGCTCATCACCATCACCATCACTCTATTTACGTTGGTAACCTG---
M A H H H H H H S I Y V G N L

Figure 5.5. Cloning of the 1.6 kb BamHI DNA fragment containing the *rbpA* gene in plasmid pALTER-1. **(A).** Restriction map of the new plasmid, named pTJB4, with the 1.6 kb BamHI DNA fragment in the correct orientation for mutagenesis. The location of relevant restriction sites are shown. **(B).** Screening of four potential pTJB4 clones by digestion with BamHI and HindIII. Digestion with BamHI confirmed that three clones (#1, #3 and #4) contained the 1.6 kb BamHI DNA fragment. Digestion with HindIII showed that only clone #3 had the insert in the correct orientation for mutagenesis. An arrow shows the location of a 0.4 kb DNA fragment confirming the correct orientation of the 1.6 kb insert. This clone was named pTJB4. The location of bacteriophage λ DNA fragments used as size markers are shown at the left.

A.


BamHI
HindIII BamHI
rbvA
EcoRI
HindIII
Tet^r
pTJB4
Amp^s
ori

B.


BamHI
HindIII
1 2 3 4 1 2 3 4
5.61
3.68
1.37
0.70

Mutagenesis with the HIS-RBPA oligonucleotide was carried out as outlined previously (Section 5.1.2). Successful mutagenesis introduced an additional 21 bp into the coding region of the gene (see Figure 5.4); this change could be visualized by electrophoresis on an agarose gel. Clones were screened for the presence of the insertion by digestion with ApoI (**Figure 5.6A**). In the wild type sequence, digestion with ApoI yielded a 402 bp DNA fragment which included the entire coding region for the *rbpA* gene (see Figure 2.5). Successful mutagenesis therefore increased the size of this fragment to 423 bp. A total of six clones were screened by digestion with ApoI (**Figure 5.6B**). Five clones contained the larger ApoI fragment indicative of a successful mutagenesis. Clone #4 was restreaked to obtain a pure clone; this plasmid was named pTJB4H.

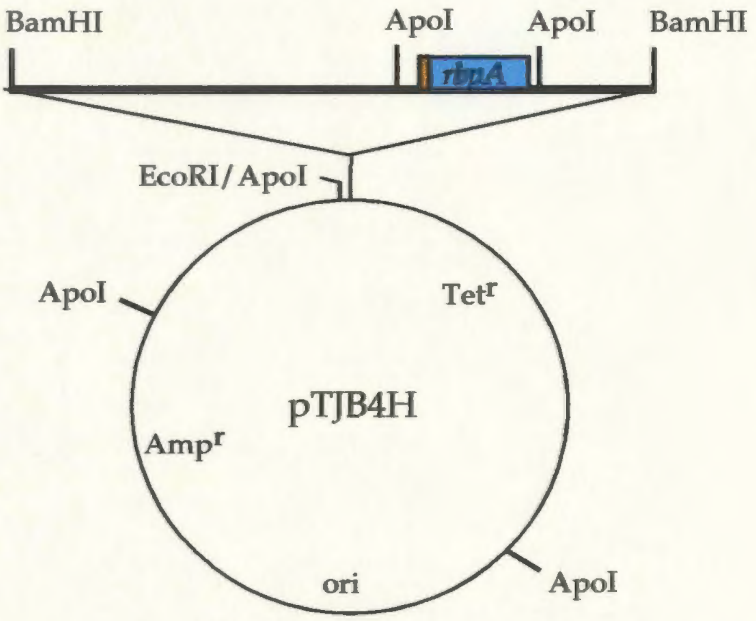
The success of the mutagenesis reaction was confirmed by dideoxynucleotide sequencing. In brief, the 1.6 kb BamHI fragment containing the mutation was removed from pTJB4H and cloned into the BamHI site of pUC19; this plasmid was named pUC4H. Sequencing was carried out using the RBPA-REV1 oligonucleotide as a primer.

5.3.2. Introduction of the NcoI Site Upstream of *rbpA* by PCR

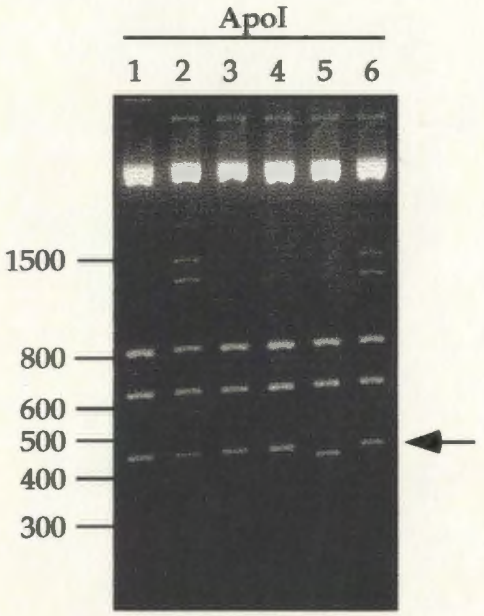
Initial attempts to express the *rbpA* gene in *E. coli* using its own predicted ribosome binding site were unsuccessful. It was therefore decided to introduce an NcoI site at the start of the *rbpA* coding region (**Figure 5.4**). This site was then used to clone the DNA fragment into an expression vector (pTRC99A) which contains a ribosome binding site. Introduction of this NcoI site altered the coding sequence for the *rbpA* gene in such a way that the

Figure 5.6. Screening of potential pTJB4H mutants with ApoI. **(A).** Restriction map of pTJB4H showing the location of BamHI, EcoRI and ApoI sites. The location of the histidine tag sequence is shown in orange at the 5' end of the *rbpA* gene (blue box). **(B).** Potential pTJB4H clones digested with ApoI and resolved on a 2% agarose gel. The location of DNA fragments used as size markers are shown at the left. The arrow indicates the location of a 402 bp DNA fragment which increases in size to 423 bp due to the insertion of the histidine tag coding sequence.

A.


BamHI
ApoI
ApoI
BamHI
ribA
EcoRI/ApoI
ApoI
Tet^r
pTJB4H
Amp^r
ori
ApoI

B.


ApoI
1 2 3 4 5 6
1500
800
600
500
400
300

serine residue at position 2 was replaced by an alanine residue. Amino-terminal sequencing of the corresponding protein (12RNP1) in *Synechococcus* 6301 revealed that this serine residue was the amino-terminal amino acid of the protein product due to removal of the methionine residue at position 1 (Sugita and Sugiura, 1994). Modification of the predicted amino-terminal amino acid of H_6RbpA was initially a concern since the half-lives of cytosolic proteins have been shown to be dependent on the nature of their amino-terminal residue (Bachmair *et al.*, 1986). However, the same report identified both methionine and alanine as "stabilizing" residues, so that the change from a methionine to alanine residue should have a minimal effect on the stability of H_6RbpA.

The NcoI site was introduced by PCR amplification using oligonucleotides RBPA-NCOI and RBPA-REV2 as primers. The PCR product (481 bp) was gel-purified; cohesive ends were generated by digestion with HindIII and NcoI and the fragment was ligated into pTRC99A. Following transformation into *E. coli* BL21, four clones were screened by digestion with HindIII and NcoI; all four clones contained the 433 bp DNA fragment containing the *rbpA* gene (data not shown). The resulting recombinant plasmid was named pH_6RBPA. The corresponding protein isolated by metal chelate affinity chromatography was named H_6RbpA.

5.3.3. Removal of the Glycine-Rich Region For Protein RbpA(ΔGly)

In addition to examining the binding properties of a full-length cyanobacterial RBP, we wished to examine the effects of removing the auxiliary glycine-rich domain on the overall RNA binding activity of the

protein. The strategy to do this is outlined in **Figure 5.7**. A 1.2 kb BamHI/HindIII DNA fragment containing the modified *rbpA* gene was cloned into pALTER-1; this plasmid was named pTJB5H (**Figure 5.8**). Site-directed mutagenesis using the RBPA-BSTEII oligo introduced a BstEII restriction site upstream of the coding sequence for the auxiliary glycine-rich domain. Digestion of the mutagenized plasmid, named pTJB5H(Bst), with BstEII removed a 235 bp BstEII DNA fragment which was totally internal to the *rbpA* gene but reduced in size relative to the 290 bp wild type sequence. Potential mutant clones were screened for the presence of this third BstEII site by digestion with BstEII (**Figure 5.9**). All clones screened contained the desired sequence mutation.

Since the goal was to express the truncated gene in *E. coli*, the next step should have involved removing the 290 bp BstEII fragment from pH_6RBPA and replacing it with the shorter 235 bp fragment derived from pTJB5H(Bst). This simple strategy was complicated by the presence of a BstEII site within the pTRC99A vector sequence itself. The alternative strategy adopted is outlined in **Figure 5.10**. The first step involved replacing the 290 bp BstEII DNA fragment in pUC4H (see Section 5.3.1) with the shorter 235 bp BstEII fragment derived from pTJB5H(Bst). Insertion of this fragment in the proper orientation restored the proper reading frame of the *rbpA* gene without the auxiliary domain; this new plasmid was named pUC4H(ΔGly). The presence of a StuI site within the sequence of the insert allowed confirmation of the proper orientation by digestion of pUC4H(ΔGly) with HindIII and StuI (data not shown).

Figure 5.7. Strategy for deletion of the auxiliary domain sequence of the *rbpA* gene. The inferred amino acid sequence of the C-terminal region of the RbpA protein in shown in outlined type below the nucleotide sequence. The two BstEII sites crucial to this mutagenesis step are shown in bold type. Mutagenesis with the RBPA-BSTEII oligonucleotide altered the sequence at the start of the auxiliary domain (**GGTGGC**) to a sequence which was recognized by BstEII (**GGTTACC**). Subsequent digestion with BstEII and re-ligation removed the coding sequence for the auxiliary domain while leaving the correct reading frame for the protein intact.

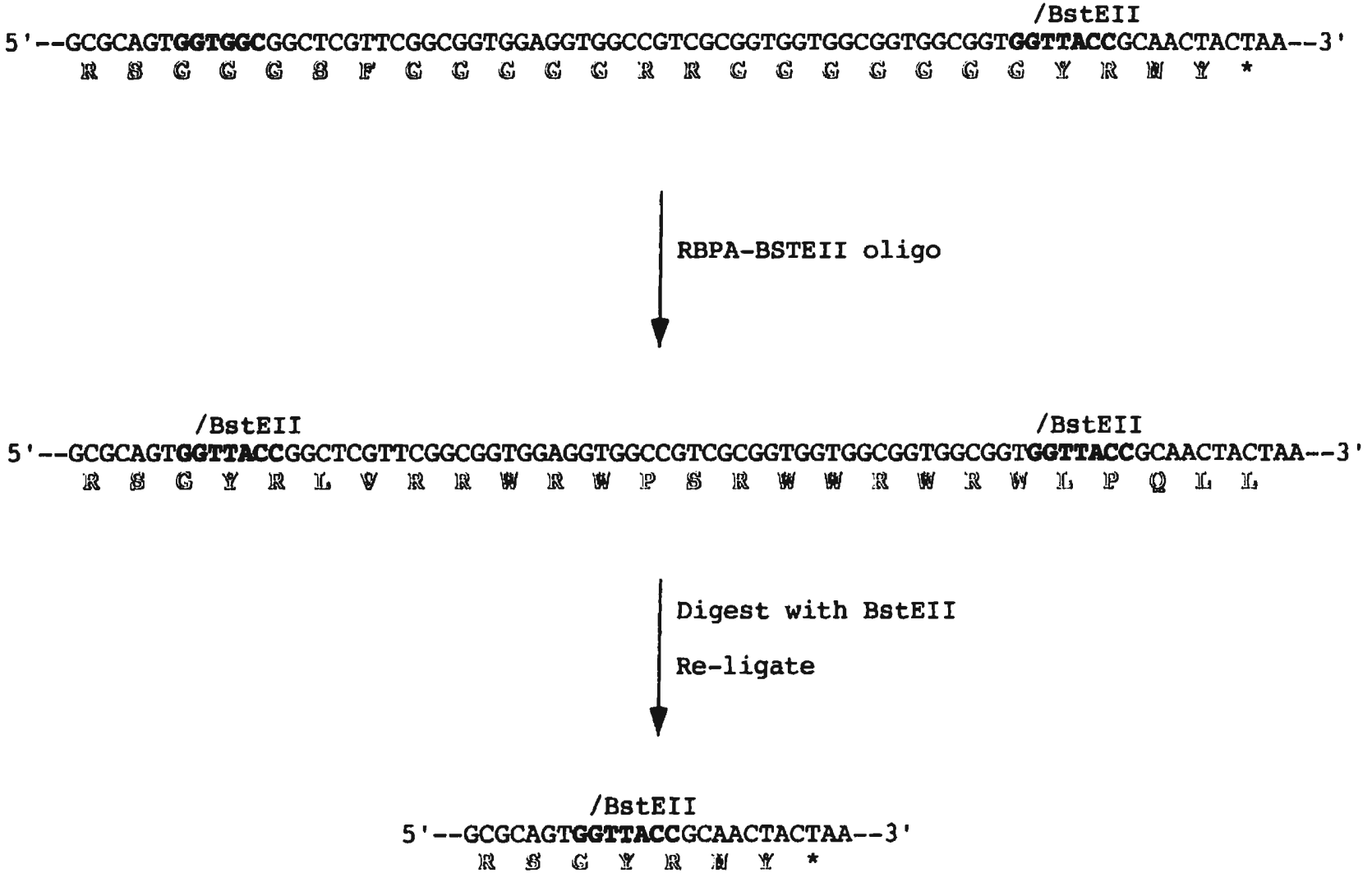
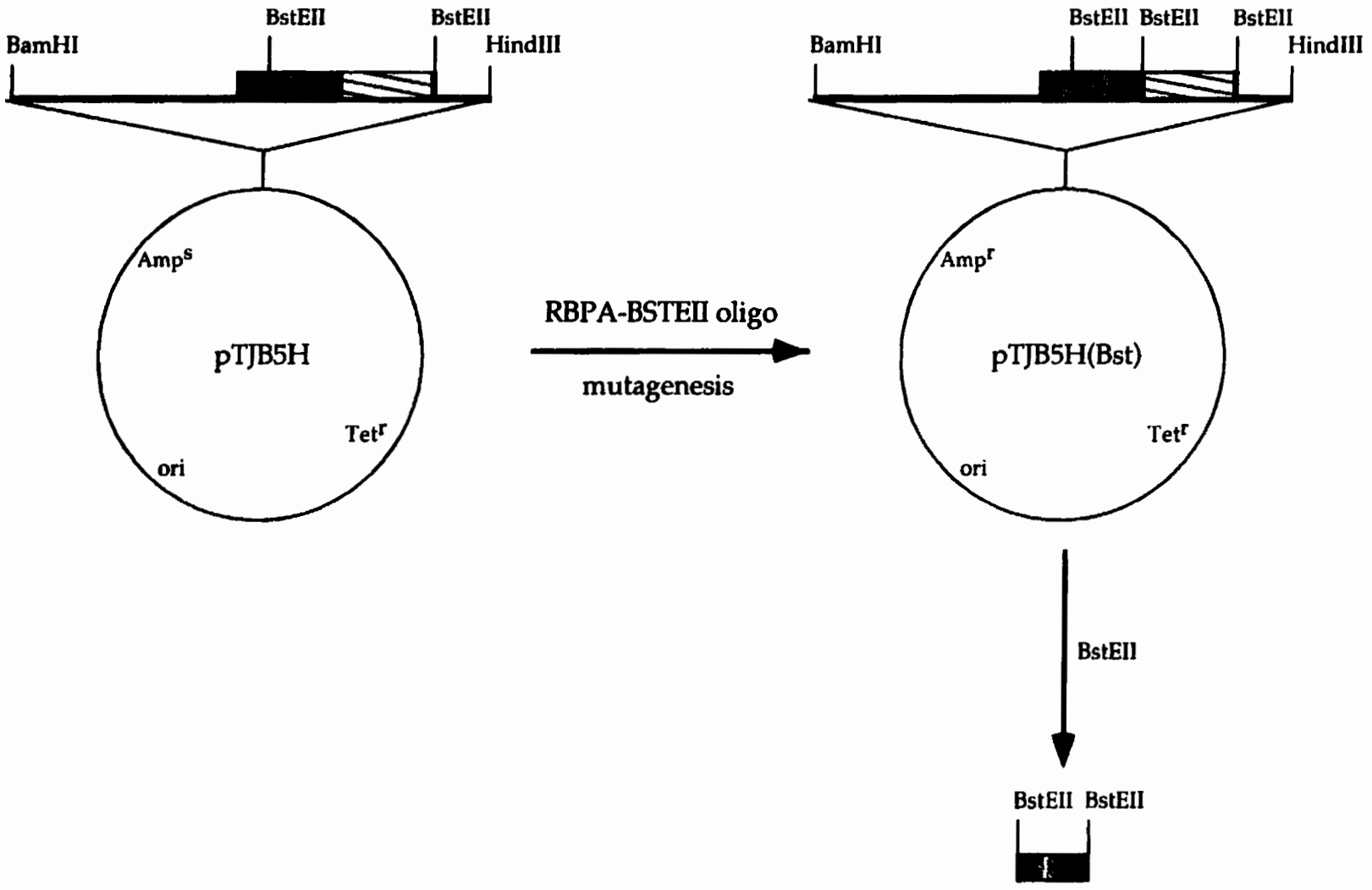

/BstEII
5'--GCGCAGTGGTGGCGGCTCGTTCGGCGGTGGAGGTGGCCGTCGCGGTGGTGGCGGTGGCGGTGGTTACCGCAACTACTAA--3'
R S G G G S F G G G G G R R G G G G G G G Y R M Y *
RBPA-BSTEII oligo
/BstEII
/BstEII
5'--GCGCAGTGGTTACCGGCTCGTTCGGCGGTGGAGGTGGCCGTCGCGGTGGTGGCGGTGGCGGTGGTTACCGCAACTACTAA--3'
R S G Y R L V R R W R W P S R W W R W R W L P Q L L
Digest with BstEII
Re-ligate
/BstEII
5'--GCGCAGTGGTTACCGCAACTACTAA--3'
R S G Y R M Y *

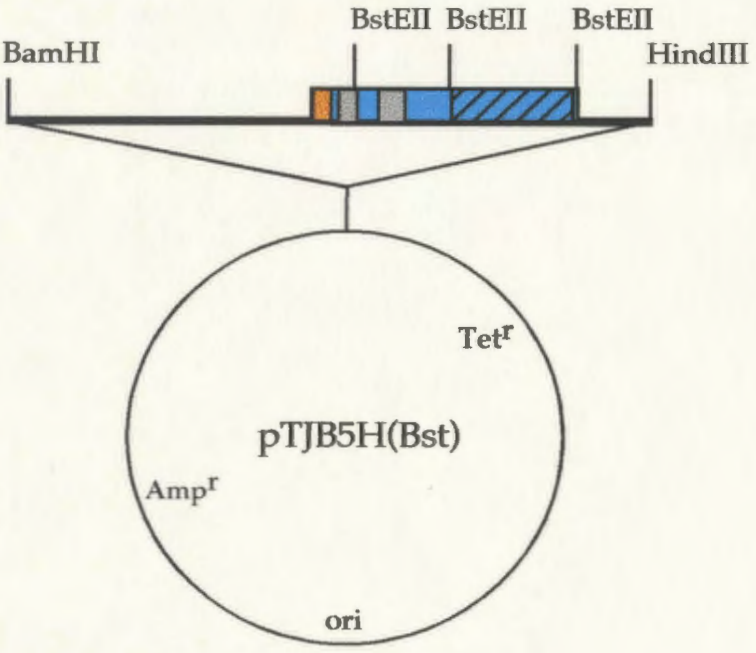
Figure 5.8. Initial step of the cloning strategy for a truncated *rbpA* gene. An expanded view of the *rbpA* gene shows the histidine tag region (orange), the RNP1 and RNP2 sequences (grey shaded) and the auxiliary glycine rich region (hatched box). Site-directed mutagenesis introduced a BstEII site at the 5' end of the coding sequence for the auxiliary domain. Digestion of the resulting mutant plasmid, named pTJB5H(Bst), with BstEII removed a 235 bp DNA fragment which contained all of the RNP1 sequence as well as part of the RNP2 sequence. This fragment was carried forward to the next step in the strategy.



Isolate 235 bp BstEII fragment

Figure 5.9. Screening of potential pTJB5H(Bst) mutants by BstEII digestion. **(A).** Restriction map of pTJB5H(Bst) showing the location of three BstEII sites. The histidine tag region is shown in orange; the RNP1 and RNP2 sequences are grey shaded; the auxiliary glycine-rich domain is shown as a hatched box. **(B).** Analysis of five potential pTJB5H(Bst) mutants by electrophoresis on a 2% agarose gel. The plasmid pUC4H was digested with BstEII and used as a negative control. The arrow shows the location of the smaller BstEII DNA fragment (235 bp) which indicated a positive result. The location of DNA fragments used as size markers are shown at the left.

A.


BamHI
BstEII
BstEII
BstEII
HindIII
Tet^r
pTJB5H(Bst)
Amp^r
ori

B.

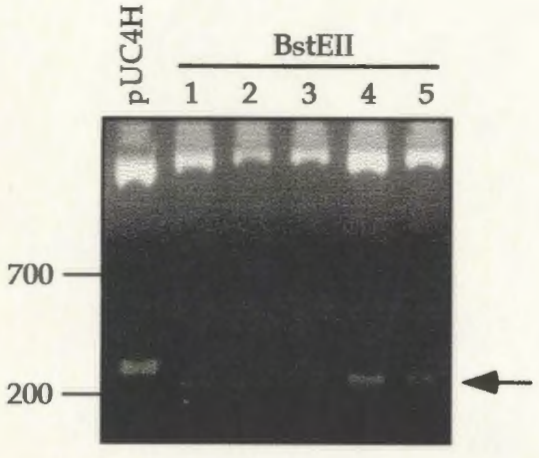
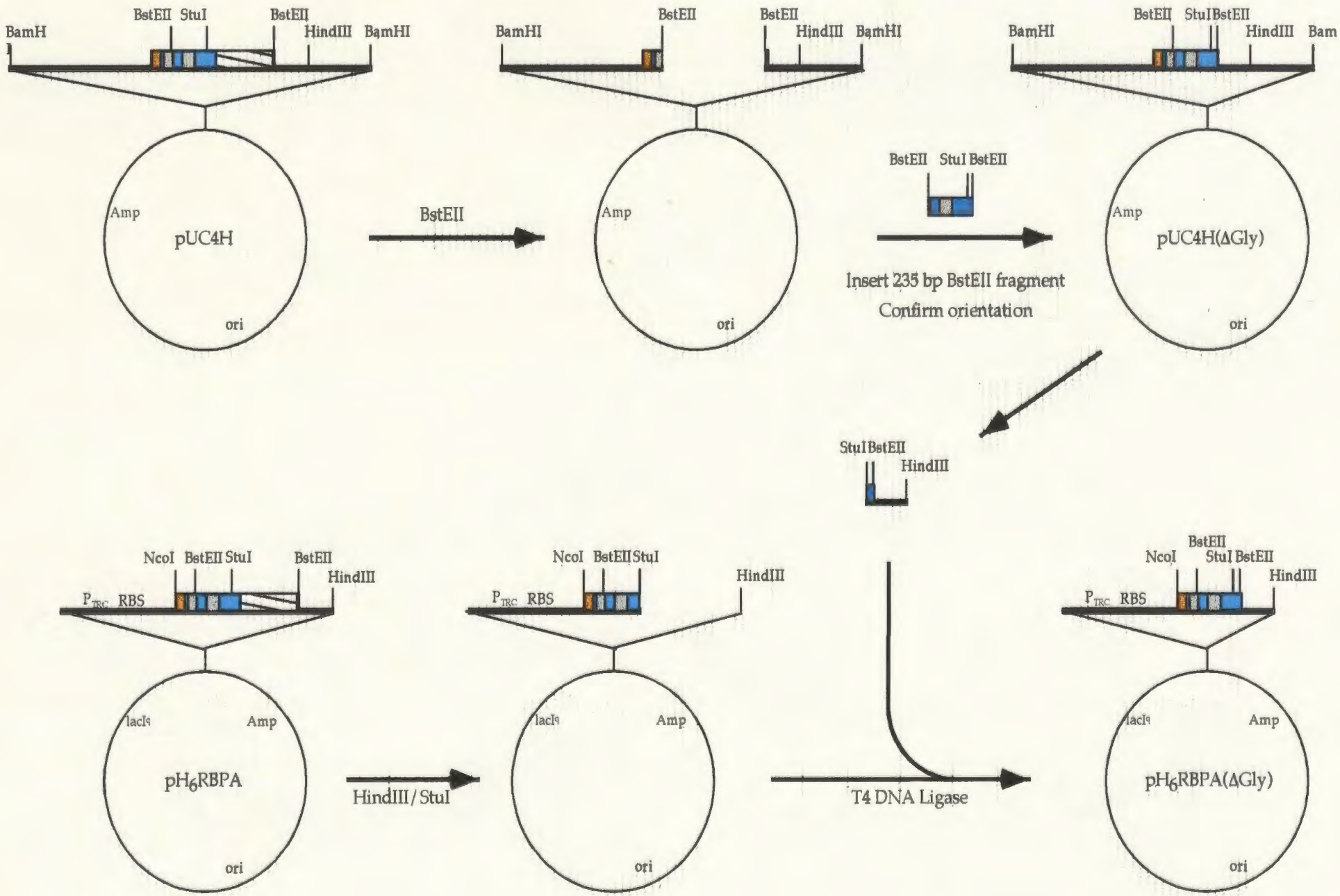

pUC4H
BstEII
1
2
3
4
5
700
200

Figure 5.10. Second step of the cloning strategy to construct a truncated *rbpA* gene. Digestion of pUC4H with BstEII removed a 290 bp DNA fragment containing most of the *rbpA* gene. This fragment was replaced with the 235 bp DNA fragment isolated from pTJB5H(Bst). This new plasmid was named pUC4H(ΔGly). The DNA fragment corresponding to the truncated C-terminal region of H_6RbpA was isolated from pUC4H(ΔGly) by digestion with HindIII and StuI. This fragment was used to replace the corresponding region in the expression vector pH_6RBPA. The result was the expression vector pH_6RBPA(ΔGly) which expresses the truncated H_6RbpA(ΔGly) protein. Abbreviations: P_{TRC}, strong promoter containing the *trp* (-35) and the *lac* UV5 (-10) region separated by 17 bp; RBS, *lacZ* ribosome binding site; $lacI^q$, *lac* repressor gene.


BamH
BstEII StuI
BstEII
HindIII
BamHI
Amp
pUC4H
ori
BstEII
BamHI
BstEII
HindIII
BamHI
Amp
ori
BstEII StuI BstEII
Insert 235 bp BstEII fragment
Confirm orientation
BamHI
BstEII StuI BstEII
HindIII
Bam
Amp
pUC4H(ΔGly)
ori
StuI BstEII
HindIII
NcoI BstEII StuI
BstEII
HindIII
P TRC RBS
lacIq
Amp
pH6RBPA
ori
HindIII / StuI
NcoI BstEII StuI
HindIII
P TRC RBS
lacIq
Amp
ori
T4 DNA Ligase
BstEII
NcoI
StuI BstEII
HindIII
P TRC RBS
lacIq
Amp
pH6RBPA(ΔGly)
ori

In the final step, pH_6RBPA was digested with HindIII and StuI in order to isolate a 205 bp fragment containing the auxiliary domain of the *rbpA* gene as well as the downstream flanking sequence. Insertion of this HindIII/StuI DNA fragment derived from pUC4H(ΔGly) restored the proper reading frame to the now truncated *rbpA* gene; the new plasmid was named pH_6RBPA(ΔGly). Clones were screened by digestion of plasmid DNA with NcoI and HindIII (**Figure 5.11**). All three clones contained the 378 bp DNA fragment containing the *rbpA(ΔGly)* gene.

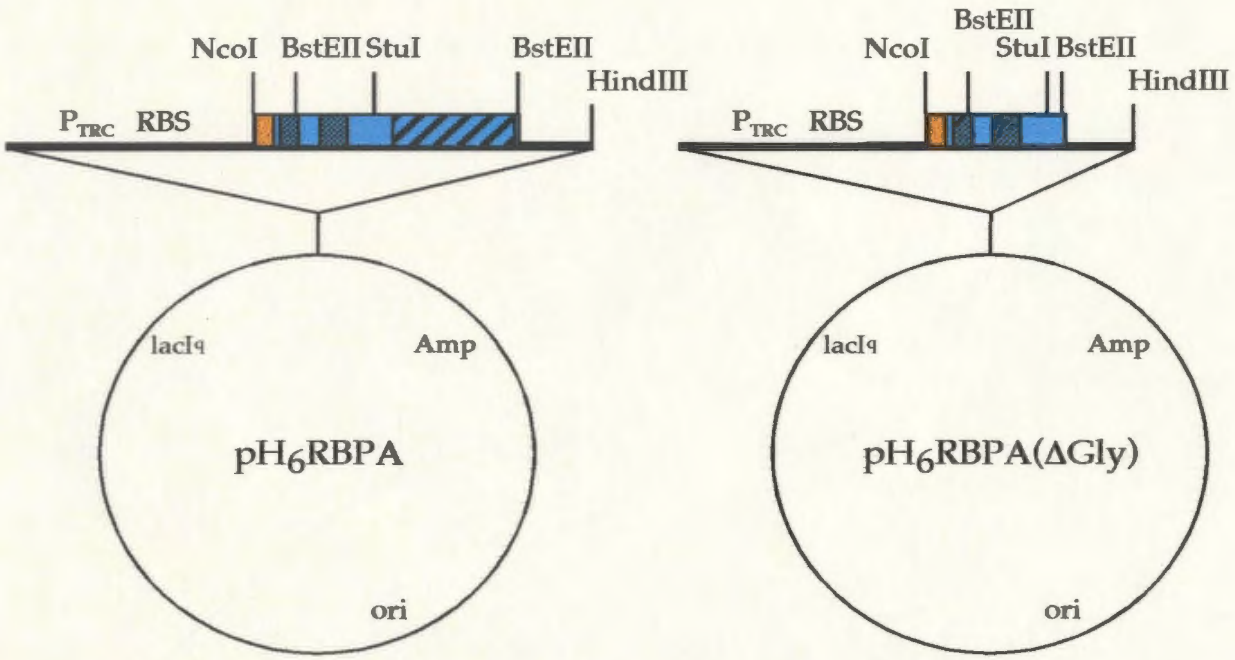
5.3.4. *In Vitro* Expression and Purification of Cyanobacterial RBPs

The molecular weight of both proteins was calculated using the inferred amino acid sequence of each recombinant protein and the "Compute pI/Mw" program on the EXPASY Molecular Biology Server (Bjellqvist *et al.*, 1993) (**Figure 5.12**). The molecular weight of H_6RbpA was calculated to be 12.2 kDa; the truncated protein H_6RbpA(ΔGly) had a theoretical molecular weight of 10.8 kDa. Both histidine-tagged RNA-binding proteins were expressed in *E. coli* BL21 and purified using the described protocol (Section 5.2.6 and 5.2.7). SDS-PAGE analysis of a preparation of both RNA-binding proteins is shown in **Figure 5.13**; a 500 mL *E. coli* culture would typically yield 0.5-1.0 mg of recombinant protein.

During the purification of these proteins, several interesting observations were made. First, SDS-PAGE electrophoresis of both proteins revealed that the proteins migrated at a rate more slowly than expected for the calculated molecular weight. The protein H_6RbpA had an apparent molecular weight of 14.4 kDa when run on an 8 - 25% gradient

Figure 5.11. (A). Restriction maps of expression plasmids pH_6RBPA and pH_6RBPA(ΔGly). Relevant restriction sites are shown. Within the *rbpA* gene, the histidine tag region is shown in orange, the RNP1 and RNP2 sequences are grey shaded and the auxiliary glycine rich domain is shown as a hatched box. **(B).** Screening of potential pH_6RBPA(ΔGly) clones by digestion with HindIII and NcoI and separation of DNA fragments on a 2% agarose gel. This digestion removed a 378 bp DNA fragment containing the entire truncated *rbpA(ΔGly)* gene as well as downstream flanking sequence. As a control, plasmid pH_6RBPA was also digested with HindIII and NcoI. This digestion removed a 433 bp DNA fragment which contained a full-length *rbpA* gene along with the downstream flanking sequence. An arrow indicates the location of the shorter 378 bp DNA fragment. The location of 100 bp DNA fragments used as size markers are indicated at the left.

A.


NcoI
BstEII
StuI
BstEII
HindIII
P TRC
RBS
lacIq
Amp
pH6RBPA
ori
BstEII
NcoI
StuI
BstEII
HindIII
P TRC
RBS
lacIq
Amp
pH6RBPA(ΔGly)
ori

B.

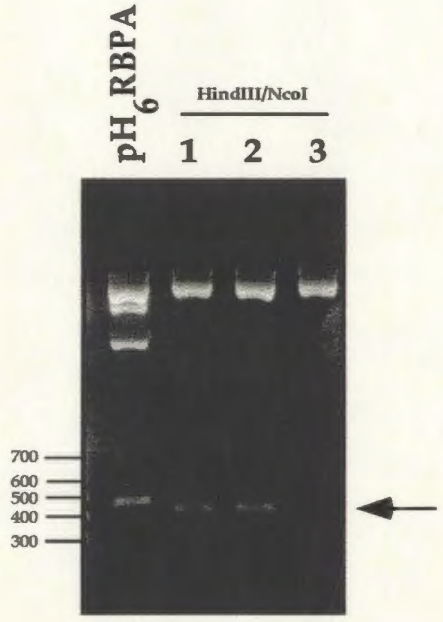

pH6RBPA
HindIII/NcoI
1
2
3
700
600
500
400
300

Figure 5.12. Inferred amino acid sequences of H_6RbpA and H_6RbpA(ΔGly). Amino acids belonging to the RNP1 and RNP2 conserved sequences are shaded. Glycine residues of the auxiliary domain are shown in bold type. Amino acids removed in the creation of protein H_6RbpA(ΔGly) are underlined.

H_6RbpA 114 amino acids Mol. Wt. 12,194

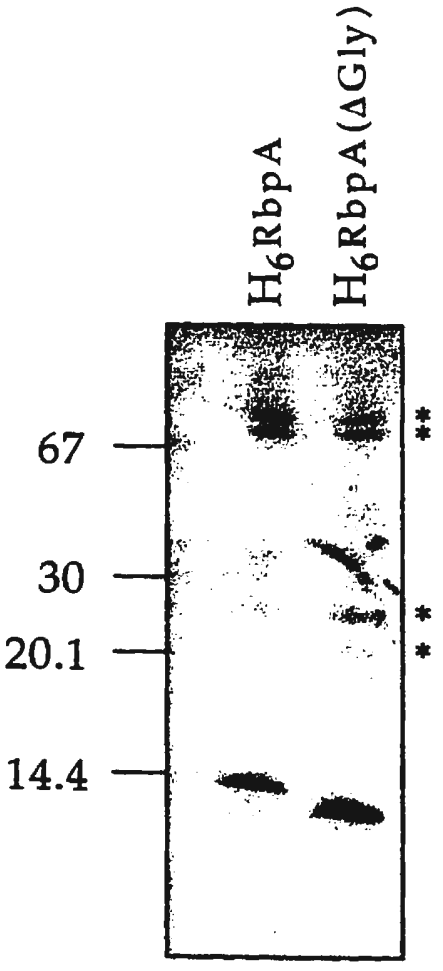
MSHHHHHHS[redacted]SYEVTEADLTAVFTEYGAVKRVQLPIDRETGRM[redacted]SADAEEDAAIAALDGAEWMGRGLRVNKAKPREERSGGGSFGGGGGRRGGGGGGGYRNY

H_6RbpA(ΔGly) 96 amino acids Mol. Wt. 10,792

MSHHHHHHS[redacted]SYEVTEADLTAVFTEYGAVKRVQLPIDRETGRM[redacted]SADAEEDAAIAALDGAEWMGRGLRVNKAKPREERSGYRNY

RNP2 RNP1

Figure 5.13. SDS-PAGE analysis of MCAC-purified fractions of H_6RbpA (lane 1) and H_6RbpA(ΔGly) (lane 2). Approximately 0.3 - 0.6 μg of recombinant protein was loaded in each lane and electrophoresed as described in Section 5.2.8. The location of molecular weight standards used as size markers are indicated at the left. Contaminating proteins are indicated to the right with an asterisk (*).


H6RbpA
H6RbpA(ΔGly)
67
30
20.1
14.4
**
**
*

polyacrylamide gel; the apparent molecular weight of the truncated H_6RbpA(ΔGly) was 13.2 kDa. These results could be explained in part by the observation that some proteins with histidine tags run more slowly on SDS-gels than equivalent untagged proteins, and therefore appear to be several kDa larger than expected (Qiagen, 1995).

Four contaminating proteins were observed in purifications of both H_6RbpA and H_6RbpA(ΔGly). Two of the co-purifying proteins had apparent molecular weights of 23 kDa and 25.5 kDa. The molecular weights of the other two proteins could not be calculated but were larger than 67 kDa. While the source of these proteins was unclear, several possibilities have been suggested (Hengen, 1995). A 21 kDa *E. coli* protein consisting of a domain homologous to FK506-binding proteins (FKBPs) had been observed as a contaminant in metal chelate affinity purification of recombinant proteins expressed in *E. coli* (Wulfing *et al.*, 1994). This protein had a histidine content of approximately 10% (20 out of 196 amino acid residues) and was known to bind tightly to Ni^{2+} and other divalent cations. This contrasted with the average *E. coli* protein which has a histidine content of approximately 1.8% (Niedhardt, 1987).

Another possible contamination in MCAC purification was the protein chloramphenicol acetyl transferase (CAT). In purifying a histidine-hexapeptide-tagged restriction endonuclease EcoRV expressed in *E. coli*, Oswald and Rinas (1996) identified a 25 kDa contaminant which co-purified with the fusion protein. N-terminal sequencing of this contaminant identified the protein as CAT, an enzyme which provides resistance to the antibiotic chloramphenicol. The enzyme itself has a high histidine content of

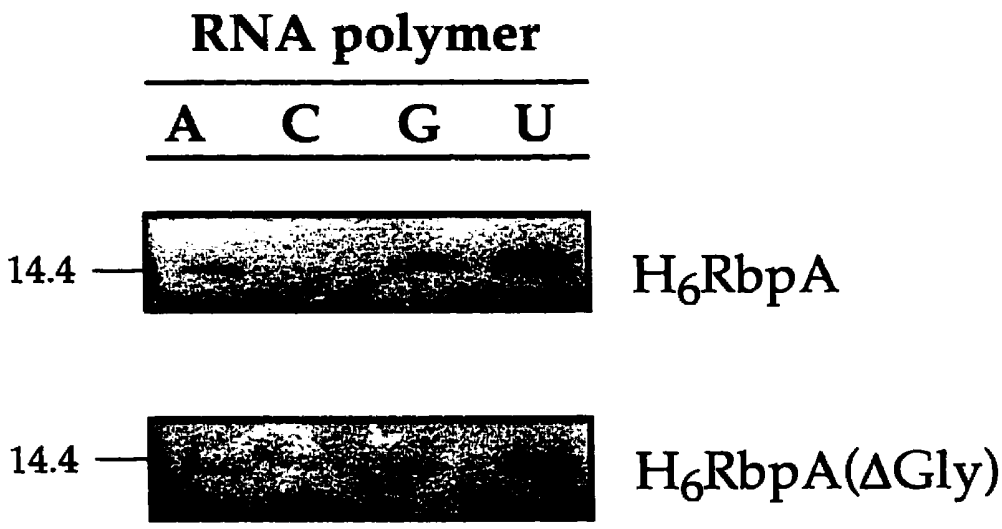
5.5% (12 of 219 amino acid residues). The gene coding for this enzyme is often used as a selection marker in expression systems. In *E. coli* BL21, the pLysS plasmid, which contains the *cat* gene, is necessary if expression of a cloned gene is directed by T7 RNA polymerase (Studier *et al.*, 1990). However, expression of *rbpA* was under the control of the P_{TRC} promoter in pH_6RBPA so that pLysS was not necessary for expression. Therefore, *E. coli* BL21 strains containing the pH_6RBPA and $pH_6RBPA(\Delta Gly)$ were cured of pLysS prior to induction by growing each on solid media in the absence of chloramphenicol until a chloramphenicol-sensitive strain was isolated. However, removal of plasmid pLysS had little effect on the presence of contaminating proteins in preparations of H_6RbpA and $H_6RbpA(\Delta Gly)$.

Another study revealed that some of the higher molecular weight protein contaminants observed in affinity purifications such as MCAC were multimeric forms of the fusion protein being expressed which had formed through disulfide bonds between cysteine residues (Stuber *et al.*, 1990). There is also some anecdotal evidence that His-tagged proteins are more prone to self-association and aggregation. This association may be mediated through trace metal ions in solution which bridge between the His tags. Other proteins have also been identified as contaminants in MCAC columns, including superoxide dismutase (Stuber *et al.*, 1990).

5.3.5. Nucleic Acid Binding Properties of H_6RbpA

RNA binding assays were carried out as described (Section 5.2.9). The results of protein binding to immobilized RNA polymers are shown in **Figure 5.14**. Both H_6RbpA and $H_6RbpA(\Delta Gly)$ were retained by columns

Figure 5.14. Binding of H_6RbpA and H_6RbpA(ΔGly) to immobilized RNA polymers. RNA binding assays were carried out as described in Section 5.2.9. The presence of protein in the sample eluted from each column was taken to indicate binding to the specific RNA polymer. The protein used for each assay is indicated at the right. The location of a 14.4 kDa molecular weight marker is shown at the left of each gel.


RNA polymer
A C G U
14.4
H6RbpA
14.4
H6RbpA(ΔGly)

which contained poly(A), poly(G) and poly(U) RNA but not poly(C) RNA. In fact, immobilized poly(U) RNA appeared to retain the greatest amount of protein suggesting that these proteins have the highest affinity for poly(U) RNA. The binding specificity is only partially consistent with that of other cyanobacterial Rbps characterized to date from *Synechococcus* 6301 (Sugita and Sugiura, 1994), *Synechoccus* 7002 (Sato *et al.*, 1997), and *Anabaena variabilis* M3 (Sato, 1995), all of which bound to poly(G) and poly(U) RNA. In studies of H_6RbpA and H_6RbpA(ΔGly), both proteins displayed an affinity for poly(A) RNA, a property not previously observed with the other cyanobacterial Rbps. Column retention of either H_6RbpA or H_6RbpA(ΔGly) in the absence of RNA polymer was found to be negligible.

Overall, these results confirmed that RbpA is a real RNA-binding protein; the protein had the ability both to bind to RNA and to discriminate between several RNA polymers. These particular binding properties did not appear to be affected by the introduction of a hexahistidine tag immediately adjacent to the RNP2 sequence. This had been a concern since it had been previously reported that histidine tags could affect the interaction of DNA binding protein NF-κB with its target sequence (Buning *et al.*, 1996). However, until the RNA target for this protein *in vivo* is identified, it cannot be ascertained whether or not binding to its true RNA substrate is affected by the modification.

5.3.6. Binding Properties of H_6RbpA(ΔGly) and the Role of the Auxiliary Glycine-Rich Domain

Specificity of binding to RNA homopolymers also did not appear to be affected by the removal of the auxiliary glycine-rich domain. The truncated protein H_6RbpA(ΔGly) showed the same specificity for binding to poly(A), poly(G) and poly(U) RNA polymers as the protein H_6RbpA. However, there did appear to be a decrease in the relative amount of H_6RbpA(ΔGly) protein retained by immobilized poly(G) RNA suggesting that the ability to bind to bind poly(G) RNA may have been partially affected by removal of the auxiliary domain (**Figure 5.14**). Nevertheless, it appears that the RNA binding specificity observed with this binding assay originates in the RRM.

As stated in Chapter 1, the presence of an RRM motif in an amino acid sequence has always been a good indicator of RNA binding activity. The RNP family consists of over 200 proteins which recognize a wide range of RNA sequences and structures. Binding studies with many members of the RNP family suggest that the presence of the RNP1 and RNP2 sequences provides a "platform" for general RNA binding while the determinants of specificity were usually located at another site within the RRM or even outside of the RRM altogether (Mattaj, 1993; Nagai and Mattaj, 1994). In the case of H_6RbpA and H_6RbpA(ΔGly), it appears that both the ability of the proteins to bind to RNA and their ability to specifically discriminate among RNA homopolymers were both properties contained in the RRM region of the proteins.

The function of the auxiliary glycine-rich domain therefore remains unknown. Protein interaction studies have provided some evidence

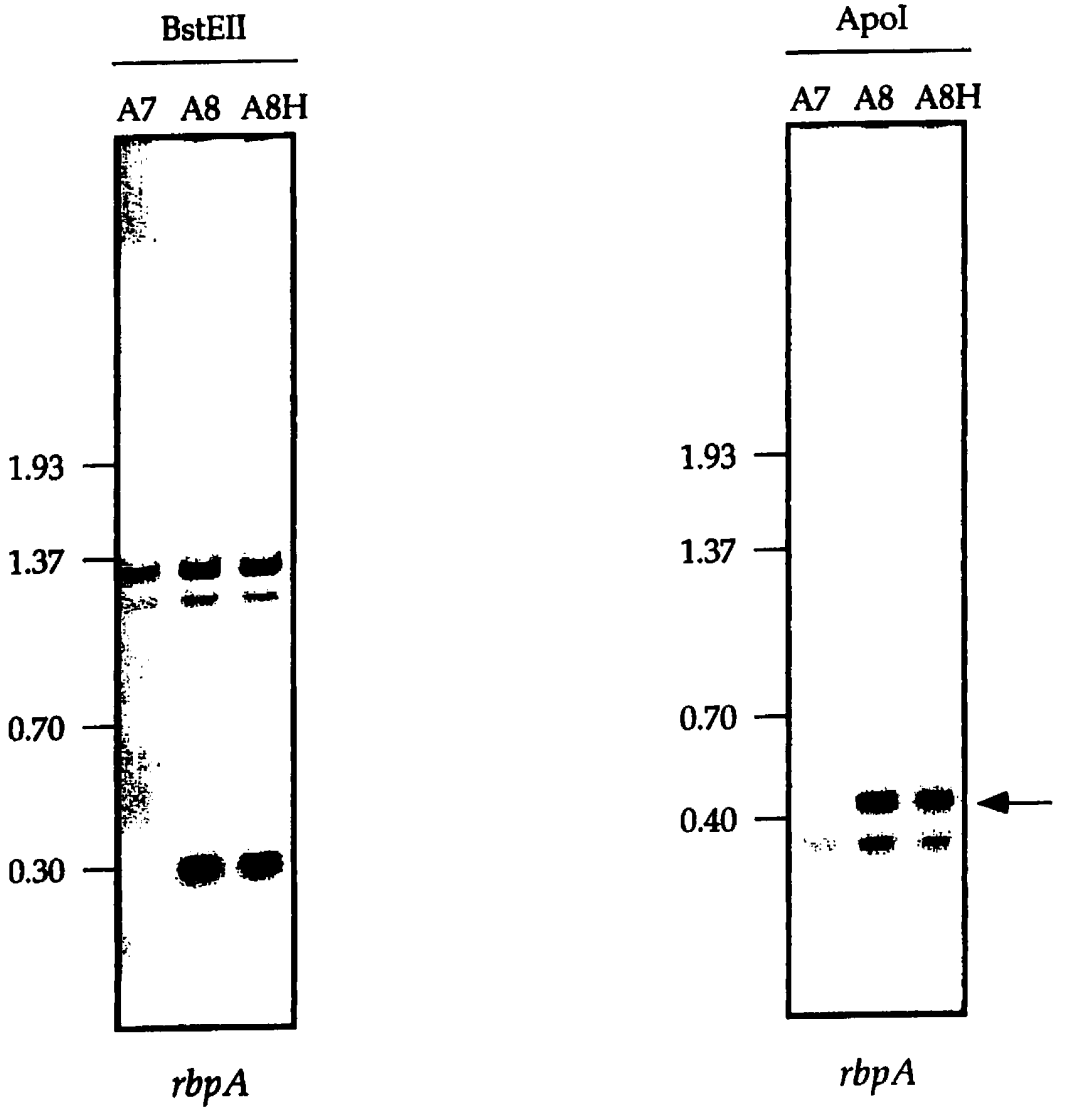
pointing to a role for this domain in protein-protein interactions (Section 1.2.3). It is therefore possible that in *Synechococcus* 7942, the RbpA protein may function in association with a complex of other factors which interact with RNA.

5.3.7. Characterization of SY-RBPA8H Mutant Phenotype

Although the incidence of histidine tags interfering with the normal biological activity of a protein are rare, at least one case has been reported for the DNA-binding protein NF-κB (Buning *et al.*, 1996). In the case of the cyanobacterial RNA-binding proteins H_6RbpA and H_6RbpA(ΔGly), only a single amino acid residue separates the histidine tag from the highly conserved RNP2 sequence. In order to test the functional integrity of a histidine-tagged RbpA product *in vivo*, the mutant SY-RBPA8H was created which contained a modified *rbpA* gene, containing the coding sequence for a hexahistidine tag, in the "neutral site" of the mutant cyanobacterial chromosome (Section 5.2.10).

The genotype of the SY-RBPA8H strain was confirmed by Southern blotting (**Figure 5.15**). Genomic DNA from strain SY-RBPA8H was digested separately with BstEII and ApoI and probed with the 290 bp BstEII *rbpA* probe outlined previously (see Figure 2.5). As expected, the pattern of hybridizing fragments observed in both BstEII-digested and ApoI-digested SY-RBPA8H genomic DNA was essentially indistinguishable from SY-RBPA8 genomic DNA. It was thought that it might be possible to distinguish the 402 bp hybridizing fragment observed in SY-RBPA8 from the larger 423 bp hybridizing fragment in SY-RBPA8H resulting from the additional coding

Figure 5.15. Analysis of SY-RBPA8 and SY-RBPA8H mutant genotypes by Southern hybridization. Genomic DNA from each strain was digested separately with BstEII and ApoI and probed with the 290 bp BstEII *rbpA* DNA probe. An arrow is used to indicate the location of the 423 bp hybridizing fragment corresponding to the modified *rbpA* gene in SY-RBPA8H. The location of BstEII-digested bacteriophage λ DNA, as well as a 100 bp DNA ladder (Gibco BRL), used as size markers are indicated at the left.


BstEII
A7 A8 A8H
1.93
1.37
0.70
0.30
rbpA
ApoI
A7 A8 A8H
1.93
1.37
0.70
0.40
rbpA

sequence for the histidine tag at the start of the *rbpA* gene. However, this was not possible.

Measurements of whole cell absorbance spectra of the mutant SY-RBPA8H indicated that this strain showed an increase in the ratio of phycobiliprotein (630 nm) to chlorophyll a (680 nm) absorbance peaks (ratio 1.04) relative to the knock-out strain SY-RBPA1 (**Figure 5.16**). Growth experiments carried out as described previously (Section 3.2.5) at 21°C indicated that the mutant SY-RBPA8H (approximate doubling time 2.5 days) grew at a rate comparable to both the wild type *Synechococcus* 7942 and the mutant SY-RBPA8 (**Figure 5.17**). This result confirmed that introduction of the modified *rbpA* gene into the neutral site of the cyanobacterial genome was sufficient to rescue the temperature-sensitive phenotype observed in SY-RBPA1. It is therefore likely that the presence of the histidine tag does not seriously impede the proper functioning of RbpA. Furthermore, the presence of a histidine tagged copy of the RbpA protein *in vivo* offers an opportunity in future experiments to isolate associated proteins which may co-purify with H_6RbpA from total protein preparations of SY-RBPA8H cells.

Figure 5.16. Representative whole cell spectra of SY-RBPA1 (green), SY-RBPA8 (purple) and SY-RBPA8H (orange) *Synechococcus* 7942 strains. As before, cells were diluted with BG-11 media to an absorbance of 0.2 prior to spectral measurement (see Figure 3.9).

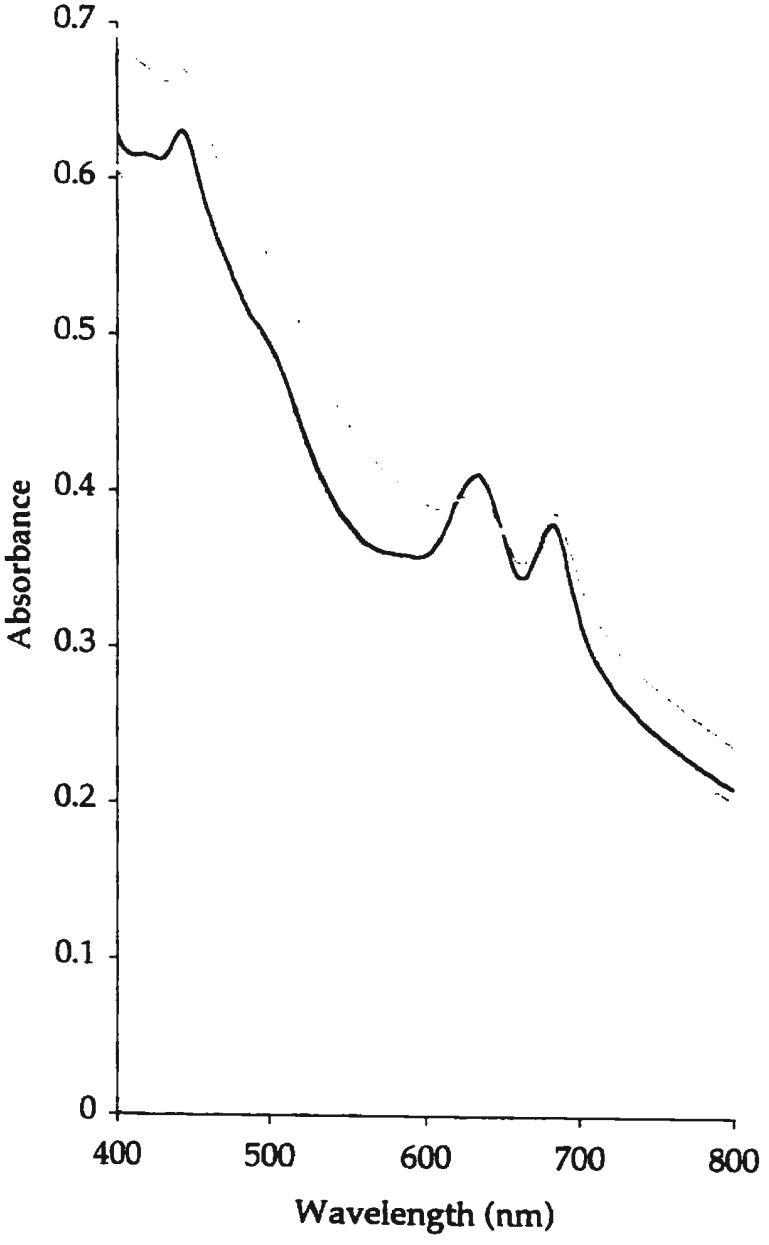
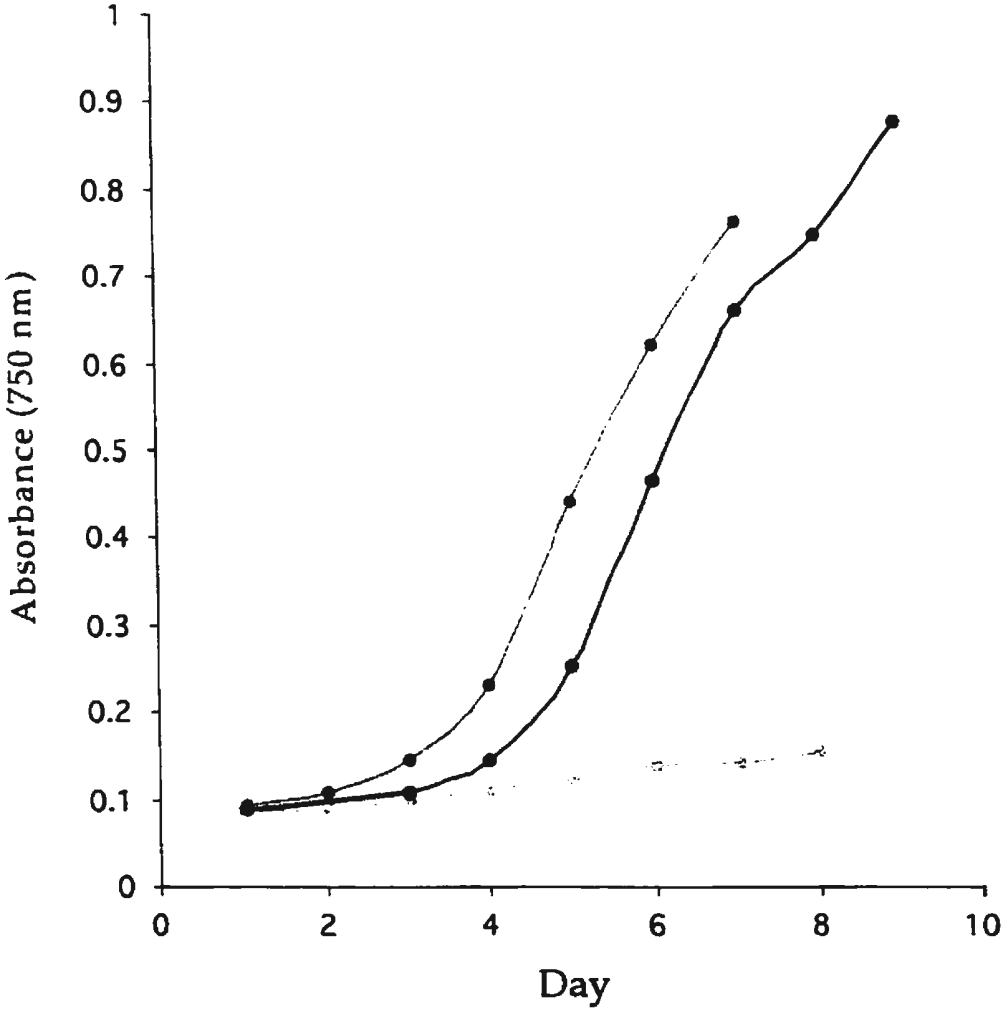

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Absorbance
400
500
600
700
800
Wavelength (nm)

Figure 5.17. Growth of SY-RBPA1 (green), SY-RBPA8 (purple) and SY-RBPA8H (orange) *Synechococcus* 7942 strain in BG-11 media. Cultures (100 mL) were incubated with shaking (200 rpm) under constant illumination at 21°C. The density of each culture was monitored by measuring light scattering at 750 nm. Each data point represents the average of absorbances taken for two cyanobacterial cultures grown in parallel under identical conditions.


Absorbance (750 nm)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
Day

CHAPTER 6. CONCLUSIONS

6.1. Possible Functions of Cyanobacterial Rbps

Evidence presented here suggests that RbpA is essential for growth at lower temperatures. Inactivation of the *rbpA* gene in mutant strain SY-RBPA1 resulted in a phenotype which grew very slowly at 21°C. This reduction in growth at 21°C was a feature of the SY-RBPA1 mutant phenotype which could be reversed by re-introduction of an intact copy of the *rbpA* gene into the neutral site of the cyanobacterial chromosome. Studies of *rbp* genes in several other strains of cyanobacteria, including *Synechococcus* 6301 (Sugita and Sugiura, 1994), *A. quadruplicatum* PR-6 (Sato *et al.*, 1997) and *A. variabilis* M3 (Sato, 1995) have also shown an increase in expression of the genes in response to colder temperatures. Secondary structural elements upstream of the cyanobacterial *rbp* genes have already been implicated in the regulation of expression of these genes by temperature (Sato and Wada, 1996). These proteins may therefore function by altering the pattern of gene expression in response to environmental stress such as temperature, either by selectively protecting mRNA transcripts from degradation or facilitating the degradation of transcripts through recruitment of other proteins. Alternatively, they may also function as "RNA chaperones" which maintain the correct secondary and tertiary structures of RNAs during a decrease in environmental temperature. They would likely function in conjunction with the cold shock-induced RNA helicase genes which have been identified in *Anabaena* 7120 and *Synechocystis* 6803 (Chamot *et al.*, 1999; Kaneko *et al.*, 1995).

The work presented in this thesis as well as the work of Vijay and Kovacs (1993) also suggest that this protein may function as a regulatory

component of the cyanobacterial phycobilisome complex. The inactivation of the *rbpA* gene in mutant SY-RBPA1 had an effect on both the phycobiliprotein composition, as determined by measurements of phycobiliprotein concentrations, and the ratio of the two phycobiliproteins (AP and PC) of *Synechococcus* 7942. Recovery of intact phycobilisome complexes by sucrose density step gradients was also decreased in the SY-RBPA1 mutant strain. The decrease in phycobiliprotein content was further confirmed by the measurements of ratios of phycobiliprotein to chlorophyll a absorbance peaks in whole cell absorbance spectra for each strain. These spectra demonstrated that the inactivation of the *rbpA* gene reduced the ratio of the phycobiliprotein (630 nm) to chlorophyll a (680 nm) absorbance peaks. Furthermore, this reduction could be reversed by introduction of an intact copy of the *rbpA* gene at an alternate "neutral" site in the cyanobacterial chromosome (mutants SY-RBPA6, SY-RBPA8 and SY-RBPA8H).

The function of many of the other cyanobacterial RNA-binding proteins is not known. However, surveys of a large number of cyanobacterial strains for the presence of *rbp* genes show that heterocyst-forming cyanobacteria have many more *rbp* genes than non-heterocyst-forming filamentous and unicellular cyanobacteria (see Figure 1.11 and Mulligan *et al.*, 1994). This suggests that cyanobacterial Rbps may play a role in RNA processing reactions during heterocyst differentiation, a period at which the entire spectrum of gene expression within a developing cell changes. Experiments with the cold-regulated RNA-binding protein (RbpA1) in *Anabaena variabilis* M3 showed that this protein was required for enhanced

repression of heterocyst initiation at low temperatures in the presence of nitrate (Sato and Wada, 1996).

6.2. RNA targets for Cyanobacterial Rbps?

Few experiments to date have been carried out to identify RNA sequences or secondary structures which could be specific targets for cyanobacterial Rbps? In immunoprecipitation experiments with total protein extracts of the cyanobacterium *Synechococcus leopoliensis*, an anti-RNP monoclonal antibody precipitated a number of proteins and several small RNAs, some of which carried the 5' $m^{2,2,7}G$ (m^3G) trimethyl cap structure (Kovacs *et al.*, 1993). This structure is usually observed only in certain species of snRNA and could be a site of interaction by an RNP-type RNA-binding protein.

More indirect experiments have also offered some clues as to possible RNA targets for cyanobacterial Rbps *in vivo*. First, cyanobacterial Rbps may act at the 3' end of mRNA transcripts. Cyanobacterial mRNA transcripts have not been well characterized to date. Few operons have the characteristic sequence of a bacterial *rho*-dependent terminator. In fact, several polyadenylated RNA transcripts have been detected in two species of cyanobacteria (Crouch *et al.*, 1983). Although the function of these poly(A) tracts (15 - 22 nt) is unclear, results showed that darkened cultures synthesized three-fold more poly(A) mRNA than cultures that were illuminated and that this RNA was more resistant to degradation by RNases. RNA processing events such as these could be facilitated by an RNP-type RNA-binding protein. It is known that *E. coli* cells contain a gene for poly (A) polymerase

(*pcnB*) and also utilize adenylation of bacterial RNA to influence decay of transcripts (reviewed in Cohen, 1995). Using as a probe a DNA fragment containing the *pcnB* gene of *E. coli*, our lab has determined that some cyanobacteria, including *Synechococcus* 7942, contain similar genes (Thomas and Mulligan, unpublished). Cyanobacterial Rbps may therefore function as a component of an RNA polyadenylation pathway.

Cyanobacterial Rbps may also act at the 5' end of mRNA transcripts. One set of candidate genes now being studied in *Synechococcus* 7942 are the *psbA* genes which encode different form of the reaction centre D1 protein of photosystem II (Golden *et al.*, 1986). The relative ratios of proteins (D1:1 and D1:2) have been demonstrated to be altered in response to different light and temperature conditions (Soitamo *et al.*, 1998). Several mRNA transcripts originating from these genes are subject to accelerated turnover when cells are exposed to high light intensities; this accelerated turnover is controlled by the untranslated leader regions of these transcripts (Kulkarni and Golden, 1997). These leader sequences may regulate mRNA stability through a mechanism which involves RNA-binding proteins.

It is also tempting to include splicing events among the possible processing reactions involving cyanobacterial Rbps. There have been no evidence for the existence of spliceosomal introns in cyanobacteria so far, although a small number of Group I and Group II introns have been identified (Ferat and Michel, 1993; Kuhsel *et al.*, 1990; Xu *et al.*, 1990). While these introns are self-splicing *in vitro*, it is possible that a cyanobacterial Rbp may facilitate this reaction *in vivo*.

6.3. Future Directions

Work presented here opens many directions for future research into the study of RbpA and cyanobacterial Rbps in general. One of the most important next steps is to identify the RNA species bound by RbpA in *Synechococcus* 7942. Our lab has begun experiments to identify RNA sequences which exhibit tight-binding to H_6RbpA using a selective amplification technique known as SELEX (Tuerk and Gold, 1990). This technique, which isolates nucleic acid sequences from a random pool based on specific properties, has been used in the study of other nucleic acid binding proteins, including the RNP-type protein U1-A (Tsai *et al.*, 1991). RNA sequences identified by this technique can then be used as probes to locate similar sequences in the cyanobacterial genome.

Another important aspect of future research is to identify the amino acid residues which are critical to the function of RbpA. Future experiments could use site-directed mutagenesis to modify specific amino acids and examine the effect of these changes on the biological activity of the protein. We can then quantitate these effects in two ways. First, we can measure the binding affinity of a mutated RbpA protein for its RNA target. Second, the "neutral-site" in the chromosome of mutant SY-RBPA1 gives us a system by which we can measure the ability of a mutated *rbpA* gene to complement the mutant phenotype. This system has already been used in mutant SY-RBPA8H to confirm that the introduction of a histidine-tag did not seriously impede the proper functioning of this protein *in vivo* (see Section 5.3.7). I am currently in the process of characterizing another *Synechococcus* 7942 mutant SY-RBPA8H(ΔGly) which contains the *rbpA(ΔGly)* gene incorporated into the

neutral site of SY-RBPA1. These results will determine whether or not the truncated form of RbpA can complement the mutant phenotype in SY-RBPA1.

Finally, it is important to identify other proteins which may interact with RbpA as part of its biological function. Many of the eukaryotic RNP-type RNA-binding proteins, such as snRNPs and hnRNPs, carry out RNA processing reactions as part of a larger complex of proteins. The Gly-rich auxiliary domains identified in RNP-type RNA-binding proteins have previously been implicated in assembly of protein complexes via protein-protein interactions (Cartegni et al., 1996; Wang *et al.*, 1997). The *Synechococcus* 7942 mutant SY-RBPA8H makes it possible to isolate the H_6RbpA protein, along with any associated proteins, from total extracts of these cells. The isolation of histidine-tagged proteins under non-denaturing conditions has previously been shown to be a convenient method for identifying protein-protein interactions (Hoffman and Roeder, 1991). Furthermore, comparison with proteins isolated from SY-RBPA8H(ΔGly) can be used to determine whether these protein-protein interactions are a property of the auxiliary domain.

REFERENCES

Adam, S. A., Nakagawa, T. Y., Swanson, M. S., Woodruff, T., and Dreyfuss, G. (1986): mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* **6**, 2932-2943.

Allain, F. H. T., Howe, P. W. A., Neuhaus, D., and Varani, G. (1997): Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.* **16**, 5764-5774.

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

Allen, M. M., and Stanier, R. Y. (1968): Selective isolation of blue-green algae from water and soil. *J. Gen. Microbiol.* **51**, 203-209.

Appel, R. D., Bairoch, A., and Hochstrasser, D. F. (1994): A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.* **19**, 258-260.

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

Bachmair, A., Finley, D., and Varshavsky, A. (1986): *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186.

Bandziulis, R. J., Swanson, M. S., and Dreyfuss, G. (1989): RNA-binding proteins as developmental regulators. *Genes Devel.* **3**, 431-437.

Baserga, S. J., and Steitz, J. A. (1993): The diverse world of small ribonucleoproteins, pp. 359-. In R. F. Gesteland, and J. F. Atkins (Eds): *The RNA World*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.

Bell, L. R., Maine, E. M., Schedl, P., and Cline, T. W. (1988): Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037-1046.

Benson, D. A., Boguski, M. S., Lipman, D. J., Ostell, J., and Ouelette, B. F. (1998): Genbank. *Nucleic Acids Res.* **26**, 1-7.

Bentley, R. C., and Keene, J. D. (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**, 1829-1839.

Bhalerao, R. P., Lind, L. K., and Gustafsson, P. (1994): Cloning of the *cpcE* and *cpcF* genes from *Synechococcus* sp. PCC 6301 and their inactivation in *Synechococcus* sp. PCC 7942. *Plant Mol. Biol.* **26**, 313-326.

Bhalerao, R. P., Lind, L. K., Persson, C. E., and Gustafsson, P. (1993): Cloning of the rod linker genes from the cyanobacterium *Synechococcus sp.* PCC 6301 and their inactivation in *Synechococcus sp.* PCC 7942. *Mol. Gen. Genet..***237**, 89-96.

Biamonti, G., and Riva, S. (1994): New insight into the auxiliary domains of eukaryotic RNA binding proteins. *FEBS Lett.* **340**, 1-8.

Biamonti, G., Buvoli, M., Bassi, M. T., Morandi, C., Cobianchi, F., and Riva, S. (1989): Isolation of an active gene encoding human hnRNP protein A1. Evidence for alternative splicing. *J. Mol. Biol.* **207**, 491-503.

Birney, E., Kumar, S., and 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**, 5803-5816.

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., and Hochstrasser, D. F. (1993): The focusing of positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14**, 1023-1031.

Boelens, W. C., Jansen, E. J. R., Van Venrooij, W. J., Stripecke, R., Mattaj, I. W., and Gunderson, S. I. (1993): The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell* **72**, 881-892.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., and Boyer, H. L. (1977): Construction of useful cloning vectors. *Gene* **2**, 95-113.

Borthakur, D., and Haselkorn, R. (1989): *Tn*5 mutagenesis of *Anabaena sp.* strain PCC 7120: Isolation of a new mutant unable to grow without combined nitrogen. *J. Bacteriol.* **171**, 5759-5761.

Breiteneder, H., Michalowski, C. B., and Bohnert, H. J. (1994): Environmental stress-mediated differential 3' end formation of chloroplast RNA-binding protein transcripts. *Plant Mol. Biol.* **26**, 833-849.

Breslauer, K. J., Frank, R., Blocker, H., and Marky, L. A. (1986): Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. USA* **83**, 3746-50.

Bugler, B., Bourbon, H., Lapeyre, B., Wallace, M. O., Chang, J. H., Amalric, F., and Olson, M. O. J. (1987): RNA bindng fragments from nucleolin contain the ribonucleoprotein consensus sequence. *J. Biol. Chem.* **262**, 10922.

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

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

Buning, H., Gartner, U., von Schack, D., Baeuerle, P. A., and Zorbas, H. (1996): The histidine tail of a recombinant DNA binding protein may influence the quality of interaction with DNA. *Anal. Biochem.* **234**, 227-230.

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

Burd, C. G., Swanson, M. S., Goerlach, M., and Dreyfuss, G. (1989): Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: A diversity of RNA binding proteins is generated by small peptide inserts. *Proc. Natl. Acad. Sci. USA* **86**, 9788-9792.

Bustos, S. A., and Golden, S. S. (1992): Light-regulated expression of the *psbD* gene family in *Synechococcus sp.* PCC 7942: Evidence for the role of duplicated *psbD* genes in cyanobacteria. *Mol. Gen. Genet.* **232**, 221-230.

Buvoli, M., Cobianchi, F., and Riva, S. (1992): Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosomal assembly. *Nucleic Acids Res.* **20**, 5017-5025.

Buvoli, M., Cobianchi, F., Biamonti, G., and Riva, S. (1990): Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligonucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res.* **18**, 6595-6600.

Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M., Ghetti, A., Riva, S., and Morandi, C. (1988): cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. *Nucleic Acids Res.* **16**, 3751-3770.

Cai, Y. P., and Wolk, C. P. (1990): Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. *J. Bacteriol.* **172**, 3138-3145.

Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Biamonti, G. (1996): hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* **259**, 337-348.

Chang, D. D., and Sharp, P. A. (1989): Regulation by HIV Rev depends upon recognition of splice sites. *Cell* **59**, 789-795.

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

Choi, Y. D., Grabowski, P. J., Sharp, P. A., and Dreyfuss, G. (1986): Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science* **231**, 1534-1539.

Choi, Y. D., and Dreyfuss, G. (1984): Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. *J. Cell. Biol.* **99**, 1997-2004.

Chomczynski, P. (1992): One-hour downward capillary transfer for blotting of DNA and RNA. *Anal. Biochem.* **201**, 134-139.

Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M., and Riva, S. (1993): Phosphorylation of human hnRNP protein A1 abrogates *in vitro* strand annealing activity. *Nucleic Acids Res.* **21**, 949-955.

Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., and Wilson, S. H. (1988): Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. *J. Biol. Chem.* **263**, 1063-1071.

Cohen, S. N. (1995): Surprises at the 3' end of prokaryotic RNA. *Cell* **80**, 829-832.

Crouch, D. H., Ownby, J. D., and Carr, N. G. (1983): Polyadenylated RNA in two filamentous cyanobacteria. *J. Bacteriol.* **156**, 979-982.

Cruz-Alvarez, M., and Pellicer, A. (1987): Cloning of a full-length complementary DNA for an *Artemia salina* glycine-rich protein. *J. Biol. Chem.* **262**, 13377-13380.

Currier, T. C., Haury, J. F., and Wolk, C. P. (1977): Isolation and preliminary characterization of auxotrophs of a filamentous cyanobacterium. *J. Bacteriol.* **129**, 1556-1562.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A., and Valerio, R. (1989): Human immunodeficiency virus I *tat* protein binds *trans*-activation-responsive region (TAR) RNA *in vitro. Proc. Natl. Acad. Sci. USA* **86**, 6925.

Dolganov, N. A. and Grossman, A.R. (1994): The gene encoding an RNA-binding protein from *Synechococcus sp.* Strain PCC 7942: Homology between cyanobacterial and nuclear-encoded chloroplast RNAS-binding proteins. Genbank Database entry L25435.

Doudna, J. A. (1997): RNA structure: a molecular contortionist? *Nature* **388**, 830-831.

Doudna, J. A., and Cate, J. H. (1997): RNA structure: crystal clear? *Curr. Opin. Struct. Biol.* **7**, 310-316.

Doudna, J. A., and Doherty, E. A. (1997): Emerging themes in RNA folding. *Fold. Des.* **2**, R65-R70.

Dreyfuss, G. (1993): hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289-321.

Dreyfuss, G., Adam, S. A., and Choi, Y. D. (1984): Physical changes in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. *Mol. Cell. Biol.* **4**, 415-423.

Dunn, M. A., Brown, K., Lightowlers, R., and 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**, 947-959.

Elhai, J., and Wolk, C. (1990): Developmental regulation and spatial pattern of expression of the structural genes for nitrogenase in the cyanobacterium *Anabaena*. *EMBO J.* **9**, 3379-3388.

Ferat, J.-L., and Michel, F. (1993): Group II self-splicing introns in bacteria. *Nature* **364**, 358-361.

Fung, P. A., Labrecque, R., and Pederson, T. (1997): RNA-dependent phosphorylation of a nuclear RNA binding protein. *Proc. Natl. Acad. Sci. USA* **94**, 1064-1068.

Garcia-Pichel, F. (1998): Solar ultraviolet and the evolutionary history of cyanobacteria. *Orig. Life Evol. Biosph.* **28**, 321-47.

Garrett, D. J. (1994): Determination of the secondary structure and folding topology of an RNA-binding domain of mammalian hnRNP A1 protein using three-dimensional heteronuclear resonance spectroscopy. *Biochemistry* **33**, 2852-2858.

Glazer, A. N. (1989): Light guides: directional energy transfer in a photosynthetic antennae. *J. Biol. Chem.* **264**, 1-4.

Glazer, A. N. (1988): Phycobilisomes. *Methods Enzymol.* **167**, 304-312.

Glazer, A. N. (1985): Light harvesting by phycobilisomes. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 47-77.

Glazer, A. N., Fang, S., and Brown, D. M. (1973): Spectroscopic properties of C-phycocyanin and of its α and β subunits. *J. Biol. Chem.* **248**, 5679-5685.

Golden, J. W., and Wiest, D. R. (1988): Genome rearrangement and nitrogen fixation in *Anabaena* blocked by inactivation of *xisA* gene. *Science* **242**, 1421-1423.

Golden, S. S., Nalty, M. S., and Cho, D. S. (1989): Genetic relationship of two highly studied *Synechococcus* strains designated *Anacystis nidulans*. *J. Bacteriol.* **171**, 24-29.

Golden, S. S. (1988): Mutagenesis of cyanobacteria by classical and gene-transfer-based methods. *Methods Enzymol.* **167**, 714-727.

Golden, S. S., Brusslan, J., and Haselkorn, R. (1986): Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J.* **5**, 2789-2798.

Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994): The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C Proteins. *J. Biol. Chem.* **269**, 23074-23078.

Gorlach, M., Wittekind, M., Beckman, R. A., Mueller, L., and Dreyfuss, G. (1992): Interaction of the RNA-binding domain of the hnRNP C proteins with RNA. *EMBO J.* **11**, 3289-3295.

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

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

Gray, M. W. (1993): Origin and evolution of organelle genomes. *Curr. Opin. Genet. Dev.* **3**, 884-890.

Gray, M. W. (1989): The evolutionary origins of organelles. *Trends Genet.* **5**, 294-299.

Habets, W. J., Sillikens, P. T. G., Hoet, M. H., Schalken, J. A., Roebroek, A. J. M., Leunissen, J. A. M., Van de Ven, W. J., and Van Venrooij, W. J. (1987): Analysis of a cDNA clone expressing a human autoimmune antigen: full length sequence of the U2 small nuclear RNA-associated B" antigen. *Proc. Natl. Acad. Sci. USA* **84**, 2421-2425.

Hamm, J., Dathan, N. A., Scherly, D., and Mattaj, I. W. (1990): Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J.* **9**, 1237-1244.

Hamm, J., Van Santen, V. L., Spritz, R. A., and Mattaj, I. W. (1988): Loop I of U1 small nuclear RNA is the only RNA sequence for binding of specific U1 small nuclear ribonucleoprotein particle proteins. *Mol. Cell. Biol.* **7**, 4030.

Hanahan, D. (1985): Techniques for transformation of *E. coli*, pp. 109-135. In D. M. Glover (Ed.): *DNA Cloning*, IRL Press Ltd., London.

Hayes, J. M. (1983): Geochemical evidence bearing on the origin of aerobiosis, a speculative hypothesis, pp. 291-300. In J. W. Schopf (Ed.): *The Earth's earliest biosphere, its origins and evolution*, Princeton University Press, Princeton.

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

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

Hengen, P. N. (1995): Purification of His-tag fusion proteins from *Escherichia coli*. *Trends Biochem. Sci.* **20**, 285-286.

Hochuli, E. (1989): Aufarbeitung von Bioproteinen: Elegant und wirtschaftlich. *Chemische Industrie* **12**, 69-70.

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988): Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbant. *Biotechnology* **6**, 1321-1325.

Hochuli, E., Dobeli, H., and Schacher, A. (1987): New metal chelate adsorbant selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatogr.* **411**, 177-184.

Hoffman, A., and Roeder, R. (1991): Purification of His-tagged proteins in nondenaturing conditions suggests a convenient method for protein interaction studies. *Nucleic Acids. Res.* **19**, 6337-6338.

Holcomb, E. R., and Friedman, D. L. (1984): Phosphorylation of the C-proteins of HeLa cell hnRNP particles. Involvement of a casein kinase II-type enzyme. *J. Biol. Chem.* **259**, 31-40.

Holden, D. C. (1995): Sequence determination and characterization of genes encoding RNA-binding proteins in the cyanobacterial species *Anabaena sp.* PCC 7120., B. Sc. (Hon.) Thesis, Memorial University of Newfoundland.

Hope, I. A., and Struhl, K. (1986): Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46**, 885-894.

Horabin, J. I., and Schedl, P. (1993): Sex-lethal autoregulation requires multiple *cis*-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site. *Mol. Cell. Biol.* **13**, 7734-7746.

Howe, P., Nagai, K., Neuhaus, D., and Varani, G. (1994): NMR-studies of U1 snRNA recognition by the N-terminal RNP domain of the human U1A protein. *EMBO J.* **13**, 3873-3881.

Inoue, M., Muto, Y., Sakamoto, H., Kigawa, T., Takio, K., Shimura, Y., and Yokoyama, S. (1997): A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino terminal RNA-binding domain of Drosophila Sex-lethal. *J. Mol. Biol.* **272**, 82-94.

Inoue, K., Hoshijima, K., Sakamoto, H., and Shimura, Y. (1990): Binding of the *Drosophila* sex-lethal gene product to the alternative splice site of transformer primary transcript. *Nature* **344**, 461-463.

Iyer, R. (1993): Merlin DNA purification system. *bionet.molbio.methds-reagnts.*

Jaeger, J. A., Turner, D. H., and Zuker, M. (1989): Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* **86**, 7706-7710.

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A., and Stunnenberg, H. G. (1991): Rapid and efficient purification of native histidine tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci. USA* **88**, 8972-8976.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., and Nagai, K. (1991): Identification of the molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, 3447-3456.

Kalla, R., Bhalerao, R. P., and Gustafsson, P. (1993): Regulation of phycobilisome rod proteins and mRNA at different light intensities in the cyanobacterium *Synechococcus* 6301. *Gene* **126**, 77-83.

Kalla, R., Lind, L. K., and Gustafsson (1989): Genetic analysis of phycobilisome mutants in the cyanobacterium *Synechococcus sp.* PCC 6301. *Mol. Microbiol.* **3**, 339-347.

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M., and Tabata, S. (1995): Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC 6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.* **2**, 153-166.

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

Kim, S. (1997): Identification of N-G-methylarginine residues in human heterogeneous RNP protein A1. *Biochemistry* **36**, 5185-5192.

Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J., and Rich, A. (1973): Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* **179**, 285-288.

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

Klaff, P., Riesner, D., and Steger, G. (1996): RNA structure and the regulation of gene expression. *Plant Mol. Biol.* **32**, 89-106.

Kormanec, J., and Farkasovsky, M. (1994): Isolation of total RNA from yeast and bacteria and detection of rRNA in Northern blots. *Biotechniques* **17**, 839-842.

Kovacs, S. A., O'Neil, J., Watcharapijarn, J., Moe-Kirvan, C., Vijay, S., and Silva, V. (1993): Eubacterial components similar to small nuclear ribonucleoproteins: Identification of immunoprecipitable proteins and capped RNAs in a cyanobacterium and a gram-positive eubacterium. *J. Bacteriol.* **175**, 1871-1878.

Kranz, J. K., Lu, J., and Hall, K. B. (1996): Contribution of the tyrosines to the structure and function of the human U1A N-terminal RNA binding domain. *Protein Sci.* **5**, 1567-1583.

Kuhlemeier, C. J., and van Arkel, G. A. (1987): Host-vector systems for gene cloning in cyanobacteria. *Methods Enzymol.* **153**, 199-215.

Kuhlemeier, C. J., Thomas, A. A., van der Ende, A., van Leen, R. W., Borrias, W. E., van den Hondel, C. A., and van Arkel, G. A. (1983): A host-vector system for gene cloning in the cyanobacterium *Anacystis nidulans* R2. *Plasmid* **10**, 156-163.

Kuhsel, M. G., Strickland, R., and Palmer, J. D. (1990): An ancient group I intron shared by eubacteria and chloroplasts. *Science* **250**, 1570-1573.

Kulkarni, R.D. and Golden, S.S. (1997): mRNA stability is regulated by a coding-region element and the unique 5' untranslated leader sequences of the three *Synechococcus psbA* transcripts. *Mol. Microbiol.* **24**, 1131-1142.

Kumar, A., and Wilson, S. H. (1990): Studies of the strand-annealing activity of mammalian hnRNP complex protein A1. *Biochemistry* **2**, 10717-10722.

Kuritz, T., Ernst, A., Black, T. A., and Wolk, C. P. (1993): High resolution mapping of genetic loci of *Anabaena* PCC 7120 required for photosynthesis and nitrogen fixation. *Mol. Microbiol.* **8**, 101-110.

Laing, L. G., Gluick, T. C., and Draper, D. E. (1994): Stabilization of RNA structure by Mg ions: specific and non-specific effects. *J. Mol. Biol.* **237**, 577-587.

Lapeyre, B., Bourbon, H., and Amalric, F. (1987): Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. *Proc. Natl. Acad. Sci. USA* **84**, 1472-1476.

Lee, A. L., Kanaar, R., Rio, D. C., and Wemmer, D. E. (1994): Resonance assignments and solution structure of the second RNA-binding domain of sex-lethal determined by multidimensional heteronuclear magnetic resonance. *Biochemistry* **33**, 13775-13786.

Legrain, P., and Robash, M. (1989): Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57**, 573-583.

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

Li, R., and Golden, S. S. (1993): Enhancer activity of light-responsive regulatory elements in the untranslated leader regions of cyanobacterial *psbA* genes. *Proc. Natl. Acad. Sci. USA* **90**, 11678-11682.

Los, D. A., Ray, M. K., and Murata, N. (1997): Differences in the control of the temperature-dependent expression of four genes for desaturases in *Synechocystis sp.* PCC 6803. *Mol. Microbiol.* **6**, 1167-75.

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

Lutz-Freyermuth, C., Query, C. C., and Keene, J. D. (1990): Quantitative determination that one of two potential RNA-binding domains of the A protein of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA* **87**, 6393-6397.

Madueno, F., Borrias, W. E., van Arkel, G. A., and Guerrero, M. G. (1988): Isolation and characterization of *Anacystis nidulans* R2 mutants affected in nitrate assimilation: Establishment of two new mutant types. *Mol. Gen. Genet.* **213**, 223-228.

Mattaj, I. W. (1993): RNA recognition: A family matter? *Cell* **73**, 837-840.

Mayeda, A., and Krainer, R. (1992): Regulation of altenative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365-375.

McMaster, G. K., and Carmichael, G. G. (1977): Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**, 4835-4838.

Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., and 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**, 3307-3313.

Michelsen, B. (1995): Transformation of *Escherichia coli* increases 260-fold upon inactivation of T4 DNA ligase. *Anal. Biochem.* **225**, 172-174.

Moore, M. J., Query, C., and Sharp, P. A. (1993): Splicing of precursors to mRNA by the spliceosome, pp. 303-357. In R. F. Gesteland, and J. F. Atkins (Eds): *The RNA World*, Cold Spring Harbour Press, New York.

Mulligan, M. E., and Belbin, T. J. (1995): Characterization of RNA-binding protein genes in cyanobacteria. *Nucleic Acids Symp. Ser.* **33**, 140-142.

Mulligan, M. E., Jackman, D. M., and 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.

Munroe, S. H., and Dong, X. (1992): Heterogeneous nuclear ribonucleoprotein A1 catalyzes RNA-RNA annealing. *Proc. Natl. Acad. Sci. USA* **89**, 895.

Murata, N., and Wada, H. (1995): Acyl-lipid desaturases and their importance in the tolerance and acclimatization to cold of cyanobacteria. *Biochem. J.* **308**, 1-8.

Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barrett, S. F., Wilson, S. H., and Williams, K. R. (1991): Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative UP1, with RNA and DNA: evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry* **30**, 2968-2976.

Nagai, K., Oubridge, C., Ito, N., Avis, J., and 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., and Mattaj, I. W. (1994): *RNA-Protein Interactions*, Oxford University Press, Oxford.

Nagai, K., and Mattaj, I. W. (1994b): RNA-protein interactions in the splicing snRNPs, pp. 150-173. In K. Nagai, and I. W. Mattaj (Eds): *RNA-Protein Interactions*, Oxford University Press, New York.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990): Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515-520.

Nagoshi, R. N., McKeown, M., Burtis, K. C., Belote, J. M., and Baker, B. S. (1988): The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. *Cell* **53**, 229-236.

Nakamura, Y., Kaneko, T., Hirosawa, M., Miyajima, N., and Tabata, S. (1998): CyanoBase, a WWW database containing the complete nucleotide sequence of the genome of *Synechocystis sp.* strain PCC 6803. *Nucleic Acids Res.* **26**, 63-67.

Niedhardt, F. C. (1987): , pp. 3-6. In F. C. Niedhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (Eds): *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, Am. Soc. Microbiol., Washington, D.C.

Noller, H. F. (1991): Ribosomal RNA and translation. *Annu. Rev. Biochem.* **60**, 191-227.

Omata, T., Andriesse, X., and Hirano, A. (1993): Identification and characterization of a gene cluster involved in nitrate transport in the cyanobacterium *Synechococcus sp.* PCC 7942. *Mol. Gen. Genet.* **236**, 193-202.

Oswald, T., and Rinas, U. (1996): Chloramphenicol resistance interferes with purification of histidine-tagged fusion proteins from recombinant *Escherichia coli*. *Anal. Biochem.* **236**, 357-358.

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

Panoff, J. M. (1993): Control of a locus that is required for growth of *Anabaena* PCC 7120 at low temperature. *Curr. Microbiol.* **27**, 273-276.

Patton, J. R., Habets, W., Van Venrooij, W. J., and Pederson, T. (1989): U1 small nuclear ribonucleoprotein particle-specific proteins interact with the first and second stem-loops of U1 snRNA, with the A protein binding directly to the RNA independently of the 70K and Sm proteins. *Mol. Cell. Biol.* **9**, 3360.

Pearson, W. R., and Lipman, D. J. (1988): Improved tools for biological sequence analysis. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.

Petty, K. J. (1987): 10.11B. Metal Chelate Affinity Chromatography. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (Eds): *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York.

Pinol-Roma, S. (1997): hnRNP proteins and the nuclear export of messenger RNA. *Semin. Cell Devel. Biol.* **8**, 57-63.

Pinol-Roma, S., and Dreyfuss, G. (1992): Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**, 730-732.

Pinol-Roma, S., Choi, M., Matunis, M., and Dreyfuss, G. (1988): Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA binding proteins. *Genes Devel.* **2**, 215-227.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975): Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**, 598-599.

Portman, D. S., and Dreyfuss, G. (1994): RNA annealing activities in HeLa nuclei. *EMBO J.* **13**, 213-221.

Prentki, P., and Krisch, H. M. (1984): *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **29**, 303-313.

Pyle, A. M., Moran, S., Strobel, S., Chapman, T., Turner, D. H., and Cech, T. R. (1994): Replacement of the conserved G•U with a G-C pair at the cleavage site of the *Tetrahymena* ribozyme decreases binding, reactivity, and fidelity. *Biochemistry* **33**, 13856-13863.

Pyle, A. M. (1993): Ribozymes: a distinct class of metalloenzymes. *Science* **261**, 709-714.

Qiagen (1995): *The QIAexpressionist: the high level expression and protein purification system*, Qiagen Inc., Chatsworth.

Query, C. C., Bentley, R. C., and Keene, J. D. (1989): A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**, 89-101.

Reed, K. C., and Mann, D. A. (1985): Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **13**, 7207-7221.

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

Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Merrill, B., Williams, K. R., Multhaup, G., Beyreuther, K., Werr, H., Henrich, B., and Schafer, K. P. (1986): Mammalian single-stranded DNA binding protein UP1 is derived from the core hnRNP protein A1. *EMBO J.* **5**, 2267-2273.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C., and Klug, A. (1974): Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546-551.

Sakamoto, T. and Bryant, D.A. (1997): Temperature-regulated messenger RNA accumulation and stabilization of fatty-acid desaturase genes in the cyanobacterium *Synechococcus sp.* strain PCC 7002. *Mol. Microbiol.* **23**, 1281-1292.

Sakamoto, T., Higashi, S., Wada, H., Murata, N., and Bryant, D. A. (1997): Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium *Synechococcus sp.* PCC 7002. *FEMS Microbiol. Lett.* **152**, 313-320.

Sakashita, E., and Sakamoto, H. (1996): Protein-RNA and protein-protein interactions of the *Drosophila* sex-lethal protein mediated by RNA-binding domains. *J. Biochem.* **120**, 1028-1033.

Sakashita, E., and Sakamoto, H. (1994): Characterization of RNA binding specificity of the *Drosophila* sex-lethal protein by *in vitro* ligand selection. *Nucleic Acids Res.* **22**, 4082-4086.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989): *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

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

Sato, N., Maruyama, K., Nishiyama, Y., and Murata, N. (1997): Identification of a cold-regulated RNA-binding protein from the marine cyanobacterium *Synechococcus sp.* PCC 7002. *J. Plant Res.* **110**, 405-410.

Sato, N., and 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 Physiol.* **37**, 1150-1160.

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

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

Sayle, R. A., and Milner-White, E. J. (1995): RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374.

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

Scherly, D., Boelens, W., Van Venrooij, W. J., Dathan, N. A., Hamm, J., and I.W., M. (1989): Identification of the RNA-binding sequence of human U1A protein and definition of its binding site on U1 snRNA. *EMBO J.* **8**, 4163-4170.

Schnuchel, A., Wiltscheck, R., Czisch, M., Herrier, M., Willimsky, G., Graumann, P., Marahiel, M. A., and Holak, T. A. (1993): Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature* **364**, 169-171.

Schopf, J. W., Hayes, J. M., and Walter, M. R. (1983): Evolution of Earth's earliest ecosystems: recent progress and unsolved problems, pp. 361-384. In J. W. Schopf (Ed.): *The Earth's earliest biosphere: its origins and evolution*, Princeton University Press, Princeton.

Schuster, G., and Gruissem, W. (1991): Chloroplast mRNA 3' end processing requires a nuclear-encoded RNA-binding protein. *EMBO J.* **10**, 1493-1502.

Schwemmle, M., Gorlach, M., Bader, M., Sarre, T. F., and Hilse, K. (1989): Binding of mRNA by an oligopeptide containing an evolutionarily conserved sequence from RNA-binding proteins. *FEBS Lett.* **251**, 117.

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

Shapira, S. K., Chou, J., Richaud, F. V., and Casadaban, M. J. (1983): New versatile plasmid vectors for expression of hybrid proteins coded by a cloned gene fused to *lacZ* gene sequences encoding an enzymatically active carboxy-terminal portion of beta-galactosidase. *Gene* **25**, 71-82.

Shapiro, M. (1995): DNAdraw: a program for preparing DNA and protein data for publication. *BioTechniques* **18**, 1064-1067.

Shestakov, S. V., and Khyen, N. T. (1970): Evidence for genetic transformation in blue-green alga *Anacystis nidulans*. *Mol. Gen. Genet.* **107**, 372-375.

Sillikens, P. T. G., Habets, W. J., Beijer, R. P., and Van Venrooij, W. J. (1987): cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO J.* **6**, 3841-3848.

Siomi, H., and Dreyfuss, G. (1995): A nuclear localization domain in the hnRNP A1 protein. *J. Cell. Biol.* **129**, 551-560.

Smart, L. B., Warren, P. V., Golbeck, J. H., and McIntosh, L. (1993): Mutational analysis of the structure and biogenesis of the photosystem I reaction center in the cyanobacterium *Synechocystis sp.* PCC 6803. *Proc. Natl. Acad. Sci. USA* **90**, 1132-1136.

Smith, M. (1985): *In vitro* mutagenesis. *Annu. Rev. Genet.* **19**, 423-462.

Soitamo, A.J., Sippola, K., and Aro, E.M. (1998): Expression of *psbA* genes produces prominent 5' *psbA* mRNA fragments in *Synechococcus sp.* PCC 7942. *Plant Mol. Biol.* **37**: 1023-1033.

Spratt, B. G., Hedge, P. J., Heesen, S., Edelman, A., and Broome-Smith, J. K. (1986): Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8 and pEMBL9. *Gene* **41**, 337-342.

Srivastava, M., McBride, O. W., Fleming, P. J., Pollard, H. B., and Burns, A. L. (1990): Genomic organization and chromosomal localization of the human nucleolin gene. *J. Biol. Chem.* **265**, 14922-14931.

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

Steitz, T. A. (1990): Structural studies of protein-nucleic acid interactions: The sources of sequence-specific binding. *Quart. Rev. Biophys.* **23**, 205-280.

Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., and Moncoeur, E. (1998): Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nat. Struct. Biol.* **5**, 60-66.

Strobel, S. A., and Doudna, J. A. (1997): RNA seeing double: close packing of helices in RNA tertiary structure. *Trends Biochem. Sci.* **22**, 262-266.

Stuber, D., Matile, H., and Garotta, G. (1990): pp. 121-152. In I. Lefkovits, and B. Pernis (Eds): *Immunological Methods*, Academic Press.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Bubendorff, J. W. (1990): Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzmol.* **185**, 60-89.

Sturm, A. (1992): A wound-inducible glycine-rich protein from *Daucus carota* with homology to single-stranded nucleic acid-binding proteins. *Plant Physiol.* **99**, 1689-1692.

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

Suhnel, J. (1996): Image library of biological macromolecules. *Comput. Appl. Biosci.* **12**, 227-229.

Sulkowski, E. (1989): The saga of IMAC and MIT. *Bioessays* **10**, 170-175.

Surowy, C. S., Van Santen, V. L., Scheib-Wixted, S. M., and Spritz, R. A. (1989): Direct, sequence-specific binding of the human U1-70K ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. *Mol. Cell. Biol.* **9**, 4179.

Swanson, M., and Dreyfuss, G. (1988a): RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J.* **7**, 3519-3529.

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

Swanson, M. S., Nakagawa, T. Y., LeVan, K., and Dreyfuss, G. (1987): Primary structure of the human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol. Cell. Biol.* **7**, 1731-1739.

Tandeau de Marsac, N., and Houmard, J. (1988): Complementary chromatic adaptation: physiological conditions and action spectra. *Methods Enzymol.* **167**, 318-328.

Tandeau de Marsac, N., Borrias, W. E., Kuhlemeier, C. J., Castets, A. M., van Arkel, G. A., and van den Hondel, C. A. M. M. J. (1982): A new approach for molecular cloning in cyanobacteria: cloning of an *Anacystis nidulans met* gene using using a *Tn*901-induced mutant. *Gene* **20**, 111-119.

Thiel, T. (1994): Genetic analysis in cyanobacteria, pp. 581-611. In D. A. Bryant (Ed.): *The Molecular Biology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

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

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tsai, D. E., Harper, D. S., and Keene, J. D. (1991): U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res.* **19**, 4931-4936.

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

Van Gelder, C. W. G., Gunderson, S. I., Jansen, E. J. R., Boelens, W. C., Polycarpou-Schwartz, M., Mattaj, I. W., and van Venroojj, W. J. (1993): A complex secondary structure in U1A pre-mRNA that binds two molecules of U1A protein is required for regulation of polyadenylation. *EMBO J.* **12**, 5191-5200.

Van Nocker, S., and Vierstra, R. D. (1993): Two cDNAs from *Arabidopsis thaliana* encode putative RNA-binding proteins containing glycine-rich domains. *Plant Mol. Biol.* **21**, 695-699.

Varani, G., and Pardi, A. (1994): Structure of RNA, pp. 1-24. In K. Nagai, and I. W. Mattaj (Eds): *RNA-Protein Interactions*, Oxford University Press Inc., Oxford.

Vijay, S., and Kovacs, S. A. (1993): snRNP-like proteins associate with thylakoid/phycobilisome fraction of cyanobacterium *Synechococcus leopoliensis*. *FASEB J.* **7**, A1090.

Wagner, S. J., Thomas, S. P., Kaufman, R. I., Nixon, B. T., and Stevens, S. E. (1993): The *glnA* gene of the cyanobacterium *Agmenellum quadruplicatum* PR-6 in nonessential for ammonium assimilation. *J. Bacteriol.* **175**, 604-612.

Wang, J., Dong, Z., and Bell, L. R. (1997): Sex-lethal interactions with protein and RNA: Roles of glycine-rich and RNA-binding domains. *J. Biol. Chem.* **272**, 22227-22235.

Wang, J., and Bell, L. R. (1994): The Sex-lethal amino terminus mediates cooperative interactions in RNA-binding and is essential for splicing regulation. *Genes Devel.* **8**, 2072-2085.

Weeks, K. M., and Crothers, D. M. (1993): Major groove accessibility of RNA. *Science* **261**, 1574-1577.

Wieghardt, F., Biamonti, G., and Riva, S. (1995): Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J. Cell Sci.* **108**, 545-555.

Wilk, H. E., Herr, H., Friedrich, D., Kiltz, H. H., and Schafer, K. P. (1985): The core proteins of 35S hnRNP complexes: characterization of nine different species. *Eur. J. Biochem.* **46**, 71-81.

Wilusz, J., Feig, D. I., and Shenk, T. (1988): The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of the polyadenylation cleavage sites. *Mol. Cell. Biol.* **8**, 4477-4483.

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G., and Mueller, L. (1992): ^{1}H, ^{13}C, and ^{15}N NMR assignments and global folding patterns of the RNA-binding domain of the human hnRNP C proteins. *Biochemistry* **31**, 6254-6265.

Wolk, C. P., Ernst, A., and Elhai, J. (1994): Heterocyst Metabolism and Development. In D. A. Bryant (Ed.): *The Molecular Biology of Cyanobacteria*, Kluwer Academic Publishers, Dordrecht.

Wolk, C. P., Cai, Y., and Panoff, J.-M. (1991): Use of a transposon with luciferase as a reporter to identify environmentally responsive genes in a cyanobacterium. *Proc. Natl. Acad. Sci. USA* **88**, 5355-5359.

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

Wulfing, C., Lombardero, J., and Pluckthun, A. (1994): An *Escherichia coli* protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif. *J. Biol. Chem.* **269**, 2895-2901.

Xu, M.-Q., Kathe, S. D., Goodrich-Blair, H., Nierzwicki-Bauer, S. A., and Shub, D. A. (1990): Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* **250**, 1566-1570.

Yamanaka, G., Glazer, A. N., and Williams, R. C. (1978): Cyanobacterial phycobilisomes. *J. Biol. Chem.* **253**, 8303-8310.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985): Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Ye, L., Li, Y., Fukami-Kobayashi, K., Go, M., Konishi, T., Watanabe, A., and Sugiura, M. (1991): Diversity of a ribonucleoprotein family in tobacco chloroplasts: two new chloroplast ribonucleoproteins and a phylogenetic tree of ten chloroplast RNA-binding domains. *Nucleic Acids Res.* **19**, 6485-6490.

Zhou, J. H., Gasparich, G.E., Stirewalt, V. L., Delorimier, R., and Bryant, D. A. (1992): The *cpcE* and *cpcF* genes of *Synechococcus sp.* PCC 7002 - construction and phenotypic characterization of interposon mutants. *J. Biol. Chem.* **267**, 16138-16145.

Zuker, M. (1989): On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48-52.