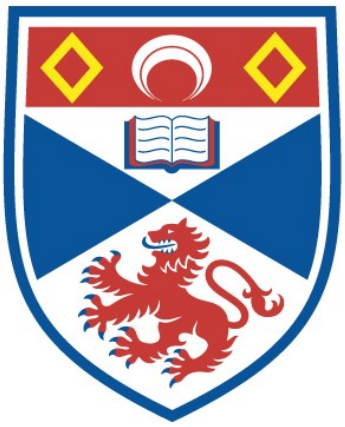
ANALYSIS OF PROTEINS BOUND TO STORED MESSENGER RNA IN 'XENOPUS' OOCYTES

Michael R. Ladomery

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1996

ANALYSIS OF PROTEINS BOUND TO STORED MESSENGER RNA IN *XENOPUS* OOCYTES

BY
MICHAEL R. LADOMERY

DEPARTMENT OF BIOLOGICAL AND MEDICAL SCIENCES

UNIVERSITY OF ST. ANDREWS

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DECLARATION

I was admitted to the faculty of Science of the University of St. Andrews under Ordinance General No. 12 on October 1st 1991, and as a candidate for the degree of Ph.D. on October 1st, 1992.

I, Michael Ladomery hereby certify that this thesis has been composed by myself, that it is a record of my own work except where stated otherwise, and that it has not been accepted in partial or complete fulfillment of any other degree or professional qualification.

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

CERTIFICATE

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate to the degree of Ph.D.

Signature of Supervisor:

Date: 19/6/95

ACKNOWLEDGEMENTS

I wish to thank Dr. John Sommerville for his almost transcendental patience in supervising this project, for his constant advice and support, and for the invariably stimulating conversations over coffee. Thank you to Dr. Jeff Graves for sharing his good sense of humour, and for putting up with my innumerable protein gels. Thank you also to all the glasswashers- Evelyn, Maureen and Joyce- for their daily contribution. I am also grateful to various Honour's students for their help: Rachael Marshall in cloning C4SR, Scott Lyons in cloning AB21 and Rob Little with the density gradients and slot blots. Thank you to Karen, Sean and Dave from the photographic unit for preparing those numerous negatives. Finally, I am deeply grateful to my family and friends for their support over the years, and in particular in the last phases of the degree, a very busy and intense time.

ABSTRACT

Regulation at the post-transcriptional level is gaining significance at a rapid pace. One example is the storage of messenger mRNA molecules in a translationally quiescent state, the so-called "masked messengers". Their existence has been known since the 1960s, but many details of their composition and structure have not yet been resolved. Masked messenger RNAs are particularly abundant in the oocytes of the African clawed toad *Xenopus laevis*. The aim of this study has been to examine the proteins bound to stored mRNAs in the oocytes, by focussing on the Y-box proteins which had already been identified as major components in mRNA masking, and by analyzing some of the other unidentified mRNP proteins.

The YB proteins were studied in greater detail, gaining fresh information about their RNA-binding properties, defining distinct binding domains. The presence of an mRNP-associated protein kinase was confirmed, and binding assays suggested that phosphorylation influences the ability of the YB proteins to bind to mRNA. cDNA expression libraries were screened both with an RNA-binding assay and with an immunoscreening method, isolating a variety of known and novel cDNAs. Peptide sequencing of mRNP proteins revealed the presence of an RNA helicase distinct from the translation initiation factor eIF4A. It is postulated that the RNA helicase, in addition to the YB proteins, will be seen to have an important role in the formation and activity of the masked messenger RNA particles.

ABBREVIATIONS

A	Absorption
A+	Polyadenylated
Amps	Ampères
AMPS	Ammonium persulphate
ATP	Adenosine triphosphate
B/A	Basic/aromatic
CAT	Chloramphenicol acetyl-transferase
CBB	Column binding buffer
cDNA	Complementary DNA synthesized from mRNA
Ci	Curie
CIP	Calf intestinal phosphatase
CMV	Cytomegalo virus
cpm	Counts per minute
CSD	Cold-shock domain
CTP	Cytidine triphosphate
DAB	3,3'-diaminobenzidine tetrahydrochloride
dH_2O	Distilled water
dATP	Deoxy adenosine triphosphate
dCTP	Deoxy cytidine triphosphate
DEPC	Diethylpyrocarbonate
DEAE	Diethylaminoethyl
dGTP	Deoxy guanosine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonucleic acid nuclease
DTT	Dithiothreitol
dTTP	Deoxy thymidine triphosphate
EBI	European Bioinformatics Institute
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
FA	Formic acid
FCS	Fetal calf serum
GTP	Guanosine triphosphate
GV	Germinal vesicle (nucleus)
h	hours
HA	Hydroxylamine

HCB	Heparin column buffer
HEPES	N-[2-Hydroxyethyl] piperazine-N'[2-sulphonic acid]
HIV	Human immunodeficiency virus
HPLC	High pressure liquid chromatography
HT	Heat-treatment
HTB	Heat-treatment buffer
HTLV	Human T-cell lymphotrophic virus
HTSN	Heat-treatment supernatant
HTP	Heat-treatment pellet
IgG	Immunoglobulin G
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilo bases
kDa	Kilo Dalton (1 Dalton=mass of 1 hydrogen atom)
λ	Lambda
LB	Luria broth
LUC	Luciferase
LTR	Long terminal repeat
(m/μ/n/p)	(milli/micro/nano/pico)
MBP	Maltose binding protein
MBT	Mid-blastula transition
MHC	Major histocompatibility
min	minutes
MOPS	3[N Morpholino] propanesulphonic acid
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
MS222	Tricane-methane sulfonate
NP-40	Nonidet P-40
NRE	Negative regulatory element
NSE	Nuclease sensitive element
°C	Degrees centigrade
OD	Optical density
ORF	Open reading frame
P	Pellet
PABP	Poly(A) binding protein
PAGE	Polyacrylamide gel electrophoresis
PB	Polysome buffer
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

pfu	Plaque-forming unit
PVDF	Polyvinylidene fluoride
PVP	Polyvinyl pyrrolidone
Riboblot	Binding of riboprobe to proteins transferred from SDS-PAGE to nitrocellulose
Riboprobe	*in vitro* labelled RNA probe
RNA	Ribonucleic acid
RNase	Ribonucleic acid nuclease
RRM	RNA recognition motif
RSV	Rous sarcoma virus
S	Svedberg units
SDS	Sodium Dodecyl Sulphate
sec	seconds
SN	Supernatant
SN10	Supernatant from a 10,000 g spin
SSC	Standard sodium citrate
TAE	Tris-acetate EDTA buffer
TBE	Tris-borate EDTA buffer
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween-20
TCA	Trichloroacetic acid
TD	Tail domain
TE	Tris EDTA buffer
TEMED	N,N,N',N'-tetramethylenediamine
TMNT	Tris-buffered saline with $MgCl_2$ and Tween-20
TNES	Tris buffered saline + EDTA and SDS, buffer for RNA extraction
UTP	Uridine triphosphate
UTR	Untranslated region
UV	Ultra violet
V	Volts
X-gal	5-Bromo-4-Chloro-3-Indoyl-β-D-Phosphate
YB	Y-Box

CONTENTS

LIST OF FIGURES

Chapter 1

Introduction

1.0 The masked message

The study of gene regulation is a fundamental area of research in molecular biology. Historically, gene regulation was first described at the transcriptional level. However, the path leading from gene transcription to the translation of messenger RNAs is a multi-step process, including transcript splicing, polyadenylation, transport, localisation, as well as translational initiation and elongation. Each of these steps can potentially be regulated. It is becoming increasingly apparent that complex and diverse regulatory mechanisms occur post-transcriptionally.

One specific example of post-transcriptional regulation is the storage and translational repression of messenger RNA, summarized in the "masked-messenger hypothesis". In general, a "masked messenger" can be defined as a fully processed mRNA whose translation is delayed. Masked mRNAs are stored in association with "masking proteins" in a cytoplasmic pool of non-polysomal, or "free" mRNP. This study will focus on the masking proteins involved in the storage of mRNA in the oocytes of the African clawed toad *Xenopus laevis*, a popular model organism. Before defining the aims of the project, it is necessary to refer to the original experiments which led to the masked messenger hypothesis; to illustrate, with a series of examples, why masked mRNAs should exist, and to describe what was known about the masking proteins at the start of the project.

1.1 Discovery of the masked message

The masked messenger hypothesis started with the work of Spirin *et al.* (1964) on embryos of the loach *Misgurnus fossilis*, a fresh-water fish. The irradiation of embryos by X-rays was observed to be intermittently lethal, depending on the developmental stage at which irradiation took place. For

example, if embryos were irradiated during phases in which the embryonic nuclei were transcriptionally active, such as between 6 and 8.5 hours after fertilization (blastula), the embryos did not develop to gastrula. However, if these embryos were irradiated between 8.5 and 14 hours after fertilization, a phase in which nuclei are inactive, they did progress to gastrula. It was suggested that during the active phase, the nuclei were producing the mRNA molecules that convey the information required for the embryo to progress to gastrulation. Whereas X-rays were lethal during this transcriptional phase, they were not once the genetic information was cytoplasmic, in the form of *stored mRNAs*. The analysis of cytoplasmic extracts fractionated on sucrose gradients revealed the presence of a broad spectrum of RNP (RNA + protein) material sedimenting between 40 and 60 S. This RNP material did not consist of ribosomal components. When combined with ribosomes *in vitro*, the 40-60 S RNP formed polysomes. The density of the RNP material, ~1.4 g/cm^3 on a CsCl gradient, suggested a protein to RNA ratio (w/w) of 3:1. These RNP, or masked messenger RNAs were named "informosomes", in the sense that they carry the information directing the synthesis of proteins necessary for development to progress; however, the term "informosome" did not catch on.

Further work described masked messengers in the sea urchin *Lytechinus pictus* (Spirin and Nemer, 1965), and in various other animal tissues such as sheep thyroid, rat brain and liver, giant silkworm epidermis and human HeLa cells (Ovchinnikov and Spirin, 1970). Whereas the masked mRNP were heterogenous in size, presumably due to the variable lengths of mRNA molecules contained within them, their density was remarkably uniform. Like the loach embryo masked mRNP, their density on CsCl gradients peaked at about 1.4 g/cm^3. This suggested that the types of proteins present and the structure of the informosomes are quite universal.

Soon after, Lukanidin *et al.* (1972) infected human FL cells with adenovirus, and studied the resultant mRNP particles by electron microscopy. The particles, with a diameter in the range of 0.2 μm, gave the impression of being composed of many similar subunits. Masked mRNP were also detected in plant cells, more specifically in wheat grain (Ajtkhozhin *et al.*, 1976).

In subsequent reviews, Spirin developed the concept of masked mRNP further. Spirin proposed that masked mRNP carry both translation initiation and elongation factors, as well as components which "repress and mask mRNAs" (Spirin, 1978 and 1980), arguing that due to the size of the eukaryotic cell, it is more efficient to group these factors together on the same mRNP, rather than rely on random collisions to bring about their association. Spirin described the concept with the Latin quote *omnia mea mecum porto*, which can be translated as "I carry all my things along".

In vivo, messenger RNA molecules do not exist free of protein in the cytoplasm (Henshaw, 1968; Ovchinnikov and Spirin, 1970; Greenberg, 1977). Does their association with proteins determine their "masked" state? When Gurdon *et al.* (1971) originally microinjected protein-free rabbit globin mRNAs into *Xenopus* oocyte cytoplasms, they observed that globin protein was actively synthesized. In contrast, many cytoplasmic mRNAs are "masked" in the *Xenopus* oocyte . This is clearly illustrated in an experiment where the occurrence of different proteins is compared in oocytes and eggs from the surf clam, starfish and *Xenopus* (Standart *et al.*, 1985; reviewed in Standart, 1992). In oocytes, *in vivo* synthesized proteins do not include ribonucleotide reductase, cyclin A and cyclin B; however, these proteins appear in eggs. When mRNP was extracted from oocytes, deproteinized, and added to a rabbit reticulocyte *in vitro* translation system, ribonucleotide reductase, cyclin A and cyclin B were produced.

Therefore the oocytes contain the mRNAs that direct the synthesis of these proteins, but the mRNAs are thought to be complexed with masking proteins, stored and prevented from translating until the appropriate developmental stage. Since the formulation of the masked messenger hypothesis in the 1960s, one of the main aims of research has been to identify and characterize the masking proteins.

1.2 Masked messages in *Xenopus* oocytes

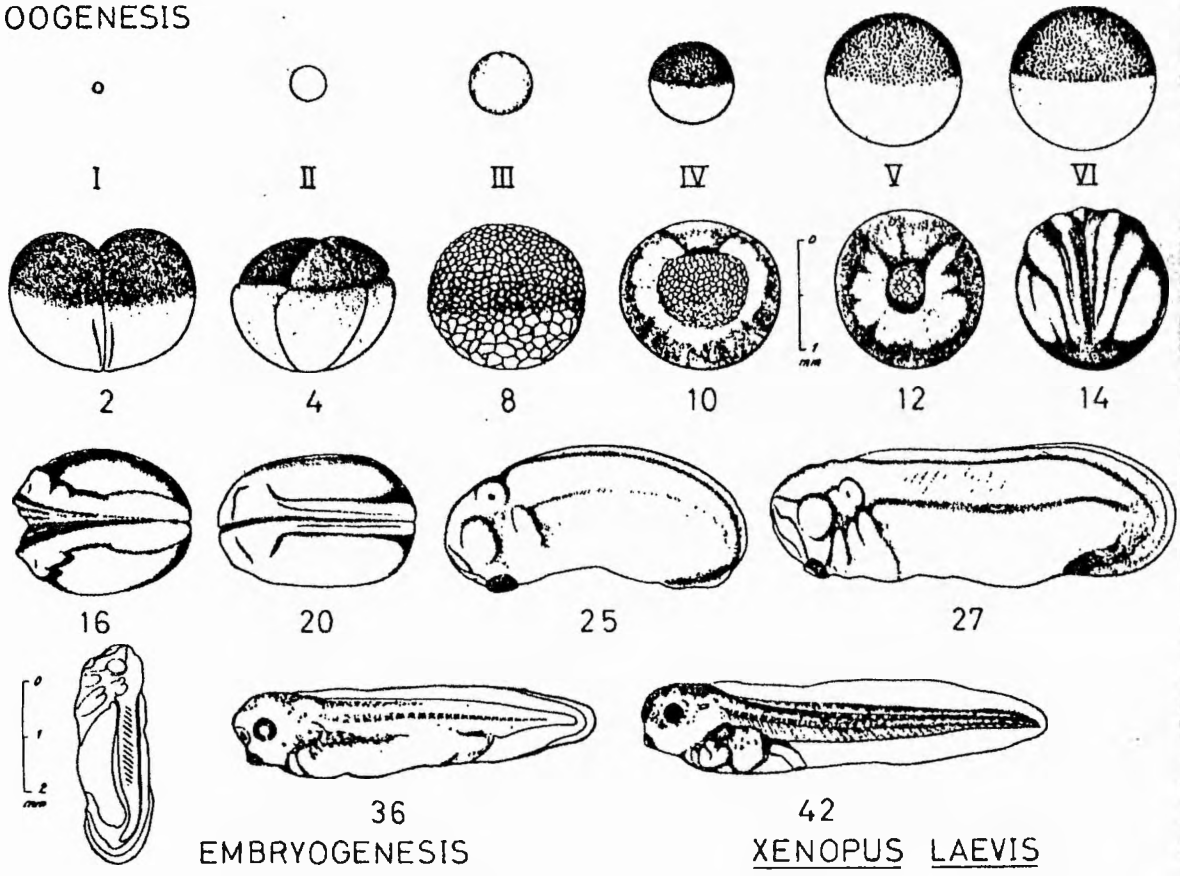
The need for masked mRNAs in *Xenopus* oocytes can be illustrated with a series of examples. In *Xenopus*, oogenesis is a lengthy process, lasting at least eight months (reviewed in Davidson, 1986; Smith *et al.*, 1991). The duration of oogenesis is influenced by hormones, food and crowding conditions. Throughout oogenesis, the diameter of the oocytes increases from about 50 μm to 1200 μm. Oogenesis has been divided into six stages, I-VI, based on the anatomy of the developing oocyte (Dumont, 1972; see **Fig. 1**). All oocyte stages are present in the ovary of a mature animal, oogenesis being asynchronous. Yolk accumulation (vitellogenesis) begins at stage II. Throughout oogenesis, a large store of macromolecules is synthesized: a staggering figure of 10^{12} ribosomes and 2×10^{11} mRNA molecules per oocyte (Davidson, 1986). Total RNA increases from 0.04 μg (stage I) to over 4 μg per oocyte (stage VI), of which only about 1% (40 ng) is polyadenylated mRNA, containing approximately 20,000 distinct mRNA sequences; the rest being mainly ribosomal (90%) and transfer RNA (4%) (Davidson, 1986). The accumulation of mRNA and mRNP proteins starts in previtellogenic oocytes, whereas ribosomal RNA and ribosomal proteins start accumulating in vitellogenic oocytes.

Despite adaptations in chromatin structure, such as the formation of lampbrush chromosomes and the use of histone variants, many months of

oogenesis are needed to transcribe all the RNA to produce the required pool of mRNA molecules. In *Xenopus*, the growing oocyte is blocked at the diplotene stage of the first meiotic division. At oocyte maturation, induced by the hormone progesterone, fundamental changes occur. Six hours after exposure to progesterone, the germinal vesicle (nucleus) breaks down and lampbrush chromosomes contract, in preparation for the first meiotic division. By ovulation, the egg is in second meiotic metaphase. Progesterone treatment is also followed by the recruitment of various stored masked messages into polysomes. Moreover, individual stored messages are recruited at different time points, depending on the cell's requirements (Dworkin *et al.*, 1985). The overall rate of translation increases three-fold at maturation, and a further two-fold after fertilization (Davidson, 1986).

A remarkable process follows: the embryonic cells divide rapidly, to the extent that within three days a swimming tadpole has developed: the original fertilized egg has given rise to 10^6 cells (Woodland, 1982). The mass of the young tadpole is comparable to that of the mature oocyte (**Fig. 1**). While the embryonic cells are dividing rapidly, going straight from the "S phase" (DNA synthesis) into the "M phase " (mitosis) of the cell cycle, there is no time to transcribe DNA which is replicating and segregating continuously. For this reason, in order to by-pass the need for transcription, the developing embryo relies on the store of maternal mRNP, ribosomes and assorted translational machinery. Zygotic gene transcription first starts at the "mid-blastula transition" (MBT), five hours into development. The pool of maternal masked messages encodes the proteins needed for rapid cell cleavage to occur, for example: DNA synthesis enzymes, such as ribonucleotide reductase; proto-oncogene mRNAs needed for cell proliferation, and histone H1, replacing its oocyte

Figure 1 **Early development in *Xenopus laevis*.** The diagrams are adapted from Nieuwkoop and Faber, 1956, and the figure prepared by Dr. Sommerville. The stages are all to scale, as shown on the diagram. Oogenesis has been divided into six stages (Dumont, 1972). Stage **I** oocytes are 50 to 100 μm in diameter, and are translucent. Stage **II** oocytes range up to 450 μm and are opaque white. In stage **III** oocytes (450-600 μm) pigment deposition begins. In stage **IV** oocytes (600-1000 μm), animal and vegetal hemispheres differentiate. In stage **V** oocytes (1000 to 1200 μm) the hemispheres are clearly delineated and the animal hemisphere appears light brown. Vitellogenesis (the deposition of yolk) occurs between stages II and V; stage **VI** oocytes (1200-1300 μm) are post-vitellogenic, ready for maturation. Various developmental stages are shown and are numbered according to Nieuwkoop and Faber, 1956: stage **2**: 1.25 h after fertilization, beginning of first cleavage; stage **4**: age 2.25 h, advanced eight cell stage; stage **8**: age 5 h, mid-blastula stage; stage **10**: age 9 h, initial gastrula stage; stage **12**: age 13.25 h, medium yolk plug stage; stage **14**: age 16.25 h, neural plate stage; stage **16**: age 18.25 h, mid neural fold stage; stage **20**: age 21.75 h, neural folds fused; stage **25**: age 1 day, 3.5 h., eyes and gills; beginning of fin formation; stage **27**: age 1 day, 7.25 h, lateral flattening of eyes, fin translucent, tail bud formation accentuated in lateral outline, 19 somites are segregated; stage **36**: age 2 days, 2 h, formation of two gill rudiments, melanophores appear on back, tail bud length three times its breadth; stage **42**: age 3 days, 8 h, yolk consumed in tail structures: swimming tadpole stage.



EMBRYOGENESIS

XENOPUS LAEVIS

variant B4. In a very short time, the products of many months of oogenesis are used up. Two typical masked messages can now be considered in more detail.

The proto-oncogene c-*mos* is a typical masked message in the oocyte. In the cascade of events that is triggered by progesterone, the maturation promoting factor (MPF) has a key role. MPF allows the oocyte to enter metaphase of the first meiotic division, and contains cyclin as well as the cell-cycle associated kinase cdk2 (reviewed in Gilbert, 1991). In general, cyclins are made during interphase and are degraded rapidly at cell division by a specific protease. In the oocyte, levels of cyclin protein are influenced by the product of c-*mos*, a 39 kDa protein kinase ($pp39^{mos}$). This kinase phosphorylates and thereby inactivates the protease which degrades cyclin. Hence the effect of active $pp39^{mos}$ is to promote the activity of MPF. Consequently, the injection of antisense c-*mos* mRNA blocks germinal vesicle breakdown, one of the consequences of oocyte maturation (Sagata *et al.*, 1988). The c-*mos* mRNA is transcribed from early oogenesis onwards, yet it is not translated until oocyte maturation (Sagata *et al.*, 1988). In contrast, $pp39^{mos}$ protein persists until fertilization, when it is itself degraded by a calcium-dependent protease, calpain II (Watanabe *et al.*, 1989). If $pp39^{mos}$ is not degraded, the oocyte stalls at metaphase II; likewise, if $pp39^{mos}$ is artificially present in embryonic cells, these cells are arrested at metaphase because the cell cycle cannot progress. In summary, c-*mos* mRNA is translated in a well defined time frame, its mRNA having been masked in the oocytes for many months.

The proto-oncogence c-*myc* encodes a nuclear transcription factor involved in cell proliferation. An abnormally high level of expression of this oncogene has been observed in various tumours. A *Xenopus* c-*myc* cDNA was isolated and its expression studied (Godeau *et al.*, 1986; Taylor *et al.*,

1986). Its mRNA is accumulated from early oogenesis onwards and, like c-*mos*, stored in a masked state. It is estimated that there are 7.4 pg of c-*myc* mRNA molecules in a stage VI oocyte, equivalent to approximately $5x10^6$ molecules. After fertilization, c-*myc* mRNA is actively degraded, with a half-life of 4 hours 20 minutes, and by gastrula there are only about ten transcripts per cell (Taylor *et al.*, 1986). The instability of c-*myc* mRNA is a more general phenomenon in somatic cells, and is mediated by the presence of AU-rich sequences in the 3' UTR, more specifically repeats of the sequence AUUUA. These are present in the 3' UTRs of a variety of unstable proto-oncogene mRNAs such as c-*fos* and c-*jun* (Kruys *et al.*, 1989; Laird-Offringa, 1992; You *et al.*, 1992). The protein $p65^{c\text{-}myc}$ is accumulated in late oogenesis, to be used during the highly proliferative phase of early development. Like c-*mos*, c-*myc* is a masked message in oocytes, but the exact timing of their translational recruitment differs.

Masked messages are also found in mammalian oocytes. Primary mouse oocytes, for example, have been shown to contain a "dormant", or masked, message encoding t-PA, tissue plasminogen activator (Strickland *et al.*, 1988). The mRNA is present in primary oocytes, but the enzyme is not synthesized until the resumption of meiosis in the hours preceding ovulation. The onset of zygotic transcription in mammals occurs at the two-cell cleavage stage, and therefore the requirement for masked maternal mRNAs is perhaps less extensive in mammalian oocytes.

1.3 Masked messages in spermatids

Spermatogenesis starts with the mitotic proliferation of spermatogonia into spermatocytes. The spermatocytes in turn give rise to haploid spermatids, which in turn develop into mature spermatozoa. To do this they need to produce a flagellum, an acrosome, and a sperm head in

which the genome is highly compacted; mitochondria are also arranged spirally around the base of the flagellum, being necessary to drive its movement (Kleene, 1989). In mice, spermiogenesis takes two weeks, and has been divided into sixteen steps based on the morphology of the spermatids. In the round spermatids (steps 1-8), the flagellum and acrosome appear while transcription declines. After step 8, the process of nuclear condensation begins, and RNA synthesis ceases. Nuclear compaction is achieved first by replacing histones with the transition proteins, and the transition proteins are subsequently replaced by the protamines. The protamines are small, basic proteins rich in arginine and cysteine. Once the genome is packaged by these protamines, transcription is no longer possible. Therefore messages encoding proteins that are required for the maturation of spermatids have to be synthesized in early, round spermatids, and stored for several days until they are required in late spermatids (Kleene *et al.*, 1984; Kleene and Smith, 1994). The masked messages encode the transition proteins 1 and 2 (Heidaran *et al.*, 1988; Morales *et al.*, 1991), the protamines 1 and 2, and the mitochondrial capsule selenoprotein (Kleene, 1993). The storage of these masked messages lasts between two and eight days in total (Kleene *et al.*, 1984).

Messenger RNP particles were isolated from testis cells from the rainbow trout, *Salmo gairdnerii* (Sinclair and Dixon, 1982), using sedimentation on sucrose gradients and binding to an oligo(dT) column which bound the particles through hybridization to their poly(A) tails. The density of the particles was found to be 1.35-1.37 g/cm^3 in a CsCl gradient, which is quite consistent with previous findings. Whereas mRNP particles derived from polysome fractions were actively translated *in vitro*, the non-polysomal particles did not translate, unless treated with 0.3 M KCl or deproteinized. However, both polysomal and non-polysomal mRNP

particles had similar densities. Sinclair and Dixon (1982) speculated that whereas the bulk of the mRNP proteins were similar, the set of translationally silent particles are bound by a specific repressor that dissociates in 0.3 M KCl.

In studying mouse spermatogenesis, Kwon and Hecht (1991) defined two highly conserved sequences in the 3' UTR of masked mRNAs, located between nucleotides +537 and +572 of protamine 2 mRNA. A protein of 18 kDa was UV-crosslinked to one of these sites. These conserved sequences were later called the Y and H elements, and are found in the 3' UTRs of all sequenced mammalian protamine mRNAs (Kwon and Hecht, 1993). The 18 KDa protein was shown to be present in isolated mRNP particles. Using an *in vitro* rabbit reticulocyte translation system, Kwon and Hecht demonstrated that when reporter constructs containing the Y and H elements were supplemented with a testis extract containing the 18 kDa protein, translation was repressed. Translational repression was dependent on the phosphorylation of the 18 kDa protein: in late spermatids, the 18 kDa repressor did not bind to the mRNAs, presumably as a consequence of its dephosphorylation. The 18 kDa protein is therefore necessary for the translational repression of these messages while they are stored. However, other proteins were shown to bind to these masked mRNAs in the testis (Kwon *et al.*, 1993; Tafuri *et al.*, 1993). These other proteins will be discussed in a later section (1.6). Their functional relationship with the 18 kDa protein is unclear.

1.4 Masked messages in somatic cells

The storage of mRNA as "masked messages" is by no means restricted to gametocytes. For example, storage of mRNA is well documented in somatic tissues in plants. In wheat grains, protein synthesis declines

during seed ripening, and a significant proportion of mRNA is increasingly found in free rather than polysomal mRNPs (Ajtkhozhin *et al.*, 1976). In *Zea mays* (maize), stored mRNAs exist in the seedling's embryo (Sánchez-Martínez *et al.*, 1986). In maize, embryogenesis lasts 30-40 days, while maturation and seed drying takes 10-20 days. When mRNPs are extracted from young maize embryos and added to an *in vitro* translation system, they give rise to proteins which normally do not appear until the mature embryo stage. This implies that these mRNAs are stored in a translationally repressed form (masked mRNAs).

In the plant *Medicago sativa* (alfalfa), mRNAs encoding the "storage proteins" are expressed in the earliest stages of the somatic embryo, during which the storage proteins are not detected (Pramanik *et al.*, 1992). Storage proteins are an important component in seed development, and are not found in vegetative tissues. Their mRNAs only become polysomal at the beginning of cotyledon development. Following a familiar pattern, after deproteinization, mRNPs obtained from younger somatic embryos were able to translate *in vitro*. In another example, long-lived mRNAs were studied in the radish *Raphanus sativus* (Raynal *et al.*, 1989). A specific cDNA was isolated, encoding a protein described as one of the "early germination polypeptides". Northern analysis showed that its mRNA starts accumulating during the dessiccation phase, reaches its peak in the dry seed, and disappears in 12 hour old seedlings.

Furthermore, in the context of light-induced chloroplast biogenesis, there is a set of mRNAs encoding photosynthesis-related proteins that are derived from the chloroplast (Danon *et al.*, 1991). These mRNAs accumulate in dark-grown plants without being translated: after exposure to light there is a dramatic increase in the synthesis of plastid proteins associated with photosynthetic membranes (Malnoë *et al.*, 1988). The plant

is able to exploit sunlight quickly, activating the translation of the relevant proteins while by-passing the need for transcription. According to the above examples, it is clear that masked messages occur in somatic tissues in plants. It is conceivable that some of the plant mRNA masking proteins and mechanisms may be related to those found in animal cells.

In mammals, the development of reticulocytes into erythrocytes involves the loss of the nucleus as well as the degradation of mitochondria. The mammalian red blood cell is terminally differentiated, with a very specialized purpose: to carry oxygen and CO_2 in hemoglobin. Messenger RNAs encoding globins are therefore very abundant in the reticulocyte. The second most abundant message encodes LOX (15-lipo-oxygenase), an enzyme involved in the degradation of mitochondria which attacks intact phospholipids (Ostareck-Lederer *et al.*, 1994). Its mRNA is transcribed in bone marrow but is only translated when the reticulocytes reach the peripheral blood, and is therefore a masked message. By using gel retardation and photocrosslinking assays, a 48 kDa protein has been identified which binds to a pyrimidine-rich sequence in the 3' UTR of LOX mRNA (Ostareck-Lederer *et al.*, 1994). p48 was shown to be required for translational repression, and is currently being cloned and characterized. However, as was the case for the protamine mRNA, another protein has also been identified in rabbit reticulocyte mRNPs, with a role in translational repression (Minich and Ovchinnikov, 1992; Evdokimova *et al.*, 1995), and will be discussed in section 1.6.

As is the case for mRNAs encoding protamines, a repressor binds to a specific site in the 3' UTR of LOX mRNA. The 3' UTR has been involved in other examples of translational regulation: for example, as was mentioned previously (section 1.2), in the mouse oocyte, tissue plasminogen activator mRNA is a masked message. The injection of antisense RNA

complementary to 103 nucleotides in the extreme 3' UTR blocked the translational recruitment of this message (Strickland *et al.*, 1988). In this case, the 3' UTR is required for translational recruitment, and the antisense RNA is acting like a repressor of translation.

5' UTR sequences have also been involved in translational regulation. For example, the various mRNAs encoding ribosomal proteins are coordinately regulated at the level of translation according to cellular demand for ribosomes, and in response to physiological cues. This coordinated translation is achieved through a protein binding to a pyrimidine-rich stretch in the 5' UTR (Cardinali *et al.*, 1993). When this pyrimidine stretch was fused to a chimeric reporter mRNA, the reporter acquired the translational pattern of the ribosomal protein mRNAs. A protein of 57 kDa bound specifically to this site. Similarly, the mRNA encoding the translation elongation factor EF1α possesses an analogous pyrimidine tract in its 5' UTR, and its translation is coordinately regulated with the ribosomal protein mRNAs (Loreni *et al.*, 1993). In *Xenopus* tissue culture supplemented with serum, mRNAs encoding EF1α as well as the ribosomal protein mRNAs, moved from a non-translating pool of masked mRNAs to polysomes, and conversely from polysomes back to masked mRNP during serum deprivation.

In another well characterized example, the mRNA encoding ferritin contains a 5' UTR element, the IRE (iron responsive element) which interacts with a repressor protein, the IRE-BP (IRE-binding protein; Klausner *et al.*, 1993). Ferritin protein binds to excess Fe^{3+} ions. When these ions are absent, the repressor binds with a high affinity to the IRE; this affinity is reversed in the presence of the ions, leading to translation of the ferritin mRNA.

In a recent experiment involving both human HeLa and yeast cells, a bacteriophage protein and a spliceosomal protein were converted into translational repressors when their binding sites were added to the 5' UTR of reporter mRNAs (Stripecke *et al.*, 1994). The authors suggested that binding of these proteins to the 5' UTR resulted in a "steric blockage" effect, in which the binding of translation initiation factors to the 5' UTR was prevented. There is no doubt that many more examples of message-specific translational regulation will be discovered.

1.5 The role of polyadenylation in mRNA recruitment

The translational recruitment of various masked maternal mRNAs in oocyte maturation has been associated with the cytoplasmic polyadenylation of these mRNAs. This is in addition to their initial nuclear adenylation (McGrew *et al.*, 1989). Two known *cis*-acting elements in the 3' UTR of various masked messages are: the hexanucleotide AAUAAA, required for polyadenylation in general, and the cytoplasmic polyadenylation element (CPE), of sequence UUUUUAU (reviewed in Wickens, 1990). More specifically, the process of polyadenylation, rather than the presence of an extended poly(A), tail has been suggested to be required for translational recruitment (Simon *et al.*, 1992). The exact timing of cytoplasmic polyadenylation has been studied with respect to the masked messages c-*mos*, and cyclins A1, B1 and B2 (Sheets *et al.*, 1994). Results suggest that both the magnitude and timing of translational stimulation depended on message-specific 3' UTRs. Recently, the *Xenopus* oocyte cytoplasmic poly(A) polymerase has been cloned (Gebauer and Richter, 1995).

The role of polydenylation has also been described in the surf clam *Spisula solidissima* (Rosenthal and Ruderman, 1987; Standart and Dale, 1993).

During meiotic maturation, masked mRNAs become polyadenylated, and eventually become deadenylated when no longer required . The 3' UTRs of mRNAs encoding ribonucleotide reductase, cyclins A and B, and histone H3 were found to contain CPE-like motifs (Standart and Dale, 1993). While the 3' ends of these messages were polyadenylated, actin mRNA, which ceases to translate at this stage, was not. The actin mRNA 3' UTR did not have a comparable CPE-like sequence.

Similarly, the marine worm *Urechis caupo* was also studied with respect to polyadenylation (Rosenthal and Wilt, 1986). As in *Spisula* and *Xenopus*, different classes of maternal mRNAs accumulate throughout oogenesis. One class has a short poly(A) tail which is adenylated following fertilisation. Another class accumulates in oocytes as fully polyadenylated mRNAs, but numbers diminish in full-grown oocytes, coupled with deadenylation. The former class are the masked messages, whereas the latter are the messages that are translated in the oocyte, encoding the proteins required for oogenesis to proceed. The translation of specific maternal mRNAs from *Urechis* has been studied more recently using cDNA probes (Rosenthal and Wilt, 1993). Translational recruitment correlates with further cytoplasmic polyadenylation, and the 3' UTR sequence of a number of these mRNAs also contains CPE-like elements.

Although these results are consistent with respect to CPE-containing mRNAs, it is not altogether clear that polyadenylation is a universal requirement for translational recruitment (see reviews by Standart, 1992, and Spirin, 1994). For example, core histone messages in somatic cells are not polyadenylated, yet they are efficiently translated. In *Xenopus* oocytes, the histone H1 mRNA is masked in a polyadenylated form, and deadenylated when translated (Ballantine and Woodland, 1985). Similarly, protamine and transition protein mRNAs are masked in an adenylated

form in the developing spermatid, and partially deadenylated when recruited during spermiogenesis (Kleene *et al.*, 1984; Kleene 1989). There are more exceptions to the rule, such as the ODC (ornithine decarboxylase) mRNA, which is unmasked during *Xenopus* oocyte meiotic maturation, without any significant change in its poly(A) tail (reviewed in Standart, 1992). The masking and unmasking of the LOX mRNA in a rabbit reticulocyte cell-free translation system is also independent of polyadenylation (Ostareck-Lederer *et al.*, 1994). Finally, a recent experiment involving the injection of CAT-reporter constructs into *Xenopus* oocytes has suggested that masking and unmasking of these constructs is unaffected by polyadenylation (Braddock *et al.*, 1994). According to these examples, the relationship between polyadenylation and translational recruitment is therefore unclear.

1.6 The masking proteins are identified

Having described the phenomenon of masked messages in a variety of contexts, we now focus on the *Xenopus* oocyte mRNA masking proteins. In 1974, Rosbash and Ford reported that in full-grown oocytes, over 90% of polyadenylated mRNA is located in free mRNP sedimenting between 30 and 120 S. In 1977 Ford *et al.* confirmed that very long-lived mRNAs exist in *Xenopus* ovaries. The search for the *Xenopus* oocyte masking proteins soon followed.

In 1981, Darnborough and Ford isolated free mRNP from *Xenopus* oocyte extracts using sucrose density centrifugation and oligo(dT) chromatography. Numerous proteins were described, including four major proteins of 50, 52, 56 and 59 kDa, and these were labelled, respectively, mRNP1, mRNP2, mRNP3 and mRNP4. There were many other less abundant proteins: for example proteins of 75 and 100 kDa

(mRNP 5 and mRNP6 respectively) and 16 and 22 kDa (mRNP7 and mRNP8). The protein composition of the mRNP did not therefore present itself as a simple problem. The sedimentation properties of the mRNP were consistent with a protein:RNA ratio of 4:1 (w/w). The major proteins $mRNP_{1-4}$ were largely cytoplasmic, and were not detected in *Xenopus* liver or reticulocytes.

A *Xenopus* oocyte protein described as "p60" was shown to be tightly bound to mRNA, heavily phosphorylated, and present in nuclear fractions (Dearsly *et al.*, 1985). Likewise, an abundant mRNP protein "p56" was also phosphorylated. These proteins were then referred to as pp56 and pp60 (pp=phosphoproteins), and correspond to mRNP3 and mRNP4. In another amphibian, the newt *Triturus cristatus*, an abundant RNP component of 60 kDa was also described, and was part of a fibrillar matrix in the nuclei (Kloetzel *et al.*, 1982). Similarly, an antiserum raised against *Triturus* pp60 cross-reacted with RNP components in *Drosophila melanogaster* spermatocyte nuclei (Glätzer *et al.*, 1986).

In vitro reconstitution experiments had earlier shown that mRNP proteins could inhibit translation *in vitro* (Richter and Smith, 1984). Using purified pp60 and an *in vitro* transcribed globin message, Kick *et al.* (1987) reconstituted mRNP, and assayed translational activity in a wheat germ lysate. It was suggested that translational repression was dependent on the phosphorylation of pp60, because dephosphorylated mRNP were able to translate. Could the phosphoprotein pp60 be the main masking protein? In 1987, Crawford and Richter used an antiserum raised against *Xenopus* pp56, the second abundant mRNP phosphoprotein, to immunoprecipitate mRNP. They reasoned that because in the oocyte certain mRNAs are masked, while others are translatable, only masked mRNAs should immunoprecipitate. They found that the mRNA XRNP6, which translates

in stage 1-2 oocytes, could not be immunoprecipitated with this antiserum. In contrast, the mRNAs XRNP1, XRNP3 and XRNP10, which are masked messages, were immunoprecipitated. Consequently, pp56 (and pp60) could be referred to as the "masking proteins". These proteins were also photocrosslinked to mRNA using 254 nm UV light, suggesting a tight interaction with the nucleic acid (Swiderski and Richter, 1988).

In addition, Cummings and Sommerville (1988) found that a kinase activity was an integral component of the mRNP particle. During the time of maximum rate of mRNA accumulation in stage I and II oocytes, the kinase activity was present in particles sedimenting between 40 and 80S. The proteins pp56 and pp60 present in these particles were phospholabelled both *in vitro* and *in vivo* after the addition of [γ-^{32}P]ATP. In more mature oocytes, when ribosomes become abundant, the kinase activity and pp56/pp60 sedimented between 80 and 110 S, suggesting that at this stage in oogenesis, the mRNP particles interact with ribosomal subunits. The nature of this interaction was not and is still not understood. It is conceivable that mRNPs are bound to ribosomes to facilitate an immediate translational recruitment at the appropriate time. The association of the mRNP with the ribosomal subunits was disrupted with EDTA, after which the phosphoproteins were found in particles with a buoyant density of 1.40 g/cm^3 in a CsCl gradient. This value is consistent with earlier findings (Spirin *et al.*, 1964; Ovchinnikov and Spirin, 1970; Sinclair and Dixon, 1982). The phosphoproteins pp56/pp60 showed a range of ionic forms in two-dimensional gels, and proteinase digestion patterns gave almost identical profiles (Cummings *et al.*, 1989). In summary, it emerged that both the masking proteins are phosphorylated by an mRNP-associated kinase activity, and that pp56 and pp60 are structurally similar, probably related proteins, their phosphorylation being

required for keeping the mRNA in a masked state. The phosphorylation of the masking proteins also appeared to enhance their binding to RNA *in vitro* (Murray *et al.*, 1991).

The properties of the associated kinase activity were investigated further (LaRovere and Sommerville, unpublished). Its characteristics, such as the ability to use both ATP and GTP as phosphate donors, the phosphorylation of casein, inhibition by heparin, and activation by spermine were consistent with its being of a casein kinase II type. A casein kinase II activity has been purified from *Xenopus* ovary (Mulner-Lorillon *et al.*, 1988), and cDNAs encoding its various subunits (α, α' and β) were later isolated (Jedlicki *et al.*, 1992).

At this point, it became a priority to clone and sequence the masking proteins pp56 and pp60. This task proved relatively straightforward, due to the abundance of these proteins in the oocytes. Deschamps *et al.* (1991) discovered that the masking proteins could be purified via "heat-treatment". After heating oocyte extracts to 80 °C, cooling and spinning them, they found that pp56 and pp60 remained in solution whereas other proteins precipitated. This material was amenable to peptide sequencing. Soon the masking proteins were identified (Deschamps *et al.*, 1992; Murray *et al.*, 1992), giving an unexpected result: pp56 and pp60 belonged to a novel class of *transcription* factors, the "Y-box proteins".

The first Y-box protein, human YB-1, was identified in 1988 by virtue of its binding to the Y-box DNA element found in the promoters of MHC class II promoters (Didier *et al.*, 1988). The *Xenopus* oocyte homologues were cloned in 1990 by Tafuri and Wolffe, who isolated FRGY1 and FRGY2, the former being expressed in somatic tissues, the latter in the oocyte. The mRNP masking protein pp60 was essentially identical to the oocyte Y-box

protein FRGY2. Confusion has since arisen with respect to terminology: pp60 is mRNP4 (Darnborough and Ford, 1981), which is the same as FRGY2 (Tafuri and Wolffe, 1990), and is referred to as p56 in Murray *et al.*, (1991)! The smaller masking protein pp56 is referred to as mRNP3 (Darnborough and Ford, 1981) and p54 (Murray *et al.*, 1991). For the sake of consistency, in this study the *Xenopus* oocyte masking proteins will be referred to as pp56 and pp60, and the Y-box proteins in general will be referred to as "YB proteins".

Subsequently, YB proteins have been identified in two other examples of mRNP particles. It is claimed that MSY1, which is equivalent to FRGY1, is a major mRNP protein in mouse spermatocytes (Tafuri *et al.*, 1993). However, as was discussed in section 1.3, the masking of messages in spermatocytes involves an 18 kDa repressor. It is not clear what the interaction between the 18 kDa repressor and MSY1 might be. Perhaps MSY1 packages the mRNP as a general masking protein, and the sequence-specific 18 kDa protein acts as a *temporal regulator* of translation. MSY1 was detected in a 60-80 S free mRNP fraction and could be photocrosslinked to mRNA (Tafuri *et al.*, 1993). Not surprisingly, given the virtual identity between FRGY1 and MSY1, this fraction cross-reacted with the antiserum anti-FRGY1.

MSY1 is equivalent to FRGY1, but not to the *Xenopus* oocyte YB protein FRGY2. The involvement of MSY1 in masked mRNA in the mouse testis strongly suggests that somatic Y-box proteins like FRGY1/YB-1 could mask mRNAs in somatic contexts in which FRGY1/YB-1 are expressed. One such example is now available. In rabbit reticulocytes, polysomal mRNPs were found to be translatable both in rabbit reticulocyte and wheat germ lysate cell-free translation systems, whereas non-polysomal mRNP were translatable in rabbit reticulocyte but not in wheat germ

lysate (Minich *et al.*, 1989). This is similar to the findings of Sinclair and Dixon (1982) with respect to the trout spermatid mRNP. In rabbit reticulocytes, a 50 kDa protein (p50) was identified as one of the major components of the free mRNP, as well as p70, believed to be the poly(A) binding protein. p50 was shown to be responsible for the translational repression of globin mRNA in the wheat germ lysate (Minich and Ovchinnikov, 1992). p50 was purified from the reticulocyte mRNP and tested for RNA-binding *in vitro* (Minich *et al.*, 1993). Like its *Xenopus* oocyte counterparts pp60 and pp56, p50 was also phosphorylated *in vitro* and *in vivo*. Finally, the identity of p50 has been revealed through peptide sequencing (Evdokimova *et al.*, 1995). It has been found to be the rabbit equivalent of human YB-1, mouse MSY1 and *Xenopus* FRGY1. This was the first example of a positive identification of a native YB protein in somatic mRNP. Using immunoblots, Evdokimova *et al.* (1995) also detected p50 in mRNP isolated from two other somatic tissues: rat liver and rabbit muscle, and suggested that the presence of YB proteins in mRNP is a widespread phenomenon.

Had the masking proteins finally been revealed (Sommerville, 1992)? The identification as YB protein transcription factors immediately raised the intriguing possibility of links between transcriptional and translational processes. Much attention has since been devoted to these proteins, and it is therefore necessary to consider the YB proteins in some detail. The next sections will describe in more detail their discovery, structure, expression, and various proposed functions.

1.7 Discovery of the YB protein family

The YB proteins are a recently discovered family of gene regulators. Their story begins in the context of MHC class II gene regulation (the major

histocompatibility class II genes, labelled HLA in humans). The expression of these genes in the immune system is regulated transcriptionally by a series of cis-acting elements including the TATA box, the octamer motif, and the W, X and Y boxes (Didier *et al.*, 1988). The Y-box sequence CTGATTGGCCAA, with an "inverted CCAAT box" underlined, is highly conserved among MHC class II gene promoters. Using a double-stranded oligo comprising the X and Y boxes, in 1988 Didier *et al.* cloned the first YB protein, human YB-1, from a Swei cell cDNA library. In the same year, Sakura *et al.* used as a ligand part of the promoter of the human oncogene c-*erbB*-2, also containing a Y-box sequence, and cloned dbpa and dbpb from a human placental cDNA library. These cDNAs encoded two YB proteins, of which dbpb was essentially identical to YB-1.

In 1990, Ozer *et al.* were attempting to clone and identify the "enhancer factor" which binds to the high affinity binding site of the RSV (Rous Sarcoma Virus) LTR (Long Terminal Repeat). The RSV LTR high-affinity binding site contains two inverted CCAAT boxes. They cloned EFI_A from a rat liver cDNA library, which turned out to be the same protein as YB-1/dbpb. As mentioned previously, Tafuri and Wolffe (1990) used a Y-box sequence-containing ligand to clone *Xenopus* oocyte transcription factors, and isolated two very similar proteins, which they called FRGY1 and FRGY2. FRGY1 was almost identical to human YB-1/dbpb/EFI_A described above. FRGY2, in contrast to FRGY1, was observed to be germ-cell specific in its expression: it is the oocyte mRNA masking protein pp60 (see previous section 1.6). In 1990, Wistow noted the similarity between a highly conserved region of the YB proteins and the *E. coli* cold-shock protein CSP7.4 described by Goldstein *et al.* (1990). This conserved region is present in all YB proteins and is now referred to as the "**cold-shock domain**" (CSD).

Subsequently, many other YB proteins have been cloned *without* the use of Y-box containing ligand, thus complicating the picture. YB-1 was cloned again using a W-box ligand derived from the human HLA-DQβ gene, as well as with an apurinated DNA ligand (Hasegawa *et al.*, 1991). Furthermore, the human YB protein NSEP-1 (nuclease-sensitive element binding protein) was cloned with a pyrimidine-rich strand ligand derived from the region of the c-*myc* promoter which has a strong purine/pyrimidine strand asymmetry, and which is predicted to give rise to a H-DNA triplex configuration (Kolluri *et al.*, 1992). Likewise, Horwitz *et al.* (1994) cloned YB-1 using a part of the human γ-globin promoter region which is also predicted to form a triplex. Grant and Deeley (1994) cloned chicken YB-1 using the promoter of the liver specific, estrogen-dependent gene *apoVLDLII*, also describing a preference for pyrimidine-rich strands. Cohen and Reynolds (1991) cloned *Xenopus* YB3 using the B-box which belongs to the core promoter of class III genes such as 5S RNA and tRNA genes. The use of these diverse ligands to clone YB proteins caused considerable confusion.

It is now apparent that the YB proteins are a ubiquitous family of gene regulators. Wolffe *et al.* (1992) cross-hybridized a FRGY1 cDNA probe to genomic DNA from rhesus monkey, rat, mouse, dog, cattle and chicken, as well as *Drosophila*. Similarly, Ozer *et al.* (1993) cross-hybridized human YB-1 to galago bush baby, mouse, rat, dog, cattle, pig and sheep genomic DNA, and cloned *Bos taurus* YB-1. It will be of interest to see more examples of YB proteins: in the vertebrate lineage, there are no examples yet from fish and reptiles. In the invertebrates, there are only two examples so far: a YB protein cloned from the neuronal tissue of the invertebrate *Aplysia californica* (Skehel and Bartsch, 1994), and a partial sequence obtained in a survey of expressed genes in *Caenorhabditis elegans*

(Waterston *et al.*, 1992). In the plant kingdom, the only example so far of a YB protein is grp-2, described in *Nicotiana sylvestris* and *Arabidopsis thaliana* (Obokata *et al.*, 1991).

What is the structure of the YB protein genes and what is their copy number per genome? Familiari *et al.* (1994) estimate that there are at least 15 copies per haploid genome in the mouse, including at least one pseudogene. Ozer *et al.* (1993) also found a pseudogene among genomic clones from *Bos taurus*. As for the chromosomal locations of the YB-1 protein genes, Spitkovsky *et al.* (1992) suggested that loci with YB-1 related sequences are present on four separate chromosomes in the mouse. Finally, the *unr* gene is of considerable interest. It encodes a YB protein consisting of a five-fold repeat of the CSD, of which the odd-numbered CSD are the best conserved, and is the only known example of its kind (Doniger *et al.*, 1992). Partial sequencing of *unr* genomic DNA suggests that the last CSD repeat is encoded by a separate exon. Perhaps in a similar way, the CSD of more common YB proteins could be encoded by an exon of ancient origin, to which auxiliary domains have been added via "exon shuffling". An alignment of ten YB proteins is presented in **Fig. 2**, highlighting the conserved CSD. The alignment suggests that the CSD is a very ancient structure which has been conserved throughout millions of years of evolution.

1.8 Structure of the YB proteins

The feature shared by all YB proteins is therefore the highly conserved CSD of about 70 amino-acids. The CSD of FRGY2 (pp60) or FRGY1 is 43% identical to the *E. coli* cold-shock protein CS7.4 (Goldstein *et al.*, 1990; Wistow, 1990). CS7.4 is, which is also referred to as CspA, is a 7.4 kDa protein whose expression is induced at ten

Figure 2 Amino-acid sequence alignment between a selection of YB proteins. **CS7.4**: *E. coli* cold-shock protein (Goldstein *et al.*, 1990); **CspB**: *Bacillus subtilis* cold-shock protein (Willimsky *et al.*, 1992); **grp2**: *Nicotiana sylvestris* glycine-rich protein (Obokata *et al.*, 1991); **MSY1**: mouse transcription factor and spermatid mRNP protein (Shaughnessy and Wistow, 1992; Tafuri and Wolffe, 1993); **YB1**: human transcription factor (Didier *et al.*, 1988); **FRGY1**: *Xenopus laevis* transcription factor (Tafuri and Wolffe, 1990); **NSEP1**: human nuclease-sensitive element binding protein (Kolluri *et al.*, 1992); **CE21E7**: *Caenorhabditis elegans*, incomplete sequence derived from a survey of expressed genes; one amino-acid is uncertain (asterisk) (Waterston *et al.*, 1992); **RYBa**: rat liver transcription factor (Ito *et al.*, 1994); **APYB**: *Aplysia californica* YB protein expressed in neuronal tissue (Skehel and Bartsch, 1994); **FRGY2**: *Xenopus* oocyte major mRNA masking protein (Tafuri and Wolffe, 1990). Note the following features: a highly conserved region corresponding to the bacterial cold-shock proteins now referred to as the "cold-shock domain" (CSD); a variable N-terminus preceding the CSD, and following the CSD, an alternating series of basic and acidic island is commonly found. Together, the charged islands consitute the "tail domain" (TD). The plant protein **grp2** has a different TD which resembles the"RGG" motifs present in various RNA-binding proteins (Burd and Dreyfuss, 1994). The alignment was obtained using the PILEUP program, GCG package version 7.2, 1992.

```
          1                                                      50
MSY1      .......... .......... MSSEAETQQP PAAP..AAAL SAADTKPGST
YB1       .......... .......... MSSEAETQQP PAAPPAAPAL SAADTKPGTT
FRGY1     .......... .......... MSSEVETQQQ QPD....... .ALEGKAGQE
NSEP1     .......... .......... MSSEAETQQP PAAPPAAPAL RPPTPSPALR
RYBa      MSEAGEATTG GTTHPQAAAD APAAAPPDPA PKSPAASGAP QAPAPAALLA
APYB      .......... .......... .......... .......... ......MADT
FRGY2     .......... .......... .......... .......MSE AEAQEPEPVP

          51                                                    100
CS7.4     .......... .......... ........MT GIVKWFNADK GFGFITPDDG
CspB      .......... .......... .......MLE GKVKWFNSEK GFGFIEV.EG
grp2      .......... .......... MAEESGQRAK GTVKWFSDQK GFGFITPDDG
MSY1      ASGAGSGGPG GLTSAAPAGG DKKVIATKVL GTVKWFNVRN GYGFINRNDT
YB1       GSGAGSGGPG GLTSAAPAGG DKKVIATKVL GTVKWFNVRN GYGFINRNDT
FRGY1     .......... ....PAATVG DKKVIATKVL GTVKWFNVRN GYGFINRNDT
NSEP1     RRRERW..PG RLTSAA.LRR DKKVIATKVL GTVKWFNVRN GYGFINRNDT
CE21E7    .......... .......... .......... .TVKWFNVKN GYGFINRTDT
RYBa      GAPSPRARPG LISPRGKRGR EKKVLATKVL GTVKWFNVRN GYGFINRNDT
APYB      EKQPEVEENQ PDQEQNEEQK EKKIIASQVS GTVKWFNVKS GYGFINRDDT
FRGY2     QPESEPEIQK PGIAAARNQA NKKVLATQVQ GTVKWFNVRN GYGFINRNDT

          101                                                   150
CS7.4     SKDVFVHFSA IQNDG....Y KSLDEGQKVS FTIESGAKGP A.AGNVTSL.
CspB      QDDVFVHFSA IQGEG....F KTLEEGQAVS FEIVEGNRGP QAA.NVTKEA
grp2      GEDLFVHQSG IRSEG....F RSLAEGETVE FEVESGGDGR TKAVDVTGPD
MSY1      KEDVFVHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKG. AEAANVTGPG
YB1       KEDVFVHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKG. EEAANVTGPG
FRGY1     KEDVFVHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKG. AEAANVTGPE
NSEP1     KEDVFVHQTA IKKNNPRKYL RSVGDGETVE FDVVEGEKG. AEAANVTGPG
CE21E7    NEDIFVHQTA IINNNPNKYL RS.GDNEEVM FDIVKGSKG. LEAA..TGPD
RYBa      KEDVFVHQTA IKKNNHVKYL RSVGDGETVE FDVVEGEKG. AEAANVTGPD
APYB      KEDVFVHQTA IVKNNPRKYL RSVGDGEKVE FDVVEGEKG. NEAANVTGPE
FRGY2     KEDVFVHQTA IKKNNPR... .......... ......EKG. AEAANVTGPG

          151                                                   200
grp2      GAAVQGGRGG GGGGGGRGGG GYGGGSGGYG GGGRGGSRGY GGGDGGYGGG
MSY1      GVPVQGSKYA ADRNHYRR.. .YPRRRGPPR NYQQNYQNSE SGEKNEGSES
YB1       GVPVQGSKYA ADRNHYRR.. .YPRRRGPPR NYQQNYQNSE SGEKNEGSES
FRGY1     GVPVQGSKYA ADRNHYRR.. .YPRRRGPPR NYQQNYQNNE SGEKAEENES
NSEP1     GVPVQGSKYA ADRNHYRR.. .YPRRRGPPR NYQQNYQNSE SGEKNEGSES
CE21E7    GGPVQGSKYA ADRAENAR.. .GRC*RGRR. .......... ..........
RYBa      GVPVEGSRYA ADRRRYRR.G YYGRRRGPPR N......... AGEIGEMKDG
APYB      GSNVQGSKYA ADRRRFRRGG WYPRFRGGGR .........G GRPRQDMDDG
FRGY2     GVPVKGSRFA PNRRRFRRRF YRPRADTAGE S......GGE GVSPEQMSEG

          201                                                   250
grp2      GGYGGGSRYG GGGGGYGGGG GYGGGGSGGG SGCFKCGESG HFARDCSQSG
MSY1      APEG.QAQQR RPYRRRRFPP YYMRRPYARR PQYSNPPVQG EVMEGADNQG
YB1       APEG.QAQQR RPYRRRRFPP YYMRRPYGRR PQYSNPPVQG EVMEGADNQG
FRGY1     APEGDDSNQQ RPYHRRRFPP YYSRRPYGRR PQYSNAPVQG EEAEGADSQG
NSEP1     APEARPNNA. .AAYAGEVPT LLHAETYGRR PQYSNPPVQG EVMEGADNQG
RYBa      VPEGAQLQVH RN..... .......... .......... ..........
APYB      APDFMPSPRG RG....RGRP YYQNRRYFGP PRRGGGR.QY LEGEGEYQLQ
FRGY2     ERGEETSPQQ RPQRRRPPPF FYR..... .......... ..........
```

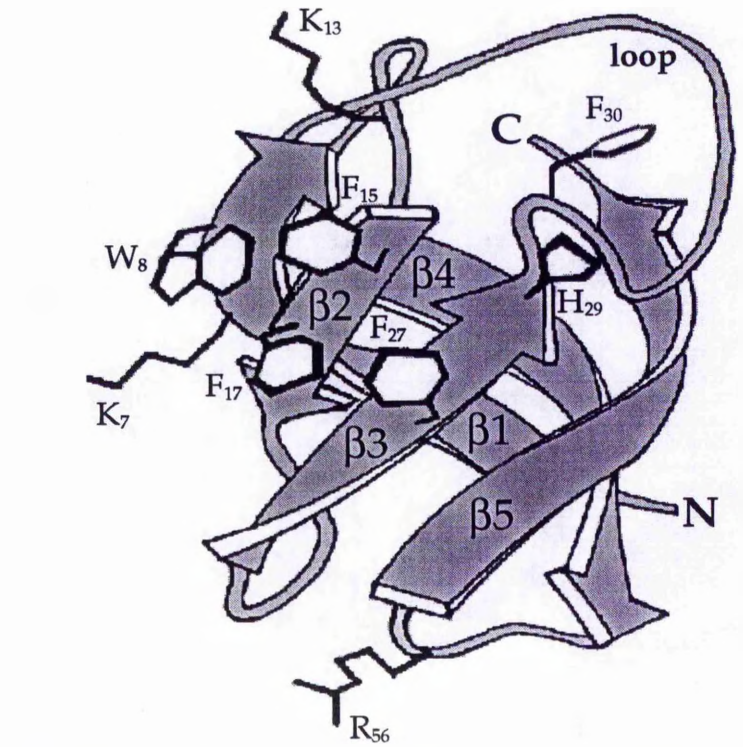
```
          251                                                       300
grp2      GGGGGGRFGG GGGGGGGGGC YKCGEDGHFA RECTSGGR.. ..........
MSY1      AGEQGRPVRQ NMYRGYRPRF RRGPPRQRQP REDGNEEDKE NQGDETQGQQ
YB1       AGEQGRPVRQ NMYRGYRPRF RRGPPRQRQP REDGNEEDKE NQGDETQGQQ
FRGY1     TDEQGRPARQ NMYRGFRPRF RRGPPRQRQP REEGNEEDKE NQGDETQSQP
NSEP1     AGEQGRPVRQ ICIGDIDHDS AGALLAKRQP REDGNEEDKE NQGDETQGQQ
RYBa      .......... ...PTYRPRF RRGPARPRPA PAIGEAEDKE NQQAANGPNQ
APYB      RDQGFRGARR PFYRPLLRTT SQGLLRRWLL R......... .LPRRTTQGR
FRGY2     ..ASGDDPQR PPPRRFRQRF RRPFRPRPAP QQTPEGGDGE TKAESGEDPR

          301                                                       350
MSY1      PPQRR.YRRN FNYRRR.RPE NPKPQDGKET KAADPPAENS SAPEAEQGGA
YB1       PPQRR.YRRN FNYRRR.RPE NPKPQDGKET KAADPPAENS RSRG......
FRGY1     PPQRR.YRRN FNYRRR.RPE NPKSQDGKET KAAETSAENT STPEAEQGGA
NSEP1     PPQAR.YRRN FNYRRR.RPE NPKPQDGKET KAADPPAENS SAPEAEQGGA
RYBa      PSARRGFRRP YNYRRRPRPL NAVSQDGKET KAGEAPTEN. PAPATEQSSA
APYB      TSQARRRERP WGLPQRQRPK PRQR...... .......... ..........
FRGY2     PEPQRQRNRP YVQRRRRQGA TQVAATAQGE GKAEPTQHPA SEEGTPSDSP

          351                  373
MSY1      E......... .......... ...
YB1       .......... .......... ...
FRGY1     E......... .......... ...
NSEP1     E......... .......... ...
RYBa      E......... .......... ...
APYB      .......... .......... ...
FRGY2     TDDGAPVQSS APDPGIADTP APE
```

degrees °C. After 1 to 1.5 hours from the initial cold shock, 13% of total protein synthesis is taken up by CS7.4 (Goldstein *et al.*, 1990). CspB, a homologue of CS7.4, was found in *Bacillus subtilis*, and is 61% identical to CS7.4 (Willimsky *et al.*, 1992). Similar proteins were also found in various other bacteria, such as *Streptomyces clavuligerus* (Avgay *et al.*, 1992), and thermophilic bacteria (Schröder *et al.*, 1993).

The structure of the bacterial cold-shock protein has been determined through X-ray crystallography and NMR spectroscopy: first the *Bacillus subtilis* CspB, a homologue of CS7.4 (Schindelin *et al.*, 1993; Schnuchel *et al.*, 1993), and later *E.coli* CS7.4, also called CspA (Schindelin *et al.*, 1994; Newkirk *et al.*, 1994). The structure of CspB and CS7.4 was found to be a "β-barrel" (see **Fig. 3**). The five antiparallel β-strands of CspB form two β-sheets: sheet-1 consisting of strands β-1 (residues 7-10), β-2 and β-3; sheet-2 consisting of strands β-1 (residues 2-5), β-4 and β-5. Highly conserved aromatic and basic side chains protrude from the solvent face of β-sheet-1, exhibiting features favouring binding to nucleic acids. The arrangement of positive charges would create an attractive potential for nucleic acids, whereas the aromatic rings would be able to stack with the bases of ssDNA or RNA (Schindelin *et al.*, 1994; Newkirk *et al.*, 1994). Close approach of protein and nucleic acid would be facilitated by adjacent glycines and residues with short side chains. These features are characteristic of many RNA-binding proteins: indeed, strand β-2 contains an RNP-1 like motif and strand β-3 contains a rudimentary RNP-2 like motif, these two motifs representing conserved RNA-binding sites in the RRM family of proteins (RNA Recognition Motif, see Kenan *et al.*, 1991; Burd and Dreyfuss, 1994). The RRM, commonly found in a wide range of RNA-binding proteins, is a 90-100 amino-acid domain which forms a four-stranded antiparallel β-sheet packed against two perpendicular α-helices




```
CspB      β1          β2          β3
     MLEGKVKWFNSEKGFGFIEV-EGQDDVFVHFSAI
      :  |  ||||||    |:|||    :   :||||||  :||
...QVQGTVKWFNVRNGYGFINRNDTKEDVFVHQTAI
pp60                RNP-1 like      RNP-2 like

          loop          β4          β5          CspB
---QGEGF-KTLEEGQAVSFEIVEGNRGPQAANVTKEA
   :    |  :::  :|   |  |::|||  :|   ||||||
KKNNPRKFLRSVGDGETVEFDVVEGEKGAEAANVTGPG...
                                          pp60
```

Figure 3 Diagram of the five-stranded β-barrel structure (β1-β5) of the cold-shock protein of *Bacillus subtilis* (CspB). (Adapted from Schindelin *et al.*, 1993). The aromatic and basic side chains exposed on the solvent face are indicated in the diagram and highlighted in the amino-acid sequence of CspB. An alignment with pp60 shows the high level of similarity in the β-strand structures. The location of RNP-1 and RNP-2 like motifs is indicated.

(the arrangement is β1/α1/β2/β3/α2/β4). Within the RRM, the RNP-1 and RNP-2 motifs are especially conserved. The consensus RNP-1 motif is (K/R)-G-(F/Y)-(G/A)-F-V-X-(F/Y) and the consensus RNP-2 motif is (L/I)-(F/V)-(V/I)-(G/K)-(N/G)-(L/M). Both the CSD and the RRM present similar conserved basic and aromatic residues on the surface of the β-sheet.

Because of the impressive sequence conservation between the cold-shock proteins and the CSD in eukaryotic YB proteins (see **Fig.s 2** and **3**), it is reasonable to assume that the eukaryotic CSD also adopts a β-barrel configuration. In *Xenopus*, the CSD has been shown to be essential for binding to Y-box containing promoters in the case of FRGY1 and FRGY2/pp60 (Tafuri and Wolffe, 1992). According to the crystallography and NMR literature, the structure of the cold-shock proteins does suggest binding to single-stranded nucleic acids, either DNA or RNA, although the structure of the CSD/DNA or CSD/RNA complex is not yet known. *In vitro* binding studies confirm that CspB binds single-stranded, but not double-stranded, oligomers containing the Y-box sequence (Schindelin *et al.*, 1993; Graumann and Marahiel, 1994; Schnuchel *et al.*, 1994). Likewise, the CSD in FRGY2/pp60 prefers ssDNA over dsDNA containing the Y-box sequence (Murray, 1994).

The bacterial protein rho, an RNA-binding transcription terminator, also contains an RNP-1 like sequence (GFGF) which can photocrosslink to RNA, but not when the phenylalanines in GFGF are mutated (Brennan and Platt, 1991). Similarly, gas phase chromatography revealed that the phenylalanine residues present in at least two GFGF sequences present in the RRM of hnRNPA1 were crosslinked to RNA (Merrill *et al.*, 1988). It is tempting to speculate that equivalent conserved residues present in the

CSD may interact with DNA in the Y-box sequence, and/or with relevant RNA sequences in mRNA. In addition to the importance of the above phenylalanines, the importance of conserved basic residues was also shown in a study where an R to Q substitution in the U1A snRNP protein RNP-1 motif abolished binding to RNA (Jessen *et al.*, 1991).

The β-barrel structure of the cold-shock proteins is also reminiscent of the OB-fold (oligosaccharide/oligonucleotide binding fold) proteins (Murzin, 1993). These proteins bind to oligosaccharides, such as the *E. coli* toxins LT (heat-labile enterotoxin) and VT1 (verotoxin-1), or oligonucleotides, such as staphylococcal nuclease and the anticodon binding domain of yeast asp-tRNA synthetase. Like the CSD, the β-barrel structure of the OB-fold proteins is formed from two antiparallel β-sheets derived from five β-strands. Like the CSD, the first three β-strands (β1-3) define the face of the β-barrel which interacts with the oligosaccharides or oligonucleotides. Although the tertiary structures of the CSD and OB-fold proteins are similar, not much overall sequence conservation is apparent.

One additional feature in the CSD is the loop between strands β3 and β4 of the CSD. In FRGY2 (pp60), this is: ...TAIKKNNPRKFLRSVGDGE..., see **Fig. 3**; it contains, among others, conserved basic and aromatic residues. In comparison, the structure of the ribosomal protein S17 from *Bacillus stearothermophilus* has been determined (Golden *et al.*, 1993). S17 is a ribosomal RNA-binding protein of 10 kDa, located in the small (30S) subunit of the prokaryotic ribosome, and has a counterpart in eukaryotes. The structure is reminiscent of the CSD and OB-fold proteins, with five β-strands giving rise to a β-barrel. However, S17 has more extensive loop regions between the β-strands, and Golden *et al.* (1993) suggest that these are involved in RNA binding. Likewise, the above loop in FRGY2 and other YB proteins may participate in nucleic acid binding.

The structure of the CSD is therefore related to various prokaryotic and eukaryotic nucleic acid binding domains. During the course of evolution, the CSD has been associated with a number of "auxiliary" domains. The notion of auxiliary binding domains has been described in the context of the RRM family of proteins (Biamonti and Riva, 1994). For example, in the plant kingdom, a CSD-containing protein, grp-2, has been cloned from *Nicotiana sylvestris* and *Arabidopsis thaliana* (Obokata *et al.*, 1991). The CSD in grp-2 is coupled to a glycine-rich domain which also contains a number of arginines and aromatic residues, and is reminiscent of the "RGG" motif present in certain RNA-binding proteins such as nucleolin, hnRNPA1 and hnRNPU (reviewed in Burd and Dreyfuss, 1994). However, except for the YB protein unr, the CSD of known vertebrate and invertebrate YB proteins is associated with a series of alternating basic and acidic charged islands. More commonly, there are four sets of alternating charged islands, but in certain YB proteins there are variations in their arrangement. Together, the charged islands will be described as the "tail domain" (TD). The modular structure of different YB proteins is presented in **Fig. 4**. The charged islands in the TD are thought to be responsible for conferring an anomalous mobility on SDS-PAGE, so that a 35 kDa YB protein such as FRGY2 (pp60) has an apparent mobility on SDS-PAGE of 60 kDa. The basic islands are rich in arginine, proline, tyrosine/phenylalanine and glutamine/asparagine and are shown in **Fig. 5**. The acidic islands are thought to adopt an α-helical configuration (Tafuri and Wolffe, 1992), and contain potential casein-kinase II phosphorylation sites. Proposed functions for the TD are to promote protein multimerization, perhaps via charge interactions between oppositely charged regions, and to stabilize RNA-binding (Tafuri and Wolffe, 1992).

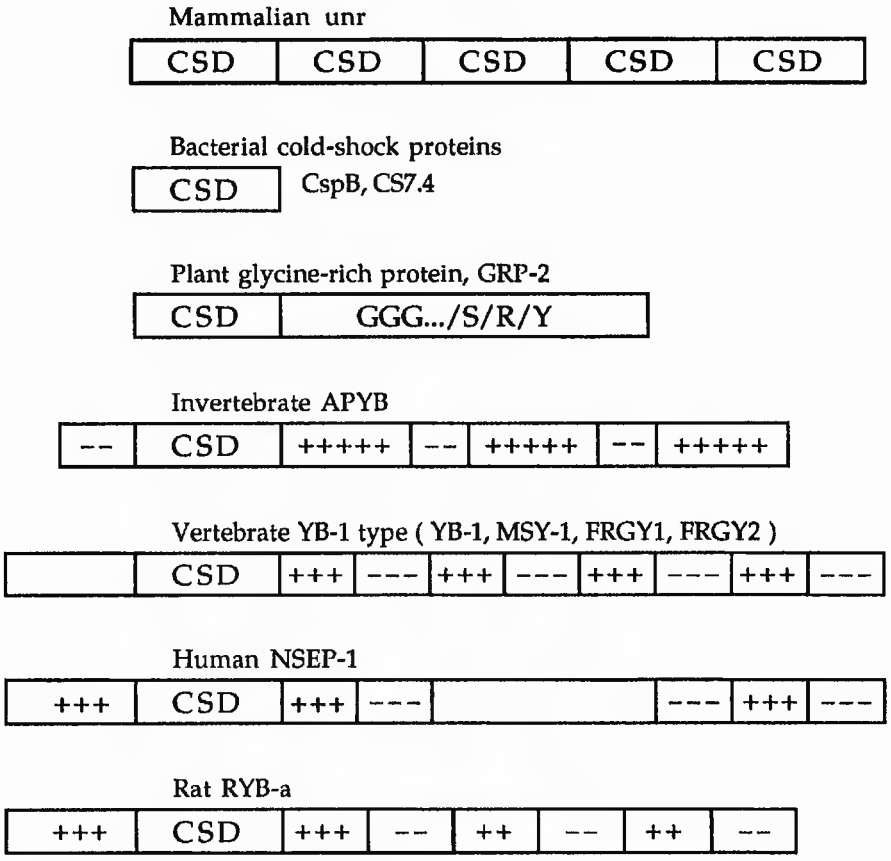


Figure 4 Schematic comparison of the structure of various YB proteins. The ancestral cold-shock domain (CSD) corresponds to the cold-shock proteins of *Bacillus subtilis* (CspB;) and *Escherichia coli* (CS7.4/CspA). In the eukaryotic YB proteins, ancillary domains have been added to the CSD. Mammalian unr (upstream of N-*ras*) consists of a five-fold repeat of the CSD, the odd-numbered CSDs being the most highly conserved. Plant GRP-2 (glycine-rich protein) has a single CSD coupled to a glycine-rich domain which contains interspersed serine, arginine and tyrosine residues. The most commonly described eukaryotic form consists of a single CSD coupled to a series of typically four alternating basic (+) and acidic (-) regions.The invertebrate *Aplysia* APYB, human NSEP-1 (nuclease sensitive element protein) and rat RYB-a have different arrangements of charged regions. The sequence N-terminal to the CSD is quite variable and its function is still undetermined.

pp60	TAIKKNNP**R**KFL**R**SVGDG	**loop**
pp56	TAIKKNNP**R**KFL**R**SVGDG	
FRGY1	TAIKKNNP**R**KYL**R**SVGDG	
pp60	KGS**R**FAPN---**RRR**F**RRR**FY**R**P**R**---------	
pp56	KGS**R**FAPN---ST**R**F**RR**QFY**R**P**R**---------	**B/A1**
FRGY1	QGSKYAAD**R**NHY**RR**YP**RRR**GPP**R**NYQQNYQNN	
pp60	PQQ**R**P-Q**RRR**PPPFFY**RRR**F**RR**GP**R**PNNQQNQ	
pp56	PQQ**R**P-Q**RRR**PPPFFY**RRR**F**RR**GP**R**PNNQQNQ	**B/A2**
FRGY1	NQQ**R**PYH**RRR**FPPYYS**RR**PYG**RR**PQYSNAPVQ	
pp60	PQ**R**PPP**RR**F**R**Q**R**F**RR**PF**R**P**R**PAPQ-QTP	
pp56	PQ**R**PPP**RR**FQQ**R**F**RR**PF**R**P**R**PPPP-QTP	**B/A3**
FRGY1	QG**R**PA**R**QNMY**R**GF**R**P**R**F**RR**GPP**R**Q**R**Q**P**R**	
pp60	P**R**PEPQ**R**Q**R**--N**R**PYVQ**RRRR**QGATQ	
pp56	---EPQ**R**Q**R**--N**R**PYVQ**RRR**AQQPP-	**B/A4**
FRGY1	TQSQPPPQ**RR**Y**RR**NFNY**RRRR**P----	
HIV-1 Tat	YG**R**KK**RR**Q**RRR**PPQGSQ	
HIV-1 Rev	T**R**QA**RR**N**RRRR**W**R**E**R**Q**R**	

Figure 5 Structure of the TD RNA-binding domains. Comparison of the four basic/aromatic islands (B/A 1-4) of the tail domain and the loop between β3 and β4 strands of the CSD, *Xenopus* oocyte pp60, pp56 and their somatic relative FRGY1 are shown. Note the abundance of arginine residues (highlighted in bold), as well as the numerous proline, asparagine/glutamine, and phenylalanine/tyrosine residues. The B/A islands have been shown to interact with RNA. Listed, for comparison, are the RNA-binding regions of HIV-1 proteins Tat (residues 47-63, Green *et al.* 1989) and Rev (residues 34-50, Kjems *et al.* 1991), emphasizing a similar abundance of arginine residues.

The preference for arginine over lysine in the basic regions of the TD is striking. Various RNA-binding proteins have been shown to contain arginine-rich motifs interspersed with aromatic residues, such as the "RGG" motif. The preference for arginine over lysine in these domains may reflect the ability of arginine to provide more hydrogen bonds. In the context of the arginine-rich regions of the HIV protein Tat, Calnan *et al.* (1991) proposed the arginine fork model, in which a single arginine could form bifurcated hydrogen bonds with two phosphate residues in the TAR element. The TAR (Tat-responsive) element is a stem-loop structure present at the very 5' end of HIV transcripts (reviewed in Gait and Karn, 1993). It is clear that the basic islands in FRGY2 (pp60) and other YB proteins prefer arginine over lysine. The arginine-rich areas in the TD are interspersed with aromatic residues which might stack between bases of the RNA, and with amide groups which might contribute hydrogen-bonds.

In summary, the CSD is the most highly conserved domain present in all YB proteins. Of ancient origin, it is thought to bind to the Y-box sequence. During evolution, the CSD has been linked to auxiliary domains. However, there is considerable confusion as to what the YB proteins can bind to: in addition to the Y-box sequence, the list includes the W-box also present in the MHC class II promoters (Hasegawa *et al.*, 1991), the B-box in present rRNA III promoters (Cohen and Reynolds, 1991), apurinic DNA (Hasegawa *et al.*, 1991), pyrimidine-rich single stranded DNA present in triplex structures (Kolluri *et al.*, 1992; Grant and Deeley, 1993; Horwitz *et al.*, 1994), and even purine-rich stretches in the RSV LTR (Kandala and Guntaka, 1994), and more to the point of masked messages, a wide range of mRNA sequences (Marello *et al.*, 1992)! Whether all of these different binding activities occur *in vivo* as well as *in vitro* remains to be shown.

One possibility is that these different binding activities are due to the presence of multiple binding domains, such as the CSD and TD.

1.9 Expression of YB protein genes

Messenger RNAs encoding YB proteins can be detected at different levels in many different somatic tissues and gametocytes. In the *Xenopus* oocyte, mRNAs encoding both FRGY1 and FRGY2 (pp60) are present, but whereas FRGY2 mRNA is translated, being required for oogenesis, FRGY1 mRNA is masked in the oocyte (Wolffe *et al.*, 1992). MSY1, which is very similar to FRGY1 (see **Fig. 2**), is highly expressed in the mouse spermatocyte, and is now believed to be present in masked mRNP (Tafuri and Wolffe, 1993). This finding suggested that although FRGY2 would appear to be the oocyte-specific YB protein in *Xenopus*, MSY1 and by analogy FRGY1 can both function as mRNA masking proteins in somatic tissues where masking is required. This was confirmed when a major rabbit reticulocyte mRNP protein, p50, was identified and cloned, and is analogous to MSY1/FRGY1 and not FRGY2 (Evdokimova *et al.*, 1995; see section 1.6).

cDNAs encoding YB proteins have been cloned from a variety of somatic tissues, including lymphocytes, liver cells, placenta and lens. The closely related group FRGY1/YB-1/EFI_A/MSY1 are the most studied so far in terms of their expression. *Xenopus* FRGY1 mRNA was detected in skin, liver, oviduct, heart, kidney, ovary and testis, whereas FRGY2, the germ-line specific form, most abundant in the oocyte, declines from oocyte stage II onwards (Tafuri and Wolffe, 1990). Spitkovsky *et al.* (1992) examined the expression of human YB-1 in 14 tissues derived form a 24-week old fetus, and classified its expression into three broad categories: (i) high expression in the cerebrum, heart, muscle, adrenal gland, lung and liver; (ii) low

expression in bone marrow, spleen, thymus and kidney (low in the adult kidney too); and (iii) very low if not undetectable expression in the bladder, testis, stomach and pancreas. In 15-day old mice, MSY1 increases in the testis and persists throughout adulthood, one hundred-fold more abundant in the testes than in other tissues examined, and is switched on in individual spermatocytes at the pachytene stage of meiosis I (Tafuri *et al.*, 1993). Human YB-1 presumably appears at a comparable stage of spermatocyte development. The detailed expression profile in the fetus, where development is in progress, is thus not necessarily reflected in the adult. It is reasonable to expect that an analogous picture exists in many other species.

In the context of the immune system, findings by Sabath *et al.* (1990) made an interesting suggestion: that YB proteins are associated with cell proliferation. They found that interleukin-2, a growth factor which induces the proliferation of B-cells in response to antigen, switches on YB-1 as well as glycolytic enzymes, cytoskeletal proteins and translational machinery components, and suggested that YB-1 may be involved in the push from the G to the S phase of the cell-cycle.

In 1993, Grant and Deeley examined the expression of chicken YB-1 in the liver. It is relatively abundant at day 7 of embryogenesis, decreases 10-fold by day 20, a further 3-5-fold by hatching, and then decreases to the adult level. It is highly expressed in the adult gizzard, testis, and fetal liver; expressed at a lower level in the kidney, intestine, spleen and heart, and at an even lower level in the brain, lung and female liver. In their study of the expression of the liver-specific, estrogen-dependent *apoVLDLII* gene, Grant and Deeley noted that YB-1 mRNA levels correlate inversely with its mRNA, suggesting that YB-1 may be a negative regulator. The levels of YB-1 mRNA correlated with a binding activity to a site in the *apoVLDLII*

gene promoter (however, the activity was not conclusively shown to be due to YB-1). In addition, they noted that in a proliferating hepatoma cell line (LMH cells), chicken YB-1 mRNA was present at several-fold higher levels than in 9-day old embryos and 30 to 40-fold higher relative to normal adult liver. Moreover, when estrogen was administered to roosters, YB-1 mRNA levels increased 8-fold by 24 hours, whereas serum albumin mRNA declined two-fold. Estrogen results in a round of DNA replication within 24 hours. Could the induction of YB-1 be associated with cell proliferation and DNA synthesis? CCl_4, which chemically induces liver regeneration, resulted in a ten-fold increase in YB-1 mRNA in 24 hours, and a six-fold increase in 48 hours. At the same time, mRNA levels of the transcription factor C/EBP declined: C/EBP is associated with non-proliferative states and transactivates the albumin promoter. Moreover, in the acute-phase or stress response induced by silver nitrate, YB-1 levels did not increase while serum amyloid A, a protein associated with the stress response, increased. Hence it was suggested that YB-1 expression is associated with cell proliferation but not with the stress response.

Further evidence that YB proteins are associated with cell proliferation comes from studies on the rat YB protein RYB-a, recently described by Ito *et al.* (1994). Like chicken YB-1, RYB-a is also highly expressed in the fetal but not in the adult liver, and its expression is induced by partial hepatectomy. RYB-a is also induced in quiescent fibroblasts in cell culture when treated with serum. Furthermore, when their progression into S-phase was inhibited by preventing cell adhesion or by adding genistein, a specific inhibitor of tyrosine-kinase, levels of RYB-a mRNA declined. The expression of rat RYB-a and its association with cell proliferation are consistent with the observed expression of chicken YB-1 described above.

So far, this information suggests that the YB proteins are involved in developmental processes starting in the developing gametocyte, and are also involved in cell proliferation. YB proteins are expressed in interleukin-2 induced proliferating B-cells; in serum-induced fibroblasts; in liver cells after partial hepatectomy, estrogen or CCl_4 treatment , and are highly expressed in a hepatoma cell line. Now we consider their target genes, discussing examples of positive and negative transcriptional regulation by the YB proteins.

1.10 Genes positively regulated by YB proteins

The bacterial cold-shock proteins are thought to activate a whole range of cold-shock induced genes that contain one or more Y-box like sequences in their promoters. The expression of these genes helps the bacterium survive in temperatures of 10 °C. Examples of such cold-shock inducible genes are *hns*, encoding the nucleoid protein H-NS (La Teana *et al.*, 1991), and *gyr*A, encoding a DNA gyrase (Jones *et al.*, 1992); both of these contain Y-box like sequences in their promoters. As was mentioned earlier, the prokaryotic cold-shock protein, *E. coli* CS7.4, has been shown to bind to these sequences at least *in vitro*, and similarly the eukaryotic CSD is required for recognition of the Y-box. A Y-box sequence is also present in the promoter of *cspa* itself, suggesting a positive feedback mechanism (Tanabe *et al.*, 1992). It is noteworthy that CS7.4 should be made so abundantly (13% of total protein synthesis). Two possible explanations are: high concentration of CS7.4 could allow a more efficient transcriptional effect; alternatively, it could have additional roles, such as RNA protection (Graumann and Marahiel, 1994).

Y-box sequences are present in many gene promoters that are active in *Xenopus* oocytes, but not necessarily in all oocyte-specific genes (Tafuri

and Wolffe, 1990). For example, two copies of the Y-box are present in the hsp70 gene, one of which is essential for oocyte transcription (Bienz, 1987), and the TFIIIA oocyte-specific gene has a sequence in its promoter that is 9/12 identical to the Y-box (Tso *et al.*, 1986). That the Y-box can enhance transcription in oocytes is suggested by the fact that only when the HSV (herplex simplex virus) thymidine kinase promoter is engineered to contain a Y-box sequence which is closer to the consensus sequence, it confers maximal transcriptional activity in oocytes (Graves *et al.*, 1986). The testis-specific histone H2B genes from sea-urchin (Barberis *et al.*, 1987) and rat (Hwang *et al.*, 1990) also contain Y-boxes. Recently, the expression of the mouse protamine-2 gene, which contains a Y-box, was shown to be positively regulated by YB proteins (Nikolajczyk *et al.*, 1995). Therefore YB proteins would appear to drive the transcription of genes encoding products that are required for gametogenesis.

Sabath *et al.* (1990), who reported that YB-1 was one of the genes induced in B-cells by the growth factor interleukin-2 (IL-2), suggested that due to the presence of Y-box sequences, YB-1 may have a role in activating the expression of proliferation associated genes such as thymidine kinase (Lipson *et al.*, 1989), PCNA (proliferating cell nuclear antigen, Travali *et al.*, 1989), DNA polymerase α (Pearson *et al.*, 1991), and c-*erbB*-2 (Sakura *et al.*, 1988). More recently, the promoter of the *Xenopus cdk2* gene (encoding cyclin-dependent kinase 2) has been cloned, and has at least two Y-box like sequences (Olive *et al.*, 1994). This enzyme is required for the progression of the cell cycle. Its mRNA is highly expressed in oocytes, but as is the case for c-*mos*, it is kept as a masked message until required.

The picture is complicated by the suggestion that YB proteins can bind alternative targets other than the Y-box sequence: for example, H-DNA. Regions of strong purine-pyrimidine strand asymmetry which can adopt

the H-DNA triplex configuration are identified as promoter elements of the human c-*myc* gene, c-Ki-*ras* gene and the EGFR gene (Firulli *et al.*, 1992; Kolluri *et al.*, 1992). In each instance, a YB protein specifically binds pyrimidine-rich strands. Although the pattern of expression of these genes fits the cell proliferation model already proposed, and indeed the stabilization of H-DNA by YB and other proteins has been suggested as a mechanism for upregulating the c-*myc* gene (Firulli *et al.*, 1994; Kinniburgh *et al.*, 1994), it must be concluded that regulation of gene expression by YB proteins might be effected through more than one type of promoter element. The binding of YB proteins to H-DNA is a distinct possibility but is not yet conclusively demonstrated *in vivo*, whereas binding to the Y-box sequence, which gives the proteins their name, is more widely accepted. A list of Y-box sequences is presented in **Fig. 6**, while the structure of the H-DNA triplex thought to form in the c-*myc* promoter is presented in **Fig. 7**.

YB-1 has also been involved in viral transcription. Indicative of this, EFI_A (YB-1) was cloned with the RSV LTR enhancer which contains a reverse CCAAT box (Ozer *et al.*, 1990). Y-box sequences are also present in HTLV and HIV promoters: for example, in HIV-1, the sequence CTGATTGGCAGA lies 370 bases upstream of the transcriptional start point (Kashanchi *et al.*, 1994). This site, named "site A", is placed at the 5' end of a negative regulatory element (NRE). A functional study involving cotransfection assays demonstrated the positive influence of YB-1 on viral transcription; a mutated Y-box abolished this effect. Because the Y-box element and the NRE are in close proximity, it may be envisaged that the Y-box proteins compete with some negative regulator. As for the induction of YB-1 expression in infected cells, the authors did not find that viral proteins, such as Tax_1, which transactivate viral gene expression and a number of cellular genes, themselves induce YB-1. Kashanchi *et al.* (1994)

Human HLA DRα	CTG**ATTGG**CCAA
Human *mdr*1	CTG**ATTGG**CTGG
Human c-*erbB*-2	GTG**ATTGG**GAGC
HSV thymidine kinase	GTC**ATTGG**CGAA
RSV LTR	CCG**ATTGG**TGGA
HIV-1 LTR	CTG**ATTGG**CAGA
Xenopus hsp70	CTG**ATTGG**CTAA
Xenopus TFIIIA	CTG**ATTG**CCAAT
Rat testis H2B	CAG**ATTGG**CTCA
Sea urchin testis H2B	CTG**ATTGG**CTAA

Figure 6 Alignment between Y-box sequence elements from a variety of eukaryotic promoters. Y-box sequences are present in the human MHC class II genes, for example the HLA DRα gene (Didier *et al.*, 1988); the human multidrug resistance gene *mdr*1 (Goldsmith *et al.*, 1993); the oncogene c-*erbB*-2 (Sakura *et al.*, 1988); HSV thymidine kinase (Graves *et al.*, 1986); RSV LTR (Ozer *et al.*, 1988) and HIV-1 LTR (Kashanchi *et al.*, 1994); various germ-cell specific genes: *Xenopus hsp70* (Bienz, 1987); *Xenopus* TFIIIA (Tso *et al.*, 1986); sea urchin testis histone H2B (Barberis *et al.*, 1987); rat testis H2B (Hwang *et al.*, 1990). Note the conservation of the "reverse CCAAT box", ATTGG, which is often preceded by CTG.

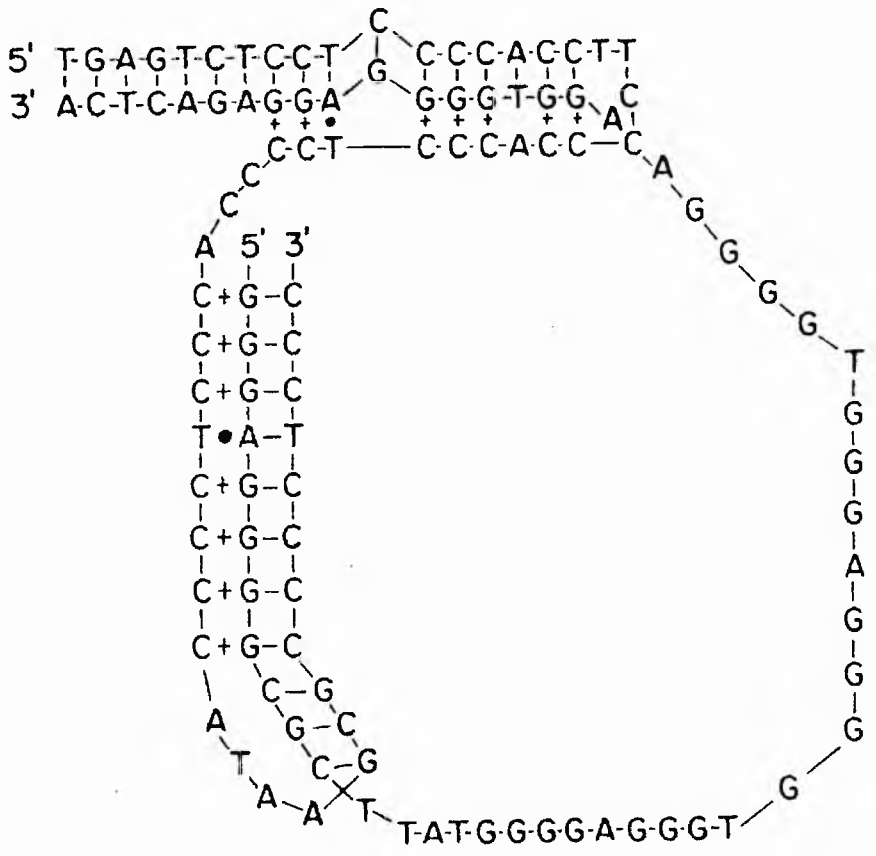


Figure 7 Tandem H-DNA structure of the c-*myc* NSE (nuclease sensitive element). The diagram is from Kinniburgh *et al.*, 1994. Watson/Crick base pairs are represented by dashes (-); TAT Hoogsten base pairs by dots (•) and CGC Hoogsten base pairs by plus (+) signs. The structure is referred to as "tandem H-DNA" because it consists of two juxtaposed DNA triplexes. The displaced single strand is purine-rich and in particular G-rich.

suggest that this induction may happen independently of viral proteins, for example via the interleukin-2 induction of B-cell proliferation which happens in response to antigen, as was observed by Sabath *et al.* (1990; see section 1.9). Perhaps significantly, Schuitemaker *et al.* (1994) have claimed that productive HIV infection is evident only in the fraction of cells which are in a proliferative state.

The human *mdr*1 gene encodes an energy-dependent drug efflux pump. It is normally highly expressed in the kidney and adrenal gland, and moderately expressed in the liver and on the apical surface of luminal cells in the colon. It is overexpressed in certain cancers which become resistant to chemotherapy, wherein the drugs are ejected from the cell by the efflux pump. Goldsmith *et al.* (1993) have studied sequences from the promoter using DNaseI footprint analysis, using nuclear extracts from a doxorubicin-resistant ovarian carcinoma cell line. They found that a region between positions -89 and -70 is protected from DNaseI digestion. By testing promoter activity via CAT assays, they found that deleting a region between -89 to -70 resulted in a marked decrease in expression. In this region, the sense-strand contains a 10 bp perfect homology with the Y-box consensus: CTGATTGGCT. A gel retardation assay showed that the binding of a nuclear factor (presumably a YB protein) to a Y-box (CTGATTGGCTGGG) was abolished with the mutated sequence (CTGAT**G**TGCTGGG). This last finding suggests that the residues in the ATTGG reverse CCAAT box are crucial for Y-box sequence recognition. Goldsmith *et al.* (1993) also noted that doxorubicin-resistant cells overexpress the oncogene c-*erbB*-2. Sakura *et al.* (1988) had used the promoter of this oncogene as a ligand to clone YB-1. The implication is that YB proteins are involved in up-regulating both the drug efflux pump, helping the cancerous cell to rid itself of noxious chemotherapeutic drugs,

while at the same time up-regulating cell proliferation genes including c-*erbB*-2. In this scenario, YB proteins would be a definite target in considering a possible "gene therapy" treatment of the condition.

1.11 Genes negatively regulated by YB proteins

The first YB protein to be cloned, as previously described, was human YB-1, using the Y-box ligand sequence from an MHC class II gene (called HLA genes is humans). Because levels of YB-1 mRNA were seen to correlate inversely witht the levels of HLA-DRβ chain mRNA, it was suggested that in that context, YB-1 is a negative regulatory factor (Didier *et al.*, 1988). The mouse homologue of YB-1 was first cloned from lens tissue, in which MHC class II genes are not expressed (Shaughnessy and Wistow, 1992). A later study provided functional evidence that YB-1 represses MHC class II gene transcription (Ting *et al.*, 1994). The activity of the IFN-γ (interferon gamma)-responsive HLA-DRα promoter was assayed using co-transfection techniques. Expression vectors containing YB-1 were co-transfected with promoter-CAT constructs, resulting in a substantially reduced IFN-γ-induced CAT expression. Control constructs containing a mutated Y-box in the HLA-DRα promoter did not result in repression of the constructs. The authors stress that many other factors bind to the cis-acting elements of the MHC class II promoters. For example, the transcription factor NF-Y binds to the CCAAT sequence present within the Y-box, and NF-Y is a positive regulator of MHC class II genes. It is conceivable that NF-Y and YB-1 may compete for the same site, and bind to the promoter with opposite transcriptional effects.

Another example of transcriptional repression mediated by YB proteins is the effect of chicken YB-1 on the expression of fetal γ-globin (Horwitz *et al.*, 1994). The γ-globin gene is active in fetal red blood cell precursors but is

silenced in the adult. Its promoter contains a region of alternating homopurine and homopyrimidine tracts that give rise to an H-DNA triplex structure. The authors claim that YB-1 binds to this structure specifically: a DNaseI footprinting study showed that within the H-DNA structure, two short polypyrimidine stretches are protected from DNaseI digestion. The protected areas are contained and underlined within the sequence TCCTCTTGGGGGCCCCTTCCCCACACT. In addition, the kinetics of binding to this site suggested positive cooperativity between proteins, which was also suggested in the context of the up-regulation of transcription by FRGY1/FRGY2 via the Y-box sequence (Tafuri and Wolffe, 1992). Horwitz *et al.* (1994) propose that YB-1 binds to the H-DNA target in the adult, preventing the association of a distal positive regulator, thus hindering the assembly of appropriate transcription complexes. In the hereditary condition HPFP (hereditary persistence of fetal hemoglobin), in which γ-globin is abnormally expressed in the adult with deleterious effects, point mutations were found in the H-DNA forming region of the γ-globin promoter.

Finally, as was mentioned previously, levels of chicken YB-1 in the liver correlated inversely with the mRNA of *apoVLDLII* (Grant and Deeley, 1993). The rat YB protein RYB-a, which has a comparable expression pattern in the liver, was also suggested to have an additional negative regulatory because its mRNA levels correlate inversely with levels of aldolase mRNA (Ito *et al.*, 1994). Further experiments will be needed to confirm the involvement of YB proteins in the repression of these genes: binding studies and co-transfection assays are not yet available.

1.12 The YB proteins are a versatile family

The examples given in the preceding sections indicate that the YB proteins are a versatile family of gene regulators. The YB proteins are in summary ubiquitous and multi-functional. They are present throughout prokaryotes and eukaryotes in diverse forms, all sharing a highly conserved CSD of ancient origin which recognizes the Y-box promoter element. During evolution, they have been coupled to a variable series of auxiliary domains which probably assist in nucleic acid binding and multimerization. The YB proteins exert their positive and negative transcriptional effects by binding to Y-box elements, H-DNA elements, and, conceivably, to a combination of both, depending on their arrangement in specific promoters. Perhaps their biochemical properties: the recognition of the Y-box, determined by the highly conserved CSD, but also of other DNA structures, such as pyrimidine-rich strands derived from H-DNA, and the ability of these proteins to multimerize cooperatively could be exploited in different promoter contexts with dramatically different outcomes. Much remains to be learned about their transcriptional effects; the structural details of sequence recognition; a possible involvement of the TD in that recognition, as well as the effects of multimerization and phosphorylation of the YB proteins. Moreover, apart from the blocking of RYB-a induction by genistein, an inhibitor of tyrosine kinase, the mechanisms regulating the expression of the YB proteins themselves are still unknown.

The YB proteins are also involved in mRNA storage, having been identified in the *Xenopus* oocyte, mouse spermatocyte and rabbit reticulocyte (section 1.6). They are "bifunctional" in that they drive the expression of certain germ-cell specific genes while at the same time packaging the set of "masked" mRNAs. Although much evidence is still sketchy and incomplete, a distinct pattern is beginning to emerge. In brief, it would appear that the YB proteins help activate germ-cell specific genes which include various cell proliferation genes,

repress certain cell-specific genes, and mask mRNAs from translation. **Fig. 8** summarizes the various proposed functions of the YB proteins.

There is an increasing list of proteins which have been described as interacting both with DNA and RNA. Three examples will be mentioned. The transcription factor TFIIIA, a nine zinc-finger protein, is required for the transcription of 5S rRNA genes, but also binds to 5S rRNA in cytoplasmic storage particles (reviewed in Tafuri and Wolffe, 1993). The RNA-binding protein hnRNPK, which contains an "RGG" domain, is now also thought to interact with DNA: it binds to the pyrimidine-rich element present in the H-DNA region of the c-*myc* promoter (Takimoto *et al.*, 1993), and in conditions in which it binds strongly to a ssDNA sequence, it binds weakly to the corresponding RNA sequence (Gaillard *et al.*, 1994). The pyrimidine tract binding protein (PTB) binds to polypyrimidines in the branchpoint upstream of exon 3 of the gene α-tropomyosin (Patton *et al.*, 1991). PTB is a 57 kDa protein, and contains the RRM (RNA-recognition motif,). In 1992, it was claimed that PTB also binds to ssDNA with high selectivity, recognizing a motif present in the liver specific enhancer of the TAT (aminotransferase) gene (Jansen-Dürr *et al.*, 1992). These examples, together with the YB proteins, suggest that there is considerable cross-talk between transcriptional and translational processes, in that the same proteins can interact with both RNA and DNA.

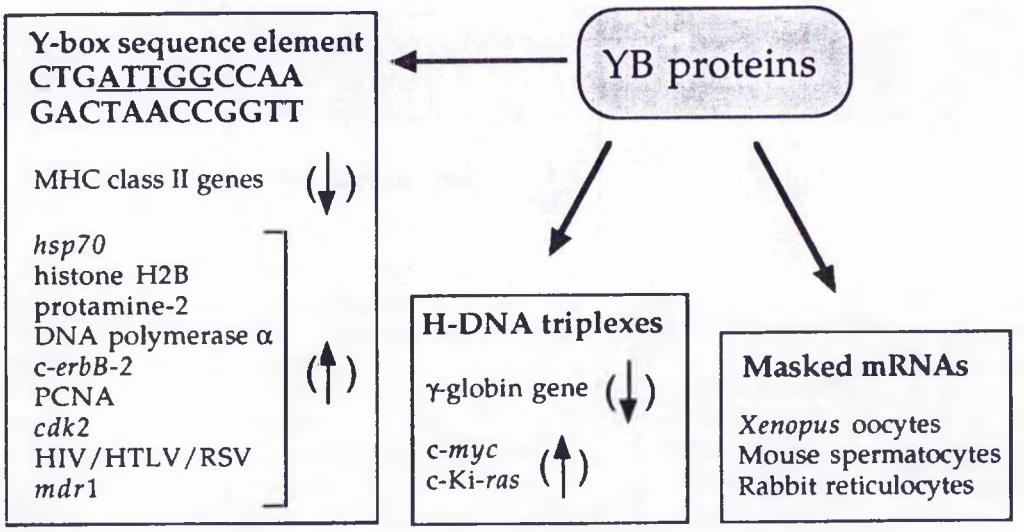


Figure 8 Proposed functions of the YB protein family. The Y-box sequence element is present in a variety of promoters: in MHC class II gene promoters; genes expressed in germ-cells (*hsp70*, histone H2B, protamine-2); cell proliferation genes (which overlap with germ-cell specific genes: DNA polymerase α, the oncogene c-*erbB*-2, PCNA, the cell-cycle dependent kinase *cdk2*); certain retroviral promoters (HIV/HTLV/RSV), and the human multidrug resistance gene (*mdr*1). Arrows denote up or down-regulation. More controversial is gene regulation via H-DNA elements: the cell proliferation genes c-*myc* and c-Ki-*ras* are up-regulated whereas fetal γ-globin is down-regulated. Finally, YB proteins have been identified in at least three masked mRNA particles: the *Xenopus* oocyte, the mouse spermatocyte and the rabbit reticulocyte. References are given in the text.

1.13 Aims of the project

This project aims to improve our understanding of the masked message in *Xenopus* oocytes, by focussing on its protein components. It is hoped that the knowledge gained would apply to other masked messages in other tissues and organisms. Two abundant mRNP proteins have already been identified as belonging to the YB protein family, a multifuctional family which are both transcriptional and translational regulators. The details of their interaction with DNA and RNA are not yet known. At least two potential RNA-binding domains have been identified in the YB proteins: the ancient cold-shock domain (CSD) and the charged tail domain (TD). How these two potential binding domains work, and what their binding specificities are, is not known. Messenger RNP particles will be isolated using existing techniques, and in particular, YB proteins will be extracted to study their interaction with RNA. The masked message particles also contain an associated protein kinase activity, believed to be essential for mRNA masking to occur. The protein kinase phosphorylates the YB proteins, thus regulating their activity; the biochemical consequence of this phosphorylation is not understood. The effect of phosphorylation on RNA-binding will also be considered. Finally, although much attention has focussed on the YB proteins, the other abundant mRNP proteins have been neglected. These additional components are likely to have an important role in the masked mRNP. Using cDNA expression libraries and peptide sequencing techniques, an attempt will be made to characterize these proteins.

Chapter 2

Materials and Methods

2.0 Animals

Xenopus laevis were obtained from Blades Biological, Edenbridge, Kent. Immature 3-6 month old post-metamorphosis females were used to obtain previtellogenic ovaries. Mature females were used to obtain ovarian tissue containing the full complement of oocyte stages, oogenesis being asynchronous in *Xenopus*. To obtain the tissues, the animals were anaesthetized in MS222 (0.2% in water). Whereas immature females were decapitated, mature females survived surgical removal of the ovary. The ovarian tissues were washed in OR2 medium (minus Ca^{2+}) to remove blood, and treated with collagenase, 0.2%, Sigma, type IV, in the same medium for 2 h at 20 °C to disperse the oocytes. The collagenase was washed off with three washes of 20 volumes of OR2 medium, then with three washes of 20 volumes of Barth's solution. Oocytes were grouped into developmental stages according to Dumont (1972) and were stored at -70 °C. Other amphibian species, the mud puppy *Necturus maculosus* and the newt *Notophthalmus viridescens* were also obtained from Blades Biological to prepare oocytes.

OR 2 medium (minus Ca^{2+})

NaCl	82.5 mM
KCl	2.5 mM
$MgCl_2$	1 mM
Na_2HPO_4	1 mM
HEPES	5 mM
PVP	0.05%
NaOH, pH 7.8	38 mM

Barth's solution

NaCl	88 mM
KCl	1 mM
$NaHCO_3$	2.4 mM
$MgSO_4$	0.82 mM
$Ca(NO_3)_2$	0.33 mM
$CaCl_2$	0.41 mM
Tris.HCl, pH 7.6	7.5 mM
Penicillin, streptomycin and kanamycin	5 units

2.1 Isolation of mRNP

In general, previtellogenic ovary was preferred as a source of mRNP. Tissues were sonicated in 2 ml of column binding buffer (CBB), centrifuging the extract at 10,000 g for 15 min to collect the supernatant ("SN10"), which was kept on ice. The SN10 was loaded onto a 1.5 ml oligo(dT) column equilibrated in CBB. The eluate was reapplied three times to ensure saturation of the column. The eluate was monitored by measuring A_{254}, which gives an indication of the RNA content. After collecting the unbound, or poly(A)$^-$ fraction, the column was washed with CBB so that A_{254} returned to background. Finally, the poly(A)$^+$ mRNP was eluted with ~2.5 ml warm dH_20. The elution of the poly(A)$^+$ mRNP was observed as an increase in A_{254}. Both the poly(A)$^+$ and poly(A)$^-$ fractions were stored at -70 °C.

Column Binding Buffer (CBB)

Tris.HCl, pH 7.5	20 mM
KCl	0.3 M
$MgCl_2$	2 mM
DTT	1 mM
NP-40	0.2% (v/v)

2.2 Heat-treatment

This technique followed the procedure reported by Deschamps *et al.* (1991) in which the *Xenopus* oocyte YB proteins were shown to remain soluble after heating whole oocyte extracts to 80 °C. Samples were similarly heated to 80 °C in HTB (see below), cooled on ice and spun in a microcentrifuge, each step for 5 min. In these conditions, the YB proteins remained in the supernatant, whereas the vast majority of other proteins were pelleted. This method was used for partial purification of the YB proteins, but also as a means to reconstitute RNP with a riboprobe (sections 2.41 and 2.42).

Heat-treatment buffer (HTB)
($MgCl_2$, included in Deschamps *et al.*, 1991, was omitted).

Tris.HCl, pH 7.5	10 mM
NaCl	50 mM

2.3 Band excision of proteins

Proteins separated on SDS-PAGE were reversibly stained using the BIO-RAD copper staining kit. Copper staining is a negative staining technique which produces a blue-green opaque background. SDS-PAGE gels were firstly washed in dH_20 and then submerged in 50-100 ml of diluted Copper Stain solution provided in the BIO-RAD kit. Bands were visualized by placing the gel against a black background. Bands of interest were cut out using a razor blade, placed in microcentrifuge tubes, and destained as described in the kit. Once destained, the polyacrylamide was allowed to dry partially (if dried excessively, polyacrylamide becomes very hard) and then macerated using a pipette tip. An equal volume of dH_20 was added to the macerated gel, allowing the proteins to diffuse overnight into the dH_20. Samples containing the proteins of interest (with some residual SDS) were pipetted out.

2.4 Heparin column chromatography

(i) **Purification of mRNP-associated protein kinase activity.** As previously observed in this lab (La Rovere and Sommerville, unpublished), the mRNP-associated kinase appears to be of a casein kinase II type and binds to heparin. Poly(A)$^+$ mRNP were prepared as described in section 2.1, but were eluted from the oligo(dT) column in 60% formamide rather than warm dH_20. This poly(A)$^+$ mRNP was dialyzed against heparin column buffer (HCB; see below) before being applied to a 1 ml column of heparin-Sepharose CL 6B (Pharmacia) equilibrated in HCB. Eluates were reloaded three times, then eluted from the heparin-Sepharose column by increasing the salt concentration up to 1 M NaCl. Protein kinase activity was measured by *in vitro* phospholabelling (section 2.30).

(ii) **Purification of YB proteins.** Following the report by Kolluri *et al.* (1992), in which the human YB protein NSEP-1 was eluted in high salt buffer (containing 0.5 M KCl) from a heparin column, a similar procedure was developed. Heat-treatment supernatant in HTB (containing YB proteins; see section 2.2) was heated to 80 °C in the presence of heparin-Sepharose CL 6B (Pharmacia), so that approximately 100 µg of RNP and 0.5 ml of resin was resuspended in HTB. After cooling with occasional vortexing, the resin was pipetted into a 1 ml column and rinsed with HTB after it had set. Bound RNA could be eluted with 8 M urea and 5 mM $MgCl_2$, but not the proteins which were eluted with 1 M NaCl or 1 M KCl. The procedure has the following potential applications: it provides a simple method of extracting YB proteins from complex mixtures, allows bound RNA to be separated from YB proteins, and could be used in binding assays in which the YB proteins are immobilized on the column.

Heparin Column Buffer (HCB)

Tris.HCl, pH 7.5	25 mM
EDTA	2 mM
2-mercaptoethanol	5 mM

2.5 Preparation of antisera

Antisera were prepared against two mRNP fractions, HTP (heat-treatment pellet) and HTSN (heat-treatment supernatant). Samples containing ~2 mg/ml of the proteins of interest were incorporated into a water-in-oil emulsion using 0.5 ml Drakeol 6VR (Pennsylvania Refining Company) containing 1/10 volume of the emulsifier Arlacel A (Atlas) as the oil phase. The material was mixed throughly through a 26-gauge syringe needle and re-emulsified in 1 ml of Tween 80 (Sigma) in 0.14 M NaCl until a free-flowing milky emulsion was generated. Emulsions were injected into Dutch rabbits as single doses. Bleeds were taken from the marginal ear vein at 6 weeks, and a booster injection identical to the first given to enhance the titre of desired antibodies. After a further 2 weeks, 10 ml of blood were collected at weekly intervals, and the antisera stored at -70 °C.

2.6 λgt11 cDNA expression library

This phage-based cloning system uses the vector λgt11 (Amersham International). λgt11 has a unique EcoRI site near the 3' end of its *lac* Z gene (encoding β-galactosidase) into which cDNAs can be inserted. These can be expressed as lacZ fusion proteins via the IPTG-inducible *lac* promoter. Fusion with lacZ protein improves the stability of foreign peptides in *E.coli*. The host strain is Y1090 [*hsd* ($r_k^-m_k^+$) *lac* U169, pro A+, *lon*⁻, *ara* D 139, *Str* A, *Sup* F *trp* C22:Tn 10 (pMC9)]. The plasmid pMC9 carries the lac repressor, which is inactive when bound by IPTG. The strain is also deficient in the lon protease, and lacks the restriction enzyme Eco K. The

phage λgt11 contains a temperature sensitive lacIq repressor whose product is inactive at 43 °C, allowing lytic growth at that temperature. A *Xenopus* ovary λgt11 cDNA expression library was kindly provided by Dr. Mark Dworkin, c/- Boehringer, Vienna.

2.7 λZAP cDNA expression library

This phage-based cloning system uses the vector λZAP (Stratagene) which has some additional advantages. It allows "directional cloning" so that cDNAs were constructed with an EcoRI site at the 5' end and a XhoI site at the 3' end. The vector has been engineered to allow the direct excision and recircularization of cloned inserts. In brief, the cDNAs were cloned directly into a pBluescript plasmid, itself already cloned into the bacteriophage genome. The plasmid is flanked by initiator and terminator sequences which are recognized by f1 bacteriophage proteins. When XL1-Blue cells were co-infected with λZAP and f1 "helper" phage, the net result was a "phagemid", in which a fully circularized pBluescript containing a cDNA sequence was packaged by f1 bacteriophage proteins. Phagemids have the ability of transforming competent cells by infecting them with pBluescript (see section 2.14). This "zapping" procedure eliminated the need to prepare bacteriophage DNA, restrict and extract the inserted cDNA, and ligate it into pBluescript as required in the λgt11 system.

The cell strain used for plating λZAP was XL1-Blue. In XL1-Blue, The F' episome contains a mutation in the *lac*Z gene required for "α-complementation" of the amino-terminus of the *lac*Z gene in the λZAP vector, so that non-recombinant background plaques remain blue in medium containing X-gal. The F' episome contains the genes required for expression of the bacterial F' pili which are required for infection by filamentous bacteriophage f1. The F' episome also contains the lacIq

repressor and the Tn10 tetracycline resistance gene. XL1-Blue stock was therefore kept on tetracycline-containing plates (12.5 μg/ml).

A cDNA library was prepared by Dr. Sommerville following the instructions in the λZAP manual. cDNAs were prepared from poly(A)$^+$ mRNAs extracted from polysomal mRNP from previtellogenic oocytes (section 2.39), the idea being to select for mRNAs encoding proteins required at a high level during this phase of development, such as mRNP proteins.

2.8 Plating bacteriophage λ

The plating cells were Y1090 for λgt11 and XL1-Blue for λZAP. Individual colonies of these plating cells were picked from the stock agar plates, inoculating 5 ml LB and growing them overnight at 37 °C. The next morning, 40 ml of fresh LB was inoculated with 1 ml of the overnight culture. Maltose and $MgSO_4$ were added to the LB to a final concentration of 0.2% (w/v) and 10 mM respectively. The cells were grown with vigorous shaking until they reached an OD_{600} of ~0.5., after which they were spun at 3,000 g at 4 °C for 5 min, and resuspended in 4 ml of ice-cold, sterile 10 mM $MgSO_4$. These plating cells could be used on subsequent days, although the plating efficiency declined.

"Plating out" began by melting top agarose (LB containing 0.8 % agarose), placing it in a 45 °C water bath, 4 ml of top agarose being required for every 90 mm plate. 100 ml of plating cells were infected with an appropriate dilution of bacteriophage λ in SM buffer. Cells were infected for 15 min at 37 °C, and then plated out by mixing them with 4 ml of top agarose cooled to 45 °C, pouring the mixture onto pre-warmed plates. It was essential that the plates were kept warm and that the top agarose was

at 45 °C, otherwise the top layer would not set evenly. The top layer was allowed to set for 15 min on the bench, after which the plates were placed inverted into a 37 °C oven in the case of λZAP, and a 43 °C oven in the case of λgt11 infections. Plaques appeared after approximately 6 h. When plating out library in the first round of screening, larger plates were preferred (150 mm diameter) in order to plate out 10,000-20,000 pfu (plaque forming units) per plate.

Luria broth (LB) per liter:

Bactotryptone™	10 g
Bacto-yeast extract™	5 g
NaCl	10 g

Dissolve and autoclave immediately

LB-agar plates

Add 15 g of Bacto-agar™ per liter of LB before autoclaving. Cool to 45 °C before pouring plates. Add the required antibiotic.

SM buffer for bacteriophage, per litre:

NaCl	5.8 g
$MgSO_4.7H_20$	2 g
Tris base	6.05 g
2% gelatin	5 ml

Adjust to pH 7.5 with HCl and sterilize by autoclaving.

2.9 Expressing cDNA fusion proteins in plaques

Nitrocellulose filters were prepared by wetting them into a sterile solution of 10 mM IPTG and dried before use. When the plaques were sufficiently developed, the filters were carefully overlayed over the plaques. The plates were returned inverted into the oven and incubated for a futher 3-4 h. This allowed the fusion proteins to be expressed. The filters were marked with a fine needle dipped in Indian ink, pushing it through the filter and agar in

three asymmetrical positions. The filters were lifted off the agar very slowly and labelled as appropriate, ready for screening. In the case of the RNA-binding screen, the lifts were placed overnight into TBST (TBS + 0.5% Tween-20; see section 2.25 for TBS), and for an immunoscreen, they were placed into TBS containing 2% filtered skimmed milk, also overnight. Meanwhile, the agar plates were stored inverted at 4 °C for the subsequent picking of positive plaques.

2.10 Riboprobe-binding screen of fusion proteins

Following the procedure described in Vinson *et al.* (1988), in which a DNA-ligand was used to isolate a DNA-binding protein expressed in a cDNA library, an RNA-binding screen was developed to select for mRNP proteins. The denaturing agent used was 6 M urea dissolved in low salt binding buffer (see below) and was based on previous experience with mRNP proteins (Dearsly *et al.*, 1985).

(i) Filter-bound proteins were denatured to release endogenously bound RNA with 6M urea dissolved in low salt buffer with 2 x 10 min washes at room temperature with gentle agitation. The denaturing solution was washed off with 2-3 abundant washes of low salt buffer.
(ii) The filters were probed use a high activity probe, for example 1-2 µl of freshly made riboprobe, (see section 2.33; using approximately 0.25 µCi / 0.1 µg RNA), carried in 10 ml low salt buffer in a single 90 mm petri-dish. The probe was incubated for 30-60 min with gentle agitation at room temperature.
(iii) The filters were washed with high salt buffer; firstly, two rapid washes to remove unbound probe, followed by two or three 5-10 min washes to reduce background binding, again in high salt buffer. Finally the filters were dried briefly for subsequent autoradiography.

Low salt binding buffer

Tris.HCl, pH 7.5	10 mM
NaCl	50 mM
EDTA	2 mM

High salt washing buffer

Tris.HCl, pH 7.5	10 mM
NaCl	500 mM
EDTA	2 mM

2.11 Picking a positive phage plaque

The immunoscreened filters (section 2.27) or the autoradiographs (section 2.10) were covered with a sheet of plastic/acetate. The outline of the filter and the position of the orientation marks was marked onto the acetate. Next, the position of the positive signals was marked, after which the acetate was aligned with the agar plates. Positive plaques were cored out using a 1 ml Gilson pipette and transferred to a 1.5 ml microcentrifuge tube containing 500 µl of SM buffer. A drop of chloroform was added for storage at 4 °C. Phage particles diffused out of the agar plug into the SM buffer; typically, a plug gave a titre of approximately 10^5 pfu/ml. To purify positive plaques, the initial stocks were plated at lower densities until a single positive isolated plaque could be picked.

To analyze the cloned cDNA: in the case of λZAP clones, pBluescript containing the cloned cDNA was excised directly (section 2.14). In the case of λgt11 clones, bacteriophage DNA was extracted using the plate lysis technique (section 2.12). The cDNA was isolated from the bacteriophage by digesting the phage DNA with EcoRI and "genecleaning" the released insert (section 2.17). The cDNAs were ligated into pBluescript vectors, also

linearized with EcoRI, and transformed into competent cells for subsequent analysis.

2.12 Preparation of bacteriophage λ DNA

Fresh plating cells (*E. coli* strain Y1090 for the λgt11 library, and XL-1 Blue in the case of the λZAP library) were infected with approximately 10^4 pfu and grown overnight, the resulting plaques being almost confluent. 4 ml of SM buffer was added to the plate lysates, which were sealed with tape and left to shake gently on a tray at 4 °C for 2-3 h. The SM lysate was transferred into a sterile tube, and the plates rinsed with a further 1 ml SM, leaving them to drain for a few minutes. The lysate was spun at 4,000 rpm for 10 min at 4 °C to remove cell debris and the supernatant collected, to be stored at 4 °C. To 4 ml of phage lysate, 4 ml of sterile 20% PEG/NaCl solution was added (containing 20 g polyethylene glycol 6000 and 11.7 g NaCl per 100 ml), mixed and left on ice for 1 h. The lysate + PEG solution was spun at 3,000 g for 20 min, draining the pellet and wiping the inside of the tube with a tissue to remove any remaining trace of PEG. The pellet was resuspended in 750 µl of LB, transferred to a microcentrifuge tube to add 750 µl of DE52 in LB (see below), and mixed by inverting 20-30 times. The slurry was microcentrifuged for 5 min and the supernatant transferred to a fresh tube. The supernatant was spun again to remove any remaining slurry. To each 1 ml of supernatant 17.5 µl of a 0.1 mg/ml solution of proteinase K and 42.5 µl of 10% SDS was added, mixing and incubating at room temperature for 5 min. Next, 173 µl of 3 M potassium acetate was added, incubating at 88 °C for 20 min to dissolve the precipitate, and then cooled on ice for 10 min (the precipitate appearing again) before spinning in a microcentrifuge for 10 min and transferring the supernatant to two fresh tubes. Next, an equal volume of cold isopropanol was added,

samples cooled to -70 °C for 10 min, warmed to room temperature and spun for 10 min in a microcentrifuge. The pellets were washed in 70 % ethanol, dried under vacuum and resuspended in a small volume of dH_2O as required.

To prepare the DE52 in LB: 10 g DEAE-cellulose was resuspended in several volumes of 0.05 M HCl, ensuring that the pH dropped below 4.5, adding concentrated NaOH to bring the pH to 7. After the resin settled, the supernatant was decanted, and the resin washed several times in LB, each time using 2 volumes of LB. Finally, the resin was resuspended in a slurry of 75% resin and 25% LB, adding sodium azide as a preservative to a final concentration of 0.1%.

2.13 pBluescript®

The pBluescript vector (Stratagene) is a 2.96 kb colony-producing plasmid. It carries the β-lactamase gene which confers resistance to ampicillin. Inserts were cloned into the polylinker region which contains 21 unique restriction sites (see **Fig. 9**). The pBluescript SK and KS series represent two different orientations of the polylinker. The polylinker sequence immediately follows the N-terminal coding region of the *lac*Z gene. An IPTG-inducible *lac* promoter upstream from the *lac*Z gene allows "α-complementation" of cells containing a deletion in their *lac*Z gene (*lacZΔM15*) to produce a functional β-galactosidase protein. This provides a blue/white colour selection system: an active β-galactosidase results in blue colonies when the substrate X-gal is cleaved, while white colonies result when an inserted sequence interrupts the *lac*Z gene. The vector can be used for restriction mapping; double-stranded DNA sequencing; fusion protein expression when the inserted cDNA is in the correct reading frame

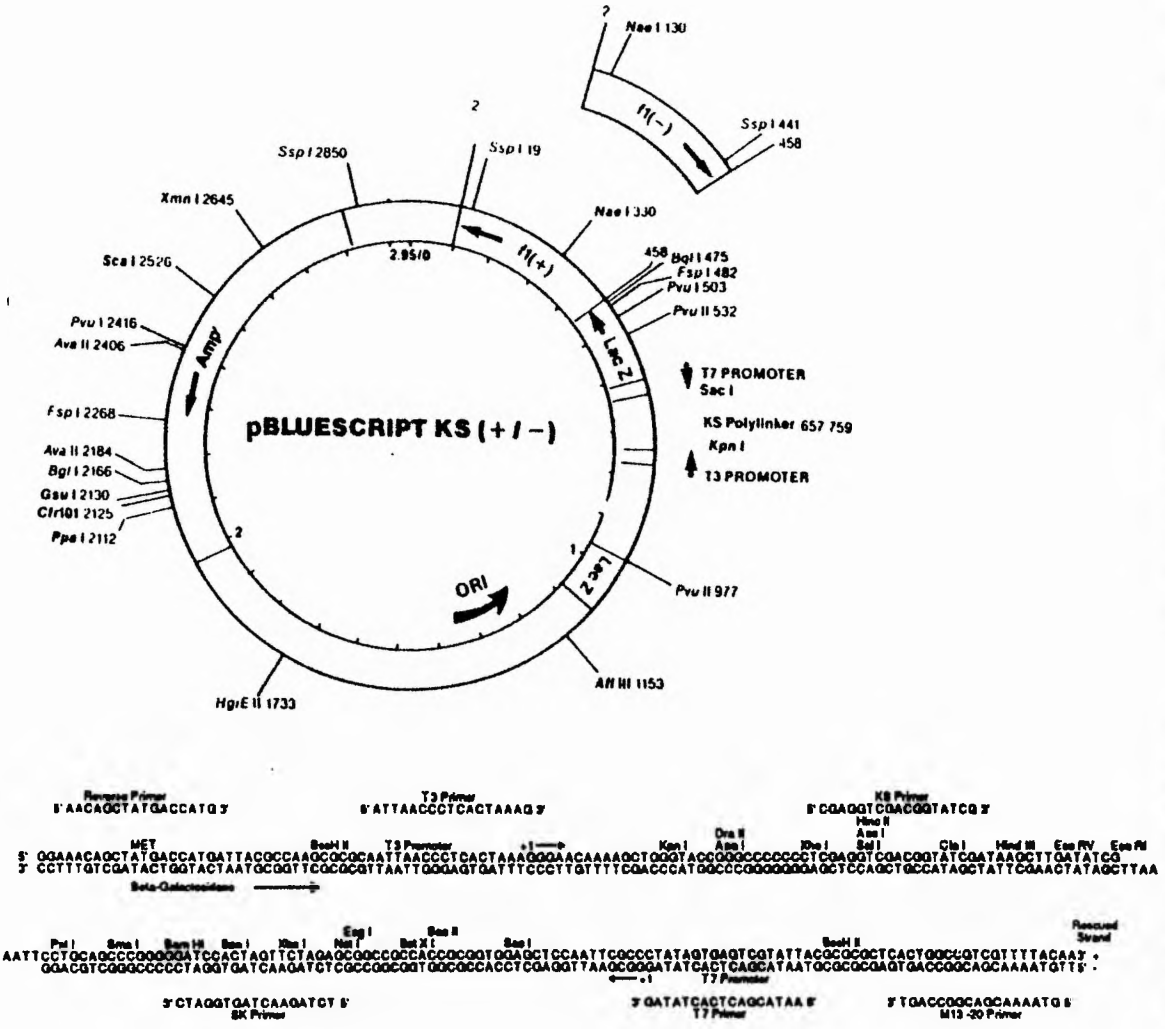


Figure 9 **pBluescript®** (Stratagene). This plasmid is a 2.96 kb colony-producing vector which confers resistance to ampicillin via the expression of the β-lactamase gene. An inducible *lac* promoter upstream from the *lacZ* gene allows the expression of fusion proteins and blue/white selection of recombinant clones. Its 21 unique restriction sites are used for restriction mapping and subcloning. T3 and T7 bacteriophage promoters allow the *in vitro* synthesis of strand-specific RNA. T3 and T7 primers are also used for double stranded sequencing. The SK and KS series represent two different orientations of the polylinker.

and *in vitro* synthesis of RNA transcripts using the T3 and T7 bacteriophage promoters.

2.14 Excision of pBluescript from λZAP

200 µl of OD_{600}=1.0 XL1-Blue cells (described in section 2.7, containing approximately 2 x 10^7 cells) were combined with 200 µl of λZAP phage stock derived from a purified plaque, containing > 1 x 10^5 phage particles, and 1 µl of R408 filamentous helper phage (> 1 x 10^3 pfu/µl). These were incubated at 37 °C for 15 min, followed by the addition of 5 ml of LB and further incubation for at least 3 h at 37 °C with vigorous shaking. Next, the tubes were heated to 70 °C for 20 min and spun at 3,000 g for 5 min. The supernatant was decanted into a sterile tube: this was the "phagemid" stock which could be stored for 1-2 months at 4 °C.

To improve the excision process, the following procedures were used: an alternative helper phage was preferred, EXASSIST™, which has been genetically modified so as to be unable to replicate. In using R408, problems arose when the titres of λZAP and R408 bacteriophage were not balanced properly, and cell lysis ensued. Hay and Short (1992) recommend the following ratios: XL1-Blue cells (10): λZAP (10): helper phage (1). The phagemids were then used to infect XL1-Blue cells by incubating 200 µl of the same OD_{600}=1.0 XL1-Blue cells with 1-200 µl of phagemid stock. A further modification to the procedure described in the λZAP manual was to add 1 ml of LB to these cells after phagemid infection, leaving them at 37 °C to allow the ampicillin resistance gene to be expressed. Finally, cells were plated out on agar plates containing 50 µg/ml ampicillin, and incubated overnight at 37 °C. Ampicillin resistant colonies were used in subsequent plasmid DNA minipreps to sequence the cloned cDNAs.

2.15 Preparation of plasmid DNA

This method was adapted from Jones and Schofield (1991). Single bacterial colonies were tooth-picked to inoculate 5 ml of LB containing 50 µg/ml ampicillin. Cultures were grown overnight at 37°C with vigorous shaking. 1.6 ml of the culture was spun in a microcentrifuge. To increase the size of the bacterial pellet, a further 1.6 ml of the overnight culture was added, spinning again. The supernatant was removed by tipping and draining. The pellets were resuspended in 180 µl of GTE solution (see below), after which 360 µl of a fresh solution of 0.2 M NaOH/ 1% SDS was added, mixing by inversion several times. After placing the lysed cells on ice for 5 min, 270 µl of 3 M potassium acetate (pH 4.8) was added, mixing by inversion and chilling on ice for a further 5 min. Samples were spun for 10 min in a microcentrifuge. The supernatant was collected carefully, wiping the outside of the pipette with a tissue. The supernatant was spun again, to remove any remaining precipitate. A phenol/chloroform extraction , which is not suggested in Jones and Schofield (1991), was included at this point to clean the "miniprep" further. One volume of ice-cold ethanol was added to precipitate the plasmid DNA, vortexing briefly and spinning in a microcentrifuge for 10 min. The pellets were washed with 1 ml of 70% ethanol, drained, dried in a vacuum pump, and resuspended in an appropriate volume, usually 20 µl, of sterile dH_20.

GTE solution

25 mM Tris-HCl, pH 8.0	25 mM
EDTA	10 mM
glucose	50 mM

2.16 Restriction of DNA and agarose gel electrophoresis

Samples of DNA were digested for 1-2 h in small volumes (10-20 µl) containing the required restriction buffer and restriction enzyme. If the removal of RNA was required, such as in the visualization of small DNA fragments, RNase was added (to a final concentration of 50 µg/ml ribonuclease A and 50 units/ml of ribonuclease T_1) to the digests in the final 20 min of their restriction. Finally, one-sixth the volume of gel loading buffer was added before proceeding to electrophoresis. The concentration of agarose in the gels varied between 0.8 and 2%, depending on the size of DNA fragments being analyzed: 0.8% agarose favoured the separation of larger fragments (> 2 kb) whereas 2% agarose favoured the separation of smaller fragments (500 bp and less). The agarose was dissolved in 30 ml (for a mini-gel) of TAE buffer by melting, and placed in a water bath at 45°C for at least 30 min to equilibrate the temperature. Next, ethidium bromide was added to a final concentration of 5 µg/ml before setting the gel. The gel was covered in TAE buffer and the restriction digests loaded (up to 15 µl per well). Samples were run at 40-100 V and the DNA bands were visualized using a 300 nm UV-light transluminator.

Tris-Acetate buffer (TAE)

Tris-acetate	0.04 M
EDTA	0.001 M

Concentrated stock solution (50 x) per litre:

Tris base	242 g
Glacial acetic acid	57.1 ml
0.5 M EDTA (pH 8.0)	100 ml

6 X Gel loading buffer for DNA (the dyes act as markers)

Bromophenol blue	0.25%
Xylene cyanol	0.25%
Glycerol	30% in dH_20

2.17 "Gene cleaning"

This procedure was generally used to purify DNA fragments excised from an agarose gel, but could, in addition, be used to desalt DNA solutions, to remove unincorporated nucleotides and to concentrate samples. The advantage of this procedure was that it gave an efficient recovery of DNA (approximately 80%, as suggested in the kit), while removing ethidium bromide. The GENECLEAN II Kit (BIO 101 Inc.) contains a silica matrix called glassmilk™ that binds to single and double stranded DNA. The first step was to excise the DNA band from the agarose gel with a clean razor blade, visualizing the bands on an ultra-violet light transluminator (wavelength 300 nm, using the lower intensity setting so as to minimise damage to DNA). To begin the extraction, three volumes of NaI solution were added to the gel. If the gel was run in TBE, 4.5 volumes of NaI and 0.5 volumes of "TBE modifier" solution were added instead. After incubating for 5 min at 45-55 °C to dissolve the agarose, the "glassmilk" suspension was added. In general, 5 µl of the suspension was added to samples containing 5 µg or less DNA, and an extra 1 µl for every 0.5 µg of DNA above 5 µg. The samples were incubated for 5 min on ice, vortexing occasionally to keep the silica matrix in suspension, and then the DNA/glassmilk complex was pelleted in the microcentrifuge for 5 sec and the supernatant removed. The pellet was washed three times in "NEW" solution, each time resuspending the glassmilk in a 0.5 ml volume of NEW solution, ensuring that all the wash solution was removed after the final wash. The DNA was eluted into the desired volume of dH_20 by resuspending the silica, heating it to 45-55 °C for 2-3 min, spinning it for 30 sec and collecting the supernatant. If 10 µl of glassmilk was used, the DNA could be eluted into a minimum of 10 µl dH_20. The exact compositions of

the NaI stock solution, TBE modifier and NEW solution were not specified in the kit.

2.18 Ligation of DNA fragments

The appropriate ratio, usually 5:1, of linearized vector to insert DNA, was mixed in a total reaction volume of 20 µl. The reaction included 2 µl of 10 X One-Phor-All PLUS buffer (100 mM Tris-acetate, 100 mM magnesium acetate, 500 mM potassium acetate, provided by Pharmacia), 0.5 to 5 Weiss units of T4 DNA ligase, and 1 mM ATP. The ligation mix was incubated overnight at 16 °C and stopped by heating it to 65 °C for 10 min, ready for transformation.

2.19 Transformation by the $CaCl_2$ method

"Competent" cells (XL1-Blue being the strain of choice, described in section 2.7) were prepared as follows (as described in Maniatis *et al.*, 1982): 5 ml LB cultures were grown overnight, the next morning inoculating a flask containing 40 ml of sterilized LB with 1 ml of the overnight culture. The culture was grown at 37 °C with vigorous shaking until an OD_{600} of 0.2 was reached (the cells being in the logarithmic phase of growth). The culture was chilled on ice for 10 min and centrifuged at 3,000 g for 5 min at 4°C. The supernatant was removed, resuspending the cells into half the original volume of an ice-cold, sterile solution of 50 mM $CaCl_2$ and 10 mM Tris.HCl, pH 8.0. The suspension was chilled on ice for 15 min and centrifuged as above. The supernatant was removed again and the cells resuspended in 1/15 the original volume of the above solution. Competent cells were stored at 4 °C, gaining maximum transformation efficiency after 12-24 h.

200 µl aliquots of the above cells were used for each transformation, using 5-10 µl of ligation mixture per transformation. Ideally, not more than 40 ng of DNA were used in each transformation, because the use of excess DNA lowers the transformation efficiency (Maniatis *et al.*, 1982). Samples were placed on ice for 30 min, and then transferred to a water bath at 42 °C for 2-3 min. Next, 1.0 ml of LB was added, and the cells incubated at 37 °C for a further 1-2 h, allowing them to express the antibiotic resistance gene. 10-200 µl of these samples were spread onto agar plates containing selective medium, usually 50 µg/ml ampicillin, and grown overnight at 37 °C. The success of this procedure varied from 1 to 10^3 transformants per plate.

2.20 Storage of cell strains and plasmid DNA

The plasmid DNA was stored in sealed microcentrifuge tubes at 4 °C for short term storage, and frozen for longer term storage. Various bacterial strains were maintained at 4 °C as colonies on an agar plate containing 50 µg/ml ampicillin, 12.5 µg/ml tetracycline or no antibiotic, as required. Stocks of bacterial strains were kept at -70 °C in 15% glycerol for several years.

2.21 DNA sequencing

Dideoxy sequencing was performed using the T7 sequencing™ kit (Pharmacia). Double-stranded templates were obtained from plasmid DNA mini-preps (section 2.15).

(i) **Annealing the primer.** 2 µl of 2M NaOH was combined with 8 µl containing 1.5-2 µg of template DNA, mixed and left at room temperature for 10 min. Next, 3 µl of 3 M NaAcetate (pH 4.5) and 7 µl of dH_20 were added, followed by 60 µl of 100% ethanol, mixed and cooled at -20 °C for 15 min. The denatured DNA was pelleted in a microcentrifuge for 10 min

and washed with 70% ethanol. The pellet was dried under vacuum and redissolved in 10 µl of dH_20. 2 µl of annealing buffer from the kit (a buffered solution containing $MgCl_2$ and DTT) and 2 µl of primer solution, diluted to give the appropriate primer:template ratio (recommended molar ratio between 5:1 and 50:1), were added to the DNA, and the sample incubated at 37 °C for 20 min to anneal the primer.

(ii) **Sequencing reactions**. T7 DNA polymerase was diluted to 1.5 units/µl and kept on ice, 2 µl of this dilution being required for each reaction. Four microcentrifuge tubes for each reaction were labelled A, C, G and T. 2.5 µl of the 'A', 'C', 'G', and 'T' mix-Short solutions were pipetted into each tube. Into the tube containing the annealed template, 3 µl of the 'labelling mix' was combined with 1 µl of [α-^{35}S]dATP (containing approximately 10 µCi) and 2 µl of diluted T7 DNA polymerase. The reaction was incubated at room temperature for 5 min while the four sequencing mixes were pre-warmed to 37 °C.

(iii) **Termination reactions**. 4.5 µl of the sequencing reaction which had been incubating for 5 min was transferred to each of the pre-warmed sequencing mixes, mixed and and incubated at 37 °C for 5 min. Then 5 µl of stop solution was added to each tube, mixed, and microcentrifuged briefly. The samples were stored at -20 °C for subsequent use. To load the sequencing gel, aliquots of 3 µl from each sample were denatured at 96 °C prior to loading. Samples were loaded in the order A, C, G and T on the sequencing gels, each loading yielding up to 200 nucleotides of sequence. Obtaining shorter and longer runs of each reaction increased the amount of information obtained.

(iv) **Sequencing gel**. To prepare 8% polyacrylamide wedge gels, the glass plates were cleaned with soap and water, rinsed with ethanol and dried. Both plates were treated with Sigmacote™ only on the side that was to be in contact with the gel. The gel was assembled using tape and wedge

spacers. 15 ml of 40% acrylamide solution was mixed with 10 ml of 10 x TBE buffer, 42 g of urea solution, dH_20 to 100 ml and de-gassed for 5 min. To initiate polymerisation, 250 µl of 10% ammonium persulphate and 50 µl of TEMED were added to the mixture. The gel was poured immediately, the plates placed at an angle, avoiding air bubbles. The plates were placed horizontally for 45-60 min, after which the combs and bottom spacer were removed, rinsing the upper surface with distilled water to remove unpolymerised acrylamide. The gel was placed into the electrophoresis apparatus to be run in 1 x TBE buffer. The gel was pre-run at 1200-1500 V for 30-60 min. To load the samples (denatured at 96 °C prior to loading), the power supply was switched off and the shark's-tooth comb inserted so that the points just touched the surface of the gel. Samples were applied and the gel run at 1200-1500 V. Electrophoresis was stopped when the bromophenol blue in the last loaded samples reached the bottom of the gel. The plates were disassembled and the gel fixed for 20 min in 10% acetic acid/10% methanol. The gel was transferred to a supporting sheet of Whatman 3MM paper, covered with plastic wrap, dried using a vacuum gel dryer and set up for autoradiography in contact with Agfa X-ray film in a cassette with an intensifying screen at -70 °C for 18-42 h.

Tris-Borate buffer (TBE)

Tris-borate	0.089 M
boric acid	0.089 M
EDTA	0.002 M

Concentrated stock solution (5 x) per litre:

Tris base	54 g
Boric acid	27.5 g
0.5 M EDTA, pH 8.0	20 ml

2.22 pMAL fusion protein system

The pMAL vectors (© New England BioLabs 1991) were used to express fusion proteins in *E. coli*. Cloned genes were inserted downstream from the *mal*E gene which encodes the 42 kDa maltose-binding protein (MBP). The vectors also carry an ampicillin gene for selection and the *lac*I^q gene encoding the Lac repressor. In the absence of IPTG, expression of the construct is low. The polylinker site contains a sequence coding for the recognition site of the protease factor Xa; the system allows purification of the fusion protein using an amylose resin which binds to MBP, followed by cleavage by factor Xa to separate the desired protein from the MBP. The vector of choice was pMalcRI, whose polylinker contains an EcoRI site which is in frame with the EcoRI site present at the 5' ends of cDNAs cloned into lambda bacteriophage (λgt11) expression libraries. Using the previously described methods, cDNAs of interest were ligated into appropriately linearized vector and transformed into *E. coli* host TB1 (ara Δ (*lac proAB*) *rpsL* (ϕ80 *lacZΔM15*) *hsdR*), and grown overnight on agar plates containing 100 μg/ml ampicillin.

To express the fusion proteins, 5 ml LB containing 100 μg/ml ampicillin was inoculated with a single ampicillin resistant colony, and grown overnight at 37 °C with good aeration. The next day, a fresh aliquot of 5 ml LB was inoculated with 1/10 volume of the overnight culture and grown at 37 °C with good aeration to an OD_{600} ~ 0.5 (2 x 10^8 TB1 cells/ml). IPTG was added to a final concentration of 0.3 mM and the cultures grown for a further 2-3 h (induction period). 1.6 ml of induced cells were spun in a microcentrifuge, resuspended in 300 μl of TBS buffer and sonicated to lyse the cells and to reduce viscosity by shearing DNA. This crude lysate could be analyzed on SDS-PAGE by adding an equal volume of SDS-PAGE sample buffer (section 2.23).

2.23 SDS-PAGE analysis of proteins

To prepare a 12 % polyacrylamide gel the following solutions were used:

	Separating gel	Stacking gel
30% acrylamide	16 ml	2 ml
1 M Tris.$S0_4$, pH 8.3	15 ml	–
0.5 M Tris.$S0_4$, pH 6.9	–	1.25 ml
dH_20	3.3 ml	6.57 ml
glycerol	5.7 ml	–
	de-gas for ten minutes	
10% SDS	0.6 ml	0.15 ml
25% AMPS	60 µl	30 µl
TEMED	60 µl	30 µl
Total volume	40 ml	10 ml

To make thicker and thinner gels, the amount of 30% acrylamide was varied accordingly, to give a final concentration of 9 % acrylamide (for a better separation of higher molecular weight proteins, above 100 kDa) and 20 % acrylamide (for a better separation of lower molecular weight proteins, below 40 kDa). Before setting, dH_20 was layered over the separating gel. The reservoir buffer contained 3 g Tris base, 14.3 g glycine and 1.5 g SDS per litre. An equal volume of sample buffer was added to the protein samples and the samples boiled for 2-3 min. If required, samples were saturated with urea to aid solubility. Molecular weight markers were obtained from Sigma and included: ("**high molecular weight**" markers: bovine erythrocyte carbonic anhydrase, 29 kDa; egg albumin, 45 kDa; bovine plasma albumin, 66 kDa; rabbit muscle phosphorylase, 97.4 kDa; *E. coli* β-galactosidase subunit, 116 kDa; rabbit muscle myosin, 205 kDa, and the "**low molecular weight**" markers: bovine milk α-lactalbumin; 14.2 kDa; soybean trypsin inhibitor, 20.1 kDa; bovine

pancreas trypsinogen, 24 kDa; bovine erythrocyte carbonic anhydrase, 29 kDa; rabbit muscle glyceraldehyde-3-phosphate dehydrogenase subunit, 36 kDa; egg albumin, 45 kDa and bovine plasma albumin, 66 kDa.) Samples were loaded onto the SDS-PAGE gel and run overnight at 40-80 V.

Protein sample buffer

SDS	2% (w/v)
2-mercaptoethanol	5% (v/v)
glycerol	10% (v/v)
Tris/SO_4, pH 6.9	10 mM

+ bromophenol blue marker as required

2.24 Coomassie staining of proteins

Proteins separated on an SDS-PAGE gel were stained using Coomassie brilliant blue. The staining solution contained:

Coomassie brilliant blue	1 g
96% ethanol	250 ml
Glacial acetic acid	50 ml
dH_2O	to 500 ml

The dye was dissolved by stirring for at least 30 min. Gels were immersed into the staining solution in a covered polypropylene box and left for 2 h with gentle agitation. The staining solution was decanted into a bottle, to be re-used once. Gels were de-stained using the solution:

96% ethanol	250 ml
Glacial acetic acid	100 ml
dH_2O	to 1000 ml

Gels were firstly rinsed with water to remove excess staining solution, then adding de-staining solution and pieces of plastic foam which have the ability to adsorb excess dye. Gels were de-stained with gentle agitation for

1 h, after which fresh de-stain was added and the gels rinsed further until adequate de-staining was obtained.

2.25 Western blotting

Mixtures of proteins were separated via SDS-PAGE and then electrophoretically transferred to nitrocellulose membrane. These "blots" could be used for ELISA ("immunoblotting", section 2.27) and "NorthWestern blots" or "riboblots" (binding of RNA probes to transferred proteins, section 2.42 part (i)). The procedure began by cutting a nitrocellulose membrane corresponding to the size of the SDS-PAGE gel being transferred and two pieces of 3 mm Whatman filter paper of the same dimensions. The polyacrylamide gel was washed three times for 5 min in transfer buffer. The gel was placed onto on a wet filter paper, and then the membrane placed over the gel. This was done slowly and with care, because even a brief contact with the gel would result in the transfer of significant amounts of protein. Bubbles trapped betweens the membrane and the gel were removed by rolling a glass pipette over the membrane. The second wet filter paper was placed over the nitrocellulose, also removing any air bubbles with the glass pipette. The transfer was placed between two pads in the holder, fitted into the transfer apparatus and covered with transfer buffer. A current of 0.3 Amps was applied for at least 4 h to overnight, the SDS-gel being on the side of the negative electrode. After the transfer, the membrane was removed carefully and placed into an appropriate "blocking" solution: for susbequent immunoblots, this was TBS-milk (TBS containing 2 % (w/v) filtered skimmed milk), and for riboblots TBST (containing 0.5 % (v/v) Tween-20 as a blocking agent).

Transfer Buffer

Tris.HCl, pH 8.3	25 mM
glycine	150 mM

Tris-Buffered Saline (TBS)

Tris.HCl, pH 7.5	10 mM
NaCl	150 mM

2.26 Amido black staining of proteins on nitrocellulose

Proteins transferred to nitrocellulose membrane could be stained on the membrane, using the method described in Wojtkowiak *et al.* (1983) which included the staining solution:

Amido black	1.5 g
Methanol	45 ml
Glacial acetic acid	10 ml
dH_2O	to 100 ml

The membrane was immersed in the staining solution for 1 h with gentle agitation. The staining solution was decanted carefully into a bottle, to be re-used once. The membrane was de-stained with gentle agitation, changing the washes as appropriate, in:

Methanol	180 ml
Glacial acetic acid	4 ml
dH_2O	to 200 ml

2.27 Immunoblotting using the ELISA technique

The ELISA (enzyme-linked immunosorbent assay) technique makes use of an enzyme covalently linked to an immunoglobulin. Antigens were firstly bound to "primary" antibody, which in this study were present in antisera obtained from rabbits. Primary antibodies were subsequently bound by goat anti-rabbit IgG "secondary" antibodies, conjugated to the enzyme

horseradish peroxidase. The antigen was finally detected using the horseradish peroxidase in a colour reaction involving the substrate DAB (diaminobenzidine). The same procedure was used to immunoblot both Western transfers of SDS-PAGE gels and plaque lifts. When blotting plaque lifts, the antisera were pre-incubated for 30 min with a 1/100 dilution of bacterial lysate, a step which reduced background signals significantly.

To bind the primary antibody, the antiserum was firstly diluted as required in TBST (TBS containing 0.05% Tween-20). The optimum dilution varied between batches of antisera and was determined empirically; however, in general, 1/2000 dilutions of antisera were adequate. The primary antibodies diluted in TBST were added after removing the blocking solution from the filters. The filters were agitated gently at room temperature for 1.5 h, then washed four times for 5 min with TBST, using twice the volume used in antibody incubation. At first, bound antibody was detected by incubating the filters with ^{125}I-labelled protein A (> 30 mCi/mg, Amersham) in TBS containing 0.5% NP-40. However, the horseradish-peroxidase conjugate system was subsequently preferred, because it gave a sharper signal and eliminated the need for autoradiography.

The secondary antibody was diluted 1/3000 in TBST and the filters incubated with it for 1.5 h, after which they were washed three times for 5 min in TBST, and the fourth wash was for 5 min in TBS without Tween-20. Sufficient aliquots of frozen DAB (0.4 ml containing 20 mg of DAB for every 50 ml of TBS to be used) were thawed, vortexing the DAB to solubilize it. The DAB and 15 µl of 30% hydrogen peroxide were mixed into 50 ml of TBS and immediately poured over the filters with the proteins facing up. Positive signals were allowed to develop (in general, this

happened between 20 sec and 5 min after substrate addition), and when the optimum signal was achieved, the reaction was stopped by washing the filters with ample distilled water. Because DAB is light-sensitive, the stained filters were stored in the dark at 4 °C.

2.28 Cleavage of proteins. (reviewed in Darbre, 1986).

(i) **Cleavage at DP** (aspartic acid-proline) sites, adapted from Nute and Mahoney (1979). Proteins were dissolved in 70 % formic acid, if necessary with the aid of 7 M guanidine.HCl, and incubated at 40 °C for 24 h. The reaction was stopped by adding an equal volume of cold, de-ionized water. The procedure was modified as follows: instead of 7 M guanidine.HCl, 0.5% SDS was added to solubilize the proteins. The samples were dried in a vacuum pump to remove the formic acid, and resuspended in an appropriate buffer.

(ii) **Cleavage at NG** (asparagine-glycine) sites, adapted from Enfield (1980). Proteins were dissolved in 6 M guanidine.HCl/2% hydroxylamine.HCl at room temperature. Concentrated NaOH was added as required to bring the reaction to pH 9. The reaction proceeded at room temperature for 4 h, while the pH was checked and maintained at pH 9. Samples were dried in a vacuum to remove the reagents, and then resuspended in an appropriate buffer.

(iii) **Cleavage at M** (methionine). The proteins were dissolved in 70% formic acid, adding solid CNBr in the fume hood. The ratio was 2 µg CNBr to 1 µg protein, so that CNBr was in at least a 20-100 fold molar excess with respect to the methionines present in the proteins. The reaction proceeded for 18-20 h at room temperature in the dark. The samples were dried in a vacuum to remove the reagent, and then resuspended in an appropriate

buffer. One point to note with respect to this reaction is that it was carried out in 70% formic acid, conditions in which DP sites were also cleaved (see above).

(iv) **V8 Protease digestion**. Samples were digested with V8 protease from *Staphylococcus aureus* (Boehringer) at 20 °C for at least 1 h in HTB or a similar buffer, and supplemented with 0.1% SDS.

2.29 Iodination of tyrosines

100 µl samples of Iodogen™ stock (Pierce Chemical Company) were dried onto the surface of a glass vial. The prepared protein samples (10-50 µg in 100 µl containing 0.2% SDS and 10 mM Tris.HCl, pH7.5) were added to the vial along with 1 µl of NaI (^{125}I, 2 mCi/20 µl) to be incubated at room temperature for 15 min with occasional swirling. Next, the following were added: 10 µl of 0.1 M DTT, a reducing agent, and 100 µl of KI which stopped the reaction by providing an excess of iodide. To store the labelled protein, 500 µl of 96% ethanol was added to each microcentrifuge tube and stored at -20 °C. Before use, the tubes were spun at 10,000 g for 10 min to pellet the protein, rinsing once with 80% ethanol + 20% KI.

2.30 Phospholabelling

Poly(A)$^+$ mRNP, germinal vesicle extracts, and fractions eluted from a heparin column were adjusted to 20 mM Tris.HCl, pH 7.5 and 2 mM $MgCl_2$, a buffer in which casein kinase II is active. 1-2 µl of [γ-^{32}P]ATP (Amersham International, 3000 Ci/mmol), containing approximately 1 µCi, was added to each sample. For riboprobe-binding experiments, poly(A)$^+$ mRNP were phosphorylated *in vitro* by adding cold ATP (10 µM), or alternatively de-phosphorylated by adding 1-2 µl of calf intestinal

phosphatase (Boehringer Mannheim, 1 unit/µl), active in "1-for-all" buffer (also Boehringer Mannheim). These reactions proceeded for 30 min at room temperature.

2.31 ATP binding

To test for ATP binding, mRNP proteins were photocrosslinked to [α-^{32}P]ATP. It was necessary to use [α-^{32}P]ATP because any hydrolysis of [γ-^{32}P]ATP would lead to loss of label. Following the succesful photocrosslinking of [α-^{32}P]ATP to the RNA helicase eIF4A (Pause *et al.*, 1993), poly(A)$^+$ mRNP was adjusted to 20 mM Tris.HCl, pH 7.5, and 2 mM $MgCl_2$. Next, 2 µl of [α-^{32}P]ATP containing approximately 1 µCi (Amersham International, 3000 Ci/mmol) was added to 100 µl poly(A)$^+$ mRNP samples, photocrosslinked on ice for 20 min, and finally digested with RNase and analysed on SDS-PAGE (see section 2.42 part (iv) for the photocrosslinking procedure).

2.32 Peptide sequencing

The HTP (heat-treatment pellet) fraction of poly(A)$^+$ mRNP was digested with CNBr to generate smaller peptides suitable for sequencing. Firstly, approximately 1 ml of poly(A)$^+$ mRNP eluate from an oligo(dT) column was adjusted to HTB and heat-treated (section 2.2). The pellet was raised in 70% formic acid and digested in CNBr to cleave at methionines (section 2.28). The digested material was sent to Dr. Kemp at the Biochemistry Department, University of St. Andrews, where the peptides were separated on an SDS-PAGE gel and transferred to PVDF membrane. The peptides were sequenced using a gas phase automated sequencer.

2.33 GCG Package computer software and electronic mail

The computer-based sequence analysis was carried out using the Wisconsin Genetics Computer Group GCG™ Package, Genetics Computer Group Inc., version 7.2, 1992. The programs used were: FASTA, used to conduct homology searches against EMBL DNA and SwissProt protein databases; TRANSLATE, which translates open reading frames into an amino-acid sequence; MAP, which translates in any specified frame(s) and finds endonuclease restriction sites: MAP also analyses amino-acid sequences and maps protease and/or chemical cleavage sites; PEPTIDESORT, which sorts peptide fragments including the whole protein according to size, calculating molecular mass, percentage amino-acid composition and estimated isolelectric point; PEPTIDESTRUCTURE and PLOTSTRUCTURE which use secondary structure prediction algorithms to predict protein secondary structure; HELICALWHEEL, which predicts the arrangement of amino-acids in an α-helix; PILEUP, which performs multiple sequence alignments. In addition, the automatic electronic mail server both at EMBL (European Molecular Biology Laboratory), Heidelberg, and EBI (European Bioinformatics Institute), Cambridge, were used to conduct nucleotide and amino-acid sequence searches (FASTA and BLITZ algorithms) on updated databases, and individual database entries could be retrieved by electronic mail. New cDNA sequences were submitted to the databases via electronic mail.

2.34 Synthesis of riboprobes

A "riboprobe" is a radiolabelled RNA molecule to be used as a molecular probe in RNA-binding assays (for example, the sense strand of cyclin B1 and a 3' end fragment of β-tubulin mRNA , see Appendices A and B).

(i) **Run-off transcripts**. Various riboprobes were prepared by synthesizing run-off transcripts from appropriately linearized cDNAs cloned into pBlueScript in the presence of [α-^{32}P]CTP. In general, unless stated otherwise, the riboprobe of choice for binding reactions was a 0.35 kb fragment of the 3' end of the sense strand of β-tubulin. The label was obtained from Amersham International at 400 Ci/mmol. Approximately 1-2 μg DNA was digested in a total volume of 7 μl using the appropriate restriction enzyme at 37 °C for 1-2 h. Next, the following were combined:

Linearised DNA digest	7 μl
5 x salts	4 μl
0.1 M DTT	2 μl
RNase inhibitor	1 μl
10 x NTP mix	2 μl
[α-^{32}P]CTP	2 μl
T7 RNA polymerase	1 μl
DEPC H_20	1 μl
Total:	20 μl

The 10 x NTP mix contained: (ATP, 10 mM; GTP, 10 mM; UTP, 10 mM; CTP, 0.5 mM). In a typical reaction, there was a 10:1 ratio of 1 nmol "cold" CTP to 0.1 nmol radiolabelled CTP. The reaction was incubated at 37 °C for 1 h, after which 1 unit of RNase-free DNase I was added and incubated for a further 10 min at 37 °C. 30 μl DEPC H_20 was added before proceeding to a spun column. Riboprobes were stored at -70 °C for subsequent use.

(ii) **5' end-labelling of RNA polymers**. The first step was to dephosphorylate the 5' ends of the polymers as follows:
20 μg of polymer (poly(C,U) and poly(A,G)) was treated with 1-2 units of CIP (calf intestinal phosphatase) in "1-for-all buffer" (Pharmacia), incubating at 45 °C for 10 min. The reaction was stopped by adding 3 μl of 0.5 M EDTA, heating to 65 °C for 10 min, extracting twice with

phenol/chloroform. 10 µl of 3 M NaCl and 320 µl of ethanol was added to 100 µl of acqueous phase, placed at -70 °C for 30 min and spun for 10 min in a microcentrifuge. The pellet was rinsed with 70% ethanol, dried under vacuum and raised in 20 µl of DEPC H_20. Next, to end-label the polymers, the following were combined and incubated at 37 °C for 40 min:

DEPC H_20	8.5 µl
RNase inhibitor	1 µl
10 x 1-for-all buffer	2 µl
Dephosphorylated polymer, 1 µg/µl	5 µl
[γ-^{32}P]ATP (~ 1 µCi)	2.5 µl
Poly-nucleotide kinase	2 µl
Total:	20 µl

To stop the reaction, 1 µl of 0.5 M EDTA, 75 µl dH_20, 5 µl of 5 M NaCl were added before proceeding to a spun column, using Sephadex G25 rather than G50. If required, the riboprobes were extracted with phenol/chloroform and re-precipitated as above. End-labelled polymers were also stored at -70 °C for subsequent use.

2.35 Spun column

Radiolabelled RNA was recovered via a "spun column" using Sephadex G-50 resin. This resin was resuspended in TE buffer and autoclaved. The column was prepared in a disposable syringe as described by Maniatis *et al.* (1982) by plugging the syringe with some flock previously boiled in TE buffer, then filling it with the Sephadex G-50 suspension. The syringe was compacted to dryness by spinning it at 3,000 g for 4 min. 50 µl of TE buffer was added and spun through similarly in order to verify that the volume recovered was equal to the volume loaded. The riboprobe labelling mix was added and spun, unincorporated nucleotides remaining in the column while the labeled probe was collected in a 1.5 ml screw-cap eppendorf

tube. The syringes and the collecting eppendorf tubes were placed in an appropriate carrier centrifuge tube. The riboprobe was aliquoted into small volumes (usually 10 µl) and stored at -70 °C.

TE Buffer	
Tris.HCl, pH 8.0	20 mM
EDTA	1 mM

2.36 RNA extraction

(i) **Extraction from oocytes.** Oocytes were collected as described in section 2.0. Pools of 5-15 oocytes (for vitellogenic stages) and 50 (for smaller previtellogenic oocytes) were placed in separate 1.5 ml microcentrifuge tubes. 500 µl of TNES buffer + 200 µg/ml proteinase K was added to each tube, mixing thoroughly by vortexing, and incubating at 50 °C for 60 min with further occasional vortexing. Samples were extracted twice with phenol:chloroform and once in chloroform. The RNA was precipitated by adding 2.5 volumes of ethanol and spinning at 10,000 g for 30 min at 0 °C. The pellet was washed in 70% ethanol, resuspended in 20 µl of DEPC dH_20 and stored at -70 °C. To extract RNA from embryos, a similar procedure was followed; however, the embryos were firstly placed for a short time in a solution of 2 % cysteine hydrochloride (pH 8.1) to remove the jelly coating and rinsed four times in one third strength Modified Barth's Solution with gentle swirling.

(ii) **Extraction from tissues**. An immature male *Xenopus* was dissected to obtain a whole brain, the two testes, part of the liver, heart and kidneys. The tissues were chopped and vortexed into 500 µl of a solution containing guanidinium thiocyanate and 1% 2-mercaptoethanol. Samples were extracted twice at 55 °C with an equal volume of phenol:chloroform. 10 µl of 3 M NaAcetate and 2.5 volumes of ethanol were added before placing

the samples at -20 °C overnight. The next day pellets were collected after spinning at 10,000 g for 30 min at 0 °C, raised in 200 µl of TNES + 200 µg/ml proteinase K and incubated at 45 °C for 2 h with occasional vortexing. Samples were extracted again in phenol:chloroform and then 2.5 volumes of ethanol were added before placing at -20 °C overnight. On the third day, the RNA was precipitated as above, and raised in 100 µl of DEPC-dH_20 plus 100 µl of 8 M LiCl. After 4 h at -20 °C, RNA was precipitated again, rinsed in 96% ethanol and air-dried. Finally the pellets were suspended in 20 µl of DEPC H_20 and stored at -70 °C.

RNA extraction buffer (TNES)

Tris.HCl, pH 8.0	100 mM
NaCl	300 mM
EDTA	10 mM
SDS	2% (w/v)

Modified Barth's Solution (MBS)

NaCl	88 mM
KCl	2 mM
$CaCl_2$	0.41 mM
$Ca(NO_3)_2$	0.33 mM
$MgSO_4$	0.82 mM
$NaHCO_3$	2.4 mM
HEPES, pH7.4	10 mM

2.37 RNA gels and Northern transfer

Extracted RNA was raised in DEPC H_20. To every 10 µl of RNA solution, 4 µl of 5 x MOPS, 7 µl of formaldehyde and 20 µl of deionised formaldehyde were added giving a total volume of 40 µl. Immediately prior to loading onto the agarose gel, 4 µl of RNA loading buffer was added. The RNA gel was prepared by melting 0.9 g of agarose into 35 ml of DEPC H_20. Once the agarose had melted, 12 ml of 5 x MOPS was added and the gel mix

cooled to 50 °C. Nextly, 13 ml of pre-warmed formaldehyde was added, and the pH monitored, to ensure that it was above pH 4. The gel was then poured, covered in 1 x MOPS buffer, loaded with 20 µl of each RNA sample, and run at 20 V overnight.

The RNA was transferred to a nylon membrane by vacuum blotting. Firstly, the gel was washed in dH_20 and soaked in 50 mM NaOH for 5-10 min to hydrolyse the RNA partially. The transfer buffer consisted of 1 litre of 10 x SSC, and was run for three hours. After the transfer, the membrane was rinsed in 2 x SSC for 5 min, air-dried, and the RNA cross-linked to the nylon membrane by baking it at 80 °C for 1 h. As an alternative, the RNA samples could be transferred onto a membrane using a "slot-blot" apparatus.

5 X MOPS

MOPS	200 mM
NaAcetate	50 mM
EDTA	2 mM

10 X Gel loading buffer for RNA

Bromophenol blue	0.25%
Xylene cyanol	0.25%
EDTA	1 mM
Glycerol	50% in dH_20

20 X Standard Sodium Citrate (SSC)

To prepare one litre:

NaCl	175.3 g
NaCitrate	88.2 g

Dissolve in 800 ml. Adjust the pH to 7 with a few drops of 10 M NaOH and adjust the volume to 1 litre.

2.38 Hybridization of antisense probes to membrane-bound RNA

The membrane was placed into 50 ml of pre-hybridisation solution containing 1% Blotto, 1 x SSC and 2% SDS prepared with DEPC dH_20, and pre-hybridized at 65 °C for 2 h. Next, the membrane was covered in 25 ml of fresh pre-hybridization solution containing the appropriate antisense riboprobe. The riboprobe was hybridized with gentle mixing overnight at 65 °C to ensure high-stringency hybridization. The next day, the filter was washed twice for 20 min at 65 °C in 0.5 x SSC/0.1% SDS, and then similarly twice in 0.2 x SSC/0.1% SDS. The membrane was finally rinsed in 100 ml of the last wash solution and set up for autoradiography.

10 % Blotto, per 100 ml:

Non-fat powdered milk	10 g
Sodium azide	0.2%
sterile dH_20 to volume	

2.39 Release of mRNAs from polysomes

Polysomal mRNA from previtellogenic oocytes was prepared in order to construct a new cDNA expression library using the λZAP vector. 500 µl PB (polysome buffer) was added to approximately 50 previtellogenic oocytes which were homogenized by pipetting them in a sterile tip. The homogenized oocytes were mixed by vortexing and centrifuged at 12,000 g for 5 min at 4 °C. EDTA was added to the supernatant to a final concentration of 20 mM. Samples were diluted with PB to 2.5 ml and layered over a 2.5 ml cushion of 20% (w/v) of sucrose in PB. The polysomes were pelleted by ultracentrifugation at 149,000 g for 2 h at 4 °C in a Beckman SW55 rotor. Non-polysomal mRNP remained in the supernatant while the polysomal mRNPs were pelleted. The pellet was

resuspended in 500 µl of TNES before proceeding to RNA extraction (section 2.36).

Polysome buffer (PB)	
Tris.HCl, pH7.5	20 mM
$MgCl_2$	2 mM
KCl	300 mM
NP-40	0.5% (v/v)

2.40 Density gradient centrifugation

Gradients containing 6-48 % Cs_2SO_4 were prepared in a buffer containing 0.2% (v/v) of the non-ionic detergent NP-40 and 50 mM sodium phosphate, pH 7.0. Samples containing ~2.5 mg of RNA-protein complexes were loaded with or without prior fixation on the top of the gradients. The gradients were centrifuged at 40,000 rpm for 16 h at 18 °C in an SW65 rotor of the Beckman L-7 ultracentrifuge. 200 µl samples were collected by careful pipetting from the top of the gradients. Density points were measured using refractometry. RNA samples were extracted by firstly diluting the samples with three volumes of dH_20, as recommended by Cardinali *et al.* (1993), precipitating the RNP samples with two volumes of ethanol. RNA was extracted as described in section 2.36 with the addition of 10 µg carrier tRNA per sample. Proteins were obtained by diluting the gradient samples with a similar volume of dH_20, but then adding 10 µg of cytochrome C as a carrier. The reddish colour of cytochrome C facilitated the visualization of protein pellets after ethanol precipitation. The RNA extraction was performed by Robert Little, who also prepared slot blots for analysis with antisense riboprobes (sections 2.37 and 2.38).

2.41 Conditions for protein:RNA binding

Three methods were used to generate protein:RNA complexes, all involving a "destabilization" step followed by a "reconstitution" step.

(i) **Destabilization by $MgCl_2$.** This strategy followed the procedure described in Marello *et al.* (1992), in which mRNP particles were destabilized with 20 mM $MgCl_2$. Complexes were allowed to re-form by gradually diluting the samples with binding buffer (usually HTB) without $MgCl_2$, bringing the concentration of $MgCl_2$ down to 2 mM or less. Various competitors were tested by adding them shortly after the riboprobe.
(ii) **Destabilization by urea.** mRNP particles were also destabilized with the addition of 8 M urea. Riboprobe was added, along with various competitors. To reconstitute the complexes, the urea was diluted at least ten-fold or alternatively removed by dialysis.
(iii) **Heat-treatment.** This technique was based on the report by Deschamps *et al.*, (1991), in which the *Xenopus* oocyte YB proteins were shown to remain soluble after treatment at 80 °C (section 2.2). Samples were heated to 80 °C, cooled on ice and spun in a microcentrifuge each step for 5 min. The riboprobe was added prior to heat-treatment. Upon cooling, the protein:RNA complexes re-formed. Deschamps *et al.* (1991) found that after heat-treatment, the YB proteins present in mRNP in a whole oocyte extract associated to tRNA and rRNA molecules. Since the YB proteins do not normally associate with these molecules *in vivo*, a dynamic exchange of bound RNA was thought to have occurred during heat-treatment.

2.42 Protein:RNA binding assays

(i) **NorthWesterns, or "riboblots".** A "riboblot" is an assay in which proteins transferred to a nitrocellulose membrane (Western transfer) are bound to riboprobe. This assay is similar to the RNA-binding screen of bacteriophage plaques described in section 2.10. The first step was to wash the Western transfers in a denaturing solution containing 6 M urea dissolved in binding buffer (usually TMNT), twice for ten minutes. The purpose of this wash was to remove unwanted SDS. Any remaining urea was washed off with binding buffer. Next, a riboprobe was added. Typically, this was 1-2 µl of freshly made riboprobe, (ca. 0.25 µCi / 0.1 µg), in a small volume of binding buffer (10-20 ml, depending on the size of the membrane). The membrane was incubated with riboprobe for 30 min after which the riboprobe-containing solution was washed off. Then the membrane was washed in the same binding buffer containing 500 mM NaCl, in order to diminish background binding. The above was the basic procedure; however, ionic conditions (both in the binding and washing buffer) could be modified, and competitors added to the binding reaction as required. Finally, riboblots were set up for autoradiography. Riboblots could be subsequently immunoblotted, the aim being to align RNA-binding with immunoreactive bands.

(ii) **Agarose gel retardation**. RNP complexes were generated in a small volume, usually not more than 10 µl, and loaded with loading dye into an agarose gel similar to those used in the analysis of DNA. The agarose gels, however, were prepared in the same binding buffer used in the binding assay, so that the ionic conditions in the gel did not alter the binding effects obtained *in vitro*. After running the gels at 40-80 V, the gels were fixed for 30 min in 10% acetic acid, dried and set up for autoradiography. "Gel

retardation" refers to the observation that protein:RNA complexes run slower than free RNA.

(iii) **Filter retention assay**. Complexes were generated in volumes ranging from 100-500 µl. The protein:RNA complexes were collected by binding to filter discs (Millipore, HAWP) while unbound RNA was washed off with 20 mM Tris.HCl, pH7.5. Protein-bound RNA was estimated by Cherenkov counting of the filters in 5 ml of dH_20.

(iv) **Photocrosslinking and SDS-PAGE**. Photocrosslinking is a treatment which results in the formation of covalent bonds between amino-acids and nucleic acids. Proteins in direct contact with RNA were covalently photocrosslinked by irradiation with ultraviolet light. Samples were kept on ice while being irradiated for up to 30 min with ultraviolet light at 254 nm, output 600 J/m²/sec. When required, unprotected RNA was digested using a mixture of 50 µg/ml RNase A and 50 units/ml of RNase T_1 at room temperature for 20-30 min before analysis on SDS-PAGE. Protein:RNA complexes could also be cleaved chemically and enzymatically *after* photocrosslinking, to isolate separate binding domains.

(v) **Photocrosslinking and phenol extraction assay**. Protein:RNA complexes were prepared as described above, usually containing approximately 2 µg YB proteins and 0.1 µg of riboprobe in a final volume of 50 µl in binding buffer (usually HTB) before being crosslinked as above. For a quantitive assessment of the extent of crosslinking, the samples were adjusted to 1% SDS, 150 mM NaCl, 5 mM EDTA and 50 mM Tris.HCl, pH7.5. The samples were extracted with an equal volume of phenol:chloroform . The radioactivity in both the phenol and acqueous phases was measured by Cherenkov counting, the free RNA partitioning into the acqueous phase, the protein:RNA complexes at the surface of the phenol phase. The percentage of initital radioactivity per sample extracted into the organic phase was calculated in each case.

TMNT binding buffer for riboblots

Tris.HCl, pH 7.5	10 mM
NaCl	50 mM
$MgCl_2$	2 mM
Tween-20	0.05%

The above RNA-binding assays are summarized in **Fig. 10**.

2.43 Tissue section immunostaining

This experiment was performed by Dr. Sommerville, briefly as follows. Previtellogenic ovary was dissected and fixed with 2% TCA for 1 h, and then 70% ethanol overnight. Sections were de-waxed and rinsed in TBST, and then incubated in 100 µl of TBST + 5% FCS as a blocking agent for 1 h. After a further rinse in TBST, the sections were incubated for 2 h in 50 µl of TBST containing 1/100 dilutions of anti-HTSN and anti-HTP antisera. After three washes in TBST, the sections were incubated in 50 µl TBST containing a 1/100 dilution of fluorescein-conjugated anti-rabbit IgG (Sigma). After a final three washes in TBST, sections were examined by fluorescence microscopy using the appropriate filters.

2.44 Electron microscopy

This experiment was performed by Dr. Sommerville briefly as follows. Band-excised proteins were mixed with (i) whole poly(A)$^+$ mRNA and (ii) synthetic cyclin B1 riboprobe to reconstitute RNP particles. The reconstituted RNP was adsorbed onto carbon coated grids, rotary-shadowed by platinum-palladium and examined by electron microscopy.

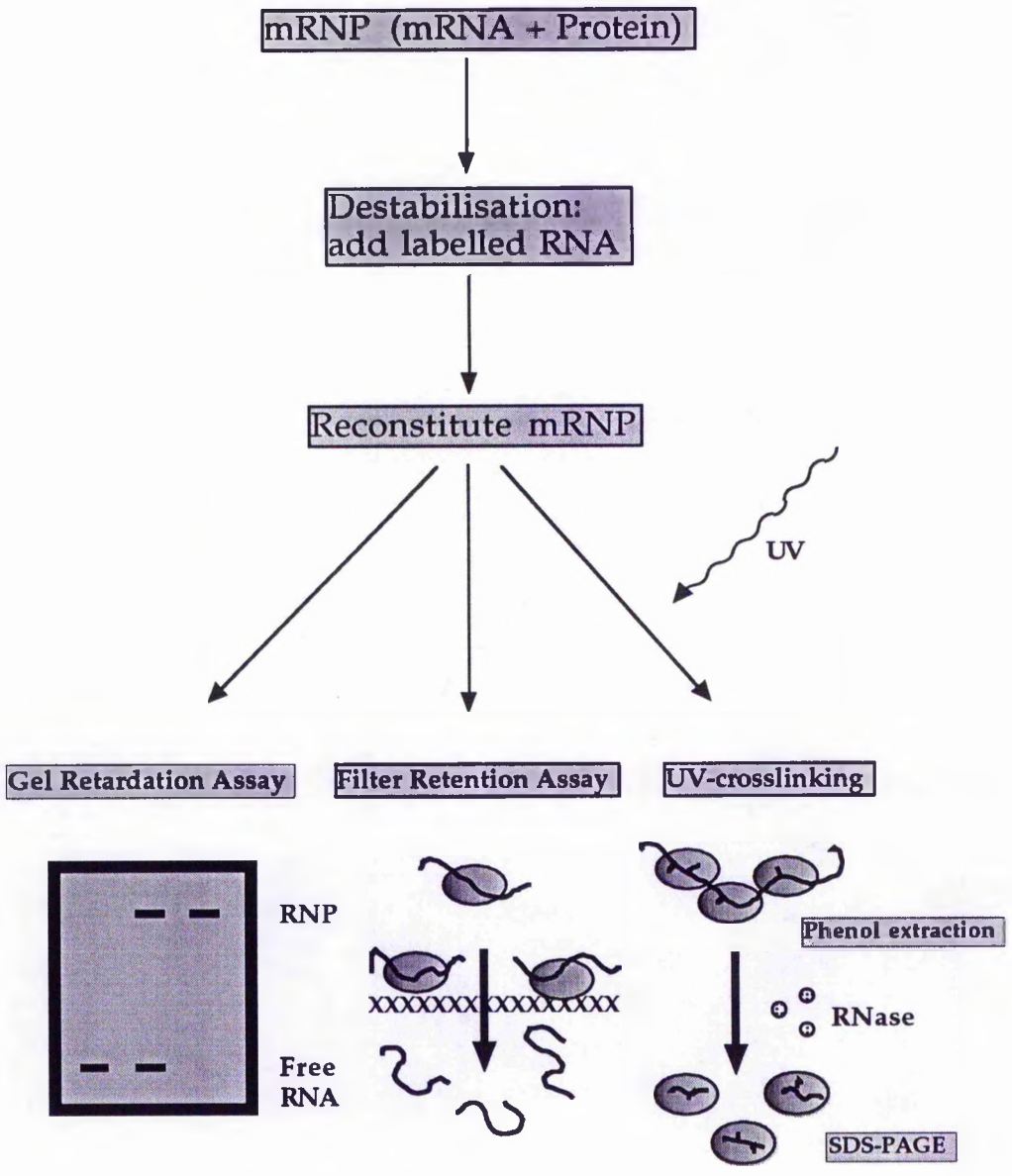


Figure 10 Conditions for protein:RNA binding. The first step is the destabilization of the pre-existing complex, and this was achieved using either 20 mM $MgCl_2$, 8 M urea or heating to 80 °C. Complexes were reconstituted by adding riboprobe, diluting the $MgCl_2$ or 8 M urea, or by cooling. Binding was measured using, for example: agarose gel retardation, filter retention, phenol extraction assays and SDS-PAGE.

Chapter 3

Composition of *Xenopus* Oocyte mRNPs

3.0 Aims

The aim of this study is to gain fresh information about the structure of the *Xenopus* oocyte "masked" mRNP; that is, the complex of mRNAs and associated proteins, which together store mRNAs for many months. The function of the mRNA-associated proteins is to package and store the mRNAs, as non-polysomal or "free" mRNPs, making them available for translation at the appropriate stage of development. The essential questions are, what are the mRNP proteins, what are their biochemical properties, and how do they work together? This chapter illustrates the experimental techniques that were used to isolate and fractionate the *Xenopus* oocyte mRNPs to investigate their composition. This study will take the approach of focussing on the most abundant mRNP proteins. Among these, the YB phosphoproteins pp60 and pp56 have so far been investigated in greatest detail: previous studies have shown them to be essential for mRNA masking (Kick *et al.*, 1987; Ranjan *et al.*, 1993; Braddock *et al.*, 1994; Bouvet and Wolffe, 1994). An attempt will be made to identify some of the other abundant mRNP proteins.

3.1 Isolation of poly(A)$^+$ mRNP by oligo(dT)-cellulose column chromatography

This technique relies on the presence of poly(A) tails at the 3' end of most mRNAs. Two disadvantages are that mRNPs without poly(A) tails are not selected, and that some non-specific binding of proteins with a natural affinity for oligo(dT) could conceivably occur. Homogenates of oocyte extracts were centrifuged at 10,000 g for 10 min to remove insoluble material. The supernatant obtained (SN10) was loaded onto an oligo(dT)-cellulose column in a loading buffer containing 300 mM KCl, which favours the hybridization of poly(A) tails to the oligo(dT) resin. In these

conditions the mRNP remains bound, and the unbound material, including ribosomes, 42S and 7S RNP particles and most cellular proteins, is washed through. **Fig. 11** is an example of an oligo(dT) mRNP isolation, and is illustrated by monitoring the absorbance of the column effluent at a wavelength of 254 nm. The absorbance is proportional to the RNA content, such that 1 OD/cm ≡ 40 μg RNA/ml. The sample was recycled through the column three times to ensure the column's maximal saturation. Next, the flow through, or unbound fraction, was collected as the "A^-" fraction. Once the A^- fraction was collected, the absorbance was washed to baseline with binding buffer, after which the mRNP was eluted in 2 ml of dH_20 at 30-40 °C, or alternatively in 60% formamide, conditions which disassociate the poly(A) tail from the oligo(dT) resin. The eluting poly(A)$^+$ mRNP generated a new peak on the trace. If the average OD in the poly(A)$^+$ peak in **Fig. 11** is taken as 0.5 OD, and the optical path length in the flow cell is 0.3 cm, ~66 μg RNA/ml is estimated to be present in the poly(A)$^+$ eluate. Assuming a protein to RNA ratio of 4:1 in the mRNP (Sommerville 1990), this corresponds to a protein concentration of ~264 μg/ml in the poly(A)$^+$ fraction.

The eluted mRNP was analyzed for protein composition on SDS-PAGE. **Fig. 12** illustrates the complex array of mRNP proteins. The most abundant mRNP proteins were described by Darnborough and Ford (1981) and labelled mRNP1 to mRNP8, but here will be referred to by their estimated molecular weights. In **Fig. 12A**, the mRNP proteins are illustrated, and include a protein band around 52 kDa (mRNP1, to be referred to as p52), a highly abundant band of 54 kDa (mRNP2, to be referred to as p54), as well as the phosphoproteins pp56 (mRNP3) and pp60 (mRNP4 or FRGY2) now identified as YB proteins (Deschamps *et al.*, 1992; Murray *et al.*, 1992). However, other abundant proteins are apparent

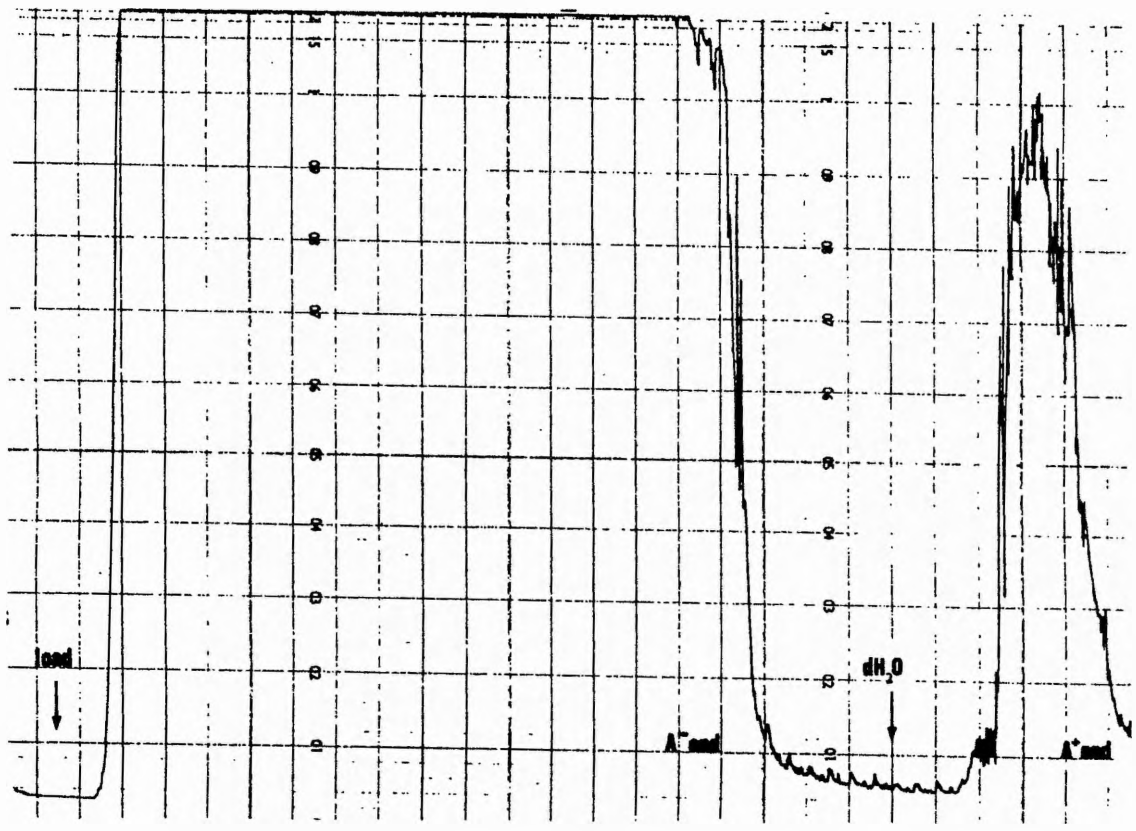


Figure 11 Elution of poly(A)$^+$ mRNP from an oligo(dT) resin. The elution of mRNP was measured via OD readings at A_{254}. An oocyte extract (SN10) was loaded (**load**) and recycled three times through the column, after which the unbound material was collected (**A$^-$ end**) while the OD dropped back to the baseline. Warm dH_2O was added to elute the poly(A)$^+$ mRNP, producing a peak OD reading of 1. Poly(A)$^+$ mRNP was eluted in a total volume of 2 ml dH_2O (**A$^+$ end**). The estimated concentrations in the eluted poly(A)+ mRNP are: ~66 µg/ml mRNA and ~264 µg/ml protein.

around 70 and 68 kDa, which may include the poly(A)-binding protein PABP (Marello *et al.*, 1992) as well as proteins of 40 and 25 kDa, and a group of proteins in the 100 to 200 kDa range. At this stage, except for the YB proteins and possibly the PABP, the identity and function of the other abundant proteins was unknown.

3.2 Further separation methods

Having obtained oocyte poly(A)$^+$ mRNP from an oligo(dT)-cellulose column, a procedure which in principle can be applied to any tissue extract, further purification strategies were considered. A useful technique, which will be used widely in later sections, is "heat-treatment" (HT). It relies on the thermostability of the YB proteins reported by Deschamps *et al.* (1991), who noted that the YB proteins remain soluble after an 80 °C treatment, whereas almost all the rest of the mRNP proteins precipitate. The procedure generates two crude fractions: the **HTSN** (heat-treatment supernatant) and the **HTP** (pellet) fractions, the former containing the YB proteins, and the latter the abundant p54/p52. These fractions are illustrated in **Fig. 12B**. Another strategy is "band excision", whereby specific protein bands, reversibly stained with a copper stain, can be cut out of an SDS-PAGE gel as accurately as possible, dehydrated, macerated and extracted by adding back dH_2O (see **Fig. 12C**).

A further separation method was developed following the report by Kolluri *et al.* (1992) that the human YB protein NSEP-1, expressed in *E.coli*, could be eluted from a heparin-Sepharose column using 1 M KCl. The *Xenopus* oocyte YB proteins were tested in a similar way using *in vitro* reconstituted RNP. The *in vitro* reconstituted RNP was prepared by heat-treatment, adding a riboprobe to native poly(A)$^+$ mRNP. Upon cooling, the YB proteins bound to the radioactive riboprobe as well as to the native

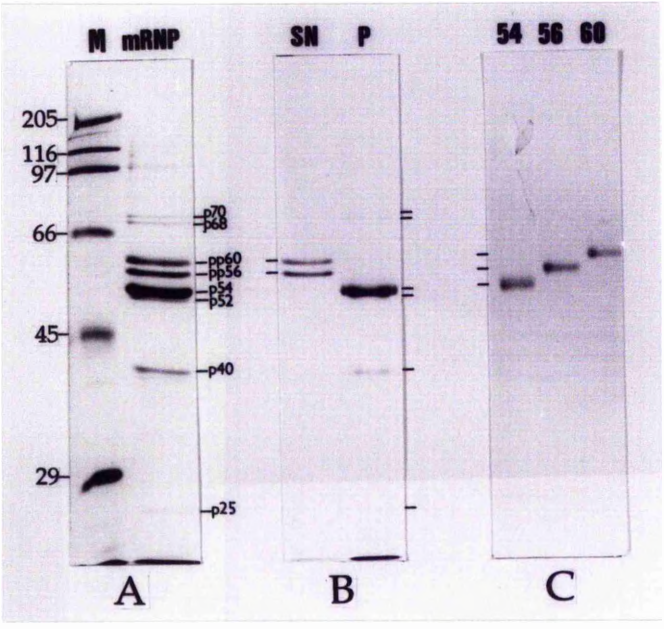


Figure 12 Protein composition of poly(A)$^+$ mRNP. Proteins separated by SDS-PAGE were stained with Coomassie blue. **(A)** 40 µl (containing approximately 12 µg) of poly(A)$^+$ mRNP is aligned with high molecular weight markers (**M**) (Sigma). Molecular sizes are indicated in kDa. **(B)** After 80 °C heat-treatment, the YB proteins pp60 and pp56 as well as some less abundant proteins of 96 and 100 kDa remained in the supernatant (**SN**), while the other mRNP proteins precipitated into a pellet (**P**). **(C)** The abundant mRNP proteins pp60, pp56 and p54 were band-purified by manual excision from an SDS-PAGE gel.

mRNA present in the mRNP. Because other mRNP proteins such as p54 were spun out in the pellet, the reconstituted RNP contained almost exclusively YB protein. This reconstituted RNP did not bind significantly to the heparin column (**Fig. 13A**). In agreement with this observation, previous trials with untreated native mRNP also suggested that the mRNP particles do not bind to the resin without any denaturation step (not shown). The elution of proteins was monitored by analyzing fractions on SDS-PAGE, and the fate of the RNA riboprobe in the reconstituted RNP was monitored by scintillation counting.

Because the reconstituted RNP did not bind to the heparin resin, it was deemed necessary to heat the reconstituted RNP to 80 °C *in the presence* of the resin. This resulted in the RNP becoming bound. Some of the riboprobe eluted in the flow-through (FT), but most eluted in 8 M urea (**Fig. 13A**). Similarly, the riboprobe could be eluted in 5 mM $MgCl_2$, while the proteins remained bound to the resin. $MgCl_2$ was chosen to elute riboprobe because in a previous study (Marello *et al.*, 1992), it had been reported to be a useful mRNP destabilizing agent. (The $MgCl_2$ elution effect will gain significance in chapter 4). In contrast, the heparin-bound YB proteins which were not removed by urea or $MgCl_2$, were instead eluted in 1 M NaCl (**Fig. 13B**), as was predicted from Kolluri *et al.* (1992). A possible interpretation is that the YB proteins bind to the poly-anionic heparin through ionic interactions due to the basic arginine-rich clusters in the tail domain. This ionic association is presumably reversed by 1 M NaCl or 1 M KCl but not by 8 M urea.

Because the YB proteins were loaded onto the heparin column via heat-treatment, the method involves an RNP sample which is already almost purified YB protein. However if RNA can be eluted by $MgCl_2$ or urea, this

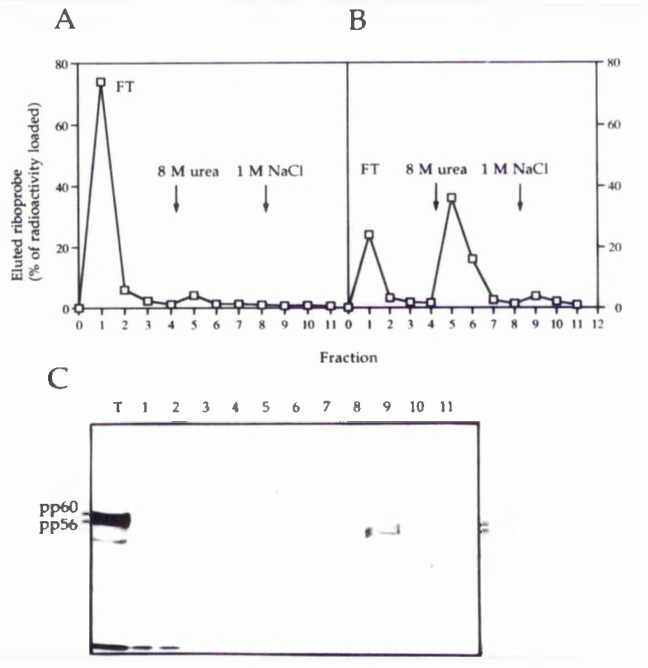


Figure 13 Elution of RNA and YB proteins from Heparin-Sepharose. **(A)** Riboprobe-protein complexes applied to the column eluted in the flow-through (**FT**) fraction. **(B)** Riboprobe-protein complexes formed after heat-treatment in the presence of heparin-Sepharose bound to the resin, the riboprobe being eluted with 8 M urea, the YB proteins being eluted with 1 M NaCl. **(C)** Proteins in the eluted fractions were detected by SDS-PAGE (Coomassie stain). Riboprobe was eluted with 5 mM $MgCl_2$ in fractions 5 and 6 as an alternative to 8 M urea; YB proteins were eluted with 1 M NaCl in fraction 9. Total protein preparation prior to chromatography is also shown (T).

method could be used as a means of separating YB proteins from the RNA to which they bind avidly. In subsequent experiments, the eluted YB proteins gave the impression that they tended to aggregate once the NaCl was diluted out. It seems likely that when RNA is not bound, the YB proteins aggregate by default via their charged tail domains. These observations are reminiscent of those of Lukanidin *et al.* (1972), who reported having to keep their mRNP in 2 M NaCl in order to maintain solubility. In summary, by combining oligo(dT) extractions with heat-treatment and heparin-Sepharose columns, YB proteins are greatly enriched. In principle, the method could applied to any tissue extract, and could be useful in a somatic tissue sample where the YB proteins are considerably rarer than in the *Xenopus* oocytes.

3.3 Electron microscopy of mRNP

The structure of reconstituted mRNP was examined using electron microscopy. Band-excised pp60 was combined with a heterogenous mixture of mRNAs extracted from native mRNP, as well as a riboprobe of defined length. The results are shown in **Fig. 14**. In panel **A**, the complexes are seen to have different sizes and shapes, due to the variable lengths of the native mRNAs. In panel **B**, in contrast, the mRNP was reconstituted with a synthetic cyclin B1 riboprobe. The three [pp60/cyclin B1] RNPs shown at a higher magnification in panel **C** suggest a degree of structural similarity, the length of the particles approximately 0.3 μm. This size is within the range of sizes of formaldehyde-fixed nuclear RNP particles described by Lukanidin *et al.* (1972). It may be suggested that the YB proteins are able to cover the RNA throughout its length, giving a beaded appearance, and that the shape of the RNP is determined by the secondary structure of the cyclin B1 RNA sequence.

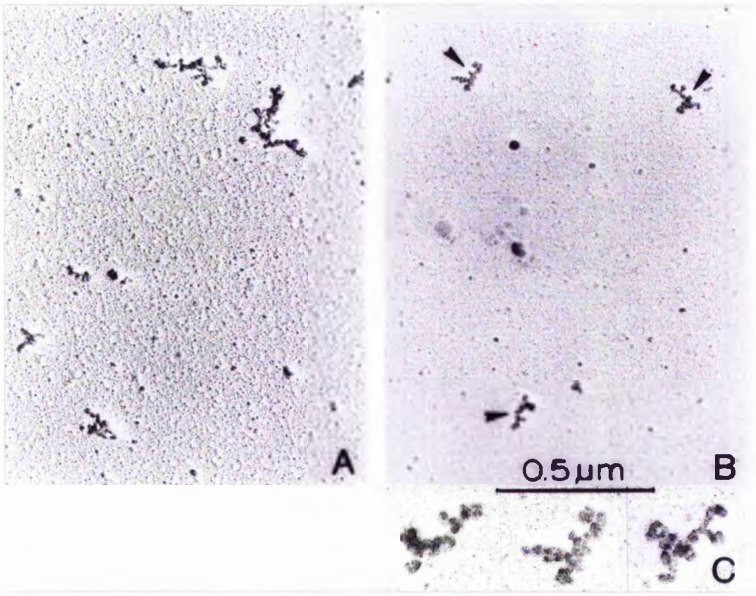


Figure 14 Electron microscopy of reconstituted riboprobe-pp60 complexes. **(A)** Band-purified pp60 was mixed with total oocyte mRNA, generating mRNP particles of different size. **(B)** pp60 was combined with *in vitro* synthesized cyclin B1 mRNA, forming particles of uniform structure, which are indicated by three arrows. **(C)** The three particles from **(B)** are enlarged, emphasizing structural similarities. The electron microscopy was done by Dr. J. Sommerville with the help of John Mackie.

3.4 Preparation of polyclonal antisera

Polyclonal antisera are a useful tool with which to attempt a number of experiments, such as Western blots and immunoscreening of cDNA expression libraries. To this end, two rabbits were injected respectively with the poly(A)$^+$ mRNP HTSN and HTP fractions generated by heat-treatment (**Fig. 12B**). The HTSN was expected to give rise to an antiserum which recognizes the the YB and any other heat-stable proteins, whereas the HTP fraction was expected to give rise to an antiserum which recognizes the remainder of the mRNP proteins.

Fig. 15 shows the reciprocal nature of the staining patterns obtained from anti-HTSN and anti-HTP. Whereas anti-HTP recognizes p54 and p52 but not the YB proteins pp56/pp60, anti-HTSN does not recognize p54 and p52, but prefers pp56/pp60 as well as two heat-stable proteins of 96 and 100 kDa. These two antisera defined two distinct fractions derived from *Xenopus* oocyte mRNP. In general, the antisera were diluted 1/1000 or 1/2000 for use on Western blots, and the secondary antibody was diluted 1/3000. Moreover, it was later found that blocking the Western blots in 10% fat-free skimmed milk dissolved in TBS, at least overnight if not longer, reduced the amount of background signal on the filters. At first, bound antibodies were detected with ^{125}I-Protein A, but subsequently a horseradish peroxidase ELISA assay was preferred which uses DAB (diaminobenzidine) and hydrogen peroxide to generate a brown colour on the Western blots. It proved more satisfactory, producing sharper bands, and became the method of choice.

Further information about the proteins recognized by anti-HTSN can be seen in **Fig. 16**, in which the native HTSN proteins are compared to the same proteins treated with calf intestinal phosphatase. Although

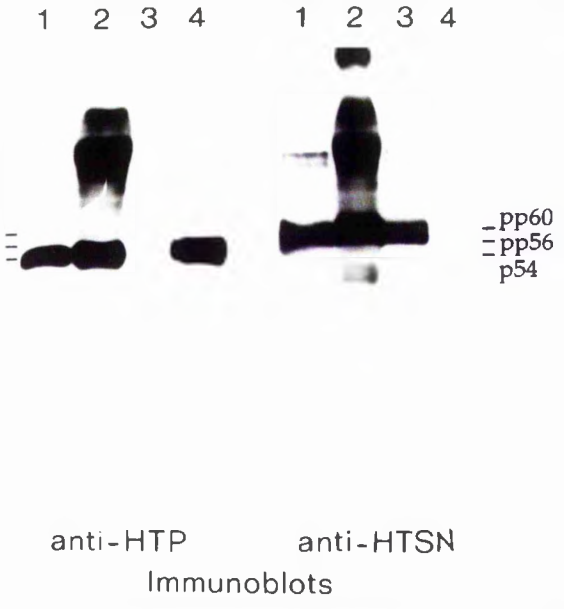


Figure 15 Immunoblot of poly(A)$^+$ mRNP with the antisera anti-HTSN and anti-HTP. Tracks **1**: 40 µl poly(A)$^+$ mRNP; **2:** 20 µl poly(A)$^-$ fraction; **3**: 40 µl poly(A)$^+$ HTSN (heat-treatment supernatant); **4**: HTP (heat-treatment pellet) fraction generated from 40 µl of poly(A)$^+$ mRNP. Protein transfers were immunoblotted with 1/2000 dilutions of antisera. Bound IgG was detected with ^{125}I-iodinated Protein A.

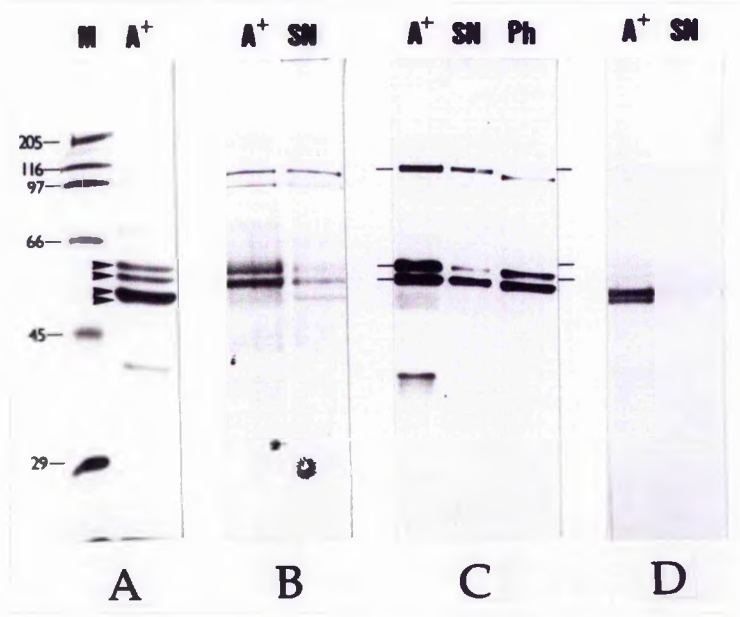


Figure 16 Immunoblot of poly(A)$^+$ mRNP using a horseradish-peroxidase detection system. **(A)** mRNP proteins were stained with Coomassie blue, including high molecular weight markers (**M**) and total poly(A)$^+$ mRNP (**A$^+$**). The YB phosphoproteins pp60, pp56, and the unidentified p54 and p52 are indicated (arrows). **(B)** Anti-HTSN immunoblot comparing poly(A)$^+$ mRNP to poly(A)$^+$ mRNP heat-treatment supernatant (**SN**). **(C)** Anti-HTSN immunoblot comparing poly(A)$^+$ mRNP, HTSN and HTSN treated with calf intestinal phosphatase (**Ph**). The clearer signal in **(C)** compared to **(B)** is due to better filter blocking. **(D)** Anti-HTP immunoblot of poly(A)$^+$ mRNP and HTSN. Both antisera were diluted 1/2000.

the protein bands are slightly sloped, suggesting a gel running artefact, mobility differences are apparent. This is clearer in the case of the 100 kDa protein, which appears to lose about 4 kDa upon dephosphorylation and will be hence be described as pp100 (phosphoprotein of 100 kDa). The YB proteins pp56 and pp60 are known from previous studies to be phosphoproteins (Cummings and Sommerville, 1988; Cummings *et al.*, 1989), however a mobility shift is less obvious in their case.

3.5 Immunoblots of proteins from different stages of development

Before this project, antisera had been generated in this lab against band-purified mRNP proteins. This approach yielded a number of antisera, including anti-pp60 and anti-p54, which were used in Western blots and with sectioned ovary. The anti-p54 antiserum was made use of in an honour's project (Robert Sykes, 1992), suggesting that p54 is present in the cytoplasms as well as in localized foci within the nuclei of tadpole-derived XTC cells. The same anti-p54 has been used to immunostain chromatin (Sommerville *et al.*, 1993); furthermore, anti-p54 has detected a cross-reactive protein in mRNP extracts from the alga *Fucus serratus* (Hetherington *et al.*, 1990). With these observations in mind, the new antisera were used to gain further information about the distribution of the mRNP proteins.

Oocytes and embryos from various stages of development were collected to determine the abundance of the main mRNP proteins through early development. **Fig. 17** shows immunoblots of extracts from various developmental stages. The immunoblot combined both anti-HTSN and anti-HTP antisera to cover the full extent of abundant mRNP proteins. Note that in this particular immunoblot, the YB protein pp56 appears as a doublet. A pp56 doublet has been observed on various occasions in

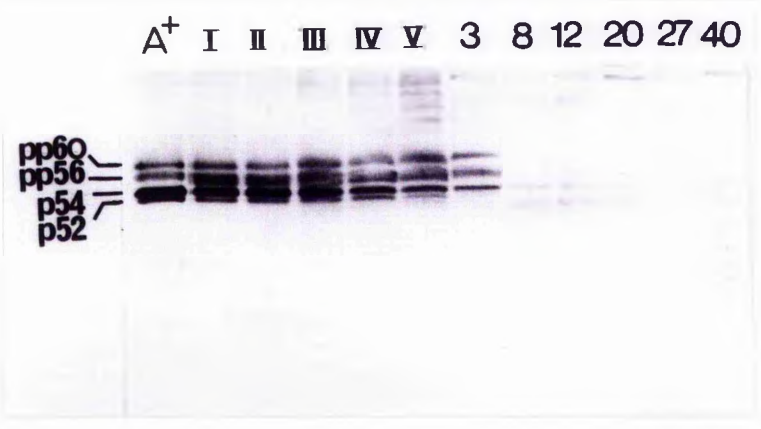


Figure 17 Immunoblot of total extracts from various developmental stages. Each track contains an SN10 from six oocytes or embryos. A combination of anti-HTP and anti-HTSN was used, both diluted 1/2000. Stages are indicated according to Dumont (1972): **A+**: 40 µl poly(A)+ mRNP from previtellogenic ovary; **I-V**: oocyte stages; **3**: 8-cell cleavage stage; **8**: blastula; **12**: gastrula; **20**: late neurula; **27**: tailbud and **40**: tadpole stages. Note the rapid decline of the major oocyte mRNP proteins between early cleavage and blastula.

material isolated from individual animals. Because both proteins in the 56 kDa doublet are heat-stable (not shown), it is presumed that they are both YB proteins, and that the presence of a pp56 doublet reflects the presence of different pp56 alleles in these individuals.

The main observations stemming from the stages immunoblot were that the abundant proteins pp60, pp56, p54 and p52 are all present throughout oogenesis, but that there is a dramatic decline of these major mRNP proteins after early cleavage stages and by blastula they have almost disappeared, except for the persistence of some p54. It is during the so-called mid-blastula transition (MBT) that zygotic gene transcription takes over, after the store of maternal mRNAs has been utilized. At that stage the oocyte mRNA "masking" proteins are no longer required. Their presence may even be deleterious in a context where efficient mRNA translation is required in the rapidly developing embryo. Given the picture in **Fig. 17**, it is possible that there is an active degradation pathway in which the mRNP proteins disappear in a matter of hours after fertilization, having been present in the oocyte for many months.

Stage I oocyte cross-sections were also tested with anti-HTSN and anti-HTP (**Fig. 18**). Whereas anti-HTSN gave a strong cytoplasmic as well as a weaker nuclear signal, anti-HTP gave an overwhelmingly cytoplasmic signal. The anti-HTP picture contrasts with the results obtained when the older anti-p54 antiserum was tested on XTC cells giving a nuclear stain (Robert Sykes, honour's project 1992), and with its staining of chromatin (Sommerville *et al.*, 1993). Either p54 is more obviously cytoplasmic in oocytes, the picture being different in other cell types such as XTC cells, or anti-p54 and anti-HTP do not recognize the same proteins. Later evidence will suggest that the abundant band of 54 kDa in oocyte mRNP comprises

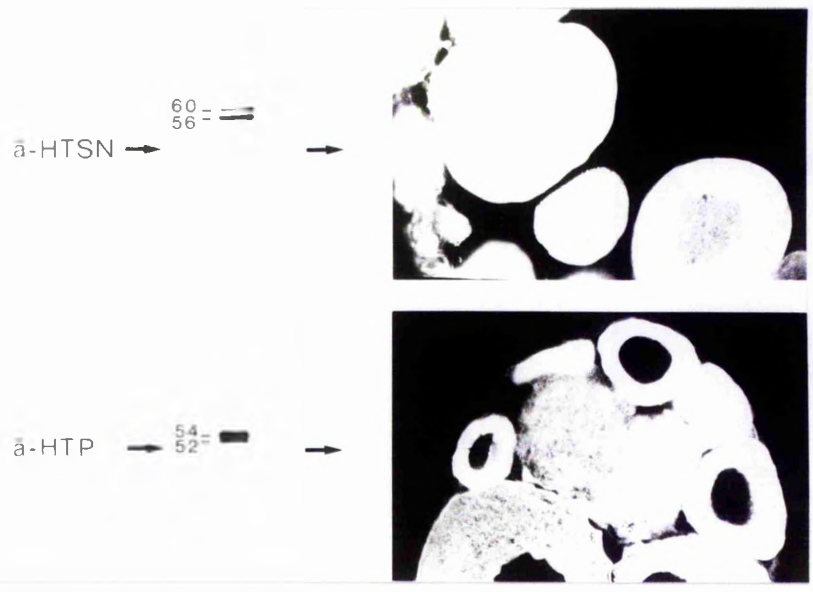


Figure 18 Immunostaining of stage I oocyte cross-sections. Anti-HTSN and anti-HTP immunostaining of stage I oocyte cross-sections, including their respective SDS-PAGE Western blots on the left for reference. The GV (germinal vesicles, or nuclei) are prominent. The sections were treated with a 1/1000 dilution of primary antisera, stained with a fluorescein-conjugated secondary antibody, and observed through microscopy. The immunostaining and microscopy were performed by Dr. J. Sommerville.

at least two or more proteins which cannot be separated efficiently by one-dimensional SDS-PAGE.

Finally, a cross-species immunoblot was attempted using oocyte extracts from two other amphibian species, *Notophthalmus viridescens* and *Necturus maculosus*. Crude oocyte extracts were heat-treated to generate HTSN and immunoblotted with anti-HTSN (**Fig. 19**). A candidate YB protein(s) of the appropriate size was detected in *Necturus* but not *Notophthalmus*. Given the higher C-value (genomic mass) of, for example, *Necturus* compared to *Xenopus*, and consequently its more prominent lampbrush chromosomes, anti-HTSN could be used to stain *Necturus* oocyte chromatin. The aim of such a study would be to investigate the distribution of YB proteins in promoter regions and at the same time their incorporation into nascent RNP.

3.6 Fractionation of poly(A)$^+$ mRNP by density centrifugation

Within the oocyte poly(A)$^+$ mRNP population, mRNAs with different translational fates exist. For example, nucleolin mRNA is translatable in previtellogenic oocytes, whereas c-*mos* mRNA is fully masked. If differences in translational fate are determined by the abundant mRNP proteins, then different mRNP classes could be purified which may be seen to have radically different protein compositions. Purifying specific mRNPs out of a complex mixture is not an easy task. On the other hand, it is possible to use simple techniques to fractionate mRNPs. Fractionation methods that use glycerol or sucrose gradients sort materials according to size, for example separating polysomes from free mRNP. Earlier unpublished data from this lab suggested that the major mRNP proteins pp60, pp56, p54, p52 and p40 go "hand-in-hand" over a range of particle sizes, broadly sedimenting between 40S-100S.

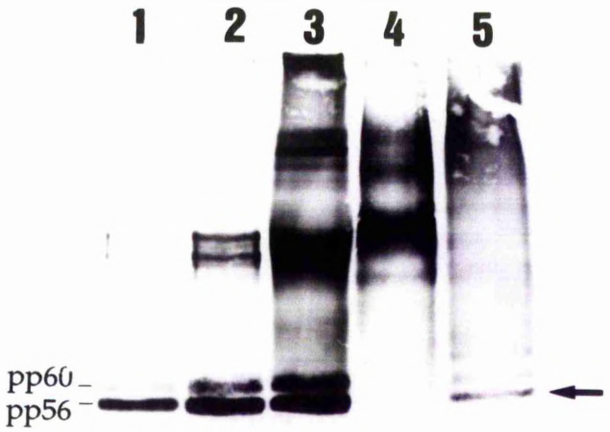


Figure 19 Anti-HTSN immunoblot detecting heat-stable proteins in oocyte extracts from other amphibian species. Tracks 1, 2, and 3: material derived from *Xenopus* oocytes. Track **1**: 40 µl poly(A)$^+$ mRNP HTSN: YB proteins are indicated (pp60 and pp56); **2**: 40 µl of whole poly(A)$^+$ mRNP; **3**: 20 µl poly(A)$^-$ fraction in which there is significant residual YB protein; **4**: 40 µl HTSN from *Notopthalmus viridescens* total oocyte SN10; **5**: 40 µl HTSN from *Necturus maculosus* total oocyte SN10. Anti-HTSN was diluted 1/1000 and a cross-reactive protein band with a mobility similar to the *Xenopus* oocyte YB proteins is indicated in track 5 (arrow).

A further separation method was attempted: density centrifugation. It exploits differences in densities between mRNPs. When separated on, for example, Cs_2SO_4 gradients, pure protein has a density of 1.20 g/cm^3 whereas pure RNA has a density of 1.62 g/cm^3, and any protein:RNA complexes have intermediate densities. In other words, mRNP will sediment according to the overall protein to RNA ratio. Because these gradients contain high salt concentrations, it is likely that some of the protein:RNA or protein:protein associations in the native particle could be disrupted. Given that YB proteins are stripped from a heparin resin using 1 M NaCl, as described previously, it is possible that the YB dissociate from the mRNP in the Cs_2SO_4 gradient. Consequently, the analysis included gradients of both fixed and unfixed mRNP.

The fixation method involved the use of either UV-crosslinking or formaldehyde. Both methods have advantages and disadvantages. Formaldehyde fixation forms permanent covalent bonds between proteins; therefore the proteins cannot be characterized on SDS-PAGE, but the mRNA can be extracted. On the other hand, UV crosslinking favours RNA:protein crosslinking, but used in high doses can result in protein:protein crosslinking (Setyono and Greenberg, 1981). After UV-crosslinking, the mRNAs are covalently bound to protein, but most of the RNA mass can be digested from the proteins by ribonuclease treatment. Therefore UV-crosslinking is to be preferred over formaldehyde fixation for SDS-PAGE analysis, but formaldehyde fixation is to be preferred for mRNA extraction.

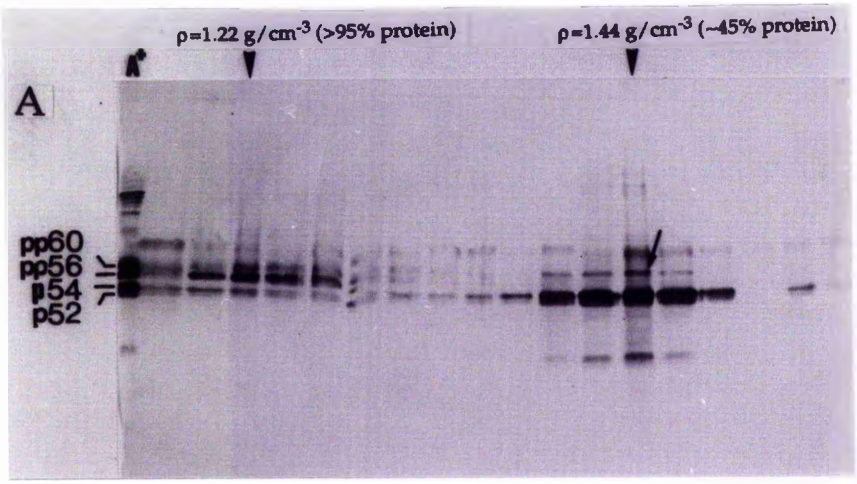
Fresh poly(A)$^+$ mRNP material was loaded onto preformed 5-50% Cs_2SO_4 gradients. After centrifugation at 36,000 rpm for 16 hours, equilibrium was reached. Protein samples were obtained by diluting the gradient fractions

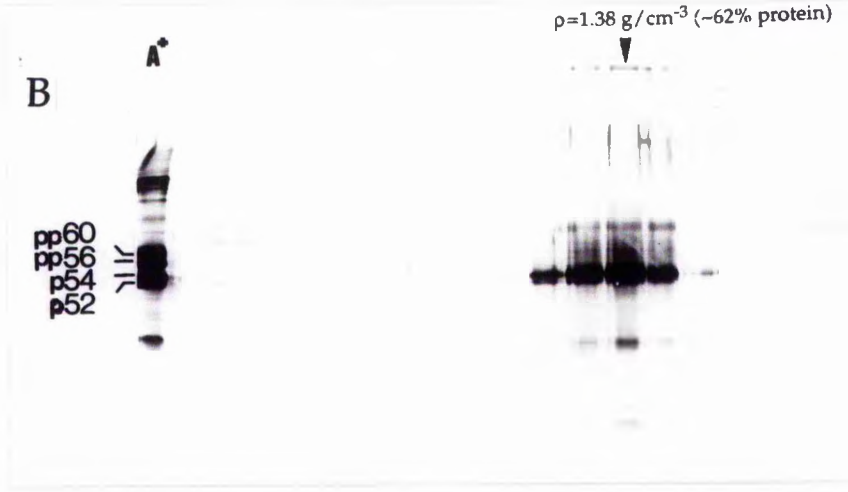
with 3 volumes of water, as described in Cardinali *et al.*, (1993), with the further addition of 10 μg of cytochrome C as carrier in order to enhance the size of the protein pellets and to colour them red. In addition, mRNA samples were obtained from gradient fractions by phenol extraction and ethanol precipitation, with the addition of 10 μg of carrier tRNA also to enhance the size of the pellets. The mRNA from the density fractions was hybridized on slot-blots with antisense probes from a variety of oocyte mRNAs so as to compare their distribution across the gradient.

Two fractionations are shown in **Fig. 20.** In the native, unfixed mRNP fractionation shown in **Fig. 20A**, some but not all YB proteins have been stripped from the particles, forming a fairly wide distribution throughout a wide range of densities towards the top of the gradient. In particular, most of the pp56, rather than the pp60, appears to have been stripped from the native particle. This difference between pp56 and pp60 will be noted and discussed in chapter 4. Apart from the YB proteins, a significant amount of p54/p52 exists in protein:RNA fractions with an average a density (ρ) of 1.44 g/cm^3, which corresponds to a protein content of 45%. These unidentified proteins resist the severe ionic conditions in the gradient and remain in protein:RNA complexes.

A different result is obtained with the UV-crosslinked material. The YB protein bands are not clearly visible: it is likely that they have been crosslinked efficiently, but not recovered as discrete proteins after ribonuclease digestion. As is shown later (Chapter 4), protein-protein crosslinking becomes a problem after several minutes of UV irradiation. Protein-protein crosslinking would account for the immunoreaction appearing as a smear (**Fig. 20B**). In this gradient, the bulk of immunoreactivity was located in the density range of ρ=1.35-1.39 g/cm^3, corresponding to an average protein content of 62%. These values are less

Figure 20 Analysis of protein composition of poly(A)$^+$ mRNP fractionated on Cs_2SO_4 density gradients. **(A)** Combined anti-HTSN and anti-HTP immunoblot of unfixed poly(A)$^+$ mRNP, fractionated on a Cs_2SO_4 gradient, then analysed by SDS-PAGE and immunoblotting. Total poly(A)$^+$ mRNP markers are present in the first track (**A$^+$**). The gradient runs from left to right in increasing density. Densities and corresponding % protein compositions are indicated. Arrows denote peaks of immunoreactivity. The position of pp60 is indicated by an oblique arrow. Anti-HTSN and anti-HTP antisera were diluted 1/2000. **(B)** A similar gradient, using UV-crosslinked poly(A)$^+$ mRNP.


ρ=1.22 g/cm-3 (>95% protein)
ρ=1.44 g/cm-3 (~45% protein)
A
pp60
pp56
p54
p52


ρ=1.38 g/cm-3 (~62% protein)
B
pp60
pp56
p54
p52

than the 1.44 g/cm^3 observed in the unfixed material in **Fig. 20B**. It is likely that the lower density value in the fixed material is due to maintaining the YB proteins in the particles thanks to UV-crosslinking.

Antisense RNA probes were hybridized to mRNA extracted from the density fractions, corresponding to the following list of *Xenopus* oocyte messages, which were divided into "translatable" (translating during oogenesis) and "masked" (translationally silent during oogenesis). The autoradiographs of the slot-blot hybridizations were prepared by Robert Little (Honour's thesis, 1994), and are not shown. All the probes were derived from cDNAs isolated in this lab, except for c-*mos* which was kindly provided by Dr. Sue Kingsman, Dept. Biochemistry, Oxford University.

Clone	Encodes	Translation
17.1	B4, an oocyte-specific histone H1 variant	Masked
c-*mos*	A proto-oncogene translated in mature oocytes	Masked
N5	FRGY1, somatic YB protein, active in many proliferating tissues	Masked
M7	Ribosomal protein S27	Translatable
AB21	Putative oocyte mRNP component	Translatable
AB12	Nucleolin, essential for ribosomal RNA synthesis	Translatable

All hybridized messages gave similar results in that they were found to be distributed in similar density fractions. In the native, unfixed mRNP

sample, all the above mRNAs peaked at a density (ρ) of 1.42 g/cm^3, and in the formaldehyde fixed mRNP sample they all peaked at 1.35 g/cm^3. These values correspond to those described in the immunoblots in **Fig. 20**: in the unfixed mRNP sample, the bulk of immunoreactive mRNP proteins, minus some dissociated YB protein, peaked at 1.44 g/cm^3, and in the UV-crosslinked material the proteins peaked at 1.35-1.39 g/cm^3.

Taken together, the observations that the abundant mRNP proteins pp60, pp56, p54, p52 and p40 co-fractionate on glycerol gradients (LaRovere and Sommerville, unpublished), that abundant mRNP proteins peak in similar density fractions, and that messages with different translational fates peak in the same density fractions, tend to suggest that a uniform, core mRNP particle exists. This interpretation would agree with the notion that the YB proteins pp60 and pp56 are in a sense "RNA-histones" (Tafuri and Wolffe, 1993b), and translational fate is determined by rarer message-specific proteins and signals. To this list of "RNA histones" we could now add the unidentified *Xenopus* mRNP proteins p54, p52 and p40.

However, the criteria of size and density alone might still not be sufficient to separate mRNPs with different "RNA histone" compositions, if their sizes and densitites are too similar to distinguish them. Pilot experiments were attempted with immunoprecipitation, but were not succesful because of the tendency of mRNPs to aggregate on incubation. In theory, if all mRNPs contain pp60, pp56, p54 and p52, both anti-HTSN and anti-HTP might be expected to co-precipitate all of the proteins together, and all classes of mRNAs should be precipitated with both these antisera. Furthermore, immunoprecipitation has already been attempted by Crawford and Richter (1987), who used a monoclonal antibody raised against *Xenopus* oocyte pp56. They found that some but not all mRNAs were precipitated. Three different mRNAs were precipitated by anti-pp56

and belonged to the "masked mRNA", or translationally repressed class. The eventual recruitment of the three bound mRNAs into polysomes coincided with a 22-fold decrease in the binding protein. Further experiments, including immunoprecipitation, will be required to clarify this unresolved issue.

3.7 Phosphorylation by an associated protein kinase

It has been known for several years that the *Xenopus* oocyte poly(A)$^+$ mRNP have an associated protein kinase activity. Simply by adding [γ-^{32}P]ATP to isolated mRNP in an appropriate buffer, the YB proteins are efficiently phospholabelled *in vitro* (Dearsly *et al.*, 1985). Furthermore, incubation of ovary in the presence of ^{32}P-phosphate leads to phosphorylation of the same proteins *in vivo* (Cummings and Sommerville, 1988). Phosphorylation of the YB proteins has been suggested to be essential for the formation of a masked mRNP (Kick *et al.*, 1987). The properties of the protein kinase activity, such as as sensitivity to heparin, and the ability to use GTP as well as ATP as a phosphate donor, are consistent with its being of the casein kinase II type (La Rovere, 1990). Many questions remain to be answered, such as the precise cellular location of the kinase activity, as well as the timing and the biochemical consequences of phosphorylation. **Fig. 21A** illustrates the mRNP-associated kinase activity. The YB proteins, along with other proteins, were clearly phospholabelled in the poly(A)$^+$ fraction. In contrast, the A$^-$ fraction, which contains significant amounts of YB protein, was not phospholabelled. This may be due to the presence of kinase inhibitors, or of competing phosphatase activities in the A$^-$ fraction.

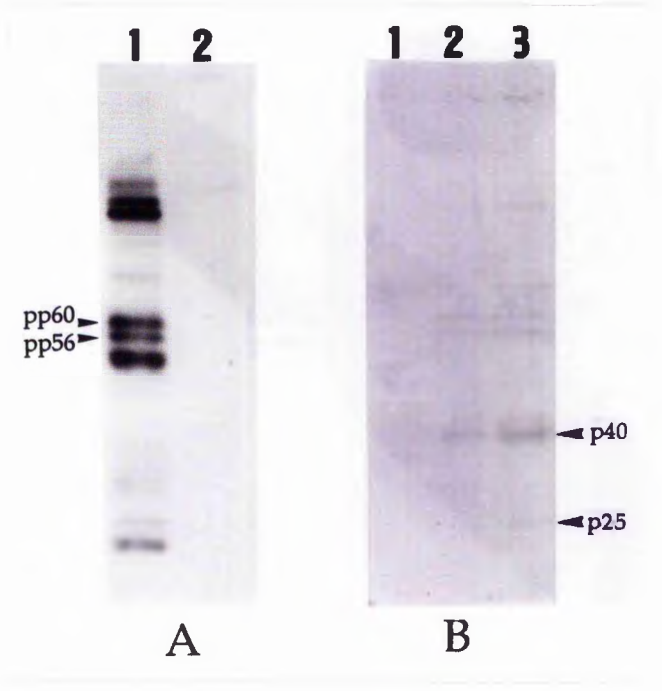


Figure 21 Detection of a protein kinase activity in the poly(A)$^+$ mRNP. **(A)** Phospholabelling: ~1 µCi [γ-^{32}P]ATP was added to 40 µl poly(A)$^+$ mRNP (track **1**) and 20 µl of the poly(A)$^-$ fraction (track **2**). **(B)** An immunoblot obtained with antisera directed against chicken casein kinase II α, α' and β subunits, each diluted 1/1000. Track **1**: 40 µl poly(A)$^+$ mRNP from very early previtellogenic ovary; **2**: 40 µl poly(A)$^+$ mRNP from previtellogenic ovary and **3**: 20 µl of the poly(A)$^-$ fraction.

Antisera raised against chicken casein kinase subunits α, α' and β were kindly provided by Dr. Erich Nigg, Swiss Institute for Experimental Cancer Research (Krek *et al.*, 1992). Although the use of anti-chicken casein kinase antisera is not ideal (anti-*Xenopus* antisera would be preferred), a strong conservation across species is apparent between cloned casein kinases, including casein kinase II subunits which have been cloned from *Xenopus* ovary (Jedlicki *et al.*, 1992). The α and α' subunits are closely related in sequence are responsible for the catalytic activity. **Fig. 21B** is an immunoblot of total poly(A)$^+$ mRNP, in which a combination of the three antisera was used. Although there was considerable background, to the extent that the YB proteins were visible, faint bands of 40 and 25 kDa are discernible. These sizes are close the sizes reported by Jedlicki *et al.* (1992: 41.5 kDa for the *Xenopus* ovarian casein kinase II α subunit, and 25 kDa for the β subunit). Moreover, in a later section, an ^{125}I-iodination experiment will show that a 40 kDa mRNP protein is strongly labelled at tyrosines. The *Xenopus* casein kinase α subunit described in Jedlicki *et al.* (1992) has 23 tyrosine residues out of a total of 350 amino-acids, which is a relatively high tyrosine content.

The poly(A)$^+$ mRNP analyzed in these studies is derived from a whole oocyte SN10, of which the bulk is cytoplasmic. However, a subset of the poly(A)$^+$ mRNP includes pre-cytoplasmic mRNP, containing certain nuclear proteins, some of which may be lost once the particle is transported into the cytoplasm. This could apply to the kinase . That the association of the kinase with mRNP is nuclear is implied by the studies of Braddock *et al.* (1994), in which it is shown that the micro-injection of anti-casein kinase II IgG into germinal vesicles (oocyte nuclei) inhibits mRNA masking. Germinal vesicles were isolated from different oocyte stages, in order to determine the cellular location of the kinase activity. **Fig. 22B**

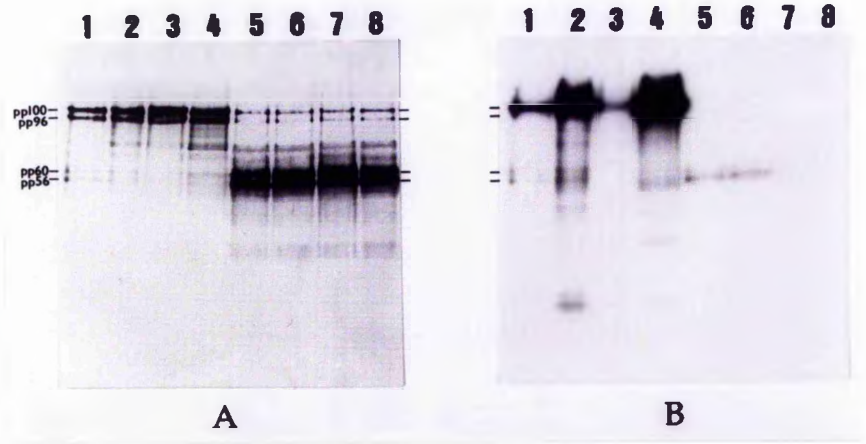


Figure 22 Comparison of the protein kinase activity present in GVs and poly(A)$^+$ mRNP. **(A)** Immunoblot, using anti-HTSN diluted 1/2000 and **(B)** autoradiograph of the same blot, showing the phospholabelled bands. Each sample was given ~1 µCi [γ-^{32}P]ATP. Track **1**: 40 GVs, stored in Barth's medium, frozen at -20 °C and thawed for use; all subsequent tracks contain GV material which was not frozen; **2**: 10 GVs + 2 mM spermine; **3**: 10 GVs + 1 µg/ml heparin; **4**: 20 GVs, no addition; **5**: 30 µl (approximately 9 µg of poly(A)$^+$ mRNP); **6**: 30 µl poly(A)$^+$ mRNP + 2 mM spermine; **7**: 30 µl poly(A)$^+$ mRNP + 1 µg/ml heparin; **8**: 30 µl poly(A)$^+$ mRNP + 5 µg/ml quercetin. Heparin and quercetin are inhibitors of casein kinase II. GVs were obtained from oocyte stages II to III. In comparing tracks 2 and 4, note that the YB proteins are phospholabelled better when 10 GVs were used with 2 mM spermine rather than 20 GVs without spermine. Spermine is an enhancer of casein kinase II activity.

shows that the kinase activity is certainly nuclear. In the germinal vesicle fractions, it is apparent that more than one protein around 100 kDa is well labelled, and that the less abundant YB proteins are not as clearly labelled. Two prominent bands of 96 and 100 kDa appear on the immunoblot of the same transfer (**Fig. 22A**), particularly in the GV samples but also in the poly(A)$^+$ mRNP samples. In contrast, in the poly(A)$^+$ mRNP fractions, the YB proteins are phospholabelled best, and are the most prominent on the immunoblot. The kinase activity was enhanced by the addition of spermine, and inhibited by heparin. These are respectively casein kinase II activators and inhibitors. When additional YB protein substrate was added to isolated GVs, which contain only a small amount of YB proteins, they were efficiently phosphorylated (**Fig. 23A**). Likewise casein, which is by definition a good substrate for casein kinase II, was also phospholabelled in the GVs. The kinase activity is not apparent in cytoplasmic extracts from oocyte stages III and IV (**Fig. 23B**). In agreement with Krek *et al.* (1992), who immunostained casein kinase II in nuclei of somatic chicken cells, casein kinase II activity is readily detected in the *Xenopus* oocyte GV (nuclei). Thus the kinase activity present in the poly(A)$^+$ mRNP fractions could be derived from the fraction of mRNP which are pre-cytoplasmic.

In an attempt to purify the kinase activity, a fresh oocyte extract SN10 was applied to an oligo(dT) column, and the eluted poly(A)$^+$ mRNP was loaded onto a heparin-Sepharose column, which binds casein kinase II. The kinase activity was eluted from the heparin-Sepharose column with 1 M NaCl. The eluted fractions were tested for kinase activity by adding [γ-^{32}P]ATP plus substrate protein. The fraction eluted in 1 M NaCl was the most active (**Fig. 24A**). A Coomassie stain or immunolot of the eluted fractions was not available, so it was not possible to confirm the isolation of putative casein kinase II subunits of the expected sizes, such as those in

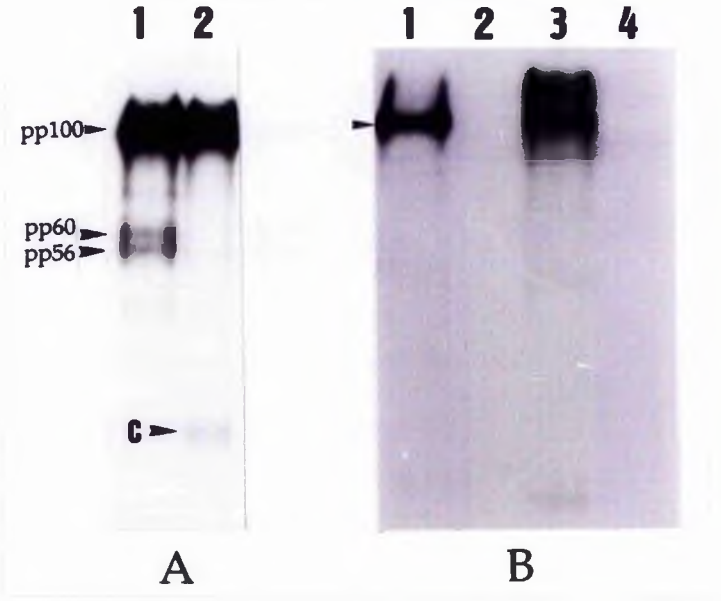


Figure 23 Comparison of protein kinase activity in GVs and cytoplasms. **(A)** Additional substrates were supplied to the GVs: track **1**: 10 stage III GVs + 5 µl of a poly(A)$^+$ mRNP HTSN; **2**: 10 stage III GVs + 10 µg casein. **(B)** The protein kinase activity was tested in cytoplasmic extracts: track **1**: 10 stage III GVs; **2**: 10 stage III cytoplasms; **3**: 10 stage V GVs; **4**: 10 stage V cytoplasms. Each sample in (A) and (B) received ~1 µCi of [γ-^{32}P]ATP.

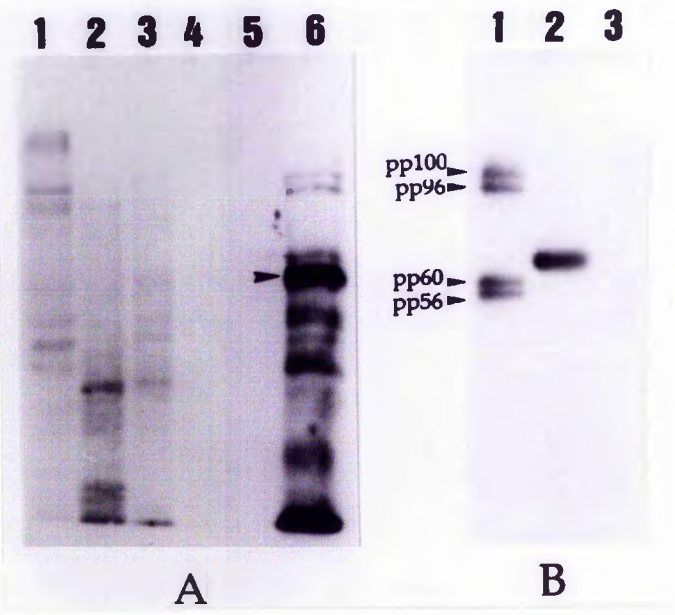


Figure 24 Attempted purification of the protein kinase activity. (A) Column chromatography. The presence of protein kinase activity in each fraction was tested by adding ~1 μCi [γ-^{32}P]ATP. Track **1**: oligo-dT column effluent (unbound fraction); **2**: 15% formamide elution; **3**: 60% formamide elution; **4**: a heparin-Sepharose column was loaded with material from track 3: column effluent; **5**: 400 mM NaCl elution from the heparin column and 6: 1 M NaCl elution. **(B)** The partially purified activity from the 1 M NaCl elution was tested with YB protein substrate. Track **1**: 30 μl poly(A)$^+$ mRNP HTSN + 5 μl kinase activity eluted in the 1 M NaCl fraction, mixed in 20 mM Tris.HCl, pH 7.5 and 2 mM $MgCl_2$, a buffer in which casein kinase II is active; **2**: the mixture shown in track 1 was heat-treated to destroy the added kinase activity, and then 10 units of calf intestinal phosphatase were added; note that the YB proteins were dephosphorylated, but the phosphatase itself was phospholabelled; **3**: control: ~1 μCi [γ-^{32}P]ATP was added to poly(A)$^+$ mRNP HTSN without any added kinase activity.

Fig. 21B. However, the activity eluted from the heparin column was tested by adding it to some mRNP HTSN. The HTSN has lost its associated kinase activity through heat-treatment, and as was shown previously, contains mostly YB proteins. When [γ-^{32}P]ATP was added to this fraction alone, no labelling occurred (**Fig. 24B**, track 3), but when some of the heparin column fraction containing protein kinase activity was added, the proteins were phospholabelled. This phospholabelling was subsequently lost after the addition of phosphatase (**Fig. 24B**, tracks 1 and 2).

Neither the immunoblot nor the purification procedure conclusively identified the expected protein bands. It is also possible that the mRNP-associated, casein kinase II-like activity is provided by divergent, specialized form of casein kinase II whose protein subunits are of different sizes. However, the difficulty in visualizing the stained bands of a predicted size may be due to the kinase's association with mRNP only in the pre-cytoplasmic mRNP. In that case, only a small fraction of cellular poly(A)$^+$ mRNP would include a detectable kinase activity, but not enough protein to be detected by these methods. According to the nuclear microinjection experiments by Braddock *et al.*, (1994), a nuclear casein kinase II activity is associated with mRNP somewhere along the path from nascent RNA transcripts to fully processed mRNA. The kinase phosphorylates the YB proteins, and it is thought that their phosphorylation enhances and is indeed essential for mRNA masking (Braddock *et al.*, 1994; Kick *et al.*, 1987). Finally, the phosphoproteins pp96 and pp100 observed in these experiments are apparent both in poly(A)$^+$ mRNP (**Fig. 22B**) and poly(A)$^+$ HTSN (**Fig. 24B**). At the same time, they appear to be more abundant in the GVs (**Fig. 22**). Their relevance to mRNP is not known. Casein kinase II is known to phosphorylate a wide spectrum of targets. After the completion of this project, nuclear phosphoproteins of

the same size (100 and 96 kDa) were crosslinked to riboprobe in GV extracts and identified by immunoblotting as nucleolin. This was shown using a monoclonal antibody specific to nucleolin, kindly provided by Prof. Dr. Ulrich Scheer, University of Würzburg. Whether or not nucleolin has a role in pre-mRNA processing or transport remains to be seen.

3.8 Summary

Xenopus oocyte mRNP was isolated using a well established technique which is based on the hybridization of polyadenylated mRNAs to oligo(dT) columns. The YB proteins can be purified further via the heat-treatment described by Deschamps *et al.*, 1991, and in addition a method is described to purify YB proteins using a heparin-Sepharose affinity column. Electron microscopy has suggested that in *in vitro* reconstituted RNP, the YB protein pp60 forms a complex with RNA, and that the beaded shape of the mRNP is dependent on the length and structure of the mRNA. Two new antisera were generated using the heat-treatment fractions: anti-HTSN and anti-HTP, which could be used to study the abundant mRNP proteins. It was found that pp60, pp56, p54, p52 and p40 are present throughout oogenesis, but decline rapidly after early cleavages and by the blastula stage they have mostly disappeared. This implies the existence of an active degradation mechanism, such that when the maternal mRNP are no longer required, they are quickly removed. The proteins recognized by anti-HTP proteins appeared to be mostly if not wholly cytoplasmic, whereas the proteins recognized by anti-HTSN appeared to be both cytoplasmic and nuclear on immunostaining of immature ovary sections. A putative YB protein homologue was detected in a heat-stable extract from *Necturus* oocytes. A fractionation of mRNP using density gradient centrifugation showed that the abundant mRNP proteins are part of an mRNP particle of a density of 1.35-1.39 g/cm^3 on a

Cs_2S0_4 gradient, which corresponds to a protein content of 62%. Both translatable and translationally repressed mRNAs were present in the same density fractions, suggesting that mRNPs with different translational fates have a similar composition with respect to the abundant mRNP proteins. Further work including immunoprecipitation studies will be needed to clarify the issue of whether the proposed masking proteins pp56 and pp60 are bound only to translationally repressed mRNAs. The density gradients draw attention to the unidentified mRNP proteins, in particular p54, p52 and p40, which are an integral part of the mRNP particle. Finally, a nuclear casein kinase II-like activity was confirmed to be associated with the mRNP. It phosphorylates the YB proteins, as well as at least two other proteins of 96 and 100 kDa which were later identified as nucleolin.

Chapter 4

Binding of *Xenopus* Oocyte mRNP Proteins to mRNA

4.0 Aims

The precise nature of the RNA:protein interactions in the *Xenopus* oocyte mRNP particles is not fully understood. This information is necessary to understand the processes of formation, maintenance, and eventual disassembly of the mRNP particles. The preceding chapter suggested that the mRNP particle is composed of a complex set of proteins, some of which make stable contacts with mRNA. Among these, the YB proteins will be considered in greater detail, because previous studies have suggested that they are essential for mRNA masking (Kick *et al.*, 1987; Ranjan *et al.*, 1993; Braddock *et al.*, 1994; Bouvet and Wolffe, 1994). Their structure includes at least two potential RNA-binding domains: the cold-shock domain (CSD) and the arginine-rich basic/aromatic islands in the tail domain (TD).

4.1 Riboprobe binds to mRNP proteins on Western transfers

A simple procedure for testing the ability of different mRNP proteins to bind RNA is to separate them on SDS-PAGE, transfer them to nitrocellulose and then incubate them with riboprobe in a suitable binding buffer. This procedure is referred to as "riboblot". The first riboblot shows riboprobe binding to poly(A)$^+$ mRNP and to poly(A)$^-$ fractions, and suggests that by modifying the salt concentrations or by adding various competitors to the binding buffers, the RNA-binding profile obtained can be affected radically (**Fig. 25**). After washing the Western transfer in 50 mM NaCl, the riboprobe remained bound to most of the protein bands, whereas a 2 M NaCl wash reduced binding to the extent that the YB proteins lost the bound riboprobe. A compromise 500 mM NaCl wash reduced background significantly but

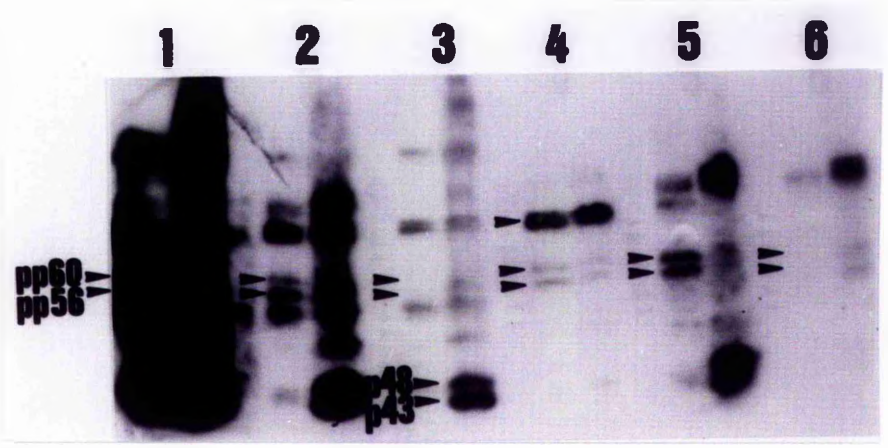


Figure 25 Binding of riboprobe to Western transfers. This assay can also be referred to as "NorthWestern" or simply "riboblot". The standard binding buffer was (10 mM Tris.HCl pH7.5; 2 mM $MgCl_2$; 50 mM NaCl; 0.05% Tween-20), abbreviated TMNT. The riboprobe was added at room temperature and left to bind for thirty minutes before washing the filter. Each separate strip (**1-6**) contains 40 µl (~12 µg) poly(A)$^+$ mRNP (left) and 20 µl of the poly(A)$^-$ fraction (right). **Strip 1**: low salt wash (50 mM NaCl); **2**: high salt wash (500 mM NaCl); **3**: very high salt wash (2 M NaCl); the next three strips were washed in 500 mM NaCl. **Strip 4**: 5 µg/ml heparin added to the binding reaction; **5**: the strip was pre-incubated for 1 hour with a 1/2000 dilution of anti-HTSN before riboprobe binding; **6**: after riboprobe binding, this strip was incubated with a mixture of 1 µg/ml RNase A and 1 unit/ml RNase T_1 for thirty minutes. Two arrows indicate the YB proteins throughout. Indicated in addition are p48 and p43 in strip 3, and two proteins of 68 and 70 kDa in strip 4.

did not prevent riboprobe binding to the YB proteins, and was chosen as a standard procedure.

Apart form the YB proteins, many other proteins bind the riboprobe using this approach. For example, in the poly(A)$^-$ fraction the riboprobe was bound to two abundant proteins: p48, an oocyte homologue of the translation factor EF1α (Mattaj *et al.*, 1987), and p43, a 5S RNA-binding protein similar to TFIIIA (Joho *et al.*, 1990). Heparin, a poly-anionic molecule, reduced riboprobe binding to the extent that binding to p48 and p43 was no longer apparent in the poly(A)$^-$ lane. Heparin also reduced binding to the YB proteins, whereas in contrast, binding to proteins of 68 and 70 kDa was unaffected. The addition of anti-HTSN (antiserum raised against the mRNP, heat-treatment supernatant) did not effectively prevent riboprobe binding to the YB proteins, suggesting that this particular antiserum does not wholly block the RNA-binding epitopes. RNase A and RNase T_1, which were added to the washing solution, removed most of the signal. This suggests that the protein-bound riboprobe is accessible to the enzymes in solution.

Another riboblot is shown in **Fig. 26**, in which the YB proteins are seen more clearly. There was a distinctly stronger binding to pp60 than to pp56. This difference, observed on a number of occasions, was not due to differences in protein concentration. In panel B, previously bound riboprobe was stripped off using 8 M urea; however, a small proportion (<10 %) of the previously bound riboprobe activity could not be removed by this means. Therefore in general, it is not advisable to recycle Western blots for a second binding assay. Nevertheless, new riboprobe was added in the presence of excess cold poly(C,U), which

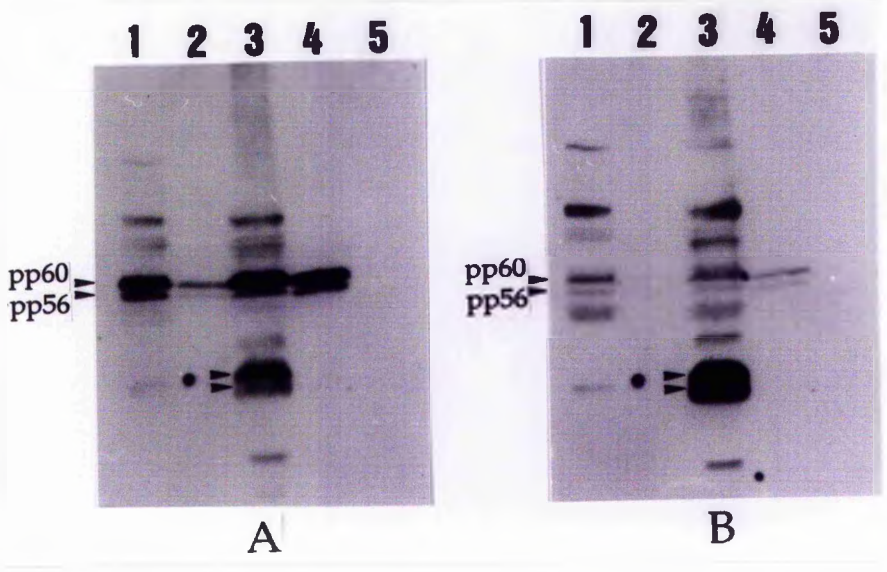


Figure 26 Further binding of riboprobe to Western transfers. (A) Cyclin B1 sense strand riboprobe was bound to transferred proteins in TMNT buffer, and the Western transfer washed in TMNT + 500 mM NaCl. **(B)** Bound riboprobe was removed from the same Western transfer using 8 M urea, and re-probed with the same cyclin B1 riboprobe in TMNT buffer with the addition of 10 µg of poly(C,U) competitor. **Track 1**: 40 µl (~12 µg) poly(A)$^+$ mRNP; **2**: 40 µl poly(A)$^+$ mRNP HTSN ; **3**: poly(A)$^-$ fraction; **4**: 20 µl poly(A)$^-$ fraction HTSN; **5**: 20 µl band-purified p54.

had been reported to be an efficient binding competitor (Marello *et al.*, 1992). When the relative binding to YB proteins and the p48/p43 markers in tracks A3 and B3 is compared, much more riboprobe has bound to p48/p43 compared to pp60/pp56 in the presence of the poly(C,U) competitor.

The riboblot technique has certain limitations. It involves denatured proteins which are separated on SDS-PAGE, transferred to nitrocellulose and washed in 8 M urea before binding. It relies on the ability of the transferred proteins to renature correctly. This cannot be assumed to work for all RNA-binding proteins, but nonetheless may be suitable for some RNA-binding proteins, such as the YB proteins. In a later section, the technique will be adapted for screening a cDNA expression library to isolate RNA-binding proteins (section 5.1).

4.2 Anti-mRNP IgG can interfere with riboprobe binding

In addition to blocking mRNA masking by injecting anti-casein kinase II IgG into germinal vesicles, the microinjection experiments of Braddock *et al.* (1994) also used anti-pp60 and anti-p54 IgG to inhibit mRNA masking *in vivo*. An attempt was made to mimic these effects *in vitro*. Poly(A)$^+$ mRNP was treated with 30 mM $MgCl_2$ as described by Marello *et al.* (1992), in order to bind riboprobe: $MgCl_2$ is thought to destabilize the mRNP, and when the $MgCl_2$ is diluted to 3 mM, the mRNP are free to "reconstitute" with the riboprobe (section 2.41). Binding to riboprobe was measured with the filter retention assay (see section 2.42). Riboprobe binding was inhibited by treating the mRNP with increasing concentrations of anti-pp60 and anti-p54 IgG, both singly and in combination, before the reconstitution step. Two riboprobes of different sizes were used (**Fig. 27**).

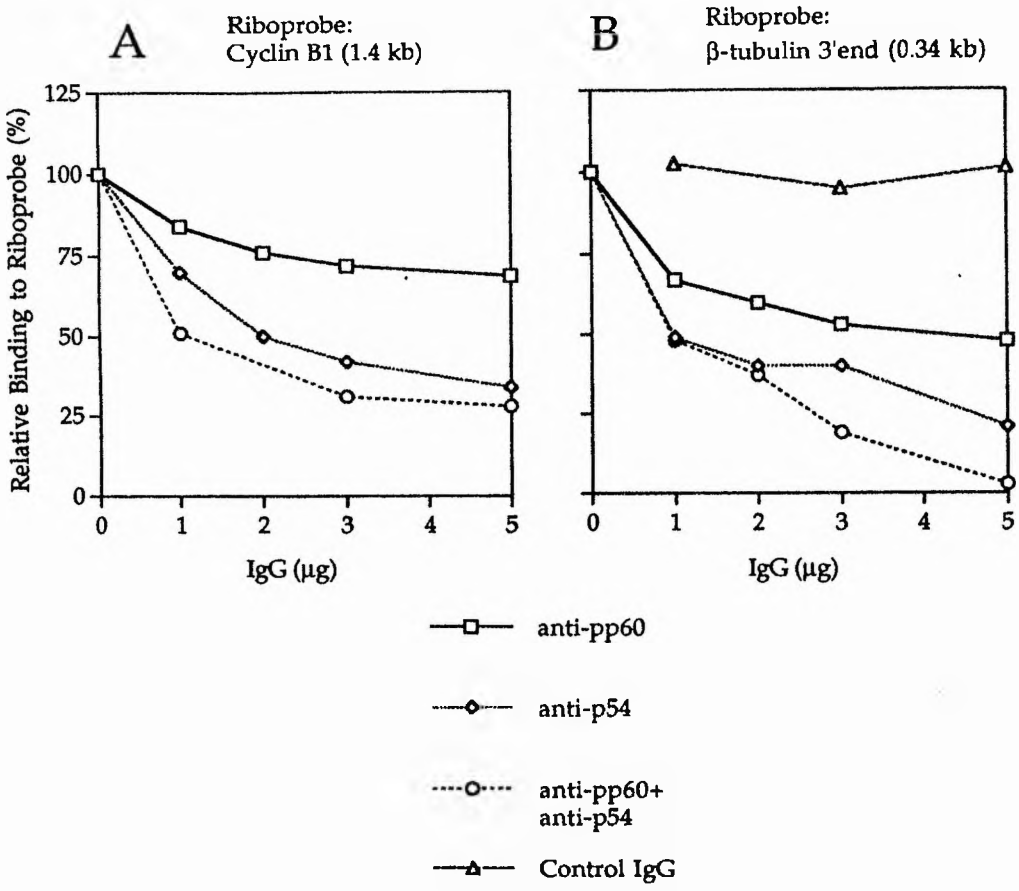


Figure 27 Anti-mRNP IgGs interfere with RNA binding *in vitro*. Riboprobe binding was measured with a filter retention assay after 30 mM $MgCl_2$ treatment of poly(A)$^+$ mRNP. IgG fractions were obtained by Dr. Sommerville from anti-pp60 and anti-p54 polyclonal antisera by chromatography through DE-cellulose. Poly(A)$^+$ mRNP was pre-incubated with increasing amounts (0-5 µg) of IgGs prior to the addition of riboprobe and $MgCl_2$. Binding values are expressed as a percentage of the binding obtained without added IgG. A control IgG was included in **(B)**: goat anti-rabbit IgG. **(A)** The riboprobe was a 1.4 kb cyclin B1 sense strand and **(B)** clone 16.2, a 0.34 kb β-tubulin 3' end fragment.

In summary, the findings were as follows: both anti-pp60 and anti-p54 interfered with RNA-binding, whereas a control IgG did not. Anti-p54 was more effective than anti-pp60, and a combination of both antibodies reduced binding even further. Binding to the shorter riboprobe was more sensitive to the antibody interference, and the following explanation is proposed: since binding was measured using a filter retention assay, in which free RNA washes through the filter and protein:RNA complexes are trapped, the binding of a single protein to a riboprobe molecule is sufficient to keep it in the bound fraction. With a longer riboprobe there was a greater chance that a single protein would bind.

It is presumed that these antibodies are able to interfere with RNA-binding by blocking RNA-binding epitopes in the mRNP proteins. Consequently, both the YB protein pp60, and the unidentified protein p54 would seem to have the ability to bind to RNA. The YB protein pp60 is known to bind to mRNA, as was confirmed in the electron micrographs presented earlier. However, the identity and function of p54 is not known. It might bind to the mRNA directly, or alternatively work together with YB proteins to facilitate the packaging of mRNA. The *in vitro* results were consistent with the nuclear microinjection results reported by Braddock *et al.* (1994).

4.3 UV-crosslinking mRNP proteins to mRNA

In this study, "UV-crosslinking" (or "photocrosslinking") refers to the formation of covalent bonds following exposure to UV irradiation. UV irradiation is used quite extensively as a technique for fixing molecular contacts between proteins and nucleic acids. UV-crosslinking is believed to be mainly but not exclusively due to the aromatic side

chains of tyrosine and phenylalanine (Smith 1969). In UV-crosslinking studies, the general assumption is that only proteins that are specifically bound to nucleic acids can be crosslinked. For example, Setyono and Greenberg (1981) UV-crosslinked the poly(A) binding protein to RNA; Van Eekelen *et al.* (1979) used high doses of UV light at 254 nm to compare the profile of hnRNA-bound proteins with that of mRNA-bound proteins, and in the *Xenopus* oocyte system, Swiderski and Richter (1988) crosslinked proteins to maternal mRNA. The latter authors reported that many oocyte proteins were UV-crosslinked to maternal mRNA. Some proteins were bound to specific mRNAs, but other proteins of 60, 56 and 40 kDa were suggested to be present on numerous if not all mRNAs. The 60 and 56 kDa proteins described by Swiderski and Richter (1988) correspond to the YB proteins pp60 and pp56.

The YB proteins possess highly conserved aromatic side-chains both in the CSD (for example, the following underlined residues are invariably present in β-strands 1 and 2: ...K**W**F**N**VRNG**Y**G**F**INR...), and numerous phenylalanines and tyrosines interspersed with arginines are present in the basic/aromatic islands of the TD. That aromatic side chains can be crosslinked to RNA is confirmed by the findings of Merrill *et al.* (1988), who observed that four phenylalanines present in the hnRNA-binding protein hnRNPA1 were the sites of covalent adduct formation when hnRNPA1 was crosslinked to [^{32}P]-oligo(dT).

UV-crosslinking was performed under similar conditions to those described in Marello *et al.* (1992). **Fig. 28** is a profile of native poly(A)$^+$ mRNP which was UV-crosslinked, RNase digested and transferred to nitrocellulose for an immunoblot. Panel A was immunoblotted with anti-HTSN, and shows the generation of novel high molecular weight

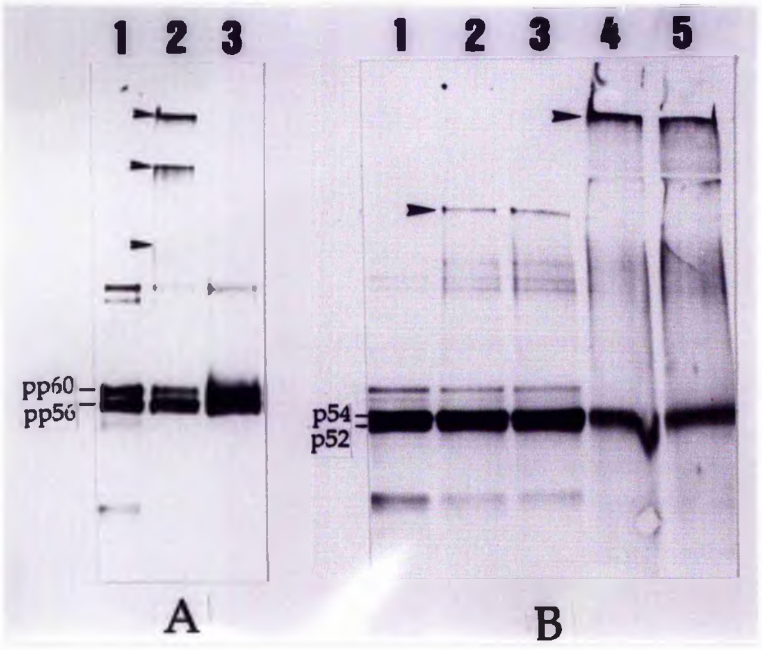


Figure 28 Immunoblots of UV-crosslinked native poly(A)$^+$ mRNP. **(A)** Anti-HTSN was diluted 1/2000. **Track 1**: 20 µg poly(A)$^+$ mRNP, no crosslinking; **2**: the same, after 30 min UV-crosslinking; **3**: material in track 2 was treated with RNase. **(B)** Anti-HTP was diluted 1/2000. **Track 1**: 20 µg poly(A)$^+$ mRNP, no crosslinking; **2**: the same, after 2 minutes crosslinking; **3**: material from track 2 was RNase treated; **4**: 20 µg poly(A)$^+$ mRNP was UV-crosslinked for 30 minutes, and **5**: material from track 4 was RNase treated. Arrows point to immunoreactive bands generated by UV-crosslinking.

complexes. These complexes were released by RNase treatment, suggesting that complex formation was achieved through multiple YB proteins crosslinking mRNA molecules. After RNase treatment, the released YB proteins in track A3 were distinctly blurred, an effect due to different lengths of residual UV-crosslinked mRNA. In panel B, the same UV-crosslinked mRNP material was immunoblotted with anti-HTP (antiserum raised against mRNP, heat-treatment pellet). After two minutes, there was evidence of complex formation (tracks B2 and B3), and a more extensive effect after 30 minutes (tracks B4+B5). In panel B, RNase treatment did not reverse the effects of UV-crosslinking, suggesting that the proteins recognized by anti-HTP are predominantly forming protein:protein crosslinks.

That the YB proteins can UV-crosslink efficiently to RNA is shown in **Fig. 29**, in which riboprobe was bound through heat-treatment, UV-crosslinked and digested with RNase. Both in the HTSN, which contains mostly YB proteins, and in the complex mixture of proteins represented in the poly(A)$^-$ fraction, the YB proteins were selectively UV-crosslinked and labelled. The radioactive labelling of the YB proteins is due to the short lengths of crosslinked riboprobe which are not wholly digested by RNase.

The ability of other, non-YB mRNP proteins to crosslink to riboprobe was considered briefly as follows. A poly(A)$^+$ mRNP HTP fraction was chosen as starting material, the removal of YB proteins being necessary because of their efficient crosslinking to riboprobe. The HTP material was redissolved in a small volume of 8 M urea. Next, it was renatured by dialysing out the urea in the presence of riboprobe. The reconstituted material was UV-crosslinked, RNase digested and run on

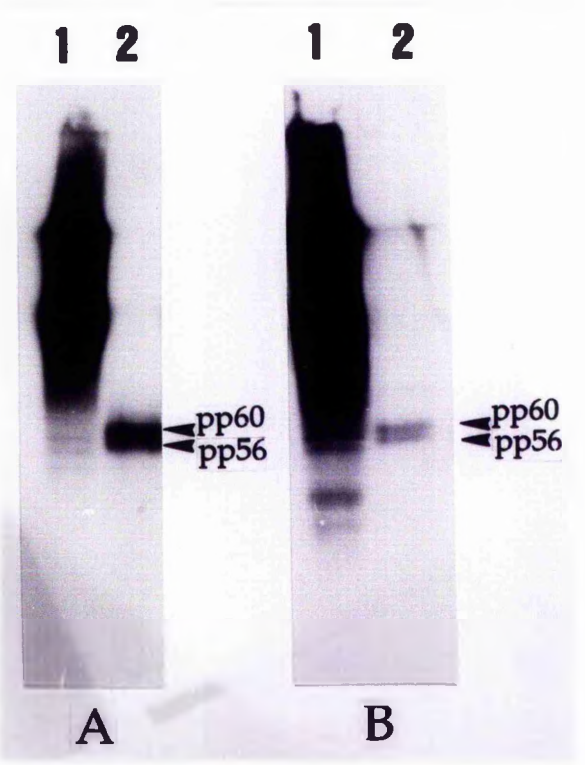


Figure 29 YB proteins present in complex mixtures can be UV-crosslinked to riboprobe. Autoradiograph of dried SDS-PAGE gel. **(A)** 40 µl of HTSN derived from 12 µg of poly(A)$^+$ mRNP was UV-crosslinked to a 16.2 sense strand riboprobe. **Track 1**: UV-crosslinked HTSN; **2**: the same, RNase treated. **(B)** 20 µl poly(A)$^-$ fraction, containing a complex mixture of proteins, was UV-crosslinked to the same riboprobe; **track 1**: undigested and **2**: RNase treated.

SDS-PAGE. In **Fig. 30**, this experiment is shown as an immunoblot (panel A) and an autoradiograph (B) of the same transfer. The starting HTP material is shown in track 1: note that there is some residual YB protein. In tracks 2 and 3, the immunoblot suggests that there is very little protein: it has been lost somewhere in the procedure, and it is possible that protein:protein crosslinking produced insoluble aggregates which were lost during dialysis. However, the autoradiograph was not blank in tracks 2 and 3, where there is a considerable radioactive signal in the loading wells, some of which was digested into protein monomers by RNase (track 3). The region where the monomers appear in track 3 corresponds to the YB proteins (pp60 and pp56) which run close to the p54/p52 group in SDS-PAGE. In contrast, the addition of excess tRNA appeared to maintain proteins in a more soluble condition after crosslinking. Most of the p54/p52 material was not crosslinked, and RNase treatment did not make much difference to the mobility of the bulk of the protein detected by anti-HTP; however, in track B5, RNase treatment released a labelled band. This labelled band was precisely aligned with p54, more specifically to the top region of the immunostained 54 kDa band. The result of this experiment was that at least another, non-YB protein of 54 kDa was UV-crosslinked to riboprobe, but only when excess tRNA maintained the material in a soluble condition. Therefore p54 can make close contacts with RNA at least *in vitro*, albeit after having been denatured. It is possible that this crosslinked 54 kDa protein may be the same protein recognized by the anti-p54 IgG, which together with the YB proteins participates in forming the masked mRNP particle (Braddock *et al.*, 1994, and section 4.2).

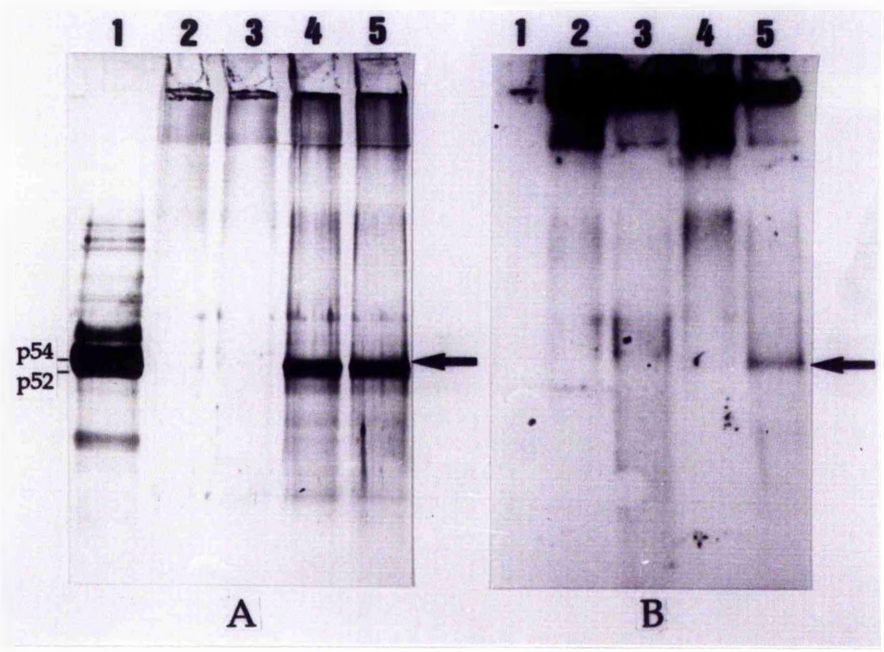


Figure 30 p54 can be UV-crosslinked to riboprobe *in vitro*. (A) Immunoblot using a 1/2000 dilution of anti-HTP and **(B)** autoradiograph of the same Western transfer. **Track 1**: Poly(A)$^+$ mRNP HTP (pellet) fraction, derived from 12 µg poly(A)+ mRNP, raised in 40 µl of 8 M urea in HTB; **2**: the same material was dialyzed into HTB without urea in the presence of ~0.5 µg 16.2 sense riboprobe, then UV-crosslinked for 30 minutes; **3**: material in track 2 was RNase treated; **4**: the same as track 2, but in addition to the riboprobe, an excess of cold tRNA (20 µg) was included; **5**: material in track 4 was RNase treated. The arrows point to a radio-labelled protein of 54 kDa which in **(B)** corresponds to the immunostained 54 kDa band in **(A)**.

4.4 RNA-binding properties of the YB proteins

In native mRNP, the YB proteins are bound to a wide range of mRNA sequences. To gain some further knowledge about the RNA-binding domains and specificities of the YB proteins, various binding competitors and different ionic environments were tested, with UV-crosslinking and heat treatment as methods of choice. Deschamps *et al.* (1991), who first described the heat treatment as a means of purifying the YB proteins, reported that after heat treatment of whole oocyte extracts, YB proteins were found associated with other cellular RNAs such as tRNA or rRNA to which they do not naturally bind *in vivo*. It is therefore conceivable that the higher levels of kinetic energy present at 80 °C allow a dynamic exchange of RNA molecules, providing a suitable context in which to introduce competitors, while assuming that the structure and RNA-binding properties of the YB proteins are regained after cooling. In the following experiments, the standard YB protein material was derived from poly(A)$^+$ mRNP HTSN. It contains YB proteins bound to a wide range of mRNA sequences present in the mRNP.

Fig. 31A illustrates data from a phenol extraction assay. This technique partitions protein:RNA complexes into the phenol phase whereas free RNA remains in the acqueous phase. The phenol extraction assay is not strictly quantitative in that only a single crosslinking event between a protein and an RNA molecule will be sufficient to bring the RNA into the phenol phase, and therefore the method does not distinguish between singly and multiply crosslinked complexes. According to the data presented in **Fig. 31A**, neither 500 mM NaCl, 4 M urea, 3 mM $MgCl_2$ nor 10 µg/ml heparin were sufficient to lower

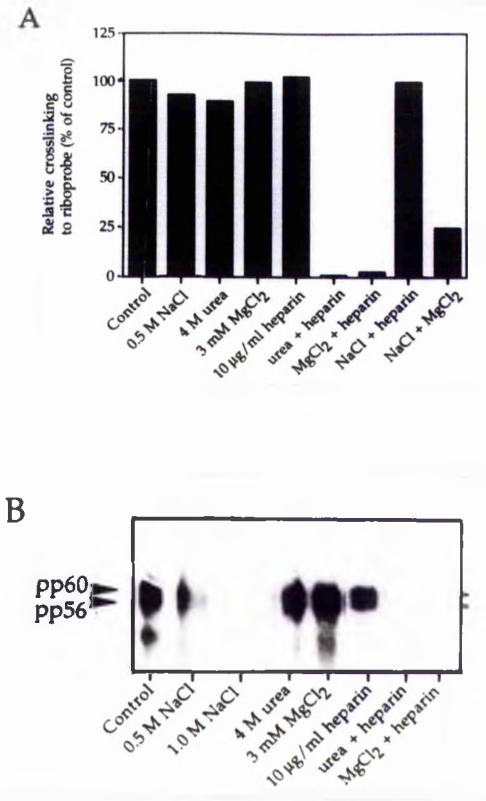


Figure 31 Effect of different agents on the UV-crosslinking of riboprobe to YB proteins. **(A)** The binding buffer (20 mM NaCl, 20 mM Tris-HCl, pH7.5) was adjusted with the agents shown prior to the binding of proteins to RNA, UV-irradiation and phenol extraction. **(B)** Confirmation of crosslinking to pp60 and pp56. In this experiment, the buffer was adjusted with the agents shown *after* the binding reaction. Next, the complexes were UV-irradiated, digested with ribonuclease and analysed by SDS-PAGE/autoradiography. Note that in **(A)**, a minimum of one crosslinking event per protein-riboprobe complex is sufficient to be recorded as maximum binding, whereas in **(B)**, the intensity of labelling is proportional to the number of crosslinking events.

crosslinking percentages; however the combinations (4 M urea + 10 μg/ml heparin) and (3 mM $MgCl_2$ + 10 μg/ml heparin) had a strong effect. The effects of the ($MgCl_2$ + heparin) combination can be related to the conditions in the riboblot in **Fig. 25, track 4**, where the binding buffer contained 2 mM $MgCl_2$ and 10 μg/ml heparin. Interestingly, a combination of 500 mM NaCl and 10 μg/ml heparin did not block binding, whereas a combination of 500 mM NaCl and 2 mM $MgCl_2$ partially blocked binding.

The visualization of RNA-binding on SDS-PAGE gives a more quantitative assessment of the amount of radioactivity UV-crosslinked to the proteins in these assays. The SDS-PAGE analysis of YB proteins crosslinked to riboprobe in **Fig. 31B** shows the effects of various competitors and ionic conditions corresponding to the histogram in **Fig. 31A**. However in the experiment shown in **Fig. 31B**, the competitors were added after heat-treatment, and therefore the results reflect the stability of the complexes formed. It is apparent that 2 M NaCl prevented binding, as was seen when with the riboblot washed with 2 M NaCl shown (**Fig. 25, track 3**). In contrast, 500 mM NaCl significantly reduced but did not totally eliminate UV-crosslinking values compared to the control, whereas in the phenol extraction assay, the value obtained for 500 mM NaCl was similar to the control value. Taken together, these observations suggest that 500 mM NaCl interferes with some but not all of the RNA-binding activity. In agreement with the phenol extraction assay, neither 4 M urea, 2 mM $MgCl_2$, nor 10 μg heparin alone had an effect, but the combinations (urea + NaCl), (urea + heparin) and ($MgCl_2$ + heparin) using the above concentrations were effective.

At this point it may be suggested that at least two distinct RNA-binding activities exist, and that each is affected differently by competitors and ionic environments. One activity is affected by NaCl and heparin, and the other by $MgCl_2$ and urea. In the phenol extraction assay, only combinations from both sets reduced crosslinking percentages, so that if only one of the RNA-binding activities was hindered, the other maintained the control values.

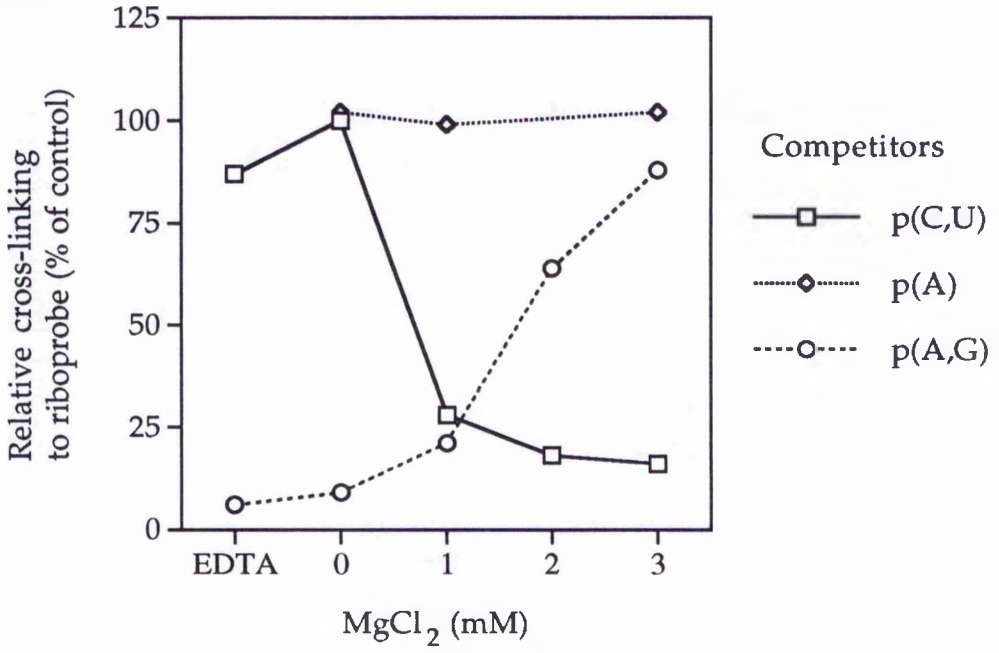
4.5 RNA-binding specificities and the effects of $MgCl_2$ on the YB proteins

Following the suggestion in Marello *et al.* (1992) that the YB proteins are competed effectively by some simple RNA sequences, notably poly(C,U) but not by poly(A) or poly(A,G), the next set of experiments considered RNA-binding specificities. In the first experiment, poly(C,U), poly(A,G) and poly(A) were used as competitors in various concentrations of $MgCl_2$. Data from phenol extraction assays is presented in **Fig. 32A**. Whereas in in the absence of $MgCl_2$ there was a clear preference for poly(A,G), and at 1 mM $MgCl_2$ an equivalent binding, at 3 mM $MgCl_2$ the situation was reversed: poly(C,U) competed but not poly(A,G) (or poly(G), not shown). Poly(A) did not compete, regardless of the presence or absence of $MgCl_2$. The competition by polypurines was unexpected given the findings in Marello *et al.* (1992). However, in Marello *et al.* (1992), the binding buffers always contained $MgCl_2$ whenever a competition by poly(C,U) was observed. It would appear that changes in the concentration of $MgCl_2$ in the range 1-3 mM alter binding specificity.

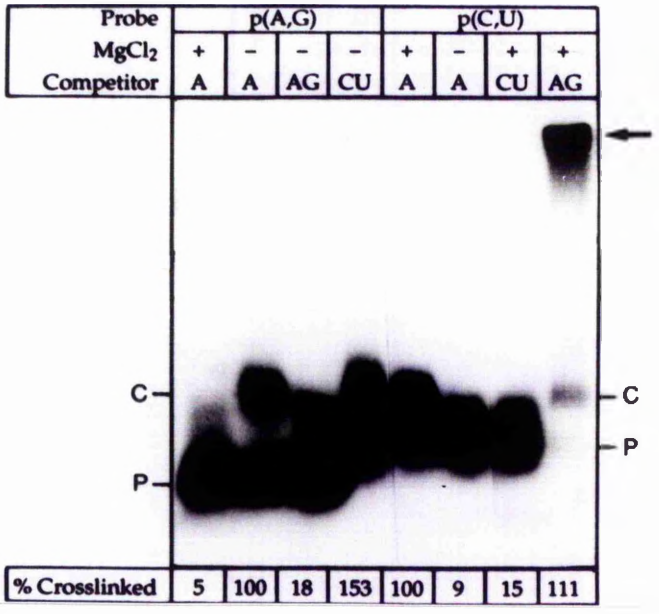
For a more direct demonstration that YB proteins can bind to both polypurines and polypyrimidines, ribopolymers were end-labelled and

Figure 32 Effects of Mg^{2+} on the interaction of YB proteins with ribopolymers. **(A)** The Mg^{2+}-induced switch in binding specificity. The YB protein-riboprobe interaction was challenged with poly(C,U), poly(A), or poly(A,G) in a 100-fold excess over riboprobe at the concentrations of $MgCl_2$ shown, or in the presence of 1 mM EDTA. **(B)** Band-shift assay showing direct binding of radiolabelled ribopolymers. The poly(A,G) probe (tracks 1-4) and the poly(C,U) probe (tracks 5-8) were bound to YB proteins in the absence (tracks 2,3,4,6) and presence (tracks 1,5,7,8) of 3 mM $MgCl_2$ and in the presence of 100-fold excess of unlabelled poly(A) (tracks 1,2,5,6), poly(A,G), (tracks 3,8) and poly(C,U) (tracks 4,7). The positions of unbound probe (P), protein-RNA complexes (C) and larger aggregates (arrow) are indicated. Confirmation that protein-RNA complexes have been formed is given by the relative crosslinking values obtained from the corresponding reactions.

A


125
100
75
50
25
0
Relative cross-linking to riboprobe (% of control)
EDTA
0
1
2
3
MgCl2 (mM)
Competitors
p(C,U)
p(A)
p(A,G)

B


Probe
p(A,G)
p(C,U)
MgCl2
+ − − − + − + +
Competitor
A A AG CU A A CU AG
C
P
C
P
% Crosslinked
5 100 18 153 100 9 15 111

used as riboprobes. Because the ribopolymers were end-labelled, UV-crosslinking and RNase treatment followed by SDS-PAGE was not feasible. As an alternative, an agarose gel mobility shift assay was used. YB proteins and riboprobes were heat-treated in the usual manner, allowed to cool, and the complexes were analysed on 1% agarose gels. Results are shown in **Fig. 32B**. Binding to poly(A,G) was observed in the absence of $MgCl_2$, and did not occur in the presence of $MgCl_2$. In contrast, $MgCl_2$ was necessary for binding to poly(C,U). The addition of self-competitors prevented binding, whereas the addition of cross-competitors enhanced binding in both cases. Moreover, it would appear that the binding of YB proteins to poly(C,U) in the presence of $MgCl_2$, favoured by the concurrent presence of poly(A,G) sequences, involves the formation of higher molecular weight complexes. The significance of this latter observation is unclear. In parallel, UV-crosslinking/phenol extraction assays gave a numerical confirmation of the observed mobility shifts.

A further confirmation of these binding specificities is presented in the SDS-PAGE analysis in **Fig. 33**. In panel A, the polynucleotide poly(A,G) competed efficiently in the absence of $MgCl_2$, whereas poly(C,U) competed best in the presence of $MgCl_2$. In panel B, $MgCl_2$ was not used; here 5 μg of heparin was unable to prevent RNA-binding. In these binding reactions, 10 μg of competitors were added to approximately 15 μg of mRNP, the mRNP containing 12 μg of protein and 3 μg of native mRNA, as well as ~0.1 μg of riboprobe. Therefore in a 100:1 excess of competitor over the riboprobe, only poly(G) competed in the absence of $MgCl_2$. In summary, competition by poly(C,U), first described in Marello *et al.* (1992), appears to be $MgCl_2$-dependent, whereas competition by poly(A,G) and more specifically poly(G)

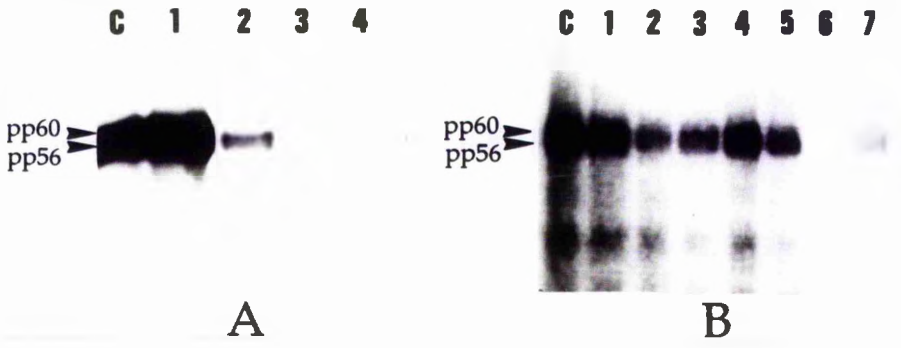


Figure 33 Binding specificities are examined on SDS-PAGE. 12 µg poly(A)$^+$ mRNP and 0.1 µg 16.2 sense strand riboprobe were used in each sample in HTB (10 mM Tris.HCl, pH 7.5, 50 mM NaCl). Riboprobe-protein complexes were formed through a cycle of heat-treatment, then UV-crosslinked and RNase treated. **(A) Track C**: control (no addition); **1**: 3 mM $MgCl_2$ was added to the binding reaction; **2**: 10 µg poly(C,U) was added; **3**: 10 µg poly(A,G); **4**: 3 mM $MgCl_2$ + 10 µg poly(C,U). **(B)** In a similar experiment, various competitors were tested in the absence of $MgCl_2$. **Track C**: control (no addition); **1**: + 0.5 µg heparin; **2**: + 5 µg heparin; **3**: + 10 µg poly(C,U); **4**: + 10 µg poly(C); **5**: + 10 µg poly(A); **6**: + 10 µg poly(G); **7**: + 10 µg tRNA.

occurs in the absence of $MgCl_2$. Poly(A), however, never competed, either in the presence or absence of $MgCl_2$ (see **Fig.s 32** and **33**). This property may be significant in that in the native mRNP, the poly(A) tail is bound by factors other than the YB proteins, such as the poly(A) binding protein (PABP).

So far, the following hypothesis can be summarised: there appear to be two distinct types of RNA-binding activity, as was suggested in section 4.4. One RNA-binding activity is sensitive to 500 mM NaCl and heparin, the sensitivity to heparin being dependent on the presence of $MgCl_2$. The other RNA-binding activity, which survives 500 mM NaCl, is sensitive to urea and $MgCl_2$. The latter activity may be responsible for binding to the polypurines poly(A,G) and poly(G) in the absence of $MgCl_2$, whereas binding to the polypyrimidine poly(C,U) is dependent on $MgCl_2$, and involves the formation of large complexes. The two RNA-binding activities can at this point be tentatively ascribed to the CSD (sensitive to urea and $MgCl_2$) and the TD (sensitive to heparin and NaCl). The formation of large complexes when binding to poly(C,U) in the presence of $MgCl_2$ is consistent with the proposed role of the TD in multimerization (Tafuri and Wolffe, 1992).

In support of these findings, a recent article has suggested that the protein unr, which consists of a five-fold repeat of the CSD without any associated TD, binds to RNA. Binding of unr to nucleic acids is sensitive to 1 mM $MgCl_2$, and is competed efficiently by poly(G) (Jacquemin-Sablon *et al.*, 1994). Secondly, the rabbit reticulocyte mRNP protein p50, now identified as a YB protein, was bound to RNA *in vitro* and was also competed most efficiently by poly(G) in a binding buffer which did not include $MgCl_2$ (Minich *et al.*, 1993). Thirdly, Murray

(1994) also differentiates between CSD and TD binding activities in the *Xenopus* oocyte YB proteins, and suggests that they have distinct properties. Fourthly, both the TD and the CSD alone have been expressed in a *Xenopus* somatic cell line after cotransfection with a CAT-reporter construct (Ranjan *et al.*, 1993), and both resulted in mRNA-masking of the CAT-reporter mRNA, implying that both these domains can bind independently to mRNA.

4.6 Ablation of the CSD binding activity

Results have hinted that two types of RNA-binding activities exist in the YB proteins. These can now be ascribed to (i) the CSD and (ii) the basic/aromatic islands in the TD. In order to uncouple the TD from the CSD activity, the CSD was cleaved with hydroxylamine (HA) in a reaction which cleaves asparagine-glycine (NG) bonds (Enfield *et al.*, 1980). There is only a single asparagine-glycine (NG) site in the oocyte YB proteins, located at the beginning of the second β-strand in the CSD, just before the RNP-1 like motif (NGYGFI). The β-barrel structure of the cold-shock domain, with its protruding aromatic and basic residues thought to interact with nucleic acids, is dependent on the integrity of all the five β-strands (see **Fig. 3**, secton 1.8). A deletion of the RNP-1 like motif in β-strand 2 of the human YB protein NSEP-1 was shown to disrupt binding to ssDNA (Kolluri *et al.*, 1992). Hence it is reasonable to suppose that cleavage at the NG site abolishes the RNA-binding activity of the CSD, and that HA-cleaved material only retains TD RNA-binding activity.

At first, the ability of HA-cleaved material to UV-crosslink to riboprobe was compared with native material using the phenol extraction assay (**Fig. 34**). Both the native and the HA-cleaved material

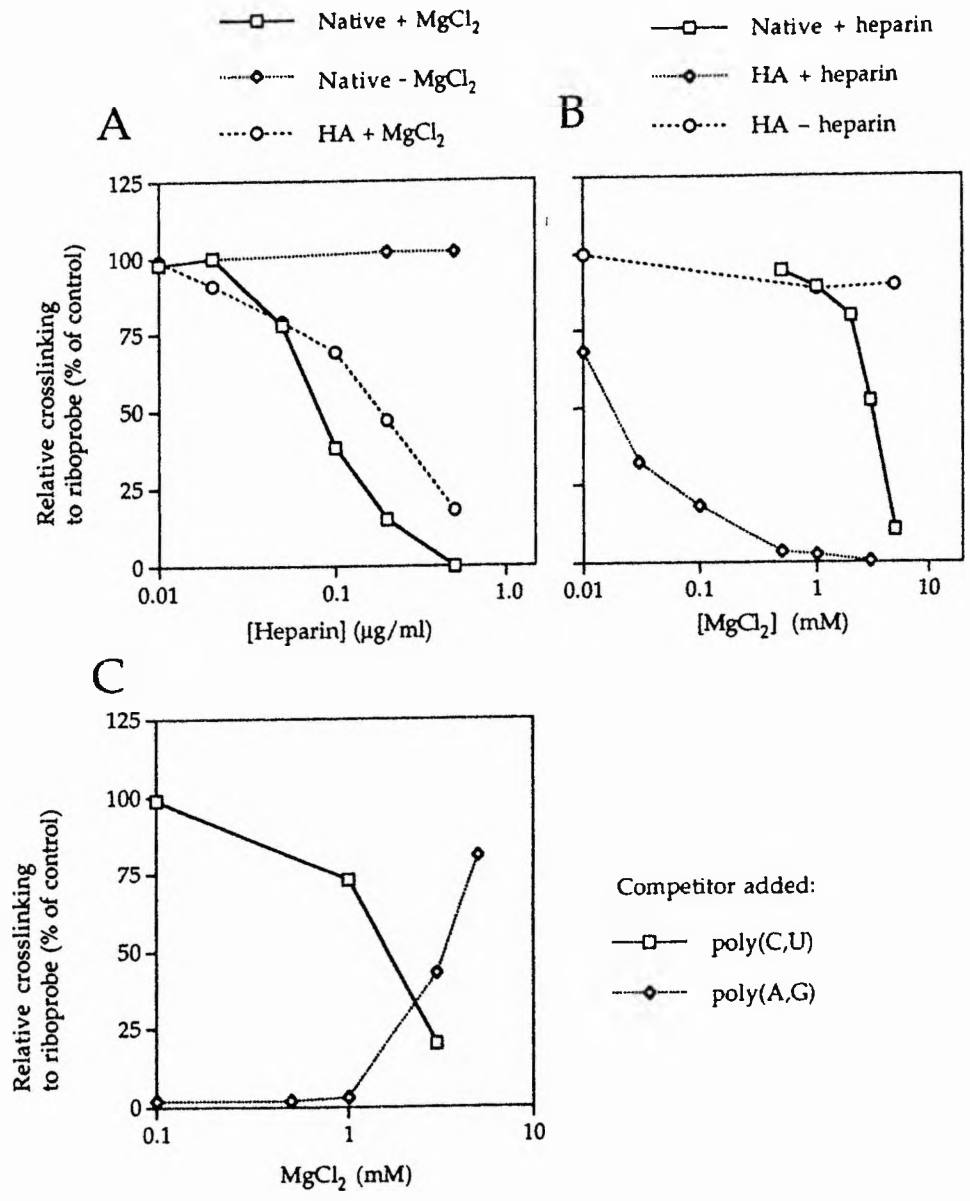


Figure 34 Effects of disrupting the CSD on the ability of YB proteins to bind to riboprobe. Binding was tested in the presence of heparin, $MgCl_2$, poly(C,U) and poly(A,G). Intact YB proteins (**Native**) and YB proteins cleaved with hydroxylamine (**HA**) were UV-crosslinked to riboprobe in the indicated conditions. **(A)** Heparin-sensitive binding activity in the presence and absence of 3 mM $MgCl_2$. **(B)** Mg^{2+}-sensitive binding activity in the presence and absence of heparin (10 µg/ml). **(C)** TD binding specificity of the HA-cleaved YB proteins is influenced by $MgCl_2$.

exhibited a similar $MgCl_2$-dependent heparin inhibition (**Fig. 34A**). Heparin was used previously to purify YB proteins, loaded onto a heparin column via heat-treatment and eluted with 1 M NaCl (**Fig. 13**), and in a riboblot to inhibit RNA-binding in the presence of $MgCl_2$ (**Fig. 25**). Because the HA-cleaved material possesses an intact TD, it should retain the ability to bind to heparin (heparin is a polyanionic molecule, and the YB protein TD contains arginine-rich stretches, see **Fig. 5**, section 1.8). Arginine-rich domains are essential components of various RNA-binding proteins; for example, the arginine-rich stretches in the HIV protein Tat are essential for binding to a stem-loop structure in HIV RNA (Green *et al.*, 1989). Arginine is preferred to lysine because it can provide more hydrogen bonds or electrostatic interactions with, for example, the negatively charged phosphate backbone of RNA. In that case, heparin might mimic the RNA's ribose-phosphate backbone. The arginine-rich stretches are also rich in aromatic residues, which could participate in binding by stacking between the bases of RNA. These aromatic residues are probably mainly responsible for UV-crosslinking to RNA. There are also numerous glutamines and asparagines in the TD, whose amide groups could also contribute hydrogen bonds. All of these postulated RNA-binding features of the TD should be maintained in the HA-cleaved material.

In **Fig. 34B**, the concentration of heparin was kept constant while the concentration of $MgCl_2$ varied. In the case of the HA-cleaved material, heparin competition was activated by very low levels of $MgCl_2$ (0.01-0.02 mM), and binding approached zero at 1 mM $MgCl_2$. In contrast, values for the native material only began to drop at 2 mM $MgCl_2$. One possible explanation for these differences is that in the intact YB proteins the CSD could still UV-crosslink to RNA in conditions where

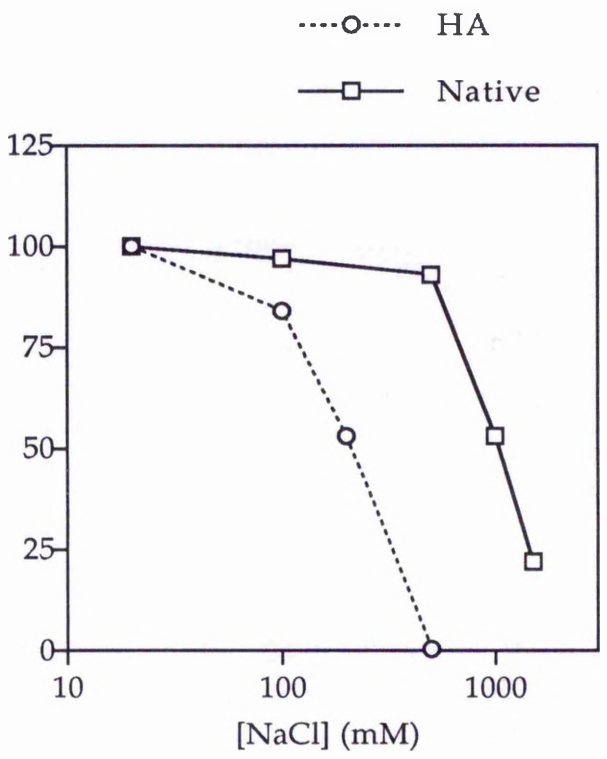
the TD was unable to, being masked by heparin. Eventually, as the concentration of $MgCl_2$ increased, Mg^{2+} also inhibited CSD binding. In 5 mM $MgCl_2$, and in the presence of heparin, there was no binding at all to the native protein. By comparison, when native YB protein was treated with only 3 mM $MgCl_2$ and no heparin, UV-crosslinking values remained high (**Fig. 31**), presumably because the TD were still active. These observations may be related to **Fig. 13** in which riboprobe was eluted in 5 mM $MgCl_2$ from a heparin column while the YB proteins remained bound. A possible interpretation is that the TD but not the CSD was bound to the heparin column, while the CSD remained free to bind to the riboprobe. Because CSD binding is sensitive to both urea and $MgCl_2$, the riboprobe was eluted in urea and $MgCl_2$. In contrast, because the TD are presumably bound to the heparin-Sepharose resin via ionic interactions mediated by the arginine-rich areas, the YB proteins were only eluted off the column in 1 M NaCl. Furthermore, the HA-cleaved material also exhibited the same shift in binding specificities depending on $MgCl_2$ concentrations (**Fig. 34C**). The binding of HA-cleaved and native YB proteins to riboprobe was challenged with NaCl (**Fig. 35**). Not surprisingly, the HA-cleaved material was much more salt-sensitive: at 500 mM NaCl there was virtually no binding, whereas the native protein survived 500 mM NaCl but not 2 M NaCl.

4.7 Further fragmentation of YB proteins

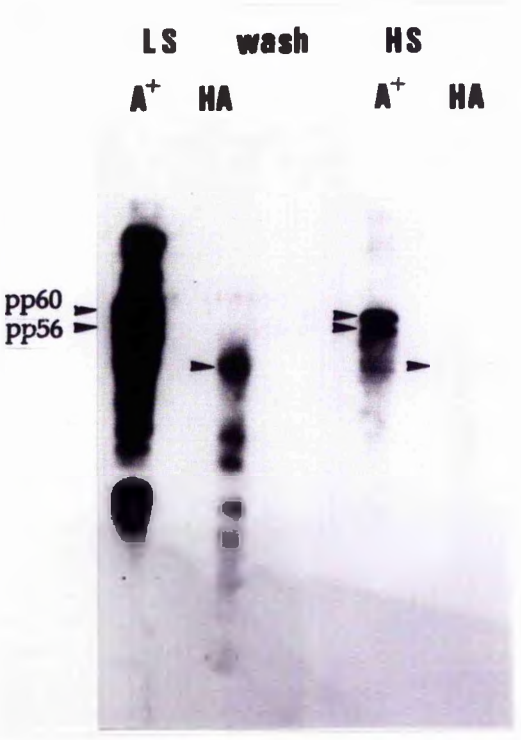
In the preceding section, HA-cleavage disrupted the CSD, and generated a protein which nonetheless possessed a functional TD. Another chemical cleavage, "mild acid cleavage" by formic acid (FA), cleaves specifically at DP (aspartate-proline) bonds. The DP sites in the YB proteins are mapped in **Fig.s 36A** and **37A**. The map shows that

Figure 35 Binding of riboprobe to HA-cleaved YB protein on a Western transfer. **(A)** Stability of binding in NaCl is measured in a filter retention assay after heat-treatment, comparing native with HA-treated YB proteins. **(B)** Riboblot comparing intact YB proteins present in total poly(A)$^+$ mRNP with HA-treated YB proteins. Binding to 0.5 μg 16.2 sense strand riboprobe was in the standard binding buffer TMNT. One part of the transfer was washed in 50 mM NaCl (**LS**) and the other in 500 mM NaCl (**HS**). **Tracks A$^+$**: 12 μg poly(A)$^+$ mRNP, arrows indicating intact YB proteins; **HA**: YB proteins previously treated with HA: the arrow indicates the major fragment of the HA-cleaved YB proteins. The RNA-binding ability of the HA-cleaved fragment, unlike the intact proteins, is salt-sensitive. Note that there is some protein breakdown in the HA-treated fraction.

A


HA
Native
125
100
75
50
25
0
10
100
1000
[NaCl] (mM)

B


LS
wash
HS
A+
HA
A+
HA
pp60
pp56

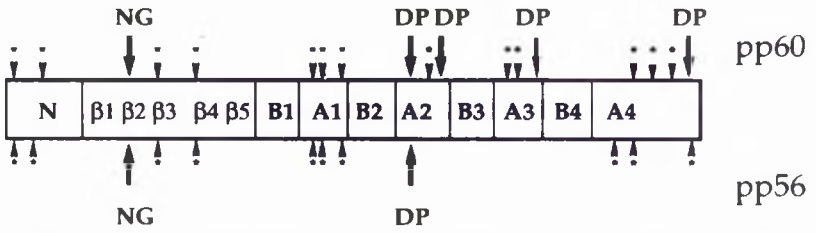
pp60 and pp56 have a common DP site in the second acidic island (A2), but pp60 has three extra DP sites: a second one in A2, and one in both A3 and A4. Hence FA-cleavage of a mixture of pp60 and pp56 will generate a mixture of fragments: a common large fragment running from the N-terminus to the first DP site in A2 (ΔA2-A4), containing an intact CSD, and a mixture of tail domain fragments: one from pp56 including basic 3 and basic 4 (B3+B4), and from pp60 there will be two smaller fragments containing only B3 and B4. In principle, cleavages can be performed *after* crosslinking to riboprobe in order to see which proteins fragments are crosslinked to the riboprobe.

The pilot experiment in **Fig. 36** shows the cleavage of proteins after they have been crosslinked to riboprobe. The intact proteins in track 1 were digested by V8 protease (tracks 2 and 3), which cleaves at aspartate or glutamate residues. Because the CSD contains numerous interspersed acidic residues, it is effectively destroyed by V8 protease. What should remain intact are the basic/aromatic islands in the TD, which were postulated in the previous section to bind and UV-crosslink to RNA. Most of the label became concentrated in small fragments below 14 kDa, which are interpreted to be the undigested basic/aromatic island fragments. In track 4, FA-cleavage generated at large fragment of 39 kDa which is interpreted as being the N-terminal ΔA2-A4 which comprises the intact CSD and the first two basic/aromatic islands up to the first DP site. ΔA2-A4 is labelled strongly. The smaller labelled material, including the smaller FA-cleavage fragments, is not very distinct, and was labelled more clearly in subsequent experiments.

Fig. 37 summarizes the UV-crosslinking/cleavage experiments. In panel B, which includes an immunoblot and a riboblot of the same

Figure 36 Proteolysis of YB proteins UV-crosslinked to riboprobe. **(A)** Diagram of pp60/pp56 showing their linear structure, consisting of: the N-terminal region (N); the CSD (β1-β5); the acidic domains (A1-A4); the basic domains (B1-B4). The position of the hydroxylamine (HA)-sensitive NG site and the positions of the formic acid (FA)-sensitive DP sites are indicated by arrows. Also shown are the potential sites of phosphorylation by the RNP-bound casein kinase II (asterisks). **(B)** Radio-labelled proteins were loaded onto a 20% acrylamide gel, favouring the separation of smaller YB protein fragments. **Track 1**: 12 μg poly(A)$^+$ mRNP, UV-crosslinked and RNase treated; low molecular weight markers are indicated; these were derived from an adjoining Coomassie stained track (not shown); **2**: radio-labelled YB proteins from track 1 were digested with protease V8 for 1 hour and **3**: digested with V8 protease for 5 hours; **4**: YB proteins were digested with FA (formic acid). The arrow points to the largest fragment of an apparent mobility of 39 kDa, which corresponds to the N-terminal fragment, ΔA2-A4: a deletion from the second acidic island onwards.

A


NG
DP DP
DP
DP
pp60
N
β1 β2 β3
β4 β5
B1
A1
B2
A2
B3
A3
B4
A4
pp56
NG
DP

B

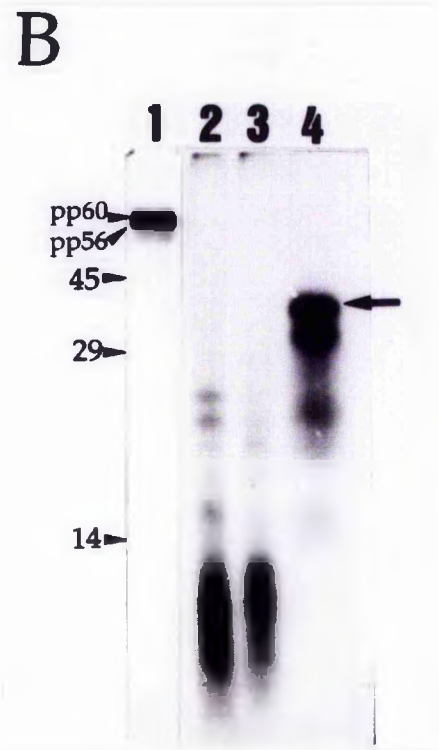
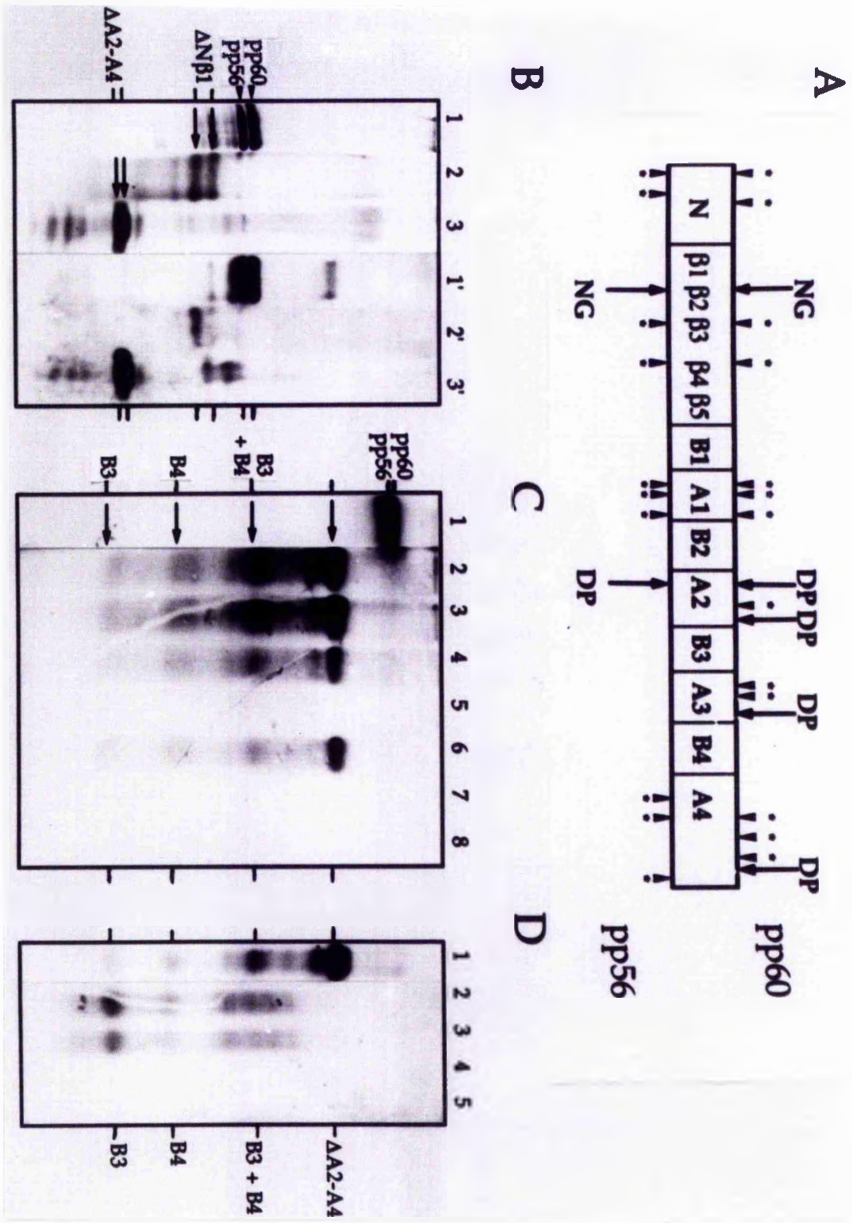

1 2 3 4
pp60
pp56
45
29
14

Figure 37 RNA binding to chemically-cleaved YB proteins. (A) The diagram of the YB proteins is shown again for clarity. **(B)** Cleaved fragments retain RNA-binding activity. YB proteins (tracks 1,1') and fragments produced by HA (tracks 2,2') and FA (tracks 3,3') treatments were separated by SDS-PAGE, transferred to nitrocellulose and either immunostained using anti-HTSN (tracks 1-3) or incubated with riboprobe in the standard binding buffer TMNT, and washed in 500 mM NaCl, to produce the autoradiograph shown (tracks 1'-3'). Positions of the major fragments C-terminal to the NG site (ΔNβ1) and N-terminal to the DP-sites in A2 (ΔA2-A4) are indicated by arrows on the immunoblot. **(C)** Points of crosslinking (contact) are established throughout much of the length of the protein. YB protein-riboprobe complexes, formed under different conditions, were crosslinked, digested with ribonuclease, cleaved with formic acid and analysed by SDS-PAGE/autoradiography. Complexes were formed from ~10 μg of protein and 0.5 μg of riboprobe in binding buffer with no addition (tracks 1,2) or with addition of: 3 mM $MgCl_2$ (track 3); 10 μg of heparin (track 4); 3 mM $MgCl_2$ plus 10 μg heparin (track 5); 20 μg of poly(C,U) in the absence (track 6) or presence (track 7) of 3 mM $MgCl_2$; 20 μg of poly(A,G) in the absence of $MgCl_2$ (track 8). The positions of: the intact proteins (pp60/pp56); the N-terminal fragments (ΔA2-A4); the C-terminal fragment from pp56 (B3+B4); the C-terminal fragments from pp60 (B3 and B4) are indicated by arrows. **(D)** Individual basic tail domains can bind RNA in the absence of a functional CSD. HA-cleaved proteins were bound to riboprobe and then crosslinked and treated with formic acid as described above. Complexes were formed in the absence (tracks 1,2) or presence (tracks 3-5) of 3 mM



$MgCl_2$ and with the addition of 2 μg heparin (track 4) or 20 μg of poly(C,U) (track 5). Fragments could be aligned with those produced by formic acid treatment alone and run on the same gel (track 1). Identity of fragments as shown for **(C)**.

Western transfer, the HA-cleaved and FA-cleaved material are compared. The intact YB proteins bound the riboprobe well after the usual 500 mM NaCl wash, but not the HA-cleaved proteins (tracks 2 and 2'), as was seen in the riboblot in **Fig. 35**. FA-cleaved proteins (tracks 3 and 3') generated the 39 kDa ΔA2-A4, seen as a doublet in the immunoblot because of minor sequence differences between pp60 and pp56. The fragment ΔA2-A4 was still bound to riboprobe after the 500 mM NaCl wash, presumably because of its intact CSD.

The autoradiographs in panels C and D were derived from denser (20%) acrylamide gels. In panel C, the YB proteins were challenged with various competitors, as in previous experiments. The findings were consistent with previous observations, in that the combinations (poly(C,U) + $MgCl_2$), (heparin + $MgCl_2$) and (poly(A,G) without $MgCl_2$) prevented protein:RNA crosslinking, but not $MgCl_2$, poly(C,U) or heparin alone. The ΔA2-A4 fragment, which contains an intact CSD, also contains the two basic islands B1 + B2, and can therefore still UV-crosslink to riboprobe in the presence of $MgCl_2$. It would be of additional interest to study the properties of a properly isolated CSD: here its binding properties can only be inferred from the data. In panel D, YB proteins UV-crosslinked to riboprobe were digested first with HA and then with FA, generating fragments which although not very sharp, can be aligned with those generated by FA-cleavage alone. The larger fragment ΔA2-A4 disappeared, as would be expected due to the cleavage of the CSD by HA. The FA/HA-cleaved material was also sensitive to ($MgCl_2$ + heparin) and ($MgCl_2$ + poly(C,U)).

In summary, chemical cleavage experiments have provided further evidence that the CSD and the tail domains have distinct RNA-binding

abilities. The Mg^{2+}-sensitive CSD RNA-binding activity is resistant to 500 mM NaCl, while the NaCl-sensitive RNA-binding activity can be attributed to the TD. In these experiments, individual basic/aromatic islands in the TD have been crosslinked to riboprobe, displaying $MgCl_2$-dependent sensitivities to poly(C,U) and heparin.

4.8 The effect of YB protein phosphorylation on RNA-binding

Previous information in the literature has promoted the idea that phosphorylation of the YB proteins in the oocytes is a pre-requisite for efficient mRNA-masking (Kick *et al.*, 1987, and the microinjection experiments of Braddock *et al.* 1994). The presence of an mRNP-associated casein kinase II-like activity was confirmed in section 3.7. The biochemical consequences of phosphorylation have not been investigated in great detail: for example, it is unclear whether or not there a direct effect on RNA-binding.

The map of the YB proteins (**Fig.s 36A and 37A**) also shows the distribution of potential casein kinase II phosphorylation sites [S/T]-X-X-[D/E], illustrating differences between the YB proteins. There are two sites in the N-terminal region before the CSD. Two phosphorylation sites are present in the CSD proper, and the remainder of the sites are in the acidic regions of the tail domains. pp60 has three extra phosphorylation sites compared to pp56, which are present in the acidic islands A2 and A3. In other words, pp60 has the potential for a more widespread phosphorylation of the TD.

A previous experiment (**Fig. 26**) has suggested that pp60 binds to riboprobe better than pp56. Might the difference might be due to the extra phosphorylation sites present in pp60? To test this hypothesis, a poly(A)$^+$ HTSN fraction was either dephosphorylated or

hyperphosphorylated, and compared to the native proteins (**Fig. 38**). Care was taken to load the same amount of protein into each track. There was a clear indication that the phosphorylated YB proteins and in particular pp60, with its three extra potential phosphorylation sites, bound the riboprobe more strongly.

The phenol extraction assay in **Fig. 39** compares untreated YB proteins with YB proteins derived from mRNP that was either untreated, treated with GTP (GTP being an acceptable phosphate donor for casein kinase II), or treated with phosphatase. **Fig. 39** suggests that the GTP-treated material was more efficiently bound to the riboprobe. At lower NaCl concentrations, phosphorylation increased binding values. As NaCl concentrations were increased, inhibiting TD contacts, binding values diminished and converged. In the absence of $MgCl_2$, binding converged to ~20%, because the CSD was presumably still active at 1 M NaCl. However in the presence of $MgCl_2$, values converged to zero in 1 M NaCl. In summary, the plots suggest that phosphorylation enhanced the stability of the protein:RNA complexes, more specifically by improving the stability of the TD-RNA interactions.

Furthermore, at the end of this project, native mRNP was incubated with [γ-^{32}P]ATP immediately after elution from an oligo(dT) column, when the protein kinase activity is at its best *in vitro*. After phosphorylation, samples of whole mRNP were UV-crosslinked, RNase digested, and analysed on SDS-PAGE (**Fig. 40**). In this experiment, not all the YB proteins were UV-crosslinked into larger complexes, as is confirmed by the Coomassie stain. Consequently some labelled monomers are apparent in the undigested tracks (no RNase treatment). More phospholabelled pp56 monomer than pp60 is

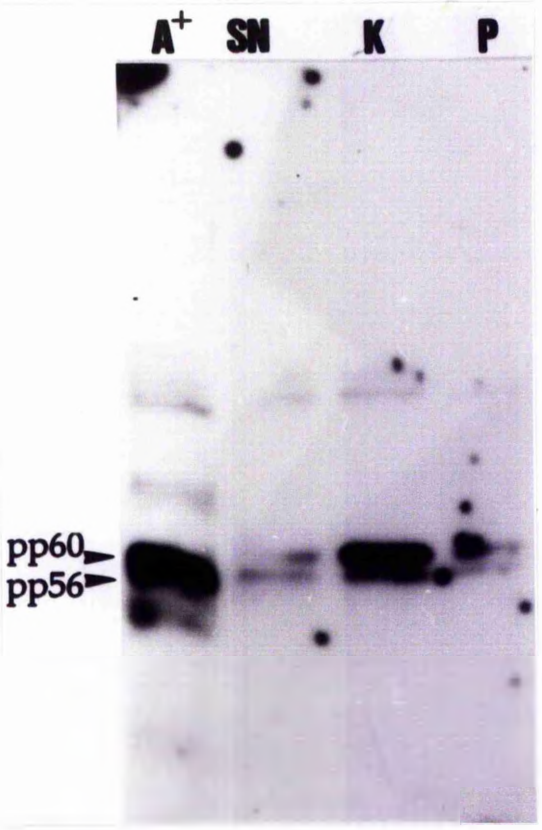


Figure 38 Binding of riboprobe to phosphorylated and dephosphorylated YB proteins on a Western transfer. Binding of 16.2 sense riboprobe was in TMNT buffer, and the wash solution contained 500 mM NaCl. Tracks: **A+**: poly(A)+ mRNP (40 µl); **SN**: poly(A)+ mRNP HTSN (heat-treatment supernatant, containing the YB proteins), 40 µl; **K**: 40 µl HTSN was treated for 30 min with 5 µl of the protein kinase activity purified earlier (section 3.7); **P**: 40 µl of HTSN was treated for 30 min with 5 units of calf intestinal phosphatase.

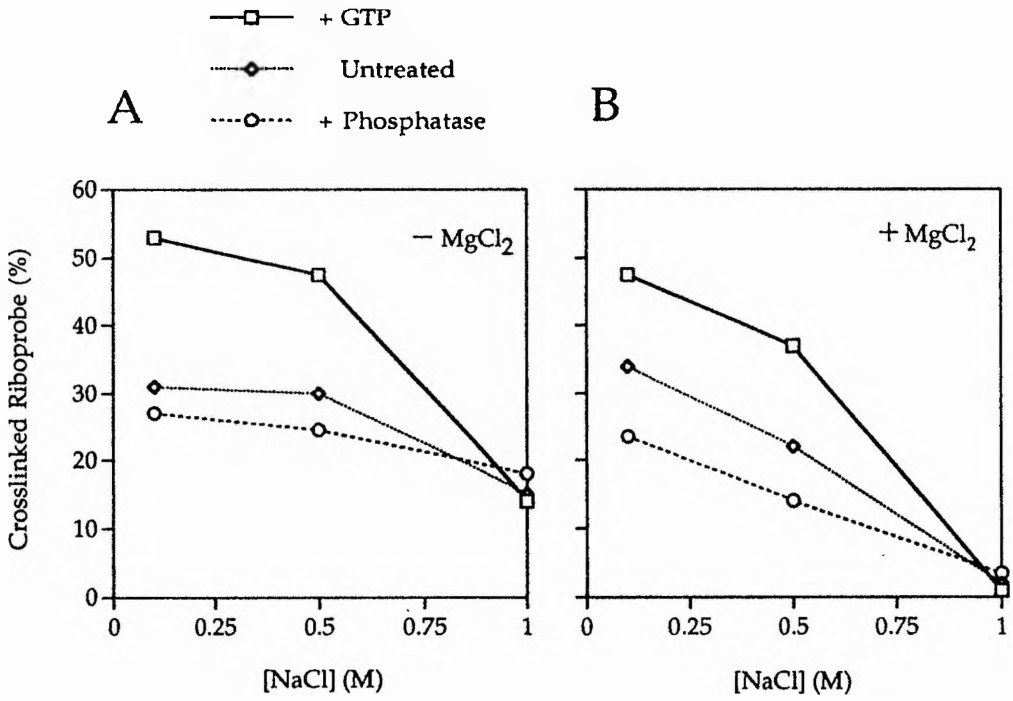


Figure 39 Effects of phosphorylation on salt stability of YB protein-riboprobe complexes. 2 µg poly(A)$^+$ mRNP was treated either with GTP or with calf intestinal phosphatase and compared with untreated material. After heat treatment, protein:riboprobe complexes were generated by UV-crosslinking in increasing concentrations of NaCl, **(A)** in the absence and **(B)** in the presence of 3 mM $MgCl_2$.

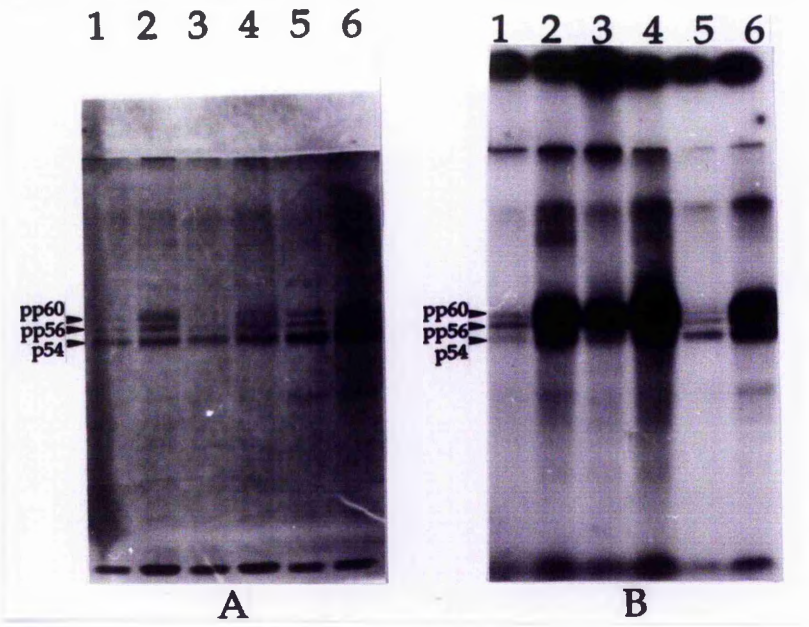


Figure 40 UV-crosslinking of *in vitro* phosphorylated poly(A)$^+$ mRNP proteins. Crosslinked material was analysed by SDS-PAGE. **(A)** A Coomassie stained dried gel and **(B)** the autoradiograph of the same gel. 12 µg poly(A)$^+$ mRNP samples adjusted to kinase buffer were supplied with ~1 µCi [γ-^{32}P]ATP before UV-crosslinking. **Track 1**: no addition; **2**: material from track 1, RNase treated; **3**: + 3 mM spermidine; **4**: material from track 3, RNase treated; **5**: + 2 mM Ap_4A; **6**: material from track 5, RNase treated. Ap_4A is a reported casein kinase inhibitor (Thoen *et al.*, 1984), but did not appear to inhibit phosphorylation significantly in this experiment.

apparent, even though pp60 has three extra phosphorylation sites in the TD. However, after RNase digestion, the situation was reversed: much more labelled monomers appeared, and there was more phospholabelled pp60 than pp56. One possible interpretation is that the hyperphosphorylated pp60 was more efficiently UV-crosslinked to mRNA, and that this greater efficiency is due to the three extra phosphorylation sites present in the acidic islands A2 and A3 of pp60.

In summary, phosphorylation appears to enhance the stability of YB protein:RNA complexes when challenged with 500 mM NaCl, as is observed both on a riboblot and in phenol extraction assays. The following observations can be grouped together: (i) riboblots suggest that pp60 binding to riboprobe is more salt resistant than pp56; (ii) pp56 is more readily stripped off native particles in the extreme ionic conditions of a Cs_2SO_4 gradient (see **Fig. 20**); (iii) phospholabelled pp60 is more efficiently UV-crosslinked to mRNA, as is suggested in **Fig. 40**, and (iv) pp60 has three more potential phosphorylation sites than pp56. The conclusion is that phosphorylation of the YB proteins enhances RNA-binding, as was suggested in Murray *et al.* (1991).

There are two possible interpretations as to how this might occur: either phosphorylation enhances the affinity for RNA directly, for instance by intramolecular conformational change, or otherwise enhances protein:protein interactions between YB proteins, which in turn may enhance the extent of RNA-binding by concentrating more YB protein on the mRNA molecules. In the TD, the phosphorylation sites are present in the acidic and not in the basic/aromatic islands which are thought to crosslink to RNA (**Fig. 36, Fig. 37**). If the function of the acidic islands is to promote protein:protein interactions between

acidic and basic islands, then by increasing electronegativity, phosphorylation would enhance such electrostatic attractions. It is not yet clear which of the two possibilities is correct.

4.9 Summary

The RNA-binding properties of mRNP were firstly explored with riboblots, for which a binding protocol was defined. Numerous proteins on Western transfers bound riboprobe in solution, including the YB proteins. *In vitro* binding of riboprobe to mRNP was inhibited by IgG derived from an anti-pp60 and an anti-p54 antisera. Of the many proteins present in clarified cell homogenates (SN10) the YB proteins, pp56 and pp60, were the ones showing most UV-crosslinking activity to riboprobe. Another mRNP protein, p54, was also UV-crosslinked to riboprobe albeit after a special denaturation/ renaturation treatment. The RNA-binding properties of the YB proteins were considered in detail, defining two separate RNA-binding activities, the CSD and the basic/aromatic islands in the TD. The TD activity, sensitive to NaCl, bound to poly(C,U) and heparin in the presence of $MgCl_2$, or to poly(A,G) in its absence. The CSD was seen to be sensitive to low concentrations of $MgCl_2$ (1-2 mM) in its binding to RNA, and competed by poly(A,G) or poly(G) but not by poly(A). Direct binding to end-labelled poly(A,G) or poly(C,U) probes was demonstrated. The YB proteins were fragmented chemically to provide further evidence for the proposed RNA-binding properties, and individual basic/aromatic islands were shown to be UV-crosslinked to RNA. Finally, the phosphorylation of YB proteins appeared to improve RNA-binding, either by enhancing protein:RNA contacts directly or by enhancing protein:protein interactions.

Chapter 5

Characterization of Additional Unidentified Proteins

5.0 Aims

The preceding section concentrated on the biochemical properties of the identified YB proteins, which are undoubtedly a crucial component of the *Xenopus* oocyte mRNP particle. However, there is very little knowledge about the identity and function of, for example, the mRNP proteins of 100, 54, 52 and 40 kDa. It is clear from Coomassie stains and immunoblots of whole mRNP, and from separations of mRNP particles on density gradients (see chapter 3), that these other abundant proteins are an integral part of the mRNP structure. However, surprisingly little is known about them. Some of these mRNP proteins may be specific to *Xenopus* oocytes or germ cell mRNPs in general, serving some specialized functions in these contexts. Alternatively, some of the proteins may be also found in somatic mRNP particles. In order to gain some fresh information about these unidentified mRNP components, the general approach was to clone cDNAs encoding mRNP proteins from expression libraries using either (i) a riboprobe ligand, or (ii) an immunoscreen using the antisera described in chapter 3. Also, for a biochemical analysis, another strategy was to label mRNP proteins with ^{125}I, cleave them with formic acid (FA) or cyanogen bromide (CNBr), and to sequence recovered peptides.

5.1 RNA-binding screen of a bacteriophage λ cDNA expression library

The general principle of this approach is to express fusion proteins from a cDNA expression library, and to screen for a biochemical or antigenic properties. The first cDNA library, prepared from mRNAs obtained from a mature ovary, and cloned into the bacteriophage vector λgt11, was kindly provided by Dr. Mark Dworkin (Boehringer, Vienna).

Because the mRNP proteins are very abundant throughout oogenesis, the selection of appropriate cDNA clones should be relatively simple. However, potential limitations should be mentioned. Firstly, because the Dworkin library was constructed from a mature ovary in which more developed oocytes are most numerous, the bulk of the cDNAs represents stored, masked maternal mRNAs, which are much more abundant than the expressed mRNAs being sought in this screen. Secondly, the λgt11 system is not directional: that is, 50% of the cDNAs are likely to be inserted into the wrong orientation, and only one sixth of the inserted cDNAs will be in frame. Thirdly, of these cDNAs, many will be incomplete, since the reverse transcriptase tends to dissociate from the mRNA template before reaching the 5' end: in general there tends to be a bias in cDNA libraries towards the 3' ends of messages. Fourthly, it cannot be taken for granted that the prokaryotic environment guarantees the appropriate polypeptide folding or post-translational modifications, such as phosphorylation, that may be essential for mRNP protein structure. With these caveats in mind, it was deemed appropriate to screen several large plates, each displaying 20-50,000 plaques, in order to generate enough positives.

Following the procedure described by Vinson *et al.* (1988), in which a labelled DNA ligand was used to screen for the DNA-binding protein C/EBP, a labelled RNA ligand was used instead to screen for RNA-binding proteins. The sense strand of the cyclin B1 message was chosen as a riboprobe. Cyclin B1 is a known stored maternal mRNA, and is therefore an appropriate substrate for the mRNP proteins. At first, the procedure closely followed that of Vinson *et al.* (1988), starting with the denaturation of the transferred proteins with 6 M guanidine-HCl. The purpose of the denaturation step is to dissociate any endogenously bound RNA, so that

the fusion proteins are ready to receive the riboprobe in an appropriate binding buffer (in Vinson *et al.*, 1998, 25 mM Tris.HCl pH 7.5; 25 mM NaCl; 5 mM $MgCl_2$; 0.5 mM DTT). After adding the riboprobe, the guanidine-HCl was diluted by stepwise addition of buffer.

With this first approach clone G2 was isolated. G2 consisted of an incomplete cDNA encoding two zinc-fingers of the 4 x cysteine variety. The clone was investigated further by an honour's student who sequenced the rest of the protein (Rachael Marshall, 1994). In addition to the zinc-fingers, the whole protein, subsequently termed C4SR, was seen to have a putative ATP-binding fold between the zinc-fingers, followed by a highly acidic stretch in the central part of the protein and an SR (serine-arginine rich) region in the C-terminal end. The protein is listed and described in greater detail in the Appendices C and D. At this stage its function is unknown, although it may be inferred to have a role in nucleic acid regulation either at the level of DNA or RNA, given the similarity of its features with various nuclear proteins.

After clone G2, the screening procedure was changed as follows: 8 M urea was used as a denaturing agent, and the binding buffer was simplified to (10 mM Tris.HCl (pH7.5), 50 mM NaCl, ie HTB). **Fig. 41** shows a typical primary screen and a final purification, emphasizing the need to wash the filter with an appropriate salt concentration after riboprobe binding. At the lower concentration of 50 mM NaCl, it is likely that basic residues present in many proteins bind non-specifically to the negatively charged ribose-phosphate backbone of the riboprobe. 500 mM NaCl was chosen because it reduced background without totally disrupting RNA:protein binding. This procedure paralleled the conditions used in the riboblots in chapter 4. Using this approach, the estimated frequency of positives in the Dworkin library was approximately one in every 5000 plaques.

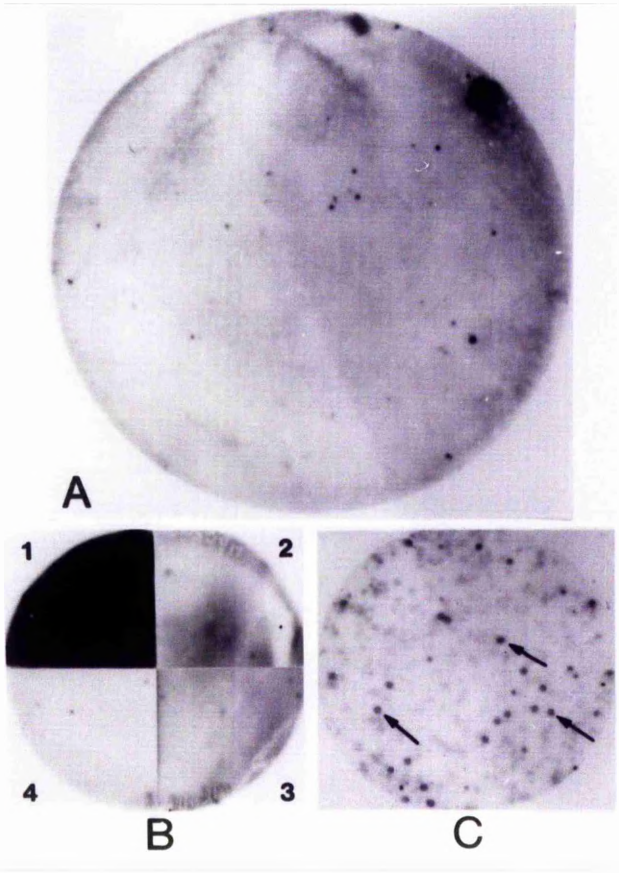


Figure 41 Screening a cDNA expression library with a riboprobe. λgt11 plaque lifts were denatured in 8 M urea before incubation with cyclin B1 sense strand riboprobe in the binding buffer (10 mM Tris.HCl pH 7.5; 50 mM NaCl). **(A)** Primary screen of ~50,000 plaques from the λgt11 library. **(B)** Effects of salt washes on background signals: **1**: 50 mM NaCl; **2**: 500 mM NaCl; **3**: 1 M NaCl; **4**: 2 M NaCl. **(C)** Final purification of clone N5 showing isolated positive plaques (arrows).

Several clones were isolated, two of which were identified unambiguosly. One of these was M7: like G2, it is another putative zinc-finger protein, identified as the *Xenopus* homologue of the rat ribosomal protein S27 (Chan *et al.*, 1993). Its amino-acid sequence is also listed and described in the Appendices E and F. Its binding to the cyclin B1 riboprobe may or may not be significant, since the function of S27 within the ribosome is unknown: although ribosomal protein S27 is not an mRNP protein, it could still bind to mRNA. For example, the ribosomal protein S6 has recently been suggested to interact, possibily directly, with the short polypyrimidine stretches present in the 5' UTRs of mRNAs encoding ribosomal proteins and translation factors: these are coordinately translated (Jefferies *et al.*, 1994).

The other identified cDNA was clone N5. It is identical to the cDNA encoding the somatic YB protein FRGY1, and includes the entire open reading frame. FRGY1 was first cloned, together with the oocyte YB protein FRGY2 (pp60), by Tafuri and Wolffe (1990). FRGY1 is transcribed but not translated in oocytes and therefore belongs to the pool of masked maternal mRNAs. Consequently it is likely that FRGY1 mRNA is itself masked by the oocyte YB proteins. It is now thought that YB proteins of the FRGY1 type also bind to mRNAs *in vivo*, and that this is a potential property of all YB proteins (Evdokimova *et al.*, 1995).

Two of the cDNAs described in this section, G2 (encoding the two zinc fingers of the novel oocyte protein C4SR) and N5 (the somatic YB protein FRGY1), were subcloned into the expression vector pMalcRI. When induced by IPTG, *E. coli* cells produced a fusion protein which included the 42 kDa maltose-binding protein (MBP), followed by the polypeptide encoded by the cDNA. The presence of the MBP potentially allows the

fusion protein to be purified via affinity to an amylose resin. The expected molecular weights of the fusion proteins were as follows: in the case of G2, the fusion protein comprised 42 kDa (the MBP) + 15 kDa (estimated molecular weight of the two zinc-fingers encoded by clone G2) and an additional 10 kDa from the β-galactosidase lacZ fragment in the pMalcRI vector (G2 remained in frame with the vector at the 3' end). This suggested a total of 67 kDa for the MBP-G2-lacZ fusion. In the case of N5, the fusion protein should consist of 42 kDa (MBP) + 35kDa (FRGY1) = 77 kDa. However, like its oocyte counterparts, FRGY1 has a higher observed mobility, in its case 50 kDa (Tafuri and Wolffe, 1992), therefore the apparent molecular weight of the MBP-N5 fusion could be as high as 92 kDa. In **Fig. 42**, both these fusion proteins were expressed using pMalcRI. Induced protein bands can be seen in total bacterial extracts, with apparent mobilities of 74 kDa for the G2 and 98 kDa for the N5 fusion proteins. These values are not too different from the predicted mobilities. The same samples were tested for RNA-binding on a riboblot. The MBP-N5 (FRGY1) fusion protein bound to riboprobe in these conditions, mirroring the properties of its oocyte equivalents, pp60 and pp56 in similar assays, but the MBP-G2-lacZ fusion did not bind the riboprobe.

5.2 Preparation of a new cDNA library and immunoscreening

The approach outlined in the previous section has the potential to isolate cDNAs encoding mRNA-binding proteins. However, other RNA-binding proteins such as the ribosomal proteins can also, in principle, be selected. More specific proteins could be targeted, for example, by modifying the binding buffer conditions, or by choosing specific RNA ligands. However, it was deemed necessary to generate a new, more focussed cDNA expression library more appropriate for the isolation of mRNAs expressed in early oocytes. The mRNAs encoding the mRNP proteins are translated

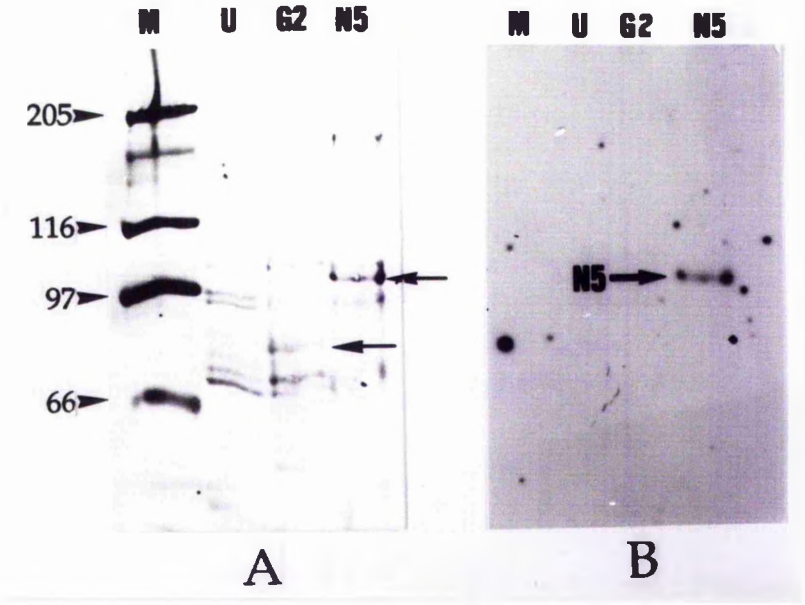


Figure 42 Fusion proteins expressed in pMALcRI. **(A)** Amido-black stain of a Western transfer. **M**: High molecular weight markers; **U**: total lysate from uninduced *E. coli* TB-1; **G2**: total lysate from cells containing pMalcRI-G2 fusion, induced with IPTG for three hours; **N5**: total lysate from cells containing pMalcRI-N5 fusion, induced in IPTG for three hours. Arrows indicate induced bands (MBP-G2-lacZ and MBP-N5). **(B)** Binding of β-tubulin 16.2 sense strand riboprobe to the same Western transfer shown in **(A)**. The RNA-binding assay preceded the amido black staining. The induced fusion proteins are indicated.

in early oocytes to generate the material necessary to package the large numbers of stored maternal mRNAs. Therefore mRNAs encoding mRNP proteins should be well represented in polysomal (translating) mRNPs of early oocytes. To that effect, a total extract from early oocytes was run on a glycerol gradient in the presence of 10 μg/ml cycloheximide, DTT, and RNase inhibitor, conditions which preserve polysomes. Polysomal mRNA was extracted and used as a template to synthesize cDNA. The cDNA was cloned into a λZAP bacteriophage vector (Stratagene), a system which has the advantage of allowing directional cloning: all cDNAs are inserted in the correct orientation. The λZAP system also allows the direct excision of pBS plasmids containing the cloned cDNAs, bypassing the need for bacteriophage λ DNA preparations and subsequent subcloning steps. **Fig. 43** shows the labelled cDNAs separated on a denaturing agarose gel, ranging in size between 500 and 2000 base pairs. (Library construction was carried out by Dr. J. Sommerville).

Having already tried riboprobe ligands with the Dworkin library, the new library was immunoscreened with anti-HTSN and anti-HTP, in the hope of detecting immunoreactive epitopes on bacterially expressed fusion proteins. Based on the Western blots discussed in chapter 3, anti-HTSN should recognize the YB proteins as well as the unidentified phosphoproteins pp96 and pp100, while anti-HTP should recognize other mRNP proteins and especially p54 and p52. Antisera were diluted 1/500, and pre-incubated with a 1/250 dilution of total bacterial lysate in order to reduce background binding to bacterial proteins. In addition, blocking the plaque lifts overnight in 5-10% skimmed milk in TBST, effectively reduced background to zero.

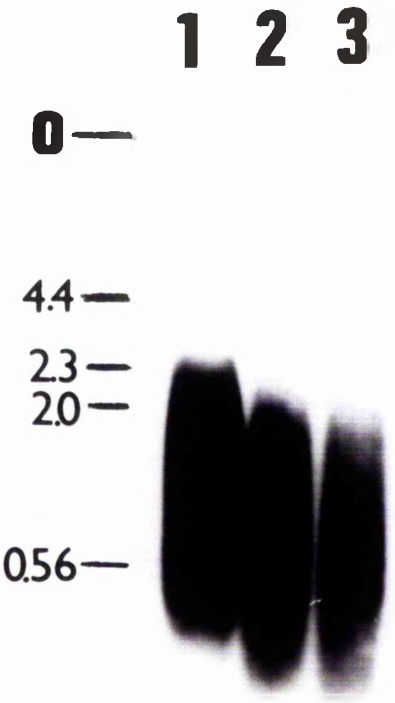


Figure 43 Preparation of a new cDNA expression library in λZAP (Stratagene). cDNA was synthesized from mRNAs extracted from a polysomal gradient fraction from previtellogenic oocytes. After second-strand synthesis in the presence of [α-^{32}P]dCTP, the cDNA was size-fractionated on a Sephacryl column and samples were analysed on an alkaline 1% agarose gel. The autoradiograph of three fractions is shown (tracks **1-3**). Fraction **1** was selected for cloning. The positions of marker DNA bands (Hind III digests of lambda DNA) run on the same gel is shown (kb). Courtesy of Dr. J. Sommerville.

The anti-HTSN antiserum yielded a higher number of positive clones. **Fig. 44A** shows the final purification of one of these, AB12. This cDNA encodes nucleolin, previously cloned by Caizergues-Ferrer *et al.* (1989). Further cDNAs, AB1A and AB1C, were selected with anti-HTSN and were sequenced: they also encoded nucleolin. Nucleolin is a multifunctional protein with roles in ribosomal RNA transcription and processing. It possesses RRM motifs (RNA Recognition Motifs) as well as acidic stretches, and in *Xenopus* oocytes is reported to be present in two forms of 96 and 100 kDa (Caizergues-Ferrer *et al.*, 1989).

It is not clear why nucleolin should have been selected. One possibility is that the polyclonal anti-HTSN cross-reacts with nucleolin. The YB proteins have certain similar structural properties which may explain cross-reactivity: for example, the presence of an RNP-1 like motif (GYGFI) in the CSD, the RNP-1 motif being part of the RRMs present in nucleolin, and the presence of highly acidic areas with casein kinase II phosphorylation sites in the TD. Casein kinase II phosphorylation sites are also present in nucleolin, a protein which is known to be one of the targets of casein kinase II *in vivo* (Csermely *et al.*, 1993; Suzuki *et al.*, 1992). A second possibility is that the poly(A)$^+$ mRNP preparations are contaminated with nucleolin, which then fractionated into the HTSN, the untested assumption being that nucleolin is heat-stable. The mobilities on SDS-PAGE of the HTSN phosphoproteins pp96 and pp100 are consistent with those reported for the *Xenopus* oocyte nucleolin forms. However, there is no current knowledge of an *in vivo* association of nucleolin with mRNP, and its presence in the mRNP preparations could be an artefact.

The anti-HTP antiserum screen yielded fewer positives, and only one clone, AB21, was purified (see **Fig. 44B**). AB21 mRNA is highly expressed

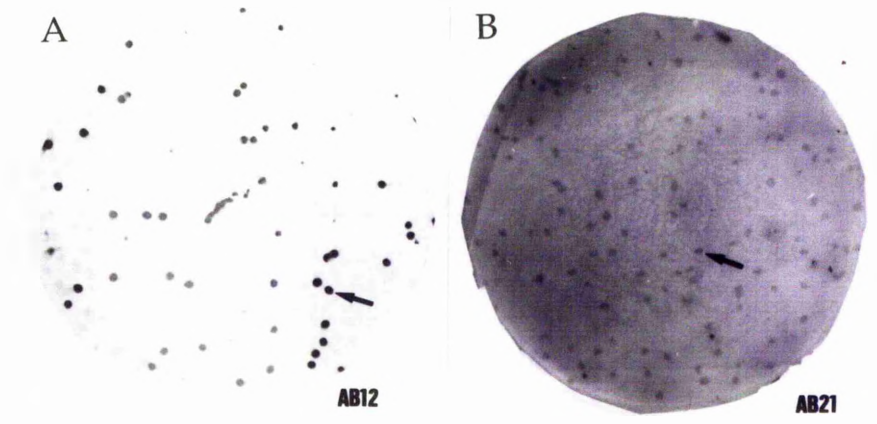


Figure 44 Screening of the λZAP cDNA library using the ELISA assay. **(A)** Tertiary screen of clone AB12, immunoselected with a 1/500 dilution of anti-HTSN. **(B)** Tertiary screen of clone AB21, immunoselected with a 1/500 dilution of anti-HTP. Arrows indicate immunoreactive plaques. Negative plaques appear in a lighter shade. Note that the positive signal in (A) is distinctly clearer than in (B) compared to the background.

in oocyte stages I and II, an expression pattern shared by pp60, which is consistent with its encoding a protein needed in early oocytes (**Fig. 45**). In stage I oocytes there was a single transcript of 2.4 kb, whereas in stages II and III, there was an additional transcript of 2.7 kb. The significance of these two detected transcripts is not known. AB21 mRNA peaked in stage II oocytes, and was not detected in somatic tissues or testis, suggesting that it encodes an oocyte-specific protein. As a control, the ribosomal protein S27 mRNA (cloned and described in section 5.1) gave a positive signal in all tracks of this same tissue Northern blot (not shown).

Most of the AB21 cDNA has been sequenced from subclones based on the available AB21 restriction sites, including the whole open reading frame (**Fig. 45**). The DNA sequence, now entered into the database, is listed in Appendix G. The first AUG in the 5' end of AB21 does not necessarily encode the first methionine in the AB21 protein. However, the known translation start site for *Xenopus* cyclin B1 is similar to the first AUG site present in the AB21 cDNA: **GAGAAA*AUG*** for cyclin B1, and **AGGAAA*AUG*** in AB21. Therefore it is possible that clone AB21 encodes the whole protein.

The AB21 protein is strikingly similar to a "global gene regulator" encoded by the gene *RPD3* from *Saccharomyces cerevisiae.* (Vidal and Gaber, 1991), and a *Caenorhabditis elegans* protein encoded by an ORF identified through genomic sequencing (database entry CEC08B11/Z46676). The amino-acid sequence of AB21 is shown in **Fig. 46**, and the sequence alignment between AB21, yeast RPD3 and the *C. elegans* homologue is shown in **Fig. 47**. The estimated molecular weight of AB21 is 54.7 kDa, which is consistent with its being one of the proteins recognised by anti-HTP.

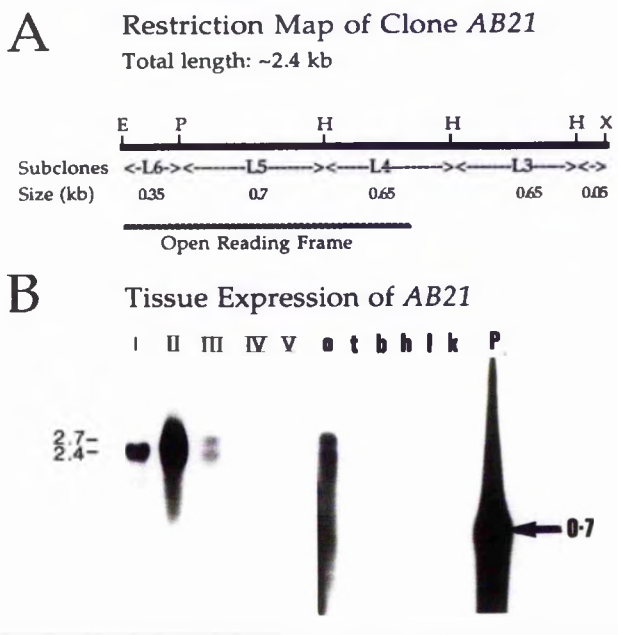


Figure 45 Tissue expression and restriction map of clone AB21. (A) Restriction map of clone *AB21* indicating four subclones, L6, L5, L4 and L3, and the ORF (open reading frame). **E**: EcoRI; **P**: PstI; **H**: Hind III; **X**: XhoI. All cDNAs in the λZAP library are cloned directionally so that the sense strand runs 5' to 3' from an EcoRI site to a XhoI site. The XhoI site follows the poly(A) tail on the cDNA. **(B)** Tissue expression of AB21 mRNA. RNA was extracted from 25 oocytes of Dumont stages I-V, and a Northern blot was probed with an antisense riboprobe generated from subclone L5. Each track has an RNA equivalent of 5 oocytes. On a separate Nothern blot, the same antisense probe was used with RNA extracted from various tissues: **o**: ovary; **t**: testis; **b**: brain; **h**: heart; **l**: liver; **k**: kidney; **P**: L5 antisense probe, 0.7 kb. Each track contains ~ 10 µg of RNA. AB21 mRNA appears in two forms in oocyte stages II and III, of 2.4 kb and 2.7 kb, but only 2.4 kb in stage I. It peaks in stage II oocytes and appears to be oocyte-specific in its expression.

AB21 protein

```
                                                                    50
MALTLGTKKKVCYYYDGDVGNYYYGQGHPMKPHRIRMTHNLLLNYGLYRK

                                                                   100
MEIFRPHKASAEDMTKYHSDDYIKFLRSIRPDNMSEYSKQMQRFNVGEDC

                                                                   150
PVFDGLFEFCQLSAGGSVASAVKLNKQQTDISVNWSGGLHHAKKSEASGF

                                                                   200
CYVNDIVLAILELLKYHQRVVYIDIDIHHGDGVEEAFYTTDRVMTVSFHK

                                                                   250
YGEYFPGTGDLRDIGAGKGKYYAVNYALRDGIDDESYEAIFKPVMSKVME

                                                                   300
MFQPSAVVLQCGADSLSGDRLGCFNLTIKGHAKCVEFIKTFNLPLLMLGG

                                                                   350
GGYTIRNVARCWTYETAVALDSEIPNELPYNDYFEYFGPDFKLHISPSNM

                                                                   400
TNQNTNEYLEKIKQRLFENLRMLPHAPGVQMQAVAEDSIHDDSGEEDEDD

                                                          _________450
PDKRISIRSSDKRIACDEEFSDSEDEGEGGRKNVANFKKVKRVKTEEEKE

possible α-helical tail _____ 480
GEDKKDVKEEEKAKDEKTDSKRVKEETKSV
```

Figure 46 The polypeptide encoded by the AB21 open reading frame. The first methionine in the AB21 open reading frame is assumed to be the first amino-acid. The total length of the polypeptide is 480 amino-acids with an estimated molecular weight of 54.7 kDa. A region predicted to form an α-helix (refer to Fig. 48) is indicated. Tyrosines are in bold; methionines are in a larger font: their presence suggested the experiment shown in Fig. 50. The amino-acid content is as follows: S+T 10.2% (49 out of 480); D+E 16.7% (80); K+R+H 16.5% (79) and F+Y+W 10.4% (50). The estimated isoelectric point is 5.58. Data was obtained using the GCG package, Genetics Computer Group Inc., version 7.2 (1992).

Figure 47 Alignment between the AB21 polypeptide and two homologous proteins. The homologous proteins are RPD3 from *Saccharomyces cerevisiae* (Vidal and Gaber, 1991) and a *Caenorhabditis elegans* homologue obtained from a cosmid genomic clone, database accession numbers CEC08B11/Z46676. Vertical bars denote identities; semicolons denote conservative substitions. The following substitutions are described as conserved: {A,G,S}, {S,T}, {D,E}, {N,Q}, {H,K,R}, {V,L,I,M) and {W,F,Y}. The lengths and estimated molecular weights of each protein are included. The alignment was obtained using the program PILEUP, GCG package, Genetics Computer Group Inc., version 7.2, 1992.

```
      1                                                          50
AB21  .......... .......... MALTLGTKKK VCYYYDGDVG NYYYGQGHPM
                            | |     |:: | |:||:||| ||||| ||||
RPD3  .......... MVYEATPFDP ITVKPSDKRR VAYFYDADVG NYAYGAGHPM
                  :  :              ||  |||:|  |||  : ||  |||
C.el  MSSDKFKLDT LFDDNDEIIE PDGADVKKRN VAYYYHKDVG HFHYGQLHPM

      51                                                         100
AB21  KPHRIRMTHN LLLNYGLYRK MEIFRPHKAS AEDMTKYHSD DYIKFLRSIR
      ||||||| |  |::|||||:| |||:| : |:   :|  :|:| :|| ||  :
RPD3  KPHRIRMAHS LIMNYGLYKK MEIYRAKPAT KQEMCQFHTD EYIDFLSRVT
      || |: :    |:: | : |  | :  :       ::  |||: :|: ||  ||
C.el  KPQRLVVCND LVVSYEMPKY MTVVESPKLD AADISVFHTE DYVNFLQTVT

      101                                                        150
AB21  PD-NMSEYSK QMQRFNVGED CPVFDGLFEF CQLSAGGSVA SAVKLNKQQT
      || |:  : :    :||||:| |||||||:|: | :|:|||:   | :||:
RPD3  PD-NLEMFKR ESVKFNVGDD CPVFDGLYEY CSISGGGSME GAARLNRGKC
      |   | |        ||:|:| ||:| ||::| |:: :|||:| || |||:
C.el  PKLGLTMPDD VLRQFNIGED CPIFAGLWDY CTLYAGGSVE GARRLNHKMN

      151                                                        200
AB21  DISVNWSGGL HHAKKSEASG FCYVNDIVLA ILELLKYHQR VVYIDIDIHH
      |::||::||| |||||||||| |||:|||||: |:|||:|| | |:||||||||
RPD3  DVAVNYAGGL HHAKKSEASG FCYLNDIVLG IIELLRYHPR VLYIDIDVHH
      |: :|: ||| |||||||||| |||:|||||| |:|||:|| | ||||||||||
C.el  DIVINWPGGL HHAKKSEASG FCYVNDIVLG ILELLKYHKR VLYIDIDIHH
```

```
      201                                                        250
AB21  GDGVEEAFYT TDRVMTVSFH KYGEYFPGTG DLRDIGAGKG KYYAVNYALR
      |||||||||| |||||| ||| ||||:||||| :|||||:| | | ||||  ||
RPD3  GDGVEEAFYT TDRVMTCSFH KYGEFFPGTG ELRDIGVGAG KNYAVNVPLR
      |||| |||   :||||| ||| ::| :|||:| : | ||| | | :|:||||
C.el  GDGVQEAFNN SDRVMTVSFH RFGQYFPGSG SIMDKGVGPG KYFAINVPLM

      251                                                        300
AB21  DGIDDESYEA IFKPVMSKVM EMFQPSAVVL QCGADSLSGD RLGCFNLTIK
      ||||| :| : :| ||: |:| | :||||||| |||:|||||| |||||||::
RPD3  DGIDDATYRS VFEPVIKKIM EWYQPSAVVL QCGGDSLSGD RLGCFNLSME
       :| |  |   :|| ||  :  | ::| |:|| |||:|||  | ||| | ||
C.el  AAIRDEPYLK LFESVISGVE ENFNPEAIVL QCGSDSLCED RLGQFALSFN

      301                                                        350
AB21  GHAKCVEFIK TFNLPLLMLG GGGYTIRNVA RCWTYETAVA LDSEIPNELP
      ||| || ::| :| :|::::| |||||:|||| | | :||::      :  :||
RPD3  GHANCVNYVK SFGIPMMVVG GGGYTMRNVA RTWCFETGLL NNVVLDKDLP
      :||  | ||| | | |:||:| |||||:|||| | |  |||::   : :| ::|
C.el  AHARAVKYVK SLGKPLMVLG GGGYTLRNVA RCWALETGVI LGLRMDDEIP

      351                                                        400
AB21  YNDYF-EYFG PDFKLHISPS NMTNQNTNEY LEKIKQRLFE NLRMLPHAPG
      ||:|| ||:| ||:|| : || || | || || |:|:   :|  ||     ||:
RPD3  YNEYY-EYYG PDYKLSVRPS NMFNVNTPEY LDKVMTNIFA NLENTKYAPS
         |  |:  |             |   |:  | |  :     |  |   : |||
C.el  GTSLYSHYFT PRLLRPNLVP KMNDANSAAY LASIEKETLA CLRMIRGAPS

      401                                                        450
AB21  VQMQA-VAED SIHDDSGEED --EDDPDKRI SIRSSDKRIA CDEEFSDSED
      ||::     | :      |||   |    |   |    | ::   | ||   |:
RPD3  VQLNH-TPRD AEDLGDVEED SAEAKDTKGG SQYARDLHVE HDNEFY....
      ||::       :::  :||      | :|    |      |    |    |:
C.el  VQMQNIVGIR LDEIEQIEEN ERLQKSSKSS IEYEVGKVSE KMEEECFVEE

      451                                                        500
AB21  EGEGGRKNVA NFKKVKRVKT EEEKEGEDKK DVKEEEKAKD EKTDSKRVKE
      :        : :  :         :: |        :          |  | |:
C.el  DSKPPSFPPG QDPRRIGQYW GYDRSGLAPP RSHSDVIEEA KYEDRDRRKD

      501
AB21  ETKSV

C.el  LNIPGIP
```

Length **AB21**: 480 estimated molecular weight: 54.8 kDa
Length **RPD3**: 433 estimated molecular weight: 48.9 kDa
Length **C.el**: 507 estimated molecular weight: 57.1 kDa
Overall identity between AB21 and RPD3: 241/421=57%
Overall conservation between AB21 and RPD3: 324/421=76.6%

The structural conservation between AB21, RPD3 and the *C. elegans* homologue is striking, considering the evolutionary distance between the three species. For example, between AB21 and RPD3 there is an overall 57% identity and at least 77% conservation: however, because a limited, arbitrary set of amino-acid substitutions is considered "conserved", the true figure may be even higher. The conservation is most impressive in the central part of the protein, while towards the C-terminus the similarity is less strong. No currently known structural motif is suggested by the amino-acid sequence. However, the sequence in the C-terminal "tail" of AB21 is very hydrophilic and is expected to form α-helical structures according to the Chou-Fasman algorithm (**Fig. 48**). When plotted on a helical-net diagram, it is apparent that there may be clusters of charged residues, in particular lysines, and acidic S/T/E/D-rich areas (**Fig. 49**).

5.3 Iodination at tyrosines

Although AB21 was cloned with anti-HTP, encodes a 54.7 kDa protein and has a homologue in yeast whose mutant phenotype could be connected to *Xenopus* oogenesis, there is still no conclusive evidence that AB21 is present in the mRNP. One characteristic of the AB21 protein is the abundance of tyrosines and methionines (see **Fig. 46**), a property which suggested the next set of experiments.

Tyrosine side chains can be labelled with the iodine isotope ^{125}I. To that end, native poly(A)$^+$ mRNP particles were iodinated, both in the presence and absence of SDS (**Fig. 50**). This resulted in various labelled proteins, in particular at 40 and 52 kDa in the absence of SDS, and 40 and 54 kDa in the presence of SDS, the positions of which corresponded to the positions of the staining bands of p40, p52 and p54. This difference might be due to the tyrosines in the 54 kDa protein being present in an mRNP structure that is

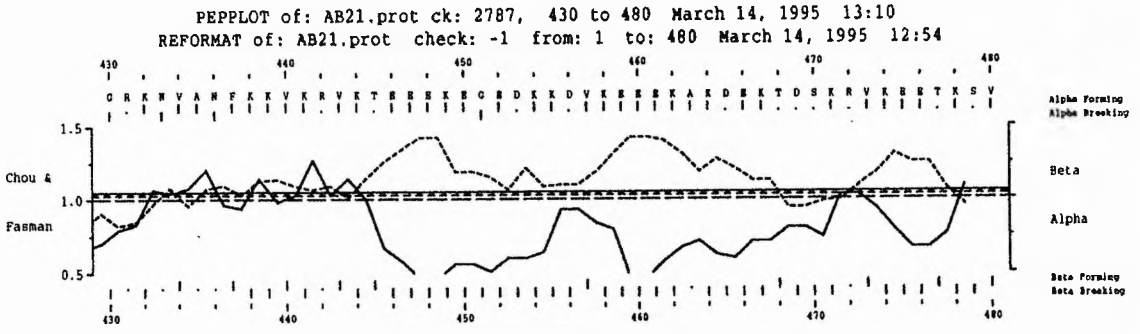


Figure 48 Peptide sequence analysis for AB21. PEPPLOT output for AB21 including basic/acidic areas, Chou-Fasman predictions for α-helices (dotted line) and β-sheets (continuous line). The software was provided by the GCG package, Genetics Computer Group Inc., version 7.2, 1992.

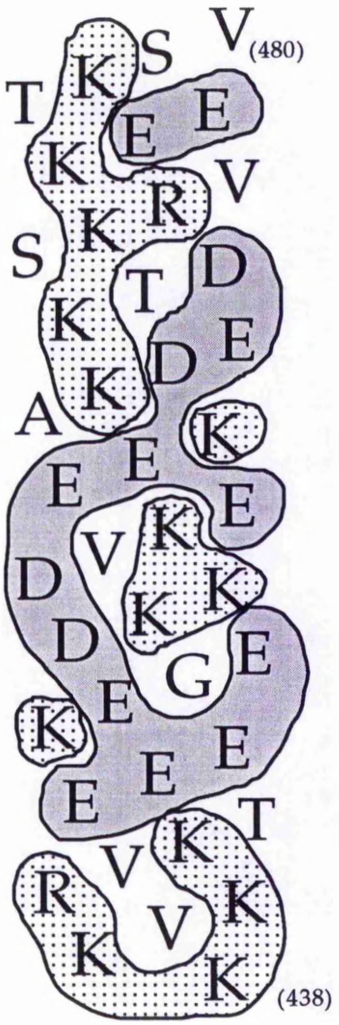


Figure 49 Helical-net diagram of the C-terminal tail of AB21. The diagram includes residues 443-480, which may form an α-helical structure (see **Figure 48**). The helix runs left to right, bottom to top. Clusters of charged residues, either E/D rich (acidic) or K/R rich (basic, mostly K) are apparent. This hypothetical structure can be described as a "sticky tail".

not easily accessed by the reagent. It is conceivable that SDS disrupts the mRNP structure, making the tyrosines in p54 more accessible, while in the native mRNP particle the p52 tyrosines are normally more accessible to the reagent. The third track in **Fig 50A** shows some band-excised p54 which was iodinated separately. After its iodination, it was digested with CNBr, a reaction which cleaves at methionines. **Fig. 50B** illustrates the digestion pattern of iodinated p54 after 6, 20 and 48 hours. Although the complexity of the banding pattern does not allow an accurate map of the position of the methionines to be deduced, there would appear to be multiple labelled fragments, some of which are better labelled than others. A strongly labelled fragment might be due to the presence of a larger number of tyrosines in the fragment, such as might be expected in the case of the N-terminal fragment of the AB21 protein: for example, the first two methionines are separated by twenty-eight amino-acids, of which six are tyrosines (**Fig. 46**).

5.4 Amino-acid sequencing reveals an RNA helicase

A more direct way to identify the mRNP proteins is to fragment them and determine the amino-acid sequence of recovered peptides. The poly(A)+ mRNP HTP (heat-treatment pellet) was used as a general source of material, so as to exclude fragments derived from the heat stable YB proteins which have already been identified. In order to generate smaller fragments, the HTP proteins were treated with CNBr. Given that the protein encoded by AB21 is rich in methionine, it was hoped that an AB21 fragment might be sequenced. Due to time constraints, it was only possible to sequence a small number of peptides. None of these corresponded to AB21. Instead, surprisingly, two of these peptide

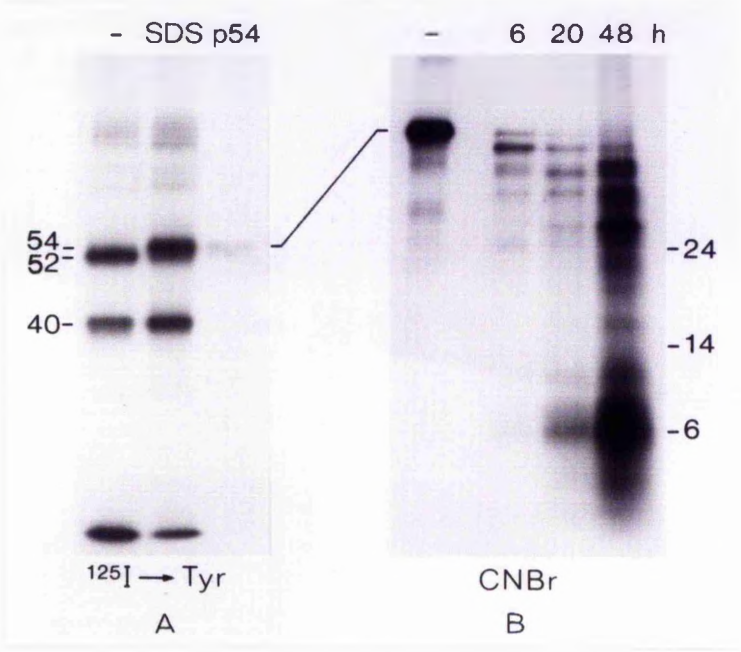


Figure 50 Iodination of poly(A)$^+$ mRNP. Proteins were iodinated at tyrosines with the isotope ^{125}I using the IODOGEN reagent (Pierce). **(A)** 12% acrylamide SDS-PAGE. Whole poly(A)$^+$ mRNP was iodinated, in the presence or absence of 0.1% SDS. Band-purified p54 was also iodinated separately. **(B)** 20% acrylamide SDS-PAGE. CNBr cleavage of iodinated p54 at methionines. Some breakdown of the undigested band-excised p54 is apparent. Samples were analysed after 6, 20 and 48 hours digestion. Size markers are shown. The SDS-PAGE gels were dried for autoradiography.

sequences were perfectly aligned with the amino-acid sequence of a novel family of RNA helicases (**Fig. 51**). The two sequenced peptides are closely aligned to three proteins: the *Drosophila* germ-cell specific RNA helicase ME31B (De Valoir *et al.*, 1991), human p54, which however is not germ-cell specific (Lu and Yunis, 1992), and the *Schizosaccharomyces pombe* protein STE13 (Maekawa *et al.*, 1994). **Fig. 51** shows the alignment between these helicases including the two *Xenopus* peptides. Throughout the alignment there is a close conservation between ME31B, human p54, STE13, and with the *Xenopus* sequence where it is known. Two other RNA helicases are included for comparison, and some identified functional domains are highlighted. A more comprehensive alignment of different RNA helicases will be presented in **Fig. 55**.

Because the sequenced peptides were among the most abundant on the protein transfer which provided the material for peptide sequencing, the RNA helicase is likely to be an abundant mRNP protein. Based on the size of the human homologue, p54, the *Xenopus* homologue may be also be present in the area around 54 kDa, and it is possible that the anti-HTP and/or anti-p54 antisera may recognize it. Since both AB21 and the RNA helicase are predicted to be in the range of 54 kDa, and both are rich in tyrosine and methionine, the previous iodination experiments might have highlighted both these proteins. One of them may be p52 and another p54; alternatively, both might be present in what is observed as the abundant band of 54 kDa. Further experiments are needed to answer these questions.

5.5 Further tests: cleavage at DP sites and ATP-crosslinking

The formic acid cleavage (FA cleavage) procedure, described in chapter 4 in the context of YB proteins, can also be applied these novel proteins. A

Figure 51 An RNA helicase is present in the poly(A)$^+$ mRNP. Alignment between the three examples of a novel subfamily of the "DEAD-box" RNA helicases, human p54 (HP54, Lu and Yunis 1992), *Drosophila* ME31B (De Valoir *et al.*, 1991), and *Schizosaccharomyces pombe* STE13 (Maekawa *et al.*, 1994). Two *Xenopus* peptide sequences are shown, which were derived from a CNBr digest of a poly(A)$^+$ mRNP HTP. Other RNA helicases are compared in a later alignment (Fig. 55), in which it is apparent that this *Xenopus* oocyte RNA helicase belongs this subfamily within the family of DEAD-box RNA helicases, due to the exact correspondence between the *Xenopus* peptides and HP54. Vertical bars denote identities; semicolons denote conservative substitions. The following substitutions are described as conserved: {A,G,S}, {S,T}, {D,E}, {N,Q}, {H,K,R}, {V,L,I,M) and {W,F,Y}. Indicated are the ATPase 'A' and 'B' motifs (AXXGXGKT and VLDEAD respectively, Walker *et al.*, 1982), and two other conserved regions, SAT and HRIGRXXR, which are known to be involved in ATP hydrolysis and helicase activity (Pause and Sonenberg, 1992; Pause *et al.*, 1993, 1994). The sequences were aligned using the PILEUP program, GCG package, Genetics Computer Group Inc., version 7.2, 1992.

```
       1                                                            50
 HP54  MSTARTENPV IMGLSSQNGQ LRGPVKPTGG PGGGGTQTQQ QMNQLKNTNT
                                                        |     |:
ME31B  .......... .......... .......... .......... MMTEKLNSGH

       51                                                           100
 HP54  INNGTQQQAQ SMTTTIKPGD D--WKKTLKL PPKDLRIKTS DVTSTKGNEF
        |  :    :  :       :| |  ||  |     :  | ||: |||  |:||||
ME31B  TNLTSKGIIN DLQIAGNTSD DMGWKSKLNC RQRTTRFKTT DVTDTRGNEF
           : ::|:  |  | |  |    :|  :        | ||  |||  ||| ||
STE13  ...MAESLIQ KLENA.NLND RESFKGQMKA QPVDMRPKTE DVTKTRGTEF

       101     Xenopus MGWEK PSPIQ              'A' motif 150
 HP54  EDYCLKRELL MGIFEMGWEK PSPIQEESIP IALSGRDILA RAKNGTGKSG
       |::|||||||  |||||  |||:  ||||||  :||  |||||:|:||  |||||||||:|
ME31B  EEFCLKRELL MGIFEKGWER PSPIQEAAIP IALSGKDVLA RAKNGTGKTG
       |::  ||||||  |||||  |:||  ||||||  :||  |||||:|:||  |||||||||:
STE13  EDYYLKRELL MGIFEAGFER PSPIQEESIP IALSGRDILA RAKNGTGKTA
```

```
      151                                                       200
HP54  AYLIPLLERL DLKKDNIQAM VIVPTRELAL QVSQICIQVS KHMGGAKVMA
      || ||:|| : |  || |||: |:||||||||| | ||||| :: ||:   :||
ME31B AYCIPVLEQI DPTKDYIQAL VMVPTRELAL QTSQICIELA KHL-DIRVMV
      |: || || : |  |  || | ::||||||||| ||||:|  |: ||:  ::|||
STE13 AFVIPSLEKV DTKKSKIQTL ILVPTRELAL QTSQVCKTLG KHM-NVKVMV

      201                                                 'B' motif
HP54  TTGGTNLRDD IMRLDDTVHV VIATPGRILD LIKKGVAKVD HVQMIVLDEA
      ||||| |:|| |:|:   | : :|||||||||| |: | || :  |  ::|||||
ME31B TTGGTILKDD ILRIYQKVQL IIATPGRILD LMDKKVADMS HCRILVLDEA
      ||||| |:|| |:|:   | : :::||||:|| |  | ||| |  |   |:|||
STE13 TTGGTTLRDD IIRLNDTVHI VVGTPGRVLD LAGKGVADFS ECTTFVMDEA

      251                                                       300
HP54  DKLLSQDFVQ IMEDIILTLP KNRQILLYSA TFPLSVQKFM NSHLEKPYEI
      ||||| ||   ::: :|| || |  ||||:|| ||||:|  ||   ||  ||||
ME31B DKLLSLDFQG MLDHVILKLP KDPQILLFSA TFPLTVKNFM EKHLREPYEI
      ||||| :|   ::: ::   | |  || |:|| |||| ||||| :|||  ||||
STE13 DKLLSPEFTP IIEQLLSYFP KNRQISLYSA TFPLIVKNFM DKHLNKPYEI

      301                                                       350
HP54  NLMEELTLKG VTQYYAYVTE RQKVHCLNTL FSRLQINQSI IFCNSSQRVE
      |||||||||| |||||:| | ||||||||||  ||:|||||||  |||||:||||
ME31B NLMEELTLKG VTQYYAFVQE RQKVHCLNTL FSKLQINQSI IFCNSTQRVE
      |||:||||:| ||||||||  |  |||||||||  ||||||||||  ||||||:|||
STE13 NLMDELTLRG VTQYYAFVDE SQKVHCLNTL FSKLQINQSI IFCNSTNRVE

      351          Xenopus M RQEHRNRVFH                         400
HP54  LLAKKISQLG YSCFYIHAKM RQEHRNRVFH DFRNGLCRNL VCTDLFTRGI
      ||||||: || | |:|||||||  | |||||||  |||:||||||  ||:||||||||
ME31B LLAKKITELG YCCYYIHAKM AQAHRNRVFH DFRQGLCRNL VCSDLFTRGI
      |||||||||| | |:| ||||  |:|||||||  ||:|:|||||  ||||| ||||
STE13 LLAKKITELG YSCFYSHAKM LQSHRNRVFH NFRNGVCRNL VCSDLLTRGI

      401                                                       450
HP54  DIQAVNVVIN FDFPKLAETY LHRIGRSGRF GHLGLAINLI TYDDRFNLKS
      |:|||||||| ||||::|||| |||||||||| ||||:|||||  ||:||| |
ME31B DVQAVNVVIN FDFPRMAETY LHRIGRSGRF GHLGIAINLI TYEDRFDLHR
      |:|||||||| ||||: |||| |||||||||| || |:||  | :: ||| | |
STE13 DIQAVNVVIN FDFPKNAETY LHRIGRSGRF GHRGLAISFI SWADRFNLYR

      451                                                       500
HP54  IEEQLGTEIK PIPSNIDKSL YV-------- ---------- --AEYHSEPV
      ||  |||||| ||| || :| ||                      | | :
ME31B IEKELGTEIK PIPKVIDPAL YV-------- ---------- --ANVGASVG
      || |||||| ||| |||:| ||                       |   |
STE13 IENELGTEIQ PIPPSIDPSL YVFPNGDYQI PRPLTASADQ VLAAQQAKGQ

      501                                                 540
HP54  EDEKP..... .......... .......... ..........
      :
ME31B DTCNNSDLNN SANEEGNVSK .......... ..........
      :   |   ||
STE13 EGYHNRPNNN RGGHPRGGGN RGGYRQSNRQ PRYRGQQKAD
```

Length **HP54**: 483 estimated molecular weight: 54.5 kDa

Length **ME31B**: 459 estimated molecular weight 52.0 kDa

Length **STE13**: 485 estimated molecular weight 54.8 kDa

poly(A)$^+$ HTP fraction was FA-digested, and is shown in **Fig. 52**. Compared to the undigested track, at least three abundant new bands appear. These are cleavage products, and could, in principle, be sequenced at the amino-acid level. In the case of AB21, the amino-acid sequence suggests the presence of a DP site towards the C-terminus (**Fig. 46**). Cleavage of the intact 54.7 kDa protein would be predicted to generate a 45.5 and a 9.2 kDa fragment. In the case of the RNA helicase, *Drosophila* ME31B has 3 DP sites, *S. pombe* STE13 has one, human p54 has none; the number of DP sites in the *Xenopus* RNA helicase is not known *a priori*. **Fig. 53** shows the cleavage of band-excised p54, which was FA-digested, separated on SDS-PAGE and immunoblotted with anti-HTP in panel C. In panels A and B, the same procedure was applied to band-excised YB proteins for a comparison. An immunoblot was necessary because of low protein yields: ideally, a protein stain is preferable because some of the cleaved products may not be antigenic. However, confirming previous suspicions, what emerges from **Fig. 53** is that within band-excised p54 there are at least two proteins. Three fragments were sized to approximately 43, 30 and 20 kDa. Together, the 30 and 20 kDa fragments almost add up to 54 kDa and could correspond to one of the proteins, whereas the 43 kDa fragment is missing 11 kDa, a small fragment not detected by the immunoblot. The 43 kDa band, with its presumed missing 11 kDa band, is not too far from the 45.6 and 9.2 kDa predicted from the AB21 map. In addition, there is some residual undigested 54 kDa protein, which either represents an incomplete FA-digestion, or corresponds to a third protein in the 54 kDa group which has no DP site. In summary, this experiment confirmed that more than one protein is present around 54 kDa in the poly(A)$^+$ mRNP.

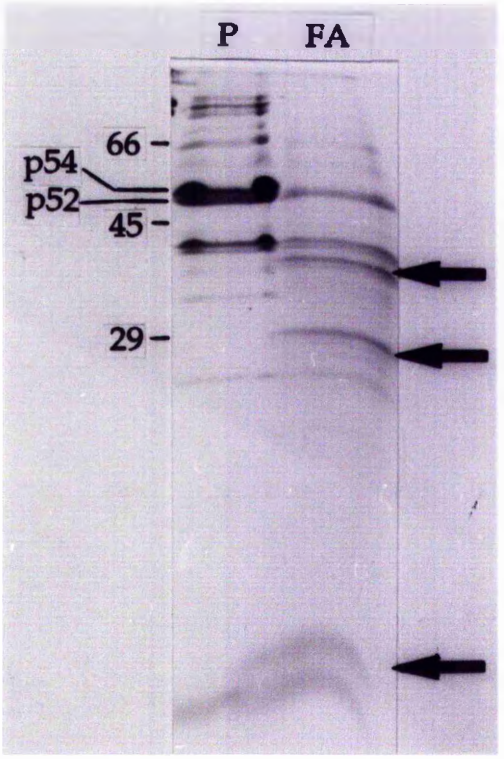


Figure 52 FA-cleavage of a poly(A)$^+$ mRNP HTP fraction. 20% acrylamide SDS-PAGE: the Coomassie stained gel shows: **P**: ~12 µg undigested poly(A)$^+$ mRNP HTP (heat-treatment pellet); **FA**: the same, but digested with FA (formic acid) cleaving at DP sites. At least three new bands appear, indicated by arrows.

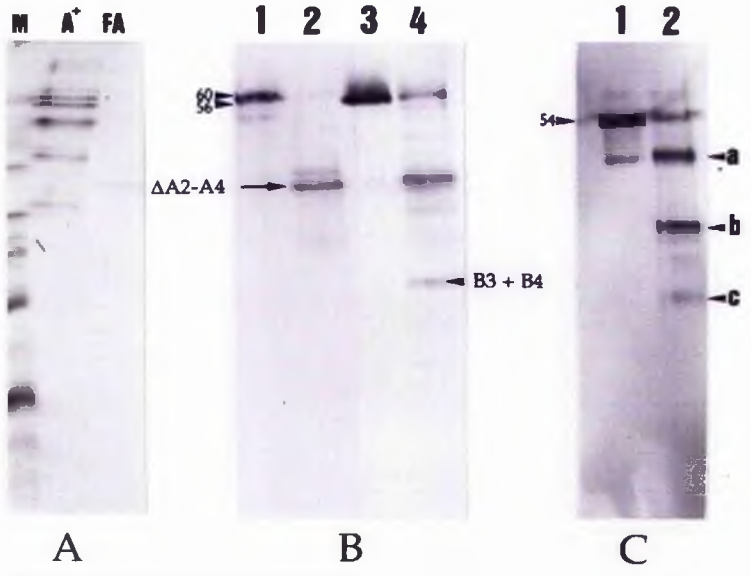


Figure 53 Immunoblot of FA-cleaved proteins. Proteins were separated on 20% acrylamide SDS-PAGE. **(A)** Coomassie stain showing **M**: molecular weight markers; **A$^+$**: 12 µg total poly(A)$^+$ mRNP, and **FA**: YB proteins were digested with FA, showing the ΔA2-A4 fragment (refer Figures 26A and 27A for a map of the YB proteins). **(B)** Immunoblot of FA-cleaved YB proteins. Anti-HTSN was diluted 1/1000. Track **1**: band-purified pp60; **2**: FA-cleaved pp60; **3**: band-purified pp56; **4**: FA-cleaved pp56. ΔA2-A4 and B3+B4 fragments derived from the YB proteins are indicated. **(C)** Immunoblot of band-purified p54 using a 1/1000 dilution of anti-HTP. **Track 1**: undigested p54; **2**: FA-cleaved p54. Note the presence of some residual undigested p54, as well as three bands, indicated **a**, **b**, and **c**, of estimated sizes of 43, 28 and 20 kDa respectively.

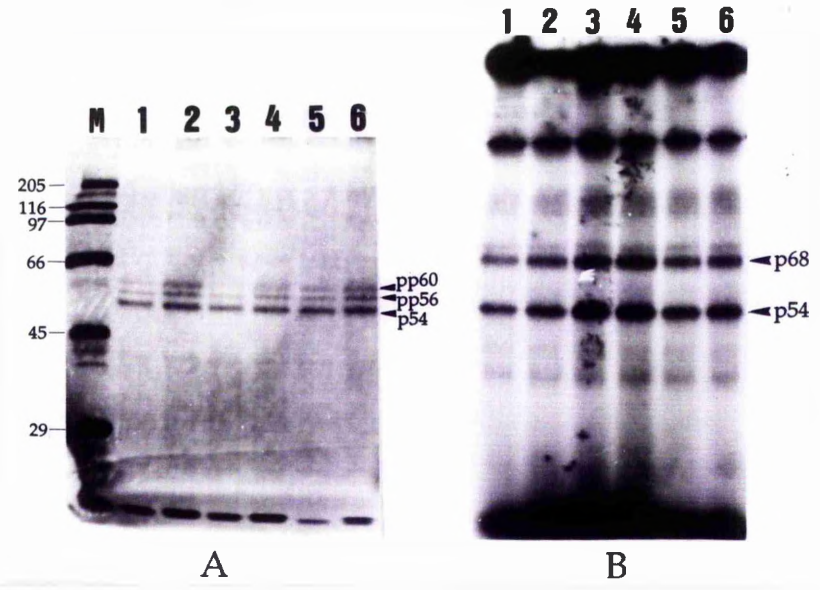


Figure 54 UV-crosslinking of [α-^{32}P]ATP to poly(A)$^+$ mRNP. 12 µg of fresh poly(A)$^+$ mRNP and ~1 µCi of [α-^{32}P]ATP was used in each sample and run on 12% acrylamide SDS-PAGE. Panel **(A)** shows the Coomassie stain of the dried gel and **(B)** the autoradiograph of the same gel. **Tracks 1**: poly(A)$^+$ mRNP, crosslinked to [α-^{32}P]ATP; **2**: material in track 1, RNase treated; **3**: + 2 mM spermine; **4**: material from track 3 was RNase treated; **5**: + 1 µg/ml heparin; **6**: material from track 5 was RNase treated. Proteins of 68 and 54 kDa were clearly crosslinked to [α-^{32}P]ATP, as well as a weaker, less distinct signal around 40 kDa. It was thought that spermine and heparin might influence the ability to UV-crosslink, but no obvious differences were apparent.

A further test was attemped, following the succesful crosslinking of ATP to the RNA helicase eIF4A reported by Pause *et al.* (1993). RNA helicases have an ATP-binding fold, to which ATP binds and is essential for driving their RNA helicase activity. [α-^{32}P]ATP, preferred to [γ-^{32}P]ATP because the labelled phosphate is protected from hydrolysis, was UV-crosslinked to fresh, native mRNP. **Fig. 54** illustrates this experiment, including a Coomassie stain and an autoradiograph showing proteins crosslinked to ATP. In all samples, proteins of 68 and 54 kDa were most clearly crosslinked to ATP. Undigested and RNase-treated samples were compared. If the ATP-binding protein had been UV-crosslinked to the native mRNA, differences between the undigested and RNase treated tracks should have been apparent, but did not appear to be the case.

5.6 Summary

The cDNA cloning strategies were succesful in isolating new clones. Using the riboprobe binding method, the following cDNAs were isolated: (i) clone G2, encoding two zinc-fingers which were later shown to be part of the the novel zinc-finger protein C4SR. It is an unusual protein in that there is nothing like it in the database with its particular arrangement of domains. Its function is unknown at this stage, but may be inferred to have a role in nucleic acid binding. (ii) The somatic YB protein FRGY1 was also cloned via its binding to riboprobe. Its counterpart in the rabbit reticulocyte has been shown to be present in mRNP (Evdokimova *et al.*, 1995). (iii) M7, a cDNA which encodes the small subunit ribosomal protein S27, was cloned. The ribosomal protein S27 may have an RNA-binding function via its zinc-fingers. With the use of antisera, the following cDNAs were obtained: (i) nucleolin was cloned on three occasions using anti-HTSN. Nucleolin was later confirmed to be present in the HTSN; however,

there is no conclusive evidence to suggest that it is present in mRNP *in vivo*. (ii) The novel protein AB21 was selected with anti-HTP. It is impressively similar to a yeast global gene regulator protein, RPD3, and to an uncharacterized *C. elegans* protein. It was not possible to demonstrate the presence of AB21 in the mRNP conclusively, but data from iodination and amino-acid cleavage experiments is consistent with the AB21 amino-acid sequence. AB21 is an oocyte-specific protein and its expression peaks in stage II oocytes. Its function remains to be determined. Finally, an RNA helicase was identified through peptide sequencing, which belongs to the subfamily of DEAD-box proteins that includes *Drosophila* ME31B, human p54, and *S. pombe* STE13. A *Xenopus* oocyte mRNP protein of 54 kDa was crosslinked to ATP, and is likely to be the *Xenopus* version of the above RNA helicases. The function of the RNA helicase is unknown: it could be an important and intriguing new component of the masked mRNP particle.

Chapter 6

Discussion

6.0 Versatility of the YB proteins

The YB proteins are a multifunctional family of gene regulators. In the context of *Xenopus* oocytes, they are an essential component of masked messenger RNAs. In order to package a wide range of mRNAs, it is necessary for the YB proteins to bind to many different mRNA sequences. The binding assays presented in chapter 4 have increased the understanding of the YB protein:RNA interactions. In essence, two distinct binding activities were observed, one mediated by the CSD (cold-shock domain) and the other by the basic/aromatic islands in the TD (tail domain).

The CSD preferred the single stranded polypurines poly(A,G) or poly(G) but not poly(A), a binding activity that was sensitive to 1-2 mM Mg^{2+}. As was mentioned in section 4.5, supporting evidence comes from studies on the protein unr, which consists of a five-fold repeat of the CSD with no associated TD. Binding of unr to nucleic acids was shown to be sensitive to 1 mM Mg^{2+}, and was competed preferentially by poly(G) (Jacquemin-Sablon *et al.*, 1994). In addition, p50, an abundant rabbit reticulocyte mRNP protein, now identified as YB-1 (Evdokimova *et al.*, 1995), prefers poly(G) in a binding buffer that lacks Mg^{2+}. (Minich *et al.*, 1993). In the context of DNA, YB-1 has also been reported to bind to purine-rich stretches present in the RSV LTR (Rous-Sarcoma Virus Long Terminal Repeat). Although this project only tackled RNA-binding properties, DNA-binding properties are also relevant in the discussion because the YB proteins bind to and regulate DNA as well as mRNA *in vivo*.

The nucleic-acid binding properties of the CSD and TD have recently been compared (Murray, 1994). By using both DNA and RNA substrates containing the Y-box sequence, it was found that the CSD binds

preferentially to ssDNA. A preference for single-stranded templates has been described in various studies (for example, Marello *et al.*, 1992; Grant and Deeley, 1993; Horwitz *et al.*, 1994). More recently, YB-1 has been shown to promote the formation of single-stranded DNA in the HLA-DRα promoter (MacDonald *et al.*, 1995). Furthermore, the the β-barrel configuration of the bacterial cold-shock proteins has been suggested to be consistent with binding to a single-stranded target, and simple model building suggests that the aromatic rings on the solvent face of the β-barrel could stack with the bases in ssDNA or RNA (Schindelin *et al.*, 1993; Newkirk *et al.*, 1994). In gel retardation assays, both the *B. subtilis* cold-shock protein CspB and its *E. coli* homologue CspA bind preferentially to ssDNA (Schnuchel *et al.*, 1993; Schindelin *et al.*, 1994; Graumann and Marahiel, 1994). What remains to be determined is the detailed structure of the CSD:DNA or CSD:RNA complex, the prediction being that the highly conserved aromatic and basic residues which protrude from the solvent face of the β-barrel interact directly with ssDNA or RNA. Eventually, the structure of the complex might explain the empirically observed preference for poly(G) in RNA-binding and for the Y-box sequence in ssDNA binding. It might also explain why Mg^{2+} and not other polycations interfere with binding, either directly, or by altering the structure of the nucleic acid target in such a way as to render it less accessible to the CSD.

The B/A (basic/aromatic) islands present in the TD of pp60, pp56, and FRGY1 were listed and compared in **Fig. 5**. Individual B/A islands were UV-crosslinked to RNA, and their binding specificity was altered by Mg^{2+}. Similar experiments were repeated with other polycations such as spermine and spermidine (not shown), resulting in a similar shift in binding specificity. In the presence of the polycations Mg^{2+}, spermine and

spermidine, the TD bound preferentially to poly(C,U) (and to poly(A,G) in their absence). This would explain the earlier report in which poly(C,U) is the most efficient competitor in binding buffers where Mg^{2+} was a standard component (Marello *et al.*, 1992).

In addition, the TD is thought to be responsible for protein:protein associations via electrostatic interactions between charged domains (Tafuri and Wolffe, 1992), and the kinetics of YB-1 binding to a polypyrimidine template are suggestive of cooperative binding effects (Horwitz *et al.*, 1994). The ability to multimerize would be a desirable property in a situation where mRNA needs to be packaged efficiently, preventing its association with translation initiation factors. Because native mRNA sequences contain arrays of both purines and pyrimidines, complex structures form in solution. The appearance of the pp60:cyclin B1 mRNA complexes in the electron micrographs (**Fig. 14**) suggested that the complexes were of a more or less uniform shape. The shape of this specific YB protein:mRNA complex would appear to depend on the structure of the specific mRNA involved.

It may be suggested here that both the CSD and the TD might cooperate *in vivo* to bind to mRNA. That both the CSD alone or the TD alone can bind to RNA and block its translation was suggested in the microinjection experiments of Ranjan *et al.* (1993), in which over-expression of constructs encoding either the CSD or the TD resulted in the translational repression of a CAT-reporter construct. By the same token, it may be suggested that both the CSD and TD might also cooperate *in vivo* in binding to DNA.

The preference for polypyrimidines in DNA binding has been reported in various other studies, and has caused an apparent dilemma: shouldn't the YB proteins bind preferentially to the Y-box sequence in the context of

DNA? It is possible that the observed affinity for polypyrimidines, such as those present in H-DNA triplex structures, was due to binding by the TD. The binding buffers used in these studies (Kolluri *et al.*, 1992; Grant and Deeley, 1993; Horwitz *et al.*, 1994) all included $MgCl_2$. According to Murray (1994), isolated B/A islands bind to RNA probes containing the Y-box sequence but *not* to a similar ssDNA probe. However, pyrimidine-rich DNA substrates were not considered in that study. It is possible that the TD prefer RNA sequences over ssDNA sequences in general, but that certain DNA sequences rich in pyrimidines are nonetheless bound by the TD, albeit with a lower affinity.

H-DNA triplex structures are thought to occur in a variety of promoters which contain regions of homopurine/homopyrimidine strand asymmetry, and are noted for their hypersensitivity to S1 nuclease (Kinniburgh *et al.*, 1994). Kinniburgh *et al.* (1994) suggest that many genes, and, particularly those that regulate growth processes, have promoter-proximal NSE (nuclease sensitive elements). For example, H-DNA is present in the mouse c-Ki-*ras* proto-oncogene promoter (Pestov *et al.*, 1991) and in the c-*myc* promoter (Firulli *et al.*, 1992). At least two protein factors, hnRNPK (Takimoto *et al.*, 1993) and the YB protein NSEP-1 (Kolluri *et al.*, 1992), are thought to bind to the *c-myc* NSE and either promote the formation of, or stabilise H-DNA (Kinniburgh *et al.*, 1994). The c-*myc* DNaseI hypersensitive site has been shown to disappear coincidentally with the cessation of c-*myc* transcription (Siebenlist *et al.*, 1988). Mutations in the H-DNA area of the c-*myc* promoter have suggested that the NSE accounts for 75-85% of the total c-*myc* transcription (Davis *et al.*, 1989; Firulli *et al.*, 1994). The proposed structure of the c-*myc* H-DNA element (see **Fig. 7**) presents a displaced purine-rich (and, perhaps significantly, G-

rich) single strand, as well as a pyrimidine-rich strand in the triplex. Could these features be suitable binding targets for the YB proteins?

Another example in which the YB proteins have been suggested to interact with H-DNA is the human γ-globin promoter (Horwitz *et al.*, 1994), An H-DNA forming region is present between -228 and -189 upstream of the γ-globin transcription start site. Horwitz *et al.* find that YB-1 binds to the homopyrimidine sequences. In the hereditary condition HPFH (hereditary persistence of fetal hemoglobin), in which fetal γ-globin persists in the infant, specific point mutations were seen to destabilise H-DNA while concurrently abolishing or reducing the binding of YB-1. Horwitz *et al.* suggest that YB-1 acts as a transcriptional repressor of γ-globin *in vivo*.

In another example, Hollingsworth *et al.* (1994) point out that the cystic fibrosis gene *CFTR* also contains a putative H-DNA forming region, and find that a 27 kDa protein binds to the purine-rich single strand in the H-DNA. Thus a number of proteins are likely to bind to or promote the formation of H-DNA *in vivo*. The binding of YB proteins to H-DNA has so far only been demonstrated *in vitro*, and some caution is therefore required in interpreting these results. The involvement of YB proteins with H-DNA *in vivo*, affecting gene regulation through H-DNA *as well as* Y-box sequences, remains to be demonstrated.

In summary, results suggest that the CSD and TD have distinct nucleic-acid binding properties. The combination of these binding domains and the influence of polycations on binding preferences might account for the observed versatility in nucleic-acid binding in the context of mRNA, but also DNA. It is tempting to speculate that both Y-box sequences and displaced strands such as those present in H-DNA may be recognized by the YB proteins *in vivo*, a recognition involving both the CSD and TD.

6.1 Phosphorylation of the YB proteins

Microinjection experiments by Braddock *et al.* (1994) have shown that nuclear phosphorylation of the YB proteins is essential for mRNA masking to occur. This is consistent with earlier experiments, in which reconstituted pp60:globin mRNA complexes were translationally inactive in a wheat germ lysate but became translationally active after dephosphorylation (Kick *et al.*, 1987). Moreover, the rabbit reticulocyte mRNP protein p50 (YB-1) is also phosphorylated *in vitro* and *in vivo*, and when reconstituted with globin mRNA, also renders globin mRNA translationally inactive in a wheat germ lysate (Minich and Ovchinnikov, 1992).

The results presented in chapter 3 confirm the presence of an mRNP-associated kinase activity, and that the YB proteins can be phosphorylated in the nuclei. In chapter 4, three different experiments suggest that phosphorylation of the YB proteins improves RNA-binding, measured in terms of an increased stability of the protein:RNA complex in NaCl, and in terms of the ability to crosslink to riboprobe. In addition, the observed differences between pp60 and pp56 in riboblots and in a density gradient separation of native mRNP particles suggest that the binding of pp60 to RNA is more stable than pp56. It is probably not coincidental that pp60 has three extra phosphorylation sites in the TD.

The role of casein kinase II in mRNA masking is not limited to *Xenopus* oocytes. For example, Thoen *et al.* (1984) have studied embryos of the the invertebrate *Artemia salina* (brine shrimp). These embryos can interrupt their development in poor environmental conditions, while storing their mRNPs in free particles. Phosphorylation of the mRNP proteins by an associated protein kinase activity was associated with the inhibition of

translation. The properties of the kinase, including activation by hemin and spermine, inhibition by heparin, its cAMP and Ca^{2+}-independence, molecular mass, and its ability to phosphorylate casein were consistent with its being casein kinase II. It may be inferred that a similar mechanism involving phosphorylation of YB proteins exists in *Artemia*. This mechanism could also be thought to occur in somatic contexts where mRNA masking is required. In other words, the phosphorylation and dephosphorylation of the YB proteins could be the one of the means by which external physiological or hormonal stimuli influence the masking and unmasking of stored mRNAs in a variety of tissues.

6.2 Future experiments involving YB proteins

The YB proteins are emerging as a versatile, multi-functional family of gene regulators. After the discovery of YB-1 as one of the most abundant mRNP proteins in rabbit reticulocyte mRNP, it is tempting to speculate that an involvement of YB proteins in mRNA masking exists in other somatic contexts. Evdokimova *et al.* (1995) used an antiserum raised against p50 to detect a cross-reactive protein in rat liver and rabbit muscle mRNP. The techniques described in chapters 3 and 4 in with respect to *Xenopus* oocyte mRNP could be applied to mRNPs from other tissue extracts: in particular, heat-treatment, the affinity for a heparin column, and efficient UV-crosslinking to RNA would be exploited to demonstrate the presence of YB proteins.

In the context of masked messenger RNAs, more precise experiments will be necessary to prove the association of YB proteins with specific mRNAs *in vivo*. For example, YB phosphoproteins might be predicted to be associated with nascent c-*mos* transcripts in the nucleus, and to remain part of the c-*mos* mRNP particle until its recruitment for translation, when

the YB proteins bound to c-*mos* would be dephosphorylated. Immunoprecipitation might be a feasible experimental strategy to associate c-*mos* mRNA (or any other masked message) with YB proteins.

Very little is known about the regulation of the YB protein genes themselves. Their expression appears to be associated with the induction of cell proliferation in somatic tissues, and it has been suggested that the YB proteins are involved in activating cell proliferation genes (Grant and Deeley, 1993; Ito *et al.*, 1994). It may be of interest to isolate genomic DNA clones to investigate the promoter sequences of YB protein genes; this information might indicate what factors regulate their expression. One possible clue comes from the work of Ito *et al.* (1994): after serum stimulation of quiescent fibroblasts (NIH/3T3 cells) expression of the rat YB protein RYB-a was rapidly induced. The induction of RYB-a was blocked by preventing cell-cell adhesion or by adding genistein, a specific inhibitor of tyrosine kinase. Thus a signalling pathway that leads to the activation of YB protein genes via a tyrosine kinase is likely to exist.

The postulated role of YB proteins in cell proliferation still needs more corroborating evidence. To that effect, a FRGY1 cDNA (clone N5, see section 5.1) has been sent to the University of Pisa, Italy, where Dr. Batistoni *et al.* are studying the regeneration of planarians (flatworms). These organisms have the remarkable ability to regenerate a full body after having been cut in half. The working hypothesis is that the planarian homologue of FRGY1 is induced, and is necessary for the ensuing cell proliferation processes needed for tissue regeneration. In addition, a research project is currently being proposed in this lab that would consider the pattern of expression of YB-1 in a human ovarian carcinoma cell line. The expression would be monitored with a nucleic acid probe. A human YB-1 cDNA would either be obtained from another group or

isolated through cross-hybridization using a FRGY1 probe (clone N5). In parallel, the expression of other cell proliferation markers, such as nucleolin and casein kinase II, would be monitored.

6.3 AB21: a protein in search of a function

Using the immunoselection technique, clone AB21 was isolated and sequenced, revealing a novel protein which is very similar to a yeast gene regulator, RPD3 (Vidal and Gaber, 1991). The estimated molecular weight of AB21 (54.7 kDa), the presence of tyrosine and methionine-rich protein(s) around 54 kDa in the mRNP fraction, and the FA-cleavage of 54 kDa material into fragments that included one of a predicted size, are consistent with the AB21 amino-acid sequence: however, the data was by no means conclusive. Whether or not AB21 is an mRNP protein, it was undoubtedly cloned from an oocyte cDNA library which represents proteins that are expressed during oogenesis, and is therefore of interest. Its presence in mRNP has not been demonstrated, and further experiments will be needed to investigate its function in gene regulation.

The only available clues as to the function of AB21 in *Xenopus* oocytes come from the phenotypic characteristic of the *rpd3* mutant in yeast. *RPD3* was originally cloned in a genetic screen designed to characterise new transcription factors. Vidal and Gaber (1991) used a model system in which a yeast strain has a deletion in the gene *TRK1*, which encodes a high-affinity K^+ transporter. The same strain, however, has a functional *TRK2* gene, encoding a low affinity K^+ transporter and its expression is essential in media containing low levels of K^+. Two mutants were isolated, *rpd1* and *rpd3* (RPD=Reduced Potassium Dependency), which restored the Trk^+ phenotype.

Vidal *et al.* (1991) showed that *RPD1* is required for the full repression and full activation of a variety of yeast genes, including cell differentiation, cell type-specific and metabolically regulated genes, and consequently describe it as a "global transcriptional regulator". The authors have shown that the *rpd1* mutation in results in the increased transcription of *TRK2*. *RPD1* was cloned; RPD1 protein contains four HLH (Helix-Loop-Helix) motifs as well as acidic, glutamine and proline-rich regions, which they propose are protein:protein interaction motifs. Vidal *et al.* (1991) suggest that RPD1 might bind to a variety of transcription factors, thus affecting their ability to bind to DNA. In a sense, RPD1 protein could be described as a "regulator of gene regulators".

The phenotype of *rpd3* is indistinguishable from *rpd1*, and the phenotypic analysis of double mutants *rpd1/rpd3* suggest that they are involved in the same pathway (Vidal and Gaber, 1991). In other words, *rpd1* and *rpd3* mutants have identical pleiotropic phenotypes, including mating defects, the inability to sporulate as homozygous diploids, and a sensitivity towards cycloheximide, an inhibitor of translation. Both *RPD1* and *RPD3* have been described as negative regulators of early meiotic genes (Bowdish *et al.*, 1993). In *Xenopus* oocytes, negative gene regulation is via the *translational* repression of mRNAs required for the progression of meiosis (in previtellogenic oocytes, mRNAs such as c-*mos* are being actively transcribed but packaged into translationally silent masked mRNPs).

There is still no biochemical information about the yeast protein RPD3. Its "global effects" on a variety of genes, including its negative effect on meiotic genes, may in principle be mediated at the level of mRNA. There are no obvious informative homologies in RPD3 or AB21 which relate them to other RNA (or DNA) binding proteins. Only one possible

structure was noticed in AB21: a putative α-helical region in the C-terminus of *Xenopus* AB21, which is not present in yeast RPD3. When plotted on a helical-net diagram (**Fig. 49**), clusters of both basic and acidic residues appear. It is tempting to speculate that the α-helix could act as a "sticky-tail" to promote protein:protein multimerization in the context of mRNP packaging.

6.4 An RNA helicase is associated with *Xenopus* oocyte mRNP

Whereas the presence of AB21 in the mRNP was not confirmed, direct peptide sequencing of mRNP proteins has revealed the presence of an RNA helicase. This RNA helicase belongs to the DEAD-box family of RNA helicases, so called because of the conserved sequence (D-E-A-D). The alignment between the helicases suggests that within the DEAD-box family, the *Xenopus* peptides are closely aligned to human p54 (Lu and Yunis, 1992), *Drosophila* ME31B (De Valoir *et al.*, 1991) and the yeast *Schizosaccharomyces pombe* protein Ste13 (Maekawa *et al.*, 1994), representing a subfamily within the DEAD-box family (**Fig. 51**). For a comparison, various DEAD-box RNA helicases are shown in **Fig. 55**, ranging from *E. coli* to mammals. A number of conserved features are clearly present in all members of the DEAD-box family, and include the ATP binding motifs "A" and "B", commonly found in various ATPases: the "A" site, consensus AXXGXGKT, and the "B" site, consensus VLDEAD (Walker *et al.*, 1982). However, there are other highly conserved regions, such as PTRELA, GG, TPGR, SAT, ARGXD and HRIGRXXR (see **Fig. 55**).

Figure 55 Alignment of various DEAD-box RNA helicases. Various RNA helicases, ranging from *E.coli* to mammals, were aligned using the PILEUP algorithm (GCG package, Genetics Computer Group Inc., version **7.2**, 1992). Areas of identity or close conservation between all examples are indicated with an asterisk. Indicated are the ATPase 'A' and 'B' motifs (AXXGXGKT and VLDEAD respectively, Walker *et al.*, 1982), and two other conserved regions, SAT and HRIGRXXR, which are known to be involved in ATP hydrolysis and helicase activity (Pause and Sonenberg, 1992; Pause *et al.*, 1993, 1994). **p68**: human RNA helicase found in nuclei of proliferating cells (Iggo *et al.*, 1991); **RM62**: *Drosophila* RNA helicase (Dorer *et al.*, 1990); **AN3**: mRNA localized to animal poles in the *Xenopus* oocyte (Gururajan *et al.*, 1991); **PL10**: RNA helicase expressed in mouse testis (Leroy *et al.*, 1989); **vasa**: *Drosophila* RNA helicase, component of the polar granules (Hay *et al.*, 1988); **p54**: Human RNA helicase (Lu and Yunis, 1992); **ME31B**: *Drosophila* germ-cell specific RNA helicase (De Valoir *et al.*, 1991); **STE13**: *Schizosaccharomyces pombe* RNA helicase involved in meiosis (Maekawa *et al.*, 1994); **eIF4A**: RNA helicase translation initiation factor from *Nicotiana sylvestris* (Owttrim *et al.*, 1994); **DBPA**: putative *E. coli* RNA helicase (Iggo *et al.*, 1990); **DeaD**: putative *E. coli* RNA helicase involved in ribosomal function (Toone *et al.*, 1991); **PRP28**: a yeast RNA helicase involved in splicing (Strauss and Guthrie, 1991). Two sequenced *Xenopus* oocyte mRNP peptides were shown in **Fig. 51**, suggesting that the *Xenopus* oocyte RNA helicase of 54 kDa belongs to the subfamily which includes human p54, *Drosophila* ME31B and yeast STE13.

```
          1                                                            50
p68       .......... .......... .......... .......... ..........
RM62      .......... .......... .......... .......... ..........
AN3       MSHVAVENVL NLDQQFAGLD LNSADAESGV A.GTKGRYIP PHLRNKEASR
PL10      MSHVAEEDEL GLDQQLAGLD LTSRDSQSGG STASKGRYIP PHLRNREAA.
vasa      MSDDWDDEPI VDTRGARGGD WSDDEDTAKS FSGERE.... GDGVGGSGGE
p54       .......... .......... .......... .......... ..........
ME31B     .......... .......... .......... .......... ..........
STE13     .......... .......... .......... .......... ..........
eIF4A     .......... .......... .......... .......... ..........
DBPA      .......... .......... .......... .......... ..........
DeaD      .......... .......... .......... .......... ..........
PR28      .......... .......... .......... .......... ..........

         51                                                           100
p68       .......... .......... .......... .......... ..........
RM62      .......... .......... .......... ........MA PHDRDFGHSG
AN3       NDSNWDSGRG GNGYINGMQD DRDGRMNGYD RGGYGSRGTG RSDRGFYDRE
PL10      .......... .......... .......... .......... ...KAFYDKD
vasa      GGGYQGGNRD VFGRIGGGRG GGAGGYRGGN RDGGGFHGGR REGERDFRGG
p54       .......... .......... .......... .......... ..........
ME31B     .......... .......... .......... .......... ..........
STE13     .......... .......... .......... .......... ..........
eIF4A     .......... .......... .......... .......... ..........
DBPA      .......... .......... .......... .......... ..........
DeaD      .......... .......... .......... .......... ..........
PR28      .......... .......... .MARPIDVSQ LIAGINKKKG LDENTSGKIS

        101                                                           150
p68       MSGYSSDR.. DR....GRDR GFGAPRFGGS .......... .......RAG
RM62      RGGRGGDRGG DD....RRGG GGGGNRFGGG GGGGDYHGIR NGAVEKRRDD
AN3       NSGWNSGRDK DAYSSFGSRG D.RG...... ..KGSLFNER .GSGSR.RTD
PL10      GSRW..SKDK DAYSSFGSRS DTRA...... ..KSSFFSDR GGSGSRGRFD
vasa      EGGFRGGQGG SRGGQGGSRG GQGGFRGGEG GFRGRLYENE DGDERRGRLD
p54       .......... .......... .......... MSTARTENPV IMGLSSQNGQ
ME31B     .......... .......... .......... .......... ..........
STE13     .......... .......... .......... .......... ..........
eIF4A     .......... .......... .......... .......... ..........
DBPA      .......... .......... .......... .......... ..........
DeaD      .......... .......... .......... .......... ..........
PR28      KPRFLNKQER SKQERLKENE ESLTPTQSDS AKVEIKKVNS RDDSFFNETN

        151                                                           200
p68       PLSGKKFGNP G......... ....EKLVKK KWNLDELPKF EKNFYQEHPD
RM62      RGGGNRFGGG GGFGDRRGGG GGGSQDLPMR PVDFSNLAPF KKNFYQEHPN
AN3       DRRQDGFDGM GNRSDKSGFG RFDRGNSRWS DDRNDEDD.. ....WSKPLA
PL10      ERGRSDYESV GSRGGRSGFG KFERGGNSRW CDKADEDD.. ....WSKPLP
vasa      REERGG..ER GERGD.GGFA RRRRNEDDIN NNNNIAEDVE RKREFYIPPE
p54       LRGPVKPTGG PGGGGTQTQQ QMNQLKNTNT INNGTQQQAQ SMTTTIKPGD
ME31B     .......... .......... MMTEKLNSGH TNLTSKGIIN DLQIAGNTSD
STE13     .......... .......... .......... ...MAESLIQ KLENA.NLND
eIF4A     .......... .......... .......... .......... ...MAGSAPE
DBPA      .......... .......... .......... .......... ..........
DeaD      .......... .......... .......... .......... ..........
PR28      DKKRNPSKQN GSKFHFSWNE SEDTLSGYDP IVSTRAIDLL WKGKTPKNAA
```

```
          201                                                       250
p68          LARRTAQEV. ......ETYR RS....KEIT VR..GHNCPK PVLNFYEANF
RM62         VANRSPYEV. ......QRYR EE....QEIT VR..GQ.VPN PIQDFSEVHL
AN3          PNDRVEQELF SGSNTGINFE KY....DDIP VEATGSNCPP HIESFHDVTM
PL10         PSERLEQELF SGGNTGINFE KY....DDIP VEATGNNCPP HIESFSDVEM
vasa         PSNDAIEIFS SGIASGIHFS KY....NNIP VKVTGSDVPQ PIQHFTSADL
p54          D..WKKTLKL PPKDLRIKTS DV....TSTK GNE.......  ....FEDYCL
ME31B        DMGWKSKLNC RQRTTRFKTT DV....TDTR GNE.......  ....FEEFCL
STE13        RESFKGQMKA QPVDMRPKTE DV....TKTR GTE.......  ....FEDYYL
eIF4A        GSQFD.....  .ARQFDAKMT EL....LGTE QEEFFTSYDE VYDSFDAMGL
DBPA         ..........  .......... .......... .......... ..........
DeaD         ..........  ......MMSY VD....WPPL ILRHTYYMAE FETTFADLGL
PR28         ESSYMGKHWT EKSLHEMNER DWRILKEDYA IVTKGGTVEN PLRNWEELNI

          251                                                 'A'motif 300
p68          ..PANVMDVI ARQNFTEPTA IQAQGWPVAL SG...LDMVG VAQTGSGKTL
RM62         ..PDYVMKEI RRQGYKAPTA IQAQGWPIAM SG...SNFVG IAKTGSGKTL
AN3          ..GEIIMGNI QLTRYTRPTP VQKHAIPIII EK...RDLMA CAQTGSGKTA
PL10         ..GEIIMGNI ELTRYTRPTP VQKHAIPIIK EK...RDLMA CAQTGSGKTA
vasa         ..RDIIIDNV NKSGYKIPTP IQKCSIPVIS SG...RDLMA CAQTGSGKTA
p54          ..KRELLMGI FEMGWEKPSP IQEESIPIAL SG...RDILA RAKNGTGKSG
ME31B        ..KRELLMGI FEKGWERPSP IQEAAIPIAL SG...KDVLA RAKNGTGKTG
STE13        ..KRELLMGI FEAGFERPSP IQEESIPIAL SG...RDILA RAKNGTGKTA
eIF4A        ..QENLLRGI YAYGFEKPSA IQQRGIVPFC KG...LDVIQ QAQSGTGKTA
DBPA         ..........  .......MTP VQAAALPAIL AG...KDVRV QAKTGSGKTA
DeaD         ..KAPILEAL NDLGYEKPSP IQAECIPHLL NG...RDVLG MAQTGSGKTA
PR28         IPRDLLRVII QELRFPSPTP IQRITIPNVC NMKQYRDFLG VASTGSGKTL
                                ** **   **          ***   *  ******
          301                                                       350
p68          SYLLPAIVHI NHQPFLE...  .......... RGDGPICLVL APTRELAQQV
RM62         GYILPAIVHI NNQQPLQ...  .......... RGDGPIALVL APTRELAQQI
AN3          AFLLPILSQI YADGPGDAMK HLQENGRYGR RKQFPLSLVL APTRELAVQI
PL10         AFLLPILSQI YTDGPGEALR AMKENGKYGR RKQYPISLVL APTRELAVQI
vasa         AFLLPILSKL LEDPHELEL.  ........GR PQVC....IV SPTRELAIQI
p54          AYLIPLLERL  .......... ........DL KKDNIQAMVI VPTRELALQV
ME31B        AYCIPVLEQI  .......... ........DP TKDYIQALVM VPTRELALQT
STE13        AFVIPSLEKV  .......... ........DT KKSKIQTLIL VPTRELALQT
eIF4A        TFCSGVLQQL  .......... ........DY SLVECQALVL APTRELAQQI
DBPA         AFGLGLLQQI  .......... ........DA SLFQTQALVL CPTRELADQV
DeaD         AFSLPLLQNL  .......... ........DP ELKAPQILVL APTRELAVQV
PR28         AFVIPILIKM SRSPPRPPSL KI........ .IDGPKALIL APTRELVQQI
             **       *                                  **  ****** **
          351                                                       400
p68          QQVAAEYC.R AC.....RLK STCIYGGAPK GPQIRDLERG VEICIATPGR
RM62         QQVATEFG.S SS.....YVR NTCVFGGAPK GGQMRDLQRG CEIVIATPGR
AN3          YEEARKFA.Y RS.....RVR PCVVYGGADI GQQIRDLERG CHLLVATPGR
PL10         YEEARKFS.Y RS.....RVR PCVVYGGADI GQQIRDLERG CHLLVATPGR
vasa         FNEARKFA.F ES.....YLK IGIVYGGTSF RHQNECITRG CHVVIATPGR
p54          SQICIQVSKH MG.....GAK VMATTGGTNL RDDIMRLDDT VHVVIATPGR
ME31B        SQICIELAKH L......DIR VMVTTGGTIL KDDILRIYQK VQLIIATPGR
STE13        SQVCKTLGKH M......NVK VMVTTGGTTL RDDIIRLNDT VHIVVGTPGR
eIF4A        EKVMRALGDY L......GVK VHACVGGTSV REDQRILQSG VHVVVGTPGR
DBPA         AGELRRLARF LP.....NTK ILTLCGGQPF GMQRDSLQHA PHIIVATPGR
DeaD         AEAMTDFSKH MR.....GVN VVALYGGQRY DVQLRALRQG PQIVVGTPGR
PR28         QKETQKVTKI WSKESNYDCK VISIVGGHSL EEISFSLSEG CDILVATPGR
                                     **          *      * ******
```

```
          401                    'B' motif                              450
p68          LIDFLECGKT NLRRTTYLVL DEADRMLDMG FEPQIRKIVD QI....RP..
RM62         LIDFLSAGST NLKRCTYLVL DEADRMLDMG FEPQIRKIVS QI....RP..
AN3          LVDMMERGKI GLDFCKYLVL DEADRMLDMG FEPQIRRIVE QDTMPPKG..
PL10         LVDMMERGKI GLDFCKYLVL DEADRMLDMG FEPQIRRIVE QDTMPPKG..
vasa         LLDFVDRTFI TFEDTRFVVL DEADRMLDMG FSEDMRRIMT HVTMRPE...
p54          ILDLIKKGVA KVDHVQMIVL DEADKLLSQD FVQIMEDIIL TL....PK..
ME31B        ILDLMDKKVA DMSHCRILVL DEADKLLSLD FQGMLDHVIL KL....PK..
STE13        VLDLAGKGVA DFSECTTFVM DEADKLLSPE FTPIIEQLLS YF....PK..
eIF4A        VFDMLRRQSL RPDHIKMFVL DEADEMLSRG FKDQIYDIFQ LL....PP..
DBPA         LLDHLQKGTV SLDALNTLVM DEADRMLDMG FSDAIDDVIR FA....PA..
DeaD         LLDHLKRGTL DLSKLSGLVL DEADEMLRMG FIEDVETIMA QI....PE..
PR28         LIDSLENHLL VMKQVETLVL DEADKMYDLG FEDQVTNILT KVDINADSAV
             *** *           *** *******    *   *  **

          451      'SAT'                                               500
p68          DRQTLMWSAT WPKEVRQLAE DFLKDYIHIN IGALELSANH NILQIVDVCH
RM62         DRQTLMWSAT WPKEVKQLAE DFLGNYIQIN IGSLELSANH NIRQVVDVCD
AN3          VRQTMMFSAT FPKEIQILAR DFLDEYIFLA VGRVG.STSE NITQKVVWVE
PL10         VRHTMMFSAT FPKEIQMLAR DFLDEYIFLA VGRVG.STSE NITQKVVWVE
vasa         .HQTLMFSAT FPEEIQRMAG EFLKNYVFVA IGIVG.GACS DVKQTIYEVN
p54          NRQILLYSAT FPLSVQKFMN SHLEKPYEIN L.MEELTLKG VTQYYAYVT.
ME31B        DPQILLFSAT FPLTVKNFME KHLREPYEIN L.MEELTLKG VTQYYAFVQ.
STE13        NRQISLYSAT FPLIVKNFMD KHLNKPYEIN L.MDELTLRG VTQYYAFVD.
eIF4A        KIQVGVFSAT MPPEALEITR KFMNKPVRIL VKRDELTLEG IKQFYVNVDK
DBPA         SRQTLLFSAT WPEAIAAISG RVQRDPLAIE IDSTDALPP. IEQQF.YETS
DeaD         GHQTALFSAT MPEAIRRITR RFMKEPQEVR IQSSVTTRPD ISQSY.WTVW
PR28         NRQTLMFTAT MTPVIEKIAA GYMQKPVYAT IGVETGSEPL IQQVVEYADN
               *  *****     *                 *           *

          501                                                          550
p68          DVEKDEKLIR LMEEI..MSE KENKTIVFVE TKRRCDELTR KM.RRDGWPA
RM62         EFSKEEKLKT LLSDIYDTSE SPGKIIIFVE TKRRVDNLVR FI.RSFGVRC
AN3          EMDKRSFLLD LLNA....TG KDSLTLVFVE TKKGADALED FL.YHEGYAC
PL10         EADKRSFLLD LLNA....TG KDSLILVFVE TKKGADSLED FL.YHEGYAC
vasa         KYAKRSKLIE ILSE....QA DG..TIVFVE TKRGADFLAS FL.SEKEFPT
p54          ERQKVHCLNT LFSRL..... QINQSIIFCN SSQRVELLAK KI.SQLGYSC
ME31B        ERQKVHCLNT LFSKL..... QINQSIIFCN STQRVELLAK KI.TELGYCC
STE13        ESQKVHCLNT LFSKL..... QINQSIIFCN STNRVELLAK KI.TELGYSC
eIF4A        EEWKLETLCD LYETL..... AITQSVIFVN TRRKVDWLTD KM.RSRDHTV
DBPA         SKGKIPLLQR LLSLH..... QPSSCVVFCN TKKDCQAVCD AL.NEVGQSA
DeaD         GMRKNEALVR FLEAE..... DFDAAIIFVR TKNATLEVAE AL.ERNGYNS
PR28         DEDKFKKLKP IVAKY..... .DPPIIIFIN YKQTADWLAE KFQKETNMKV
               *   *   *               ***          *

          551                                                          600
p68          MGIHGDKSQQ ERDWVLNEFK HGKAPILIAT DVASRGLDVE DVKFVINYDY
RM62         GAIHGDKSQS ERDFVLREFR SGKSNILVAT DVAARGLDVD GIKYVINFDY
AN3          TSIHGDRSQR DREEALHQFR SGKSPILVAT AVAARGLDIS NVKHVINFDL
PL10         TSIHGDRSQR DREEALHQFR SGKSPILVAT AVAARGLDIS NVKHVINFDL
vasa         TSIHGDRLQS QREQALRDFK NGSMKVLIAT SVASRGLDIK NIKHVINYDM
p54          FYIHAKMRQE HRNRVFHDFR NGLCRNLVCT DLFTRGIDIQ AVNVVINFDF
ME31B        YYIHAKMAQA HRNRVFHDFR QGLCRNLVCS DLFTRGIDVQ AVNVVINFDF
STE13        FYSHAKMLQS HRNRVFHNFR NGVCRNLVCS DLLTRGIDIQ AVNVVINFDF
eIF4A        SATHGDMDQN TRDIIMREFR SGSSRVLITT DLLARGIDVQ QVSLVINYDL
DBPA         LSLHGDLEQR DRDQTLVRFA NGSARVLVAT DVAARGLDIK SLELVVNFEL
DeaD         AALNGDMNQA LREQTLERLK DGRLDILIAT DVAARGLDVE RISLVVNYDI
PR28         TILHGSKSQE QREHSLQLFR TNKVQIMIAT NVAARGLDIP NVSLVVNFQI
               **   *   *      **        ** *  *  *****   *  ****
```

```
          601        'HRIGRXXR motif'                                650
p68          PNSSEDYIHR IGRTARSTKT GTAYTFFTPN .NI...KQVS DLISVLREAN
RM62         PQNSEDYIHR IGRTGRSNTK GTSFAFFTKN .NA...KQAK ALVDVLREAN
AN3          PSDIEEYVHR IGRTGRVGNL GLATSFFN.E KNI...NITK DLLDLLVEAK
PL10         PSDIEEYVHR IGRTGRVGNL GLATSFFN.E RNI...NITK DLLDLLVEAK
vasa         PSKIDDYVHR IGRTGRVGNN GRATSFFHPE KDR...AIAA DLVKILEGSG
p54          PKLAETYLHR IGRSGRFGHL GLAINLITYD DRFNLKSIEE QLGTEIKPIP
ME31B        PRMAETYLHR IGRSGRFGHL GIAINLITYE DRFDLHRIEK ELGTEIKPIP
STE13        PKNAETYLHR IGRSGRFGHR GLAISFISWA DRFNLYRIEN ELGTEIQPIP
eIF4A        PTQPENYLHR IGRSGRFGRK GVAINSVTKD DERMLFDIQK FYNVVIEELP
DBPA         AWDPEVHVHR IGRTARAGNS GLAISFCAPE E......... ..........
DeaD         PMDSESYVHR IGRTGRAGRA GRALLFVENR ERRLLRNIER TMKLTIPEVE
PR28         SKKMDDYIHR IGRTGRAANE GTAVSFVSAA ED...ESLIR ELYKYVRKHD
                 * **** ******     * *

          651                                                        700
p68          QAINPKLLQL VEDRGSGRSR GRGGMKDDR. RDRYSAGKRG GFNTFRDREN
RM62         QEINPALENL .......... .......AR. NSRYDGGGGP S.........
AN3          QEVPSWL... ...ENMAYEQ HHKSSSRGRS KSRFSGGFGA K....DYRQS
PL10         QEVPSWL... ...ENMAFEH HYKGGSRGRS KSRFSGGFGA R....DYRQS
vasa         QTVPDFLRTC GAGGDGGYSN QNFGGVDVRG RGNYVGDATN V....EEEEQ
p54          SNIDKSLYV. .......... .........A EYHSEPVEDE KP........
ME31B        KVIDPALYV. .......... .........A NVGASVGDTC NNSDLNNSAN
STE13        PSIDPSLYVF PNGDYQIPRP LTASADQVLA AQQAKGQEGY HNRPNNNRGG
eIF4A        ANVADLL... .......... .......... .......... ..........
DBPA         .......... .........AQ RANIISDMLQ IK..LNWQTP PANSSIATLE
DeaD         LPNAELLGKR RLEKFAAKVQ QQLESSDLDQ YRALLSKIQP TAEGEELDLE
PR28         PLNSNIFSEA VKNKYNVGKQ LSNEIIY... .......... ..........

          701                                                        750
p68          YDRGYSSLLK RDFGAKTQNG VYSAANYTNG SF..GSNFVS AGIQTSFRTG
RM62         .......... .RYGGGGGGG RFGGGGFKKG SLSNGRGFGG GG.......G
AN3          SGA.GSSFGS SRGGRS.... SGHGGSRGFG GG.YGGFYNS DGYGGNYGGS
PL10         SGASSSSFSS GRASNSRSGG GSHGSSRGFG GGSYGGFYNS DGYGGNY.SS
vasa         WD........ .......... .......... .......... ..........
p54          .......... .......... .......... .......... ..........
ME31B        EEGNVSK... .......... .......... .......... ..........
STE13        HPRGGGNRGG YRQSNRQPRY RGQQKAD... .......... ..........
eIF4A        .......... .......... .......... .......... ..........
DBPA         .......... .......... .......... .......... ..........
DeaD         TLAAALLKMA QGERTLIVPP DGPMRPKREF RDRDDRGPRD RNDRGPRGDR
PR28         .......... .......... .......... .......... ..........

          751                                                        800
p68          NPTGTYQNGY DSTQQYGSNV PNMHNGMNQQ AYAYPATAAA PMIGYPMPTG
RM62         GGGEGRHSRF D......... .......... .......... ..........
AN3          SQVDWWGN.. .......... .......... .......... ..........
PL10         QGVDWWGN.. .......... .......... .......... ..........
vasa         .......... .......... .......... .......... ..........
p54          .......... .......... .......... .......... ..........
ME31B        .......... .......... .......... .......... ..........
STE13        .......... .......... .......... .......... ..........
eIF4A        .......... .......... .......... .......... ..........
DBPA         .......... .AEMATLCID GGKKAKMRPG DVLGALTGDI GLDGADIGKI
DeaD         EDRPRRERRD VGDMQLYRIE VGRDDGVEVR HIVGAIANEG DISSRYIGNI
PR28         .......... .......... .......... .......... ..........

          801                                                   839
p68          YSQ....... .......... .......... .........
DBPA         AVHPAHVYVA VRQAVAHKAW KQLQGGKIKG KTCRVRLLK
DeaD         KLFASHSTIE LPKVCRVKCC NTLRALAF.. STSR.....
```

One of the best studied RNA helicases is eIF4A, which is part of the translation initiation factor eIF4F which includes the cap-binding protein eIF4E. The addition of mutant eIF4A to a rabbit reticulocyte lysate strongly inhibited translation of all tested mRNAs, including naturally uncapped mRNAs (Pause *et al.*, 1994). The RNA helicase activity of eIF4A is believed to facilitate the binding of mRNAs to the $tRNA^{met}$-activated 40S ribosomal subunit by unwinding highly structured 5' UTRs which otherwise impede translation initiation (Pause *et al.*, 1994). The unwinding activity of eIF4A is driven by ATP hydrolysis, and requires the presence of a further factor, eIF4B, which is an RNA-binding protein containing RRM (RNA recognition) motifs. eIF4A has been crosslinked to ATP as well as RNA (Pause *et al.* 1993). A series of mutational analyses combined with biochemical assays have suggested that: the ATPase A motif, AXXGXGKT, is essential for ATP binding; the ATPase B motif, VLDEAD, is essential for ATP hydrolysis, and the SAT and HRIGRXXR regions are also involved in ATP hydrolysis (Pause and Sonenberg, 1992). It was later shown that the HRIGRXXR region is also involved in the RNA-binding and helicase activity, because mutations in any of the three arginines in HRIGRXXR, including conservative substitutions to lysine, reduced crosslinking to RNA and abrogated helicase activity (Pause *et al.*, 1993; Pause *et al.*, 1994).

Apart from eIF4A, which is strongly conserved across species and functions as a general translation initiation factors, other RNA helicase subfamilies exist, some of which are developmentally significant. The RNA helicase encoded by the *Drosophila* maternal-effect posterior group gene *Vasa* is localised to the posterior pole of the oocyte, and is sequestered by the pole cells which are the progenitors of the germ line (Lasko and Ashburner, 1990). Females with both copies of *Vasa* deleted are sterile: they fail to complete oogenesis and cannot lay eggs (Lasko and

Ashburner 1988). The putative *Xenopus* homologue of Vasa, *XLVG1* (*Xenopus* Vasa-Like Gene) has recently been cloned (Komiya *et al.*, 1994). It is expressed in male and female germ cells but not detected in somatic tissues, and its function remains to be determined.

As in the case of *Vasa*, *Drosophila ME31B* mRNA is only detected in the adult female germ line. However, the authors do not exclude that a more sensitive assay might detect somatic expression: the closely related human RNA helicase p54 is detected in a variety of tissues (Lu and Yunis, 1992). *ME31B*, like *Vasa*, is expressed in oocytes, the transcripts persisting until early embryos. De Valoir *et al.* (1991) propose that the RNA helicase encoded by *ME31B* has a specialized role in the translational regulation of specific mRNAs that are under translational control during late oogenesis and embryogenesis in *Drosophila*, such as the mRNAs encoding ribosomal proteins or important developmental regulators such as *bicoid*. The function of human p54 is unknown at this stage. Its size is consistent with the size of the major protein that was crosslinked to ATP in *Xenopus* mRNP (see **Fig. 54**).

With respect to the novel *Xenopus* oocyte mRNP RNA helicase, a yeast homologue can also be described. Yeast cells enter the meiotic cycle in conditions of poor nutrition, forming haploid spores that are stored in the ascus. For example, when the fission yeast *Schizosaccharomyces pombe* is starved for nitrogen, the mating reaction is induced: cells are arrested in the G_1 phase of the cell cycle, and either remain stationary in a resting state or start sexual differentiation (Kitamura *et al.*, 1990). Various mutants have been isolated through genetic analysis identifying at least thirteen genes, including *STE13* (Kitamura *et al.*, 1990). The *ste13* mutation is not lethal, but the yeast fails to undergo G_1 arrest during nitrogen starvation and fails to enter meiosis (Maekawa *et al.*, 1994). *STE13* was cloned using

functional complementation, and found to encode a DEAD-box RNA helicase which belongs to the same subfamily as *Drosophila* ME31B, human p54 and the *Xenopus* oocyte mRNP protein. Maekawa *et al.* (1994) propose that the RNA helicase encoded by *STE13* has a role in regulating the translation of mRNAs encoding proteins required for meiosis. They found that the *ste13* mutant could not be complemented with a *Drosophila* cDNA encoding the RNA helicase Vasa, but instead could be complemented with a *Drosophila ME31B* cDNA to which *STE13* is more closely related. Therefore there is some specialised property of this as opposed to other RNA helicases which enables it to perform its particular function in meiosis.

The VLDEAD and HRIGRXXR motifs are essential for STE13 function: site directed mutagenesis changed DEAD to DPAD, and HRIGR to HGIGR, both abolishing the function of STE13 (Maekawa *et al.*, 1994). These mutations disrupted the RNA helicase activity. The major differences between RNA helicases lie outside the highly conserved catalytic regions, and probably confer the different cellular localisations and biochemical specificities upon the different members of the family. For example, in the N-terminus of *Drosophila* Vasa and its *Xenopus* homologue XLVG1 there is an arginine/glycine rich area (known as the "RGG box") which is present in various RNA-binding proteins. The RGG box, originally found in hnRNPU, is also present in hnRNPA1 and nucleolin; in general, sets of between six and eighteen RGG boxes have been found in various proteins, often interspersed with aromatic residues (Burd and Dreyfuss, 1994). This characteristic presumably enhances the ability of Vasa/XLVG1 to bind to RNA, whereas other helicases, such as eIF4A, require auxiliary factors, such as eIF4B. Due to the conservation in the key regions responsible for ATPase RNA helicase activity, all the DEAD-box proteins are expected to

share the general ability to bind ATP and unwind RNA: other domains such as RNA-binding domains have been added to specific helicases.

Further experiments are planned in order to gain more information about the *Xenopus* oocyte helicase p54: (i) oligonucleotide primers have been prepared to clone it via PCR. A PCR product has been generated after the completion of this project and is being used to isolate full length cDNAs from the oocyte cDNA library, in order to determine the entire amino-acid sequence and to study its expression. (ii) Purification of the RNA-helicase is being attempted using column chromatography with the resin Blue-Sepharose, used by Pause *et al.* (1993) to purify eIF4A, and/or ATP-Sepharose. Once the RNA helicase is purified, its biochemical activity will be characterized. (iii) Eventually, a purified RNA helicase could be used to generate a more specific antiserum, providing a useful experimental tool.

6.5 The mRNA masking process

When mRNAs are injected into the *Xenopus* oocyte cytoplasms, they are translated. This has been known since 1971, when Gurdon *et al.* originally injected globin mRNA into oocytes. However, more recently it was noted that when reporter constructs were injected into the *Xenopus* oocyte nucleus rather than cytoplasm, they were not be translated (Braddock *et al.*, 1990). Experiments have begun to address where and how mRNA masking occurs using microinjection techniques, and there is now some agreement that mRNA masking is initiated in the cell nucleus (Ranjan *et al.*, 1993; Braddock *et al.*, 1994; Bouvet and Wolffe, 1994).

Bouvet and Wolffe have studied the translational fate of a specific masked mRNA in the *Xenopus* oocyte, the histone H1 mRNA (Bouvet and Wolffe, 1994). In *Xenopus* oocyte chromatin, histone H1 is replaced by an alternative form, B4 (Smith *et al.*, 1988). At the same time, histone H1

mRNA is transcribed and masked, to be recruited for translation by the developing embryo. Bouvet and Wolffe (1994) measured the translational efficiency of injected or *in vivo* synthesized H1 mRNA by adding labelled amino-acids. It was found that when *in vitro* synthesized H1 mRNA was microinjected into the cytoplasms, it was able to translate; when injected into the nuclei, it was efficiently transported to the cytoplasm and also translated. The situation changed radically when a DNA construct containing the histone H1 gene was injected into the nuclei. It was found that *in vivo* transcribed histone H1 mRNA was masked from translation, that the masking effect was independent of the type of promoter used to drive its expression, and that masking was independent of the presence or absence of an intron in the construct. Masked histone H1 mRNPs were extracted and injected into the cytoplasms of a second oocyte, and did not translate; however when deproteinized, the mRNAs were able to translate. In other words, the histone H1 mRNA was only packaged into translationally silent mRNPs when transcribed *in vivo* from a microinjected DNA construct. The authors therefore propose that mRNA masking is transcription-dependent. However, they also show that microinjected radio-labelled histone H1 mRNA synthesized *in vitro* could be crosslinked to YB proteins, suggesting that although masking from translation is transcription dependent, binding to YB proteins is not.

At the same time, Braddock *et al.* (1994) have addressed similar issues. Braddock *et al.* injected fully capped and poly-adenylated mRNAs using CAT or LUC reporter constructs with which they could study translation. They found that when *in vitro* synthesized reporter mRNAs were injected into the nuclei, they entered the mRNA masking pathway. This contrasts with the fate of histone H1 mRNA, microinjected into nuclei (Bouvet and Wolffe, 1994; see above). In addition, the masked mRNA pathway was

avoided when an mRNA construct containing an intron was used. A single point mutation in the splice donor site reversed this effect. Braddock *et al.* propose that in the absence of an intron, or in the presence of a faulty intron, mRNAs enter the masking pathway, and they point out that the mRNAs for c-*mos*, the core histones and histone H1, which are masked in the oocyte, do not have introns. These results are at variance with the findings of Bouvet and Wolffe (1994). It is not clear whether or not all masked mRNAs are devoid of introns. Although the presence or absence of introns may influence the translational fate of an mRNA, other mRNA-specific effects might govern translatability as well. Further experiments are needed to clarify these issues.

Braddock *et al.* (1994) also considered the role of YB proteins and their phosphorylation in the nuclear mRNA masking process. They found that masking could be reversed by the co-injection of anti-pp60 and anti-p54 IgG, but not by control IgG delivered to the nuclei. The same antisera were used in an *in vitro* binding assay to block RNA-binding (**Fig. 27**). It follows that one of the 54/52 kDa proteins is involved in the nuclear mRNA masking process in addition to the YB proteins.

Messenger RNA masking was also blocked by delivering into nuclei either chemical inhibitors of casein kinase II or IgG raised against casein kinase II subunits. Moreover, Braddock *et al.* (1994) found that treating the oocytes with progesterone also reversed masking, and resulted in the translation of the previously masked reporter constructs. Progesterone acts *in vivo* to promote the maturation of oocytes, which involves the recruitment of various mRNAs for translation (mRNA recruitment starts at oocyte maturation and continues throughout fertilization and early embryogenesis). The un-masking process was blocked by okadaic acid, an inhibitor of phosphatase, in a dose-responsive curve. This is in agreement

with the findings that the phosphorylation of YB proteins is necessary to maintain mRNPs translationally repressed. Messenger RNA un-masking was independent of poly-adenylation, because the addition of a terminal cordycepin molecule, an analogoue of adenosine which blocks further poly-adenylation, did not affect the translational recruitment of the reporter constructs after progesterone treatment. This control was necessary, because the recruitment of various maternal mRNAs at oocyte maturation coincides with further cytoplasmic polyadenylation (McGrew *et al.*, 1989; Simon *et al.*, 1992; Sheets *et al.*, 1994). The precise relationship between mRNA masking by the YB proteins and polyadenylation remains unclear.

Masking has also been studied in a somatic context in *Xenopus*, using the cell line A6 (Ranjan *et al.*, 1993). These authors studied the expression of an hsp70-CAT DNA construct which was co-transfected together with a CMV-FRGY2 construct. The *hsp70* promoter contains a Y-box sequence, as is the case for other promoters active in germ cells (Tafuri and Wolffe, 1990). As might be expected, they found that the co-transfection resulted in a 3-5 fold increase in the transcription of the hsp70-CAT reporter. At the same time, there was a concurrent four-fold decrease in its translatability. The drop in translation was also achieved when the Y-box sequence in the *hsp70* promoter was mutated, suggesting that the transcriptional effects mediated by the action of FRGY2 on the Y-box sequence are uncoupled to its mRNA-packaging translational effects.

In addition, Ranjan *et al.* (1993) co-injected constructs only expressing the CSD or the TD of FRGY2. They found that the CSD alone was able to enhance transcription, just as the intact protein does. This is in accordance with the ability of the prokaryotic cold-shock protein to promote the transcription of genes with Y-box like sequences in their promoters (La

Teana *et al.*, 1991; Jones *et al.*, 1992). In contrast, the TD alone had no transcriptional effect, but was able to repress translation of the construct. This is consistent with the biochemical data presented in chapter 4, according to which the TD binds efficiently to mRNA. Ranjan *et al.* (1993) found that FRGY2 protein and hsp70-CAT mRNA co-fractionated in Nycodenz gradients in the mRNP fraction, which is consistent with FRGY2 packaging the hsp70-CAT mRNA construct thus determining its translational repression.

At this point the question arises: are the YB proteins indeed "RNA-histones", as has been claimed by Tafuri and Wolffe (1993b)? Following the identification of YB proteins in masked mRNP, Tafuri and Wolffe (1993) studied the distribution of YB proteins and specific mRNAs in *Xenopus* oocyte mRNPs using Nycodenz gradients. The mRNAs were divided into two categories: TFIIIA and FRGY2 mRNAs are translatable in the oocytes, whereas histone H1 and FRGY1 mRNAs belong to the masked message pool. They found that FRGY2 co-fractionated with all the above mRNAs, and suggested that the YB proteins are "RNA histones", in the sense that they package *all* mRNAs in the same way that histones package essentially the entire DNA genome. According to this model, the assembly of mRNP particles containing YB proteins is a default state, and the translational recruitment of mRNAs is mediated by proteins bound to the message-specific sequences in the 5' or 3' UTR (see reviews by Standart, 1993; Spirin, 1994). In confirmation of this idea, the separation of *Xenopus* oocyte mRNPs on Cs_2SO_4 density gradients presented in chapter 3 suggests that the major mRNP proteins, including the YB proteins, peak in the same fractions as translatable mRNAs, such as nucleolin, as well as masked mRNAs, such as c-*mos*.

How are these findings reconciled with the observation that YB proteins alone, *in vitro*, can prevent translation (Kick *et al.*, 1987) and that mRNPs immunoprecipitated with anti-pp60 contained some but not all mRNAs (Crawford and Richter, 1987)? Both rabbit reticulocyte p50 and *Xenopus* pp60 appear to achieve a translational block when reconstituted with the globin mRNA *in vitro* (Kick *et al.*, 1987; Minich and Ovchinnikov, 1992). The packaging of mRNA by YB proteins in these examples might be enough to hinder the progression of translation, for example by preventing the binding of translation initiation factors. However, *in vivo* various other mRNP proteins are likely to be involved: other, less abundant regulators, might bind to specific classes of mRNA and target them for immediate or delayed translation. However, by default, the exclusive packaging by YB phosphoproteins might result in translational repression.

There is a further development with respect to what determines mRNA masking in the nucleus. In an earlier study by Braddock *et al.*, (1990) an initial observation was made that an HIV-1 promoter fused to a CAT or LUC reporter could drive transcription, but translation of the construct was repressed. However, when a CMV promoter was used instead, the reporters were able to translate. This suggested that differences between promoters could determine differences in the translatability of the transcripts. In contrast, Bouvet and Wolffe (1994) found that different promoters do not influence the masking of H1 mRNA in oocytes; however, their survey of promoter sequences was by no means exhaustive. In recent experiments, Gunkel *et al.* (1995), have investigated this issue further. By using similar HIV-1 promoter/CAT constructs, they have defined an area -340 from the transcription start site, which appears to be responsible for directing transcripts into the nucleus-dependent

translational repression pathway. More specifically, they find that a three nucleotide substitution at position -340, which coincides with a sequence that can bind the haematopoietic transcription factor GATA-1, abolishes the translational block. Using gel-mobility shift assays, Gunkel *et al.* (1995) confirm the presence of an oocyte factor that binds to this sequence, and that binding to this site correlates with the ability to inhibit translation. They confirm again that anti-pp60 and anti-p54 IgG, and a specific inhibitor of casein kinase II, can reverse the translational block which originates in the nucleus, as was shown earlier (Braddock *et al.*, 1994).

Thus Gunkel *et al.* (1995) propose that a GATA-binding activity is present in oocytes and is necessary for mRNA masking by somehow influencing the recruitment of YB phosphoproteins onto transcripts. The DNA-binding protein GATA-1 is a positively acting transcription factor, regulating most erythroid genes (Evans *et al.*, 1990). GATA-1, as well as related proteins GATA-2, GATA-3 and GATA-4, have been cloned from *Xenopus*, and were detected in larval and adult tissues from early gastrula onwards (Zon *et al.*, 1991). Members of this family contain two cysteine X 4 zinc fingers, the second of which has been implicated in the recognition of the DNA target sequence (Yang and Evans, 1992). The crystallographic structure of the complex between the second zinc finger of chicken GATA-1 and its dsDNA target (AGATAAAC) has been determined (Omichinski *et al.*, 1993). Binding is described in terms of specific interactions with both strands of the DNA target sequence, between the protein and the major and minor grooves, as well as the sugar/phosphate backbone. If the GATA-binding activity detected by Gunkel *et al.* (1995) is not due to *Xenopus* GATA-1 protein (only expressed in embryos and adults), could there be a related protein possessing similar zinc fingers?

The oocyte protein C4SR (Appendix D) has two cys X 4 zinc fingers which were selected using a riboprobe binding screen (see section 5.1). The second zinc finger in C4SR has the following similarities with the second (DNA-binding) zinc finger in GATA-1:

```
GATA-1      ...CSNCQTSTTTLW-RRSPMGDPVCNAC--GLYYKLHQ...
                 |   |        | |||    :   || |     | ||
C4SR        ...CKTC---GNVNWARRS---E--CNMCNTPKYAKLEE...
```

C4SR also contains a highly acidic stretch, which is also found in gene regulators such yeast GAL4. In GAL4, the acidic region enhances the transcriptional effects mediated by a zinc-finger domain (Lin *et al.*, 1988). At the C-terminus, C4SR contains an SR-rich area, which is found in a number of hnRNP and splicing factors in which it enhances protein:protein interactions and RNA-binding (Biamonti and Riva, 1994). If C4SR, which is an oocyte-specific factor (Sommerville *et al.*, unpublished observations), has the above features, might it correspond to the GATA-binding, mRNA masking factor described by Gunkel *et al.* (1995)? No biochemical data is available yet about C4SR. The immediate test would be to synthesize the protein using a standard expression vector system, and to test its binding to dsDNA oligonucleotides containing the GATA sequence.

Whatever the identity of the GATA-binding mRNA masking factor, Gunkel *et al.* (1995) have opened a new and unexpected line of research, showing that *promoters* can influence translational control. They present the notion of "productive" and "non-productive" transcription. In "productive" transcription, nascent transcripts would recruit positive factors required for translation, whereas "non-productive" transcription

would result in a default masking pathway involving the YB phosphoproteins.

One immediate prediction is that promoter sequences of genes expressed in oocytes, whose mRNAs are masked, contain a similar GATA binding site. One example may be considered: the promoter of the enzyme cdk2 (cyclin-dependent kinase 2). This enzyme is involved in regulating entry into the S-phase of the cell cycle, and its promoter has recently been cloned (Olive *et al.*, 1994). The mRNA encoding cdk2 is a masked message, and is only recruited for translation when oocyte maturation and fertilisation are followed by the intensive cell proliferation activity in early embryogenesis. Its promoter was shown to be very active in oocytes. Inspection of the promoter sequence reveals the following features. (i) There are seven putative E2F-binding sites. E2F is a transcription factor known to be involved in the positive regulation of various cellular genes involved in cell proliferation, such as c-*myc*, c-*myb*, cyclin A, DNA polymerase, thymidine synthetase and thymidine kinase (Olive *et al.*, 1994). (ii) There are also two putative Y-box sequences: as has been discussed, YB proteins are thought to be involved in the positive regulation of germ cell as well as cell proliferation genes. (iii) A TATA-box is absent, as is the case in other cell proliferation genes such as c-Ki-*ras* (Pestov *et al.*, 1991) or PCNA (proliferating cell nuclear antigen, Travali *et al.*, 1989). (iv) Moreover, at position -442, there is the sequence GGATAA; at position -164, GGATAT, and at position -10, the sequence TGATAA. These sequences, and especially the -10 sequence, compare favourably with the GATA-binding consensus (A/T)-G-A-T-A-(A/G) (Omichinski *et al.*, 1993). It may be significant that this sequence is very close to the transcriptional start site, immediately following one of the putative Y-box sequences. Likewise, a Y-box site is also present at position -370 in the

NRE (negative regulatory element) of the HIV promoter, only thirty bases upstream of the GATA site described by Gunkel *et al.* (1995).

Finally, having discussed the biochemistry of the YB proteins, the presence and function of an associated casein kinase II, as well as the identification of an associated RNA helicase, simple models are proposed that attempt to summarise what is known about the *Xenopus* oocyte mRNA masking process. It needs to be stressed that many other unidentified proteins are probably involved, and the scheme is therefore likely to be an oversimplification. Some of these other proteins will be mRNA-specific regulators, while others might be more general mRNP packaging proteins. **Fig. 56** proposes the involvement of promoters in the mRNA masking process by illustrating two possible scenarios, and **Fig. 57** suggests the cytoplasmic process of mRNA storage and un-masking.

6.6 Concluding remarks

Having obtained mRNP particles, means of isolating YB proteins were developed. The YB proteins in particular were characterised in terms of their RNA-binding properties, differentiating between the CSD and the TD. The CSD is suggested to bind to polypurines and to be sensitive to Mg^{2+}. Individual basic/aromatic islands within the TD were UV-crosslinked to riboprobe, and their binding specificity was influenced by the presence of polycations. This variability in binding specificities would explain the ability of the YB proteins to bind to a wide range of mRNA sequences. The presence of an mRNP-associated kinase activity, thought to be of the casein kinase II type, was confirmed. Phosphorylation of the YB proteins was shown to improve the stability of protein/RNA complexes.

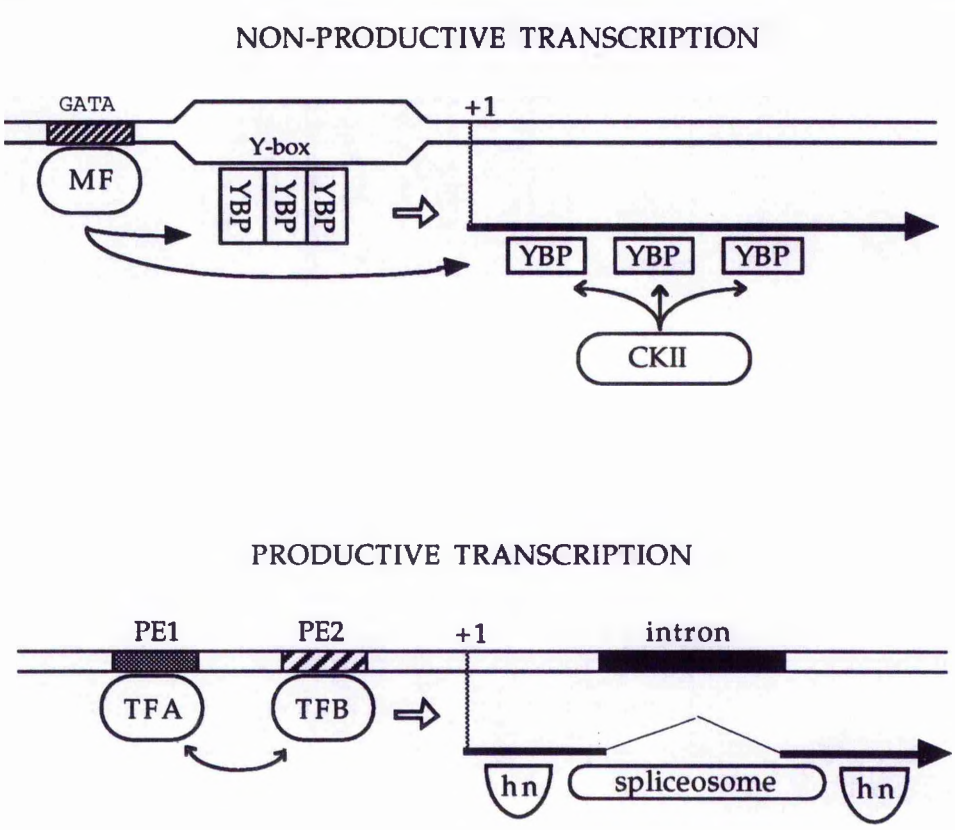
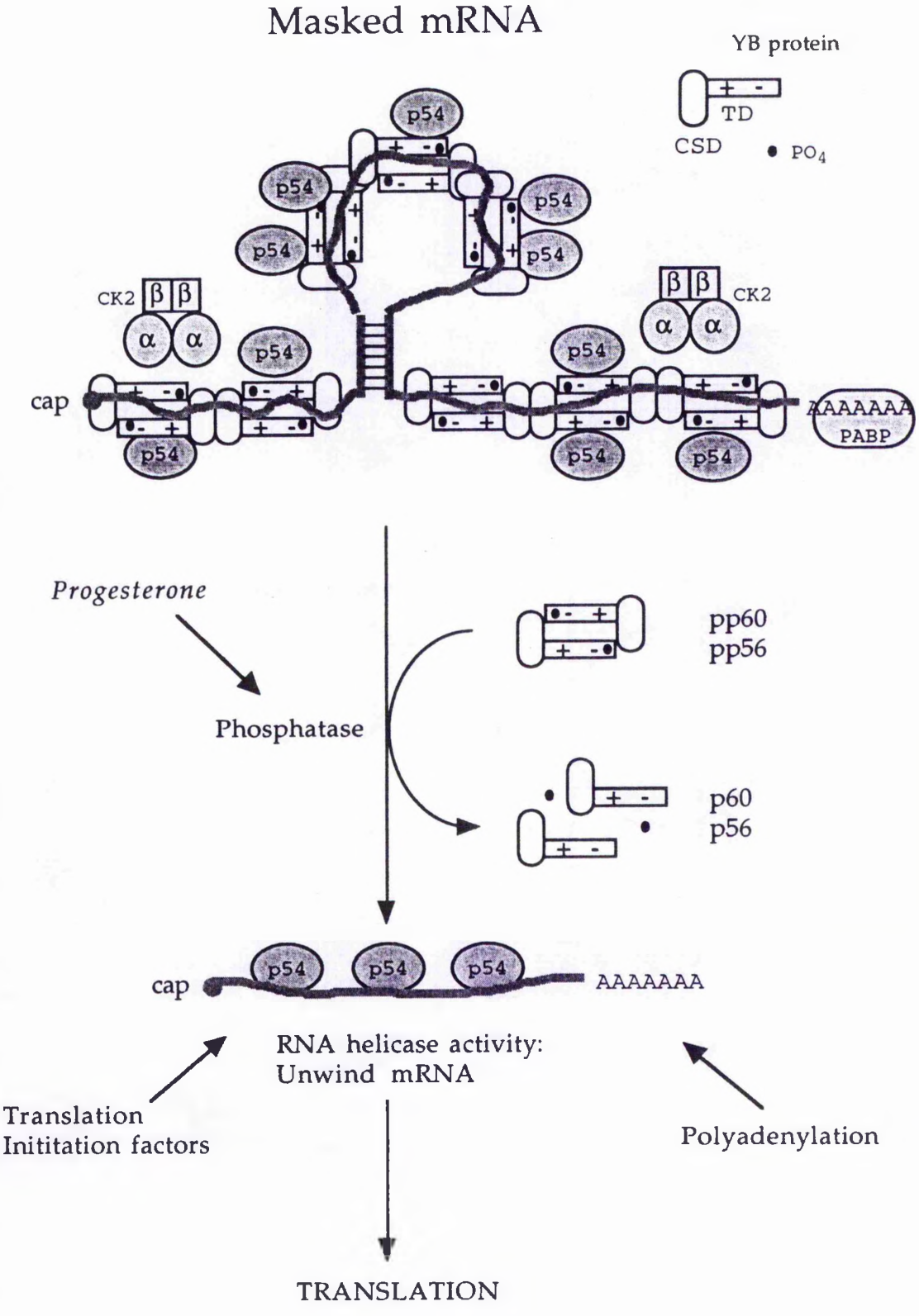


Figure 56 Promoter-driven mRNA masking. Two transcription scenarios are suggested. In "non-productive transcription", Gunkel *et al.* (1995) suggest that a GATA-binding factor is required for mRNA masking (MF, masking factor). Binding of YB proteins (YBP) to the same promoter is shown, but is not necessarily required for mRNA masking (Bouvet and Wolffe, 1994). Binding of the YB proteins to mRNA *is* however necessary for masking, and requires casein kinase II phosphorylation (Kick *et al.*, 1987; Braddock *et al.*, 1994). In "productive transcription", different transcription factors TFA and TFB bind to promoter elements PE1 and PE2, activating transcription. Splicing components and hnRNP proteins bind immediately to the nascent transcript which is processed and enters the translation pathway. The presence of faulty splice sites, or the absence of an intron, could drive the transcript into the masking pathway (Braddock *et al.*, 1994).

Figure 57 The process of mRNA un-masking. Masked mRNAs can be stored for many months in the cytoplasm. The key mRNA masking proteins are the YB proteins, consisting of a cold shock domain (CSD) and tail domain (TD). Efficient packaging by YB proteins is initiated in the nucleus and requires phosphorylation by an associated casein kinase II, CKII (Braddock *et al.*, 1994; Gunkel *et al.*, 1995; Sommerville *et al.*, unpublished). The YB proteins can multimerize (Tafuri and Wolffe, 1992; Horwitz *et al.*, 1994). The *Bacillus subtilis* protein CspB dimerizes via the β4 strand of the CSD (Schnuchel *et al.*, 1993; Schindelin *et al.*, 1993), and *Xenopus* oocyte phospholabelled YB proteins can also be crosslinked to mRNA, treated with RNase and recovered as dimers (Marello *et al.*, 1992). Other proteins are found in the mRNP, most of which are still unidentified, and are not shown here. Some are likely to be specific regulators of specific classes of mRNAs, while others may be more abundant, general mRNP proteins (the "RNA histones", Tafuri and Wolffe, 1993b). Here "p54" refers to the associated RNA helicase. At oocyte maturation, promoted by progesterone, a phosphatase activity dephosphorylates the YB proteins, thereby reversing the packaging of mRNA by YB proteins (Braddock *et al.*, 1994). The RNA helicase could intervene at this point to prepare the mRNAs for translation by unwinding structures that would hinder translation, while at the same polyadenylation and various translation initiation factors come into play.


Masked mRNA
YB protein
TD
CSD
PO4
p54
CK2
β β
α α
cap
AAAAAAA
PABP
Progesterone
Phosphatase
pp60
pp56
p60
p56
RNA helicase activity:
Unwind mRNA
Translation
Initiation factors
Polyadenylation
TRANSLATION

Two cDNA cloning techniques were used to isolate a variety of clones, including an intriguing new protein, AB21. Its yeast homologue RPD3 has a general role in gene regulation which has not yet been characterised biochemically. There was no conclusive evidence that AB21 is an mRNP protein: however, it is presumed to have a significant role in oogenesis at the level of gene regulation. On the other hand, sequencing of mRNP peptides has revealed the presence of a putative RNA helicase belonging to the DEAD-box family. Work is under way to clone it, and characterise it biochemically so as to define its role in the mRNA masking and/or un-masking process.

It appears that the process of mRNA masking is initiated in the nucleus. More biochemical data relating to the nuclear masking process is required: current crosslinking experiments are confirming that YB proteins can be crosslinked to riboprobe in nuclear extracts, and that this process is influenced by the nuclear casein kinase II activity. These experiments will, it is hoped, complement the recently reported nuclear microinjection experiments (Ranjan *et al.*, 1993; Bouvet and Wolffe, 1994; Braddock *et al.*, 1994; Gunkel *et al.*, 1995). The mRNA masking process has not yet been understood in full, but research continues at a rapid pace.

Aberrations in the mRNA masking process, wherever it occurs, could result in the inappropriate translational regulation of certain mRNAs. YB proteins have been involved in cell proliferation as well as the transcription of retroviruses: it may be significant that productive HIV infection has been associated with cell proliferation (Schuitemaker *et al.*, 1994). Although the translational masking of reporter constructs whose transcription is driven by the HIV-1 promoter is described in *Xenopus* oocytes (Braddock *et al.*, 1990, 1994; Gunkel *et al.*, 1995), masking of HIV transcripts could occur *in vivo* in T-cells. If the YB proteins, as well as

newly identified mRNP proteins, are significant players in these processes, they could become the targets of pharmacological and gene therapy.

In summary, the results have expanded our understanding of the structure of the *Xenopus* oocyte masked mRNPs, but there is clearly much more to be learned. Many mRNP proteins are still to be identified, and the ones that have been identified need to be characterised further. The phenomenon of mRNA masking no doubt occurs in other tissues, and it is hoped that these findings will also apply to them. The words "masked mRNP" (meaning stored, translationally repressed mRNAs + associated proteins) could be replaced by a simpler term, such as "maskosome" (as opposed to Spirin's "informosome").

Appendices

Appendix A

Xenopus laevis cyclin B1 mRNA
Length: 1395 bp

```
1     CCACTTAGTG AGACGTCTCT TTCAGACTGG GGGGCTGCAG TGTGACTTGT
51    GGGTACAAAT AGTAAAGCTG ACTGCAGGTT TGCGCTTGAG AAAATGTCGC
101   TACGAGTCAC CAGAAACATG CTGGCAAATG CAGAAAACAA TGTGAAAACC
151   ACTTTGGCTG GAAAGAGGGT TGTTGCTACC AAACCAGGGT TGAGACCTCG
201   TACAGCCTTG GGAGACATTG GAAACAAGGC AGAGGTGAAA GTGCCAACAA
251   AAAAGGAATT AAAGCCAGCA GTAAAAGCTG CCAAGAAGGC AAAACCTGTT
301   GACAAATTGT TGGAGCCTCT TAAAGTGATA GAAGAGAATG TTTGCCCTAA
351   ACCTGCTCAG GTTGAACCCA GCTCACCAAG CCCAATGGAA ACATCTGGTT
401   GCCTCCCTGA TGAGCTCTGC CAGGCTTTCT CTGATGTCCT CATTCACGTT
451   AAAGATGTTG ATGCTGATGA TGATGGCAAC CCAATGCTGT GCAGTGAATA
501   TGTCAAGGAC ATTTATGCTT ACCTGAGGAG CCTTGAGGAT GCACAAGCAG
551   TCAGACAAAA CTACCTTCAT GGACAGGAAG TCACAGGCAA CATGCGTGCC
601   ATTTTGATTG ACTGGCTGGT CCAGGTGCAA ATGAAATTCC GTCTACTGCA
651   GGAGACAATG TTCATGACTG TTGGCATAAT TGACCGCTTT CTGCAGGAAC
701   ATCCAGTTCC CAAAAACCAG CTACAGCTTG TGGGGGTCAC GGCTATGTTC
751   CTTGCTGCTA AATATGAAGA GATGTACCCA CCAGAAATTG GAGACTTTAC
801   ATTTGTAACT GATCACACAT ACACAAAGGC TCAAATTCGG GACATGGAAA
851   TGAAGATACT TAGGGTGCTA AAGTTTGCAA TTGGCCGACC CTTACCCCTG
901   CACTTTCTTC GGAGAGCTTC TAAAATTGGA GAGGTAACTG CTGAACAGCA
951   TAGTTTAGCC AAATATTTGA TGGAACTTGT GATGGTGGAT TATGATATGG
1001  TACATTTCAC GCCTTCCCAA ATAGCAGCTG CTTCCTCCTG CTTGTCTCTC
1051  AAAATCTTAA ATGCAGGTGA CTGGACCCCA ACACTCCATC ACTATATGGC
1101  TTACTCTGAA GAAGATCTAG TCCCTGTTAT GCAGCATATG GCCAAGAACA
1151  TCATCAAGGT GAACAAAGGA CTAACCAAGC ATCTGACTGT TAAGAACAAG
1201  TATGCTAGCA GCAAACAAAT GAAGATCAGC ACGATTCCAC AGCTGAGGTC
1251  AGATGTTGTT GTGGAAATGG CCCGCCCACT CATGTGAAGG ACTACGTGGC
1301  ATTCCAATTG TGTATTGTTG GCACCATGTG CTTCTGTAAA TAGTGTATTG
1351  TGTTTTTAAT GTTTTACTGG TTTTAATAAA GCTCATTTTA ACATG
```

Appendix B

Xenopus laevis β-tubulin clone 16.2, 3' end fragment

Length: 295 bp

EcoRI fragment: from an interanl site in the β-tubulin mRNA to the EcoRI site in the λgt11 linker sequence (italicized, after the poly(A) tail). Note the presence of purine-rich as well as pyrimidine-rich regions.

```
1    GAATTCACTG AGGCCGAGAG CAACATGAAC GACCTGGTGT CTGAGTACCA
51   ACAGTACCAG GATGCCACGG CTGAGGAGGA GGGAGAGTTT GAGGAAGGGG
101  AAGAGGAGGA AAATGCCTAA AGCTCCTTAT CACTTGTAAA TTATTCATCC
151  ATTTCTGCTC GTTTGTTCCA TTTATTTGTC ACCTGCATTT CTCTTCTCTC
201  CAGTTTCAAT GTTACCAGTT GTACAGAACG TTCCGTTCAT TAAAAGCATT
251  TTTCATACTG AAAAAAAAAA AAAAAAAAAA GGCGGCCGCG AATTC
```

Appendix C

Database entry XLZNF (for C4SR)

LOCUS XLZNF 365 bp RNA VRT 03-MAY-1993
DEFINITION X.laevis mRNA for zinc finger.
ACCESSION X70647
KEYWORDS zinc finger.
SOURCE clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Amphibia; Lissamphibia; Anura; Archeobatrachia; Pipoidea; Pipidae.
REFERENCE 1 (bases 1 to 365)
AUTHORS Ladomery,M.R. and Sommerville,J.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1993) M.R. Ladomery, St Andrews University, Dept of Biology, Bute Buildings, St. Andrews, Fife, KY16 9TS, Scotland, UK
STANDARD full automatic
COMMENT
The sequence is an internal fragment including two putative C4 zinc-fingers, cloned as a fusion protein by virtue of RNA-binding. The estimated full transcript size is about 2.4kb. The mRNA is expressed at low level throughout oocyte development and is possibly a maternal message. The protein shows some similarities with a variety of steroid receptors.
NCBI gi: 65274
FEATURES Location/Qualifiers
source 1..365
/organism="Xenopus laevis"
/tissue_type="total ovarian tissue"
/clone_lib="full ovarian cDNA library"
BASE COUNT 143 a 47 c 100 g 75 t
ORIGIN

```
  1  tgactggatc tgtccggata aaaagtgtgg gaatgtaaac tttgccagac
 51  gaaccagttg caacagatgt gggcgagaaa aaacaactga agctaaaatg
101  atgaaagctg gtggaactga aataggaaaa accttagccg agaaaagccg
151  tggattgttt agtgctaatg actggcaatg caaaacatgc gggaatgtaa
201  attgggccag aagatcagaa tgtaatatgt gcaatacacc aaagtatgca
251  aaactagagg aaagaacagg ttatggagga ggtttcaatg aacgtgaaaa
301  tgtagaatat atagaacgcg aggaatctga tggagagtat gatgagtttg
351  gccggaaaaa aaaaa
```

Appendix D

C4SR amino-acid sequence

Length: 337 amino-acids
Molecular weight: 37.9 kDa
Isoelectric point: 10.20

Features: two cys x 4 zinc-fingers (cysteines are in bold) are followed by a highly acidic region and an SR-rich region. The sequence DLSKY is repeated twice (underlined; positions 163 and 192) in the context of the acidic region.

```
1    MSTKNFRVSD GDWICSDKKC GNVNFARRTS CNRCGREKTT EAKMMKAGGT
51   EIGKTLAEKS RGLFSANDWQ CKTCGNVNWA RRSECNMCNT PKYAKLEERT
101  GYGGGFNERE NVEYIERDES DGEYDEFGRK KKKYRGKPVA PKSVSKGDEK
151  EGEAEEEEEE DQDLSKYKLD DDEEERKDEE DDDDDEEEDD GDLSKYNLAS
201  EDEDDSKNKK STRSSRSKSR SSHSSSHTSS RSRSRSRSRS SSSSKSSSRG
251  HSRSPGSKSR SSSRSHRGSS SPRKRSYASS RSSSSPERGK KRSRSRSSSG
301  DRKKKRSRSR STERRRGSSS GSSHSGSHSS NSKKKQN
```

Appendix E

Database entry XLS27H (*Xenopus laevis* ribosomal protein S27)

LOCUS XLS27H 406 bp RNA VRT 03-MAY-1993
DEFINITION X.laevis mRNA homologous to rat ribosomal protein S27.
ACCESSION X71350
KEYWORDS ribosomal protein S27 homologue.
SOURCE clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Amphibia;
Lissamphibia; Anura; Archeobatrachia;Pipoidea; Pipidae.
REFERENCE 1 (bases 1 to 406)
AUTHORS Ladomery,M.R. and Sommerville,J.
TITLE Direct Submission
JOURNAL Submitted (05-APR-1993) M.R. Ladomery,
University of St. Andrews, School of Biological and
Medical Sciences, Bute Building, St. Andrews,
Fife, KY16 9TS Scotland, UK
STANDARD full automatic
COMMENT
The protein is highly homologous to the rat ribosomal protein S27
(acc# X59375). It posseses clusters of charged residues and may
have a zinc-finger like structure of the C4 type.
NCBI gi: 297124
FEATURES Location/Qualifiers
source 1..406
/organism="Xenopus laevis"
/germline
/tissue_type="ovarian"
/cell_type="oocyte"
/clone_lib="X. laevis oocyte cDNA expression library"
CDS 17..271
/gene="XENLA.S27homologue"
/note="putative zinc-finger protein; NCBI gi: 297125"
/codon_start=1
/product="ribosomal protein S27 homologue"
/translation="MPLAKDLLHPTPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH"
BASE COUNT 106 a 109 c 104 g 87 t
ORIGIN

```
1   cgcagctccg gagaacatgc cactcgctaa ggatctcctg cacccaactc ccgaggagga
61  gaagaggaaa cacaagaaga agcgcctggt ccagagccca aactcctatt tcatggatgt
121 caaatgccca ggttgttaca agatcaccac agtgttcagt cacgcgcaaa ctgtcgtcct
181 gtgtgtcggc tgctccaccg tcctgtgtca gcccacgggg ggcaaagctc ggctcacaga
241 aggttgttcc ttcagaagaa agcaacacta atgttcctgg ctggagatac aatggggagt
301 aagctgtgag taaggactgg cccatgggaa ccattattac agggtgactc ctatgtggac
361 cctgtagcct caccctgtc tcaataaatt ttggatacaa gttgct
```

Appendix F

Alignment between *Xenopus* and *Rattus* ribosomal protein S27

XS27: *Xenopus*
RS27: *Rattus*
Cysteines are in bold; (| = identity); (: = conserved substitution)

```
      1                                                    40
XS27  MPLAKDLLHP TPEEEKRKHK KKRLVQSPNS YFMDVKCPGC
      ||||:|||||  : ||||:||| ||||||||||  ||||||||||
RS27  MPLARDLLHP SLEEEKKKHK KKRLVQSPNS YFMDVKCPGC

      41                                                   84
XS27  YKITTVFSHA QTVVLCVGCS TVLCQPTGGK ARLTEGCSFR RKQH
      ||||||||||  ||||||||||  ||||||||||  ||||||||||  ||||
RS27  YKITTVFSHA QTVVLCVGCS TVLCQPTGGK ARLTEGCSFR RKQH
```

Rattus S27:

Length:	84
Molecular weight	9.477 kDa
Isoelectric point	10.14

Xenopus S27:

Length	84
Molecular weight	9.475 kDa
Isoelectric point	10.14

Overall: 98.8 % identity, 95.2% conservation

Appendix G

Database entry XLAB21 (for AB21)

```
ID  XLAB21     standard; RNA; VRT; 2305 BP.
XX
AC  X78454;
XX
DT  29-MAR-1994 (Rel. 39, Created)
DT  12-APR-1995 (Rel. 43, Last updated, Version 10)
XX
DE  X.laevis AB21 mRNA for RPD3 homologue
XX
KW   AB21 gene; RPD3 homologue.
XX
OS  Xenopus laevis (clawed frog)
OC  Eukaryota; Animalia; Metazoa; Chordata; Vertebrata; Amphibia;
OC  Lissamphibia; Anura; Archeobatrachia; Pipoidea; Pipidae.
XX
RN  [1]
RA  Ladomery M.R., Lyons S., Sommerville J.;
RT  ;
RL  Unpublished.
XX
RN  [2]
RP  1-1040
RA  Ladomery M.R.;
RT  ;
RL  Submitted (25-MAR-1994) to the EMBL/GenBank/DDBJ databases.
RL  M.R. Ladomery, University of St. Andrews, Dept. Biology, Bute
RL  Buildings, St. Andrews Fife, Scotland KY16 9TS, UK
XX
RN  [3]
RP  1-1624
RA  Ladomery M.R.;
RT  ;
RL  Submitted (13-DEC-1994) to the EMBL/GenBank/DDBJ databases.
RL  M.R. Ladomery, University of St. Andrews, Dept. Biology, Bute
RL  Buildings, St. Andrews Fife, Scotland KY16 9TS, UK
XX
FH  Key          Location/Qualifiers
FH
FT  source       1..2305
FT               /organism="Xenopus laevis"
FT               /dev_stage="previtellogenic oocyte"
FT               /clone_lib="lambda ZAP, Stratagene; polysomal cDNA library,
FT               Sommerville & Ladomery 1992"
FT  CDS          12..1454
FT               /partial
FT               /gene="AB21"
FT               /product="yeast RPD3 homologue"
FT  polyA_signal   2264..2269
FT  polyA_site     2285..2305
```

SQ Sequence 2305 BP; 679 A; 410 C; 544 G; 672 T; 0 other;

```
   1 gcggaaggaa aatggcgctg actctaggaa caaagaagaa agtgtgctac
  51 tactatgatg gtgatgttgg aaattattat tatggtcaag ggcatcccat
 101 gaaacctcat agaattcgca tgacacacaa cctgctgctc aactatggac
 151 tttaccgaaa aatggaaatc tttaggcccc acaaagccag cgccgaggat
 201 atgacaaagt atcacagtga tgattatatc aaattcctgc gctccatacg
 251 accagacaat atgtccgaat acagtaaaca gatgcagaga tttaatgttg
 301 gtgaggactg tcctgtgttt gatggcctat ttgagttctg ccagctctct
 351 gcaggggggtt ctgtagcaag tgctgttaaa ctaaacaaac agcagactga
 401 catttcagtg aactggtctg gtggccttca tcatgcaaag aaatctgagg
 451 catctggttt ttgttatgtc aacgatattg tccttgccat cctggaacta
 501 ctaaagtatc accagagagt tgtgtatatt gatatagaca ttcaccacgg
 551 tgatggtgtt gaggaggcat tttacacaac cgatagggtt atgactgtgt
 601 ccttccataa gtatggagag tattttcctg gaactggaga tctgagagat
 651 attggtgcag ggaaaggcaa atactatgct gtaaattatg ccttacggga
 701 tgggattgac gatgagtcct atgaagcaat ttttaaacca gtaatgtcca
 751 aagttatgga aatgtttcag cccagtgcag tggtcttaca gtgcggagca
 801 gattcattat ctggggatag actgggatgc ttcaatttga ccattaaggg
 851 acatgcgaag tgtgtggagt ttataaagac ctttaacttg ccactgttga
 901 tgttaggagg tggaggttac actatccgga atgtggctcg ttgctggaca
 951 tatgaaacag ctgtggctct ggactctgag atccccaatg agcttccata
1001 taatgattat tttgaatatt ttggtccgga cttcaagctt cacatcagcc
1051 catccaacat gactaatcag aacactaatg aatatctgga gaaaattaag
1101 cagcgcctct ttgagaactt gcgcatgctc ccccatgctc ctggagttca
1151 gatgcaagcc gttgcagagg actccataca cgatgacagt ggtgaagaag
1201 atgaagatga tcccgacaag cgtatttcaa ttcggtcatc agataaaagg
1251 attgcctgtg atgaggagtt ctcagattct gaggatgaag gggagggagg
1301 tcgcaaaaac gtggccaatt tcaaaaaagt aaaacgggtt aaaactgaag
1351 aggaaaagga aggagaggac aagaaagatg ttaaagaaga ggagaaagct
1401 aaagatgaga agacggatag caaacgggta aaagaagaga ccaaatcagt
1451 ctgatccttc aactatgggg agaaaatccg aagaccaaac taattctcat
1501 ggttttatat tttgtatatg ccctgtacag agccctacta tgaaatataa
1551 gtccacacat tctaaattat ttctgtccca ctggttgagg gggggtgaag
1601 tggtcgctgt agtggattaa gcttcacatc tgttaccttt ttttaagatt
1651 cacatctgtt acctttttac cagatgtttc cagctctttg gctttttttt
1701 ttttttgac caaaaacttt ccatgttttc ctgtgcctct gtaatcttcg
```

1751 gtggtgcaat gcattacgga tttatttccc tgctcccttc tatacacact
1801 ttgctgtcag actacagact tttgctacag tacatgaaat atgtacactt
1851 atgctcagga tcaggcatat gtacacttat gctcaggatc aggcagtgag
1901 aaggaggtgg gttccagctg tcttccaaat gaatttgaga gggttacctt
1951 gagggatgga agggggaagc tgaagctcct cttaaactaa actattcagg
2001 gattccctgt tcacttaatg ctgctaaccc tcctccagat tagttcatga
2051 agcagatttt tagatgtgtg gaaacctggt ccacagttac cttataatgg
2101 ggattgtggg gatttgcaat ttgggtttct gcctttaatc ttagtgggtt
2151 ggagagtgtc tggattcatg gagtgaagaa aatggagaat ttttatgtct
2201 aatttttgtg atgggaaatt tctttttttt ttttttatg gttgagttgt
2251 agaaaagctt tgtaataaaa tctggtactt atacaaaaaa aaaaaaaaaa
2301 aaaaa

Bibliography

AJTKHOZHIN, M. A., DOSCHANOV, K. I. and AKHANOV, A. U. (1976). Informosomes as a stored form of mRNA in wheat embryos. *FEBS Lett.* **66**, pp. 124-126.

AV-GAY, Y., AHARONOWITZ, Y. and COHEN, G. (1992). *Streptomyces* contain a 7.0 kDa cold shock like protein. *Nucleic Acids Res.* **20**, pp. 5478-5479.

BALLANTINE, J. E. M. and WOODLAND, H. R. (1985). Polyadenylation of histone mRNA in *Xenopus* oocytes and embryos. *FEBS Lett.* **180**, pp. 224-228.

BARBERIS, A., SUPERTI-FURGA, G. and BUSSLINGER, M. (1987). Mutually exclusive interaction of the CCAAT-binding factor and of a displacement protein with overlapping sequences of a histone gene promoter. *Cell* **50**, pp. 347-359.

BIAMONTI, G. and RIVA, S. (1994). New insights into the auxiliary domains of eukaryotic RNA binding proteins. *FEBS Lett.* **340**, pp. 1-8.

BIENZ, M. (1987). A CCAAT box confers cell-type-specific regulation on the *Xenopus hsp70* gene in oocytes. *Cell* **46**, pp. 1037-1042.

BOUVET, P. and WOLFFE, A. P. (1994). A role for transcription and FRGY2 in masking maternal messenger RNA. *Cell* **77**, pp. 931-941.

BOWDISH, K. S., and MITCHELL, A. P. (1993). Bipartite structure of an early meiotic upstream activation sequence from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, pp. 2172-2181.

BRADDOCK, M., MUCKENTHALER, M., WHITE, M. R. H., THORBURN, A. M. and SOMMERVILLE, J. (1994). Intron-less RNA injected into the nucleus of *Xenopus* oocytes accesses a regulated translation control pathway. *Nucleic Acids Res.* **22**, pp. 5255-5264.

BRADDOCK, M., THORBURN, A. M., CHAMBERS, A., ELLIOTT, G. D. and ANDERSON, G. J. (1990). A nuclear translational block imposed by the HIV-1 U3 region is relieved by the Tat-TAR interaction. *Cell* **62**, pp. 1123-1133.

BRENNAN, C.A., and PLATT, T. (1991). Mutations in an RNP-1 consensus sequence of Rho protein reduce RNA binding affinity but facilitate helicase turnover. *J. Biol. Chem.* **266**, pp. 17296-17305.

BURD, C. G. and DREYFUSS, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, pp. 615-621.

CAIZERGUES-FERRER, M., MARIOTTINI, P., CURIE, C., LAPEYRE, B., GAS, N., AMALRIC, F., and AMALDI, F. (1989). Nucleolin from *Xenopus laevis*: cDNA cloning and expression during development. *Genes & Dev.* **3,** pp. 324-333.

CALNAN, B. J, TIDOR, B., BIANCALANA, S., HUDSON, D., and FRANKEL, A. D. (1991). Arginine-mediated RNA recognition: the arginine fork. *Science* **252,** pp. 1167-1171.

CARDINALI, B., DI CRISTINA, M., and PIERANDREI-AMALDI, P. (1993). Interaction of proteins with the mRNA for ribosomal protein L1 in *Xenopus*: structural characterization of *in vivo* complexes and identification of proteins that bind *in vitro* to its 5' UTR. *Nucleic Acids Res.* **21,** pp. 2301-2308.

CHAN, Y. L., SUZUKI, K., OLVERA, J. and WOOL, I. G. (1993). Zinc finger-like motifs in rat ribosomal-proteins S27 and S29. *Nucleic Acids Res.* **21,** pp. 649-655.

COHEN, I., and REYNOLDS, W. F. (1991). The *Xenopus* YB3 protein binds the B box element of the class III promoter. *Nucleic Acids Res.* **19,** pp. 4753-4759.

CRAWFORD, D. R., and RICHTER, J. D. (1987). An RNA-binding protein from *Xenopus* oocytes is associated with specific message sequences. *Development* **101,** pp. 741-749.

CSERMELY, P., SCHNAIDER, T., CHEATHAM, B., OLSON, M. O. J. and KAHN, C. R. (1993). Insulin induces the phosphorylation of nucleolin - a possible mechanism of insulin-induced RNA efflux from nuclei. *J. Biol. Chem.* **268,** pp. 9747-9752.

CUMMINGS, A., BARRETT, P., and SOMMERVILLE, J. (1989). Multiple modifications in the phosphoproteins bound to stored messenger RNA in *Xenopus* oocytes. *Biochim. Biophys. Acta* **1014,** pp. 319-326.

CUMMINGS, A., and SOMMERVILLE, J. (1988). Protein kinase activity associated with stored messenger ribonucleoprotein particles of *Xenopus* oocytes. *J. Cell Biol.* **107,** pp. 45-56.

DANON, A. and MAYFIELD, S. P. Y. (1991). Light regulated translational activators- identification of chloroplast gene specific messenger RNA binding proteins. *EMBO J.* **10,** pp. 3993-4001.

DARBRE, A. (Editor) (1986). Practical Protein Chemistry- A Handbook. © John Wiley & Sons.

DARNBOROUGH, C. H, and FORD, P. J. (1981). Identification in *Xenopus laevis* of a class of oocyte-specific proteins bound to messenger RNA. *Eur. J. Biochem.* **113,** pp. 415-424.

DAVIDSON, E. H. (1986). Gene Activity in Early Development. Second Edition. © Academic Press Inc. 1976.

DAVIS, T. L., FIRULLI, A. B. and KINNIBURGH, A. J. (1989). Ribonucleoprotein and protein factors bind to an H-DNA-forming c-*myc* DNA element: possible regulators of the c-*myc* gene. *Proc. Natl. Acad. Sci. USA* **86,** pp. 9682-9686.

DEARSLY, A. L., JOHNSON, R. M., BARRETT, P., and SOMMERVILLE, J. (1985). Identification of a 60-kDa phosphoprotein that binds stored messenger RNA of *Xenopus* oocytes. *Eur. J. Biochem.* **150,** pp. 95-103.

DESCHAMPS, S., VIEL, A., DENIS, H., and LE MAIRE, M. (1991). Purification of two thermostable components of messenger ribonucleoprotein particles (mRNPs) from *Xenopus laevis* oocytes, belonging to a novel class of RNA-binding proteins. *FEBS Lett.* **282,** pp. 110-114.

DESCHAMPS, S., VIEL, A., GARRIGOS, M, DENIS, H., and LE MAIRE, M. (1992). mRNP4, a major mRNA-binding protein from *Xenopus oocytes* is identical to transcription factor FRGY2. *J. Biol. Chem.* **267,** pp. 13799-13802.

DE VALOIR, T., TUCKER, M. A., BELIKOFF, E. J., CAMP, L. A., BOLDUC, C. and BECKINGHAM, K. (1991). A second maternally expressed *Drosophila* gene encodes a putative RNA helicase of the "DEAD box" family. *Proc. Natl. Acad. Sci. USA* **88,** pp. 2113-2117.

DIDIER, D., SCHIFFENBAUER, J., WOULFE, S., ZACHEIS, M., and SCHWARTZ, B. D. (1988). Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl. Acad. Sci. USA* **85,** pp. 7322-7326.

DONIGER, J., LANDSMAN, D., GONDA, M., and WISTOW, G. (1992). The product of *unr*, the highly conserved gene upstream of *N-ras*, contains multiple repeats similar to the cold-shock domain (CSD), a putative DNA-binding motif. *The New Biologist* **4,** pp. 389-395.

DORER, D. R., CHRISTENSEN, A. C. and JOHNSON, D. H. (1990). A novel RNA helicase gene tightly linked to the Triplo-lethal locus of *Drosophila*. *Nucleic Acids Res.* **18,** pp. 5489-5494.

DUMONT, J. N. (1972). Oogenesis in *Xenopus laevis* (Daudin). *J. Morphology,* **136,** pp. 153-180.

DWORKIN, M. B., SHRUTKOWSKI, A., and DWORKIN-RASTL, E. (1985). Mobilization of specific maternal mRNA species into polysomes after fertilization in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* **82,** pp. 7636-7640.

ENFIELD, D. L., ERICSSON, L. H., FUJIKAWA, K., WALSH, K. A., NEURATH, H. and TITANI, R. (1980). Amino-acid sequence of the light chain of bovine factor X_1 (Stuart Factor). *Biochemistry* **19**, pp. 659-667.

EVANS, T., FELSENFELD, G. and REITMAN, M. (1990). Control of globin gene transcription. *Annu. Rev. Cell Biol.* **6**, pp. 95-124.

EVDOKIMOVA, V. M., WEI, C., SITIKOV, A. S., SIMONENKO, P. N., LAZAREV, O. A., VASILENKO, K. S., USTINOV, V. A., HERSHEY, J. W. B. and OVCHINNIKOV, L. P. (1995). The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. *J. Biol. Chem.* **270**, pp. 3186-3192.

FAMILIARI, M., ALMOUZNI, G. and WOLFFE, A. P. (1994). Isolation of a potentially functional Y-box protein (MSY-1) processed pseudogene from mouse: evolutionary relationships within the $EF1_A$/dbpB/YB-1 gene family. *Gene* **141**, pp. 255-259.

FIRULLI, A. B., MAIBENCO, D. C. and KINNIBURGH, A. J. (1992). The identification of a tandem H-DNA structure in the c-*myc* nuclease sensitive promoter element. *Biochem. Biophys. Res. Comm.* **185**, pp. 264-270.

FIRULLI, A. B., MAIBENCO, D. C. and KINNIBURGH, A. J. (1994). Triplex forming ability of a *c-myc* promoter element predicts promoter strength. *Arch. Biochem. Biophys.* **310**, pp. 236-242.

FORD, P. J., MATHIESON, T. and ROSBASH, M. (1977). Very long-lived messenger RNA in ovaries of *Xenopus laevis*. *Devl. Biol.* **57**, pp. 417-426.

GAILLARD, C., CABANNES, E. and STRAUSS, F. (1994). Identity of the RNA-binding protein K of hnRNP particles with protein H16, a sequence-specific single strand DNA-binding protein. *Nucleic Acids Res.* **22**, pp. 4183-4186.

GAIT, M. J., and KARN, J. (1993). RNA recognition by the human immunodeficiency virus Tat and Rev proteins. *Trends in Biochem. Sci.* **18**, pp. 255-259.

GEBAUER, F. and RICHTER, J. D. (1995). Cloning and characterization of a *Xenopus* poly(A) polymerase. *Mol. Cell. Biol.* **15**, pp. 1422-1430.

GILBERT, S. F. (1991). Developmental Biology (Third Edition). Sinauer Associates Inc., 1991.

GLATZER, K. H. and KLOETZEL, P-M. (1986). Differential chromosomal distribution of ribonucleoprotein antigens in nuclei of *Drosophila* spermatocytes. *J. Cell Biol.* **103**, pp. 2113-2119.

GODEAU, F., PERSSON, H., GRAY, H. E. and PARDEE, A. B. (1986). c-*myc* expression is dissociated from DNA synthesis and cell division in *Xenopus* oocyte and early embryonic development. *EMBO J.* **5**, pp. 3571-3577.

GOLDEN, B. L., HOFFMAN, D. W., RAMAKRISHNAN, V. and WHITE, S. W. (1993). Ribosomal protein S17: characterization of the three-dimensional structure by ^{1}H and ^{15}N NMR. *Biochemistry* **32**, pp. 12812-12820.

GOLDSMITH, M. E., MADDEN, M. J., MORROW, C. S. and COWAN, K. H. (1993). A Y-box consensus sequence is required for basal expression of the human multidrug resistance (*mdr1*) gene. *J. Biol. Chem.* **268**, pp. 5856-5860.

GOLDSTEIN, J., POLLIT, S., and INOUYE, M. (1990). Major cold shock protein of *Escherichia coli. Proc. Natl. Acad. Sci. USA* **87**, pp. 283-287.

GRANT, C. E., and DEELEY, R. (1994). Cloning and characterization of chicken YB-1: regulation of expression in the liver. *Mol. Cell. Biol.* **13**, pp. 4186-4196.

GRAUMANN, P. and MARAHIEL, M. A. (1994). The major cold shock protein of *Bacillus subtilis* CspB binds with high affinity to the ATTGG- and CCAAT sequences in single stranded oligonucleotides. *FEBS Lett.* **338**, pp. 157-160.

GRAVES, B. J., JOHNSON, P. F. and McKNIGHT, S. L. (1986). Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. *Cell* **44**, pp. 565-576.

GREEN, M., ISHINO, M. and LOEWENSTEIN, P. M. (1989). Mutational analysis of HIV-1 Tat minimal domain peptides: identification of trans-dominant mutants that suppress HIV-LTR-driven gene expression. *Cell* **58**, pp. 215-223.

GREENBERG, J. R. (1977). Isolation of messenger ribonucleoprotein in cesium sulfate density gradients: evidence that polyadenylated and non-polyadenylated messenger RNAs are associated with protein. *J. Mol. Biol.* **108**, pp. 403-416.

GUNKEL, N., BRADDOCK, M., THORBURN, A. M., MUCKENTHALER, M., KINGSMAN, A. J., and KINGSMAN, S. M. (1995). Promoter control of translation in *Xenopus* oocytes. *Nucleic Acids Res.* **23**, pp. 405-412.

GURDON, J. B., LANE, C. D., WOODLAND, H. R. and MARBAIX, G. (1971). Use of frog eggs and oocytes for the study of messenger RNA and its translation in living cells. *Nature (London)* **233**, p. 177.

GURURAJAN, R., PERRY-O' KEEFE, H., MELTON, D. A. and WEEKS, D. L. (1991). The *Xenopus* localized messenger RNA An3 may encode an ATP-dependent RNA helicase. *Nature* **349**, pp. 717-719.

HASEGAWA, S. L., DOETSCH, P. W., KAMILTON, K. K., MARTIN, A. M., OKENQUIST, S. A., LENZ, J., and BOSS, J. M. (1991). DNA binding properties of YB-1 and dbpA: binding to double-stranded, single-stranded and abasic site containing DNAs. *Nucleic Acids Res.* **19**, pp. 4915-4920.

HAY, B., JAN, L. Y. and JAN, Y. N. (1988). A protein component of *Drosophila* polar granules is encoded by *vasa* and has extensive sequence similarity to ATP-dependent helicases. *Cell* **55**, pp. 577-587.

HAY, B. and SHORT, J. M. (1992). ExAssist helper phage and SOLR for lambda ZAP II excisions. *Strategies in Molecular Biology* **5**, pp. 16-18.

HEIDARAN, M. A., SHOWMAN, R. M. and W. S. KISTLER (1988). A cytochemical study of the transcriptional and translational regulation of nuclear transition protein 1 (TP1), a major chromosomal protein of mammalian spermatids. *J. Cell Biol.* **106**, pp. 1427-1433.

HENSHAW, E. C. (1968). Messenger RNA in rat liver polyribosomes: evidence that it exists as ribonucleoprotein particles. *J. Mol. Biol.* **36**, pp. 401-411.

HETHERINGTON, A. M., SOMMERVILLE, J., MASTERS, A. K. and MITCHELL, A. G. (1990). Evidence which supports the presence of stored messenger ribonucleoprotein (mRNP) in the unfertilised eggs of *Fucus serratus.NATO ASI Series,* **Vol. H45,** Mechanism of Fertilization, pp. 654-661, ed. B. Dale, Springer-Verlag Berlin Heidelberg 1990.

HOLLINGSWORTH, M. A., CLOSKEN, C., HARRIS, A., MACDONALD, C. D., PAHWA, G. S. and MAHER, L. J. (1994). A nuclear factor that binds purine-rich, single-stranded oligonucleotides derived from S1-sensitive elements upstream of the *CFTR* gene and the *MUC1* gene. *Nucleic Acids Res.* **22**, pp. 1138-1146.

HORWITZ, E. M., MALONEY, K. A. and LEY, T. J. (1994). A human protein containing a "cold shock" domain binds specifically to H-DNA upstream from the human γ-globin genes. *J. Biol. Chem.* **269**, pp. 14130-14139.

HWANG, I., LIM, K. and CHAE, C-B. (1990). Characterization of the S-phase specific transcription regulatory elements in a DNA replication

independent testis specific H2B (tH2B) histone gene. *Mol. Cell. Biol.* **10**, pp. 585-592.

IGGO, R. D., JAMIESON, D. J., MACNEILL, S. A., SOUTHGAITE, J., MCPHEAT, J. and LANE, D. P. (1991). p68 RNA helicase - identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. *Mol. Cell. Biol.* **11**, pp. 1326-1333.

IGGO, R., PICKSLEY, S., SOUTHGAITE, J., MACPHEAT, J. and LANE, D. P. (1990). Identification of a putative RNA-helicase in *Escherichia coli*. *Nucleic Acids Res.* **18**, pp. 5413-5417.

ITO, K., TSUTSUMI, K., KUZUMAKI, T., GOMEZ, P. F., OTSU, K. and ISHIKAWA, K. (1994). A novel growth-inducible gene that encodes a protein with a conserved cold-shock domain. *Nucleic Acids Res.* **22**, pp. 2036-2041.

JACQUEMIN-SABLON, H., TRIQUENEAUX, G., DESCHAMPS, S., LE MAIRE, M., DONIGER, J. and DAUTRY, F. (1994). Nucleic acid binding and intracellular localization of unr, a protein with five cold shock domains. *Nucleic Acids Res.* **22**, pp. 2463-2650.

JANSEN-DURR, P., BOSHART, M., LUPP, B., BOSSERHOFF, A., FRANK, R. W. and SCHUTZ, G. (1992). The rat poly pyrimidine tract binding protein (PTB) interacts with a single-stranded DNA motif in a liver-specific enhancer. *Nucleic Acids Res.* **20**, pp. 1243-1249.

JEDLICKI, A., HINRICHS, M. V., ALLENDE, C. C. and ALLENDE, J. E. (1992). The cDNAs coding for the α- and β- subunits of *Xenopus laevis* casein kinase II. *FEBS Lett.* **297**, pp. 280-284.

JEFFERIES, H. B. J., REINHARD, C., KOZMA, S. C. and THOMAS, G. (1994). Rapamycin selectively represses translation of the polypyrimidine tract messenger-RNA family. *Proc. Natl. Acad. Sci. USA* **91**, pp. 4441-4445.

JESSEN, T. H., OUBRIDGE, C., HIANG TEO, C., PRITCHARD, C., and NAGAI, K. (1991). Identification of molecular contacts between the U1 A small nuclear ribonucleoprotein and U1 RNA. *EMBO J.* **10**, pp. 3447-3456.

JOHO, K. E., DARBY, M. K., CRAWFORD, E. T. and BROWN, D. D. (1990). A finger protein structurally similar to TFIIIA that binds exclusively to 5S RNA in *Xenopus*. *Cell* **61**, pp. 293-300.

JONES, P. G., KRAH, R., TAFURI, S. R. and WOLFFE, A. P. (1992). DNA gyrase, CS7.4, and the cold-shock response in *Escherichia coli*. *J. Bacteriol.* **174**, pp. 5798-5802.

JONES, S. D. C. and SCHOFIELD, J. P. (1991). A rapid method for isolating high quality plasmid DNA suitable for DNA sequencing. *Nucleic Acids Res.* **18**, p. 7463

KANDALA, J. C. and GUNTAKA, R. V. (1994). Cloning of Rous sarcoma virus enhancer factor genes. *Virology* **198**, pp. 514-523.

KASHANCHI, F., DUVALL, J. F., DITTMER, J., MIRESKANDARI, A., REID, R. L., GITLIN, S. D. and BRADY, J. N. (1994). Involvement of transcription factor YB-1 in human T-cell lymphotrophic virus type I basal gene expression. *J. Virology* **68**, pp. 561-565.

KAY, B. K. and BENJAMIN PENG, H., Editors (1991). *Xenopus laevis:* Practical Uses in Cell and Molecular Biology (Volume 36, "Methods in Cell Biology" series). © Academic Press Inc. 1991.

KENAN, D. J., QUERY, C. C. and KEENE, J. D. (1991). RNA recognition: towards identifying determinants of specificity. *Trends in Biochem. Sci.* **16**, pp. 214-220.

KICK, D., BARRETT, P., CUMMINGS, A., and SOMMERVILLE, J. (1987). Phosphorylation of a 60 kDa polypeptide from *Xenopus* oocytes blocks messenger RNA translation. *Nucleic Acids Res.* **15**, 4099-4109.

KINNIBURGH, A. J., FIRULLI, A. B. and KOLLURI, R. (1994). DNA triplexes and regulation of the c-*myc* gene. *Gene* **149**, pp. 93-100.

KITAMURA, K., NAKAGAWA, T. and SHIMODA, C. (1990). Novel sterile mutants of the fission yeast *Schizosaccharomyces pombe* which are defective in their response to starvation. *Curr. Genet.* **18**, pp. 315-321.

KJEMS, J., FRANKEL, A. D. and SHARP, P. A. (1991). Specific regulation of mRNA splicing in vitro by a peptide from HIV-1 Rev. *Cell* **67**, pp. 169-178.

KLAUSNER, R. D., ROUAULT, T. A. and HARFORD, J. B. (1993). Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* **72**, pp. 19-28.

KLEENE, K. C. (1989). Poly(A) shortening accompanies the activation of translation of five mRNAs during spermiogenesis in the mouse. *Development* **106**, pp. 367-373.

KLEENE, K. C. (1993). Multiple controls over the efficiency of translation of the mRNAs encoding transition proteins, protamines, and the mitochondrial capsule selenoprotein in late spermatids in mice. *Devl. Biol.* **159**, pp. 720-731.

KLEENE, K. C., DISTEL, R. J., and HECHT, N. B. (1984). Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. *Developmental Biology* **105,** pp. 71-79.

KLEENE, K. C. and SMITH, J. (1994). Translational activity of mouse protamine 1 messenger ribonucleoprotein particles in the reticulocyte and wheat germ cell-free translation systems. *Mol. Rep. and Dev.* **37,** pp. 12-20.

KLOETZEL, P-M., JOHNSON, R. M., and SOMMERVILLE, J. (1982). Interaction of the hnRNA of amphibian oocytes with fibril-forming proteins. *Eur. J. Biochem.* **127,** pp. 301-308.

KOLLURI, R., TORREY, T.A., and KINNIBURGH, A. (1992). A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucleic Acids Res.* **20,** pp. 111-116.

KOMIYA, T., ITOH, K., IKENISHI, K. and FURUSAWA, M. (1994). Isolation and characterization of a novel gene of the DEAD box protein family which is specifically expressed in germ cells of *Xenopus laevis*. *Devl. Biol.* **162,** pp. 354-363.

KREK, W., MARIDOR, G. and NIGG, A. (1992). Casein kinase II is a predominantly nuclear enzyme. *J. Cell Biol.* **116,** pp. 43-55.

KRUYS, V., MARINX, O., SHAW, G., DESCHAMPS, J. and HUEZ, G. (1989). Translational blockade imposed by cytokine-derived UA-rich sequences. *Science* **245,** pp. 852-855.

KWON, Y. K. and HECHT, N. B. (1991). Cytoplasmic protein binding to highly conserved sequences in the 3' untranslated region of mouse protamine 2 messenger RNA. *Proc. Natl. Acad. Sci. USA* **88,** pp. 3584-3588.

KWON, Y. K. and HECHT, N. B. (1993). Binding of a phosphoprotein to the 3' untranslated region of the mouse protamine 2 messenger RNA temporally represses its translation. *Mol. Cell. Biol.* **13,** pp. 6547-6557.

LAIRD-OFFRINGA, I. A. (1992). What determines the instability of c-*myc* proto-oncogene mRNA? *BioEssays* **14,** pp. 119-124.

LA ROVERE, J. (1990). Regulation in expression of the mRNA-binding protein pp56 during early development of *Xenopus*. University of St. Andrews Master's Thesis, library code Th/QP625.N87L2.

LASKO, P. F. and ASHBURNER, M. (1988). The product of the *Drosophila* gene *vasa* is very similar to eukaryotic initiation factor-4A. *Nature* **335,** pp. 611-617.

LASKO, P. F. and ASHBURNER, M. (1990). Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. *Genes & Dev.* **4**, pp. 905-921.

LA TEANA, A., BRANDI, A., FALCONI, M., SPURIO, R., PON, C. L., and GUALERZI, C. O. (1991). Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS. *Proc. Natl. Acad. Sci. USA* **88**, pp. 10907-10911.

LEROY, P., ALWARI, P., SASSOON, D., WOLGEMUTH, D. and FELLOUS, M. (1989). The protein encoded by a murine male germ cell-specific transcript is a putative ATP-dependent RNA helicase. *Cell* **57**, pp. 549-559.

LIN, Y. S., CAREY, M. F., PTASHNE, M. and GREEN, M. R. (1988). GAL4 derivatives function alone and synergistically with mammalian activators *in vitro*. *Cell* **54**, pp. 659-664.

LIPSON, K. E., CHEN, S. T., KONIECKI, J., KU, D. H. and BASERGA, R. (1989). S-phase-specific regulation by deletion mutants of the human thymidine kinase promoter. *Proc. Natl. Acad. Sci. USA* **86**, pp. 6848-6852.

LORENI, F., FRANCESCONI, A., and AMALDI, F. (1993). Coordinate translational regulation in the synthesis of elongation factor 1α and ribosomal proteins in *Xenopus laevis*. *Nucleic Acids Res.* **21**, pp. 4721-4725.

LU, D. and YUNIS, J. J. (1992). Cloning, expression and localization of an RNA helicase gene from a uman lymphoid cell line with chromosomal breakpoint 11q23.3. *Nucleic Acids Res.* **20**, pp. 1967-1972.

LUKANIDIN, E. M., ZALMANZON, E. S., KOMAROMI, L., SAMARINA, O. P. and GEORGIEV, G. P. (1972). Structure and function of informofers. *Nature New Biology* **238**, pp. 193-197.

MACDONALD, G. H., ITOHLINDSTROM, Y. and TING, J. P. Y. (1995). The transcriptional regulatory protein, YB-1, promotes single-stranded regions in the DRA promoter. *J. Biol. Chem.* **270**, pp. 3527-3533.

MAEKAWA, H., NAKAGAWA, T., UNO, Y., KITAMURA, K. and SHIMODA, C. (1994). The *ste13*$^+$ gene encoding a putative RNA helicase is essential for nitrogen starvation-induced G1 arrest and initiation of sexual development in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet.* **244**, pp. 456-464.

MALNOE, P., MAYFIELD, S. P. and ROCHAIX, J. D. (1988). Comparative analysis of the biogenesis of photosystem-II in the wild-type and Y-1 mutant of *Chlamydomonas reinhardtii*. *J. Cell Biol.* **106**, pp. 609-616.

MANIATIS, T., FRITSCH, E. F. and SAMBROOK, J. (1982). Molecular Cloning: A Laboratory Manual. © Cold Spring Harbor Laboratory 1982.

MARELLO, K., LA ROVERE, J., and SOMMERVILLE, J. (1992). Binding of *Xenopus* oocyte masking proteins to mRNA sequences. *Nucleic Acids Res.* **20**, pp. 5593-5600.

MATTAJ, I. W., COPPARD, N. J., BROWN, R. S., CLARK, B. F. C. and DE ROBERTIS, E. M. (1987). 42S p48 - the most abundant protein in previtellogenic *Xenopus* oocytes resembles elongation-factor 1-alpha structurally and functionally. *EMBO J.* **6**, pp. 2409-2413.

McGREW, L. L., DWORKIN-RASTL, E., DWORKIN, M. B. and RICHTER, J. D. (1989). Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes. Dev.* **3**, pp. 803-815.

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 heterogenous nuclear ribonucleoprotein. *J. Biochem.* **263**, pp. 3307-3313.

MINICH, W. B., KORNEYEVA, N. L. and OVCHINNIKOV, L. P. (1989). Translational active mRNPs from rabbit reticulocytes are qualitatively different from free mRNA in their translatability in cell-free system. *FEBS Lett.* **257**, pp. 257-259.

MINICH, W. B., MAIDEBURA, I. P. and OVCHINNIKOV, L. P. (1993). Purification of the major 50-kDa repressor protein from cytoplasmic mRNP of rabbit reticulocytes. *Eur. J. Biochem.* **212**, pp. 633-638.

MINICH, W. B. and OVCHINNIKOV, L. P. (1992). Role of cytoplasmic mRNP proteins in translation. *Biochimie* **74**, pp. 477-483.

MORALES, C. R., KWON, Y. K. and HECHT, N. B. (1991). Cytoplasmic localization during storage and translation of the mRNAs of transition protein 1 and protamine 1, two translationally regulated transcripts of the mammalian testis. *J. Cell Science* **100**, pp. 119-131.

MULNER-LORILLON, O., MAROT, J., CAYLA, X., POUHLE, R., and BELLE, R. (1988). Purification and characterization of a casein-kinase-II-type enzyme from *Xenopus laevis* ovary. Biological effects on the meiotic cell division of full-grown oocyte. *Eur. J. Biochem.* **171**, pp. 107-117.

MURRAY, M. T. (1994). Nucleic acid-binding properties of the *Xenopus* oocyte Y-box protein $mRNP_{3+4}$. *Biochemistry* **33**, pp. 13910-13917.

MURRAY, M. T., KROHNE, G., and FRANKE, W. W. (1991). Different forms of soluble cytoplasmic mRNA binding proteins and particles in *Xenopus laevis* oocytes and embryos. *J. Cell Biol.* **112**, pp. 1-11.

MURRAY, M. T., SCHILLER, D. T., and FRANKE, W. W. (1992). Sequence analysis of cytoplasmic mRNA-binding proteins of *Xenopus* oocytes identifies a family of RNA-binding proteins. *Proc. Natl. Acad. Sci. USA* **89**, pp. 11-15.

MURZIN, A. G. (1993). OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **21**, pp. 861-867.

NEWKIRK, K., FENG, W., JIANG, W., TEJERO, R., DONALD EMERSON, S., INOUYE, M. and MONTELIONE, G. T. (1994). Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. *Proc. Natl. Acad. Sci. USA* **91**, pp. 5114-5118.

NIKOLAJCZYK, B. S., MURRAY, M. T. and HECHT, N. B. (1995). A mouse homologue of the *Xenopus* germ cell-specific ribonucleic acid/deoxyribonucleic acid-binding proteins p54/p56 interacts with the protamine 2 promoter. *Biology of Reproduction* **52**, pp. 524-530.

NUTE, P. E. and MAHONEY, W. C. (1979). Complete amino-acid sequence of the γ-chain from the major fetal hemoglobin of the pig-tailed macaque, *Macaca nemestrina*. *Biochemistry* **18**, pp. 467-472.

OBOKATA, J., OHME, M., and HAYASHIDA, N. (1991). Nucleotide sequence of a cDNA encoding a putative glycine-rich protein of 19.7 kDa in *Nicotiana sylvestris*. *Plant Mol. Biol.* **17**, pp. 953-955.

OLIVE, M., THEZE, N., PHILIPPE, M., LE PENNEC, J-P. and LERIVRAY, H. (1994). Cloning of the *Xenopus laevis cdk2* promoter and functional analysis in oocytes and during early development. *Gene* **151**, pp. 81-88.

OMICHINSKI, J. G., CLORE, G. M., SCHAAD, G. M., FELSENFELD, G., TRAINOR, C. and APPELLA, E. (1993). NMR structure of a specific DNA complex of Zn-containing DNA-binding domain of GATA-1. *Science* **261**, pp. 438-446.

OSTARECK-LEDERER, A., OSTARECK, D. H., STANDART, N. and THIELE, B. J. (1994). Translation of 15-lipoxygenase mRNA is inhibited by a protein that binds to a repeated sequence in the 3' untranslated region. *EMBO J.* **13**, pp. 1476-1481.

OVCHINNIKOV, L. P. and SPIRIN, A. S. (1970). Ribonucleoprotein particles in cytoplasmic extracts of animal cells. *Naturwissenschaften* **57**, pp 514-570.

OWTTRIM, G. W., MANDEL, T., TRACHSEL, H. THOMAS, A. A. M. and KUHLEMEIER, C. (1994). Characterization of the tobacco eIF-4A gene family. *Plant Mol. Biol.* **26**, pp. 1747-1757.

OZER, J., CHALKLEY, R., and SEALY, L. (1993). Isolation of the CCAAT transcription factor subunit EFI_A cDNA and a potentially functional EFI_A processed pseudogene from *Bos taurus*: insights into the evolution of the EFI_A/dbpB/YB-1 family. *Gene* **124**, pp. 223-230.

OZER, J., FABER, M., CHALKLEY, R., and SEALY, L. (1990). Isolation and characterization of a cDNA cone for the CCAAT transcription factor EFI_A reveals a novel structural motif. *J. Biol. Chem.* **265**, pp. 22143-22152.

PATTON, J. G., MAYER, S. A., TEMPST, P. and NADAL-GINARD, B. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes & Dev.* **5**, pp. 1237-1251.

PAUSE, A., METHOT, N. and SONENBERG, N. (1993). The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis. *Mol. Cell. Biol.* **13**, pp. 6789-6798.

PAUSE, A., METHOT, N., SVITKIN, Y., MERRICK, W. C. and SONENBERG, N. (1994). Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. *EMBO J.* **13**, pp. 1205-1215.

PAUSE, A. and SONENBERG, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J.* **11**, pp. 2643-2654.

PEARSON, B. E., NASHEUER, H. P. and WANG, T. S. F. (1991). Human DNA polymerase-alpha gene - sequences controlling expression in cycling and serum-stimulated cells. *Mol. Cell. Biol.* **11**, pp. 2081-2095.

PESTOV, D. G., DAYN, A., SIYANOVA, E. Y., GEORGE, D. L. and MIRKIN, S. M. (1991). H-DNA and Z-DNA in the mouse c-*Ki-ras* promoter. *Nucleic Acids Res.* **19**, pp. 6527-6532.

PRAMANIK, S. K., KROCHKO, J. E. and BEWLEY, J. D. (1992). Distribution of cytosolic mRNAs between polysomal and ribonucleoprotein complex fractions in alfalfa embryos. *Plant Physiol.* **99**, pp. 1590-1596.

RANJAN, M., TAFURI, S., and WOLFFE, A. (1993). Masking mRNA from translation in somatic cells. *Genes & Dev.* **7**, pp. 1725-1736.

RAYNAL, M., DEPIGNY, D., COOKE, R. and DELSENY, M. (1989). Characterization of a radish nuclear gene expressed during late seed maturation. *Plant Physiol.* **91**, pp. 829-836.

RICHTER, J. D. and SMITH, L. D. (1984). Reversible inhibition of translation by *Xenopus* oocyte-specific proteins. *Nature (London)* **309**, pp. 378-380.

ROSBASH, M. and FORD, P. J. (1974). Polyadenylic acid containing RNA in *Xenopus laevis* oocytes. *J. Mol. Biol.* **85**, pp. 87-101.

ROSENTHAL, E. T. and RUDERMAN, J. V. (1987). Widespread changes in the translation and adenylation of maternal messenger RNAs following fertilization of *Spisula* oocytes. *Devl. Biol.* **121**, pp. 237-246.

ROSENTHAL, E. T, and WILT, F. H. (1986). Patterns of maternal messenger RNA accumulation and adenylation during oogenesis in *Urechis caupo*. *Devl. Biol.* **117**, pp. 55-63.

ROSENTHAL, E. T, and WILT, F. H. (1993). Regulation of maternal messenger RNA translation during oogenesis and embryogenesis in *Urechis caupo*. *Devl. Biol.* **155**, pp. 297-306.

SABATH, D. E., PODOLIN, P. L., COMBER, P. G. and PRYSTOWSKY, M. B. (1990). cDNA cloning and characterization of interleuikin 2-induced genes in a cloned T-helper lymphocyte. *J. Biol. Chem.* **265**, pp. 12671-12678.

SAGATA, N., OSKARSSON, M., COPELAND, T., BRUMBAUGH, J., and VANDE WOUDE, G. F. (1988). Function of c-*mos* proto-oncogene product in meiotic maturation in *Xenopus* oocytes. *Nature* **335**, pp. 519-525.

SAKURA, H., MAEKAWA, T., IMAMOTO, F., YASUDA, K., and ISHII S. (1988). Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology. *Gene* **73**, pp. 499-507.

SANCHEZ-MARTINEZ, D., PUIGDOMENECH, P. and PAGES, M. (1986). Regulation of gene expression in developing *Zea mays* embryos. *Plant Physiol.* **82**, pp. 543-549.

SCHINDELIN, H., JIANG, W., INOUYE, M. and HEINEMANN, U. (1994). Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **91**, pp. 5119-5123.

SCHINDELIN, H., MARAHIEL, M. A., and HEINEMANN, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the *Bacillus subtilis* major cold-shock protein. *Nature* **364**, pp. 164-168.

SCHNUCHEL, A., WILTSCHEK, R., CZISCH, M., HERRIER, M., WILLIMSKY, G., GRAUMANN, P., MARAHIEL, M. A., and HOLAK, T. (1993). Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature* **364**, pp. 169-171.

SCHRODER, K., ZUBER, P., WILLIMSKY, G., WAGNER, B. and MARAHIEL, M. A. (1993). Mapping of the *Bacillus subtilis cspB* gene and cloning of its homologs in thermophilic, mesophilic and psychrotrophic bacilli. *Gene* **136**, pp. 277-280.

SCHUITEMAKER, H., KOOTSTRA, N. A., FOUCHIER, R. A. M., HOOIBRINK, B. and MIEDEMA, F. (1994). Productive HIV-1 infection of macrophages restricted to the cell fraction with proliferative capacity. *EMBO J.* **13**, pp. 5929-5936.

SETYONO, B., and GREENBERG, J.R. (1981). Proteins associated with poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. *Cell* **24**, pp. 775-783.

SHAUGHNESSY, M. and WISTOW, G. (1992). Absence of MHC gene expression in lens and cloning of dbpb/YB-1, a DNA-binding protein expressed in mouse lens. *Current Eye Res.* **11**, pp. 175-181.

SHEETS, M. D., FOX, C. A., HUNT, T., VAN DE WOUNDE, G. and WICKENS, M. (1994). The 3'-untranslated regions of c-*mos* and cyclin mRNAs stimulate translation by regulating cytoplasmic polyadenylation. *Genes & Dev.* **8**, pp. 926-938.

SIEBENLIST, U., BRESSLER, P. and KELLY, K. (1988). Two distinct mechanisms of transcriptional control operate on c-*myc* during differentiation of HL60 cells. *Mol. Cell. Biol.* **8**, pp. 867-874.

SIMON, R., TASSAN, J. P., and RICHTER, J. D. (1992). Translational control by poly(A) elongation during *Xenopus* development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes & Dev.* **6**, pp. 2580-2591.

SINCLAIR, G. D. and DIXON, G. H. (1982). Purification and characterization of cytoplasmic protamine messenger ribonucleoprotein particles from rainbow trout testis cells. *Biochemistry* **21**, pp. 1869-1877.

SKEHEL, P. A. and BARTSCH, D. (1994). Characterization of a Y-box factor from *Aplysia californica*. *Gene* **145**, pp. 231-235.

SMITH, K. (1969). Photochemical addition of amino acids to ^{14}C-uracil. *Biochem. Biophys. Res. Comm.* **34**, pp. 354-357.

SMITH, R. C., DWORKINRASTL, E. and DWORKIN, M. B. (1988). Expression of a histone H1-like protein is restricted to early *Xenopus* development. *Genes & Dev.* **2**, pp. 1284-1295.

SMITH, L. D., XU WEILONG, and VARNOLD, R. L. (1991). Oogenesis and oocyte isolation. From Kay and Peng (1991), Editors: Methods in Cell Biology, Volume **36**, *Xenopus laevis*: practical uses in cell and molecular biology, © Academic Press Inc., 1991, Chapter 4, pp. 45-61.

SOMMERVILLE, J. (1990). RNA-binding phosphoproteins and the regulation of maternal mRNA in *Xenopus*. *J. Reprod. Fert. Suppl.* **42**, pp. 225-233.

SOMMERVILLE, J. (1992). RNA-binding proteins: masking proteins revealed. *BioEssays* **14**, pp. 337-339.

SOMMERVILLE, J., BAIRD, J. and TURNER, B. M. (1993). Histone-H4 acetylation and transcription in amphibian chromatin. *J. Cell Biol.* **120**, pp. 277-290.

SPIRIN, A. S. (1978). Eukaryotic messenger RNA and informosomes. *FEBS Lett.* **88**, pp. 15-17.

SPIRIN, A. S. (1980). Messenger ribonucleoproteins (informosomes). *Reprinted from* (IUPAC) Frontiers of Bioorganic Chemistry and Molecular Biology, ed. S. N. ANANCHENKO, Pergamon Press Oxford and New York 1980, pp. 285-290.

SPIRIN, A. S. (1994). Storage of messenger RNA in eukaryotes: envelopment with protein, translational barrier at 5' side, or conformational masking by 3' side? *Mol. Rep. and Dev.* **38**, pp. 107-117.

SPIRIN, A. S., BELITSINA, N. V. and AITKHOZHIN, M. A. (1964). Messenger RNA in early embryogenesis. *Zhurnal Obschchei Biologii* (Russian) **25**, pp. 321-329. English translation (1965) *Fed proc.* **24**, pp. T907-T922.

SPIRIN, A. S. and NEMER, M. (1965). Messenger RNA in early sea-urchin embryos: cytoplasmic particles. *Science* **150**, pp 214-217.

SPITKOVSKY, D., ROYER-POKORA, B., DELIUS, H., KISSELJOV, F., JENKINS, N. A., GILBERT, D. J., COPELAND, N. G., and ROYER, H. (1992). Tissue restricted expression and chromosomal localization of the YB-1 gene encoding a 42kD nuclear CCAAT binding protein. *Nucleic Acids Res.* **20**, pp. 797-803.

STANDART, N. M. (1992). Masking and unmasking of maternal mRNA. *Seminars in Devl. Biol.* **3**, pp. 367-379.

STANDART, N. M. (1993). The RNA-protein partners in mRNP. *Mol. Biol. Reports* **18**, pp. 135-142.

STANDART, N. M., BRAY, S. J., GEORGE, E. L., HUNT, T. and RUDERMAN, J. (1985). Small subunit of ribonucleotide reductase is encoded by one of the most abundant translationally regulated maternal RNAs in clam and sea urchin eggs. *J. Cell Biol.* **100**, pp. 1968-1976.

STANDART, N. and DALE, M. (1993). Regulated polyadenylation of clam maternal mRNAs *in vitro*. *Devl. Genetics* **14**, pp. 492-499.

STRAUSS, E. J. and GUTHRIE, C. (1991). A cold-sensitive messenger-RNA splicing mutant is a member of the RNA helicase gene family. *Genes & Dev.* **5**, pp. 629-641.

STRICKLAND, S., HUARTE, J., BELIN, D., VASSALLI, A., RICKLES, R. J. and J-D. VASSALLI (1988). Antisense RNA directed against the 3' noncoding region prevents dormant mRNA activation in mouse oocytes. *Science* **241**, pp. 680-684.

STRIPECKE, R., OLIVEIRA, C. C., McCARTHY, J. E. G. and HENTZE, M. W. (1994). Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. *Mol. and Cell Biol.* **14**, pp. 5898-5909.

SUZUKI, N., SUZUKI, T., UCHIDA, A., THOMPSON, E. A. and HOSOYA, T (1992). Effect of dexamethasone on nucleolar casein kinase-II activity and phosphorylation of nucleolin in lymphosarcoma P1798 cells. *J. Steroid Biochem. Mol. Biol.* **42**, pp. 305-312.

SWIDERSKI, R. E., and RICHTER, J. D. (1988). Photocrosslinking of proteins to maternal mRNA in *Xenopus oocytes*. *Devl. Biol.* **128**, pp. 349-358.

TAFURI, S. R., FAMILIARI, M., and WOLFFE, A. P. (1993). A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes. *J. Biochem.* **268**, pp. 12213-12220.

TAFURI, S. R., and WOLFFE, A. P. (1990). *Xenopus* Y-box transcription factors: molecular cloning, functional analysis and developmental regulation. *Proc. Natl. Acad. Sci. USA* **87**, pp. 9028-9032.

TAFURI, S. R., and WOLFFE, A. P. (1992). DNA binding, multimerization, and transcription stimulation by the *Xenopus* Y-box proteins *in vitro*. *The New Biologist* **4**, pp. 1-11.s

TAFURI, S. R. and WOLFFE, A. P. (1993). Dual roles for transcription and translation factors in the RNA storage particels of *Xenopus* oocytes. *Trends in Cell Biol.* **3**, pp. 94-98.

TAFURI, S. R. and WOLFFE, A. P. (1993b). Selective recruitment of masked maternal mRNA from messenger ribonucleoprotein particles containing FRGY2 (mRNP4). *J. Biol. Chem.* **268**, pp. 24255-24261.

TAKIMOTO, M., TOMONAGA, T., MATUNIS, M., AVIGAN, M., KRUTZSCH, H., DREYFUSS, G. and LEVENS, D. (1993). Specific binding of heterogenous ribonucleoprotein particle protein K to the human c-*myc* promoter *in vitro*. *J. Biol. Chem.* **268**, pp. 18249-18258.

TANABE, H., GOLDSTEIN, J., YANG, M., and INOUYE, M. (1992). Identification of the promoter region of the *Escherichia coli* major cold shock gene, *cspA*. *J. Bacteriol.* **174**, pp. 3867-3873.

TAYLOR, M. V., GUSSE, M., EVAN, G. I., DATHAN, N. and MECHALI, M. (1986). *Xenopus myc* proto-oncogene during development: expression as a stable maternal mRNA uncoupled from cell division. *EMBO J.* **5**, pp. 3563-3570.

THOEN, C., VAN HOVE, L., PIOT, E., and SLEGERS, H. (1984). Purification and characterization of the messenger ribonucleoprotein-associated casein kinase II of *Artemia salina* cryptobiotic gastrulae. *Biochim. Biophys. Acta* **783**, 105-113.

TING, J. P-Y., PAINTER, A., ZELEZNIK-LE, N. J., NacDonald, G., Moore, T. M., BROWN, A. and SCHWARTZ, B. D. (1994). YB-1 DNA-binding protein represses interferon γ activation of class II major histocompatibility complex genes. *J. Exp. Med.* **179**, pp. 1605-1611.

TOONE, W. M., RUDD, K. E. and FRIESEN, J. D. (1991). DeaD, a new *Escherichia coli* gene encoding a presumed ATP-dependent RNA helicase, can suppress a mutation in *rspb*,the gene encoding ribosomal protein-S2. *J. Bacteriol.* **173**, pp. 3291-3302.

TRAVALI, S., KU, D. H., RIZZO, M. G., OTTAVIO, L., BASERGA, R. and CALABRETTA, B. (1989). Structure of the human gene for the proliferating cell nuclear antigen. *J. Biol. Chem.* **264**, pp. 7466-7472.

TSO, J. Y., VANDENBERG, D. J. and KORN, L. J. (1986). Structure of the gene for *Xenopus* transcription factor TFIIIA. *Nucleic Acids Res.* **14**, pp. 2187-2200.

VAN EEKELEN, C. A., RIEMEN, T., and VAN VENROOIJ, W. J. (1979). Specificity in the interaction of hnRNA and mRNA with proteins as revealed by *in vivo* cross-linking. *FEBS Lett.* **130**, pp. 223-226.

VIDAL, M., and GABER, R. F. (1991). *RPD3* encodes a second factor required to achieve maximum positive and negative states in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **11**, pp. 6317-6327.

VIDAL, M., STRICH, R., ESPOSITO, R. E., and GABER, R. F. (1991). *RPD1* (*SIN3/UME4*) is required for maximal activation and repression of diverse yeast genes. *Mol. Cell. Biol.* **11**, 6306-6316.

VINSON, C.R., LAMARCO, K.L., JOHNSON, P.F., LANDSCHULZ, W.H., and McKNIGHT, S.L. (1988). *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. *Genes & Dev.* **2**, pp. 801-806.

WALKER, J. E., SARASTE, M., RUNSWICK, M. J. and GAY, N. J. (1982). Distantly related sequences in the alpha-subunits and beta-subunits of ATP synthase, myosin kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, pp. 945-951.J

WATANABE, N., VAN DE WOUDE, G. F., IKAWA, Y. and SAGATA, N. (1989). Specific proteolysis of the c-*mos* proto-oncogene product by calpain on fertilization of *Xenopus* eggs. *Nature* **342**, pp. 505-517.

WATERSTON, R., MARTIN, C., CRAXTON, M., HUYNH, C., COULSON, A., HILLIER, L., DURBIN, R., GREEN, P., SHOWNKEEN, R., HALLORAN, N., METZSTEIN, M., HAWKINS, T., WILSON, R., BERKS, M., DU, Z., THOMAS, K., THIERRYMIEG, J. and SULSTON, J. (1992). A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genetics* **1**, pp. 114-123.

WICKENS, M. (1990). In the beginning is the end: regulation of poly(A) addition and removal during early development. *Trends in Biochem. Sci.* **15**, pp. 320-324.

WILLIMSKY, G., BANG, H., FISCHER, G., and MARAHIEL, M. (1992). Characterization of *cspB*, a *Bacillus subtilis* inducible cold shock gene affecting cell viability at low temperature. *J. Bacteriol.* **174**, pp. 6326-6335.

WISTOW, G. (1990). Cold shock and DNA binding. *Nature* **344**, pp. 823-824.

WOJTKOWIAK, Z., BRIGGS, R. C. and HNILICA, L. S. (1983). A sensitive method for staining proteins transferred to nitrocellulose sheets. *Anal. Biochem.* **129**, pp. 486-489.

WOLFFE, A., TAFURI, S., RANJAN, M., and FAMILIARI, M. (1992). The Y-box factors: a family of nucleic acid binding proteins conserved from *Escherichia coli* to man. *The New Biologist* **4**, pp. 290-298.

WOODLAND, H. (1982). The translational control phase of early development. *Bioscience Reports* **2**, pp. 471-491.

YANG, H-Y. and EVANS, T. (1992). Distinct roles for the two cGATA-1 finger domains. *Mol. Cell. Biol.* **12**, pp. 4562-4570.

YOU, Y., CHEN, C. Y. and SHYU, A. (1992). U-rich sequence binding proteins (URBPSs) interacting with a 20-nucleotide U-rich sequence in the 3' untranslated region of c-*fos* mRNA may be involved in the first step of c-*fos* mRNA degradation. *Mol. and Cell. Biol.* **12**, pp. 2931-2940.

ZON, L. I., MATHER, C., BURGESS, S., BOLCE, M. E., HARLAND, R. M. and ORKIN, S. H. (1991). Expression of GATA-binding proteins during embryonic development of *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* **88**, pp. 10642-10646.

A role for Y-box proteins in cell proliferation

Michael Ladomery and John Sommerville

Summary

Members of the Y-box (YB) family of transcription factors are expressed in a wide range of cell types and are implicated in the regulation of a rapidly increasing number of genes. Although the biological activities of YB proteins appear to be varied, distinct patterns, relating to the timing of their expression and the identity of their target genes, are beginning to emerge. A recent report by Ito *et al.*[1] focusses attention on cell proliferation and adds support to earlier suggestions[2,3] that a primary function of YB proteins is to help activate growth-associated genes.

YB proteins

The Y-box (YB) proteins comprise a family of gene regulators, each containing a nucleic acid-binding domain which is highly conserved from bacteria to humans[4,5]. The archaetypal form is represented by the bacterial cold-shock proteins, which consist of a five-stranded β-barrel structure: the three N-terminal β-strands presenting a face with exposed aromatic and basic side-chains to interact with the nucleic acid (see ref. 6). Eukaryotic YB proteins consist of this conserved cold-shock domain (CSD) plus a series of alternating basic and acidic domains, which may specify further features of nucleic acid binding and protein-protein interaction. To date, most studies have focussed on the binding of YB proteins to the Y-box DNA sequence CTGATTGGYYUU, which contains a reverse CCAAT-box. The Y-box is present in the promoter region of a variety of eukaryotic genes and binding of YB proteins can result in either activation or repression of transcription. Tissue-specific expression of YB proteins indicates that they are involved in the regulation of distinct sets of genes. One set, already described, includes various vertebrate germ-cell-specific genes[4]: a second set, now becoming apparent, are genes involved in cell proliferation. Recent studies have highlighted cell types that are stimulated to express high levels of YB protein, which in turn may activate growth-associated genes and suppress tissue-specific genes.

Growth-inducible YB protein genes

The first indications of the involvement of YB proteins in cell proliferation came from studies by Sabath *et al.* on T-lymphocytes stimulated with interleukin 2 (IL2). Analysis of the IL2-induced genes in a cloned T helper lymphocyte line (L2 cells) revealed that one of these genes encoded the mouse YB protein, mYB1[2], the corresponding human protein having been shown previously to bind specifically the Y-box sequences in promoters of major histocompatibility complex class II (MHC II) genes[7] and in the enhancer region of the epidermal growth factor receptor (EGFR) gene[8]. Sabath *et al.* suggested that YB1-type proteins have a general function in growth-associated gene expression[2].

A general observation is that mRNA encoding YB proteins is expressed at markedly different levels in different tissues[9] and at different stages of development[4,5]. Grant and Deeley examined the expression of chicken YB1 mRNA in liver cells and showed that levels are high in the early embryo but decrease steadily throughout embryogenesis and post-hatching until low levels are reached in the adult liver[3]. However, they also found that chkYB-1 mRNA levels are 30- to 40-fold higher in the proliferating hepatoma cell line LMH and could be increased to levels 10-fold higher than the adult level after administering CCl_4, an agent which chemically induces liver regeneration[3]. These results have been reinforced by the recent studies of Ito *et al.* who have cloned from rat liver the gene encoding RYB-a, a YB protein similar to YB1[1]. Again, levels of RYB-a mRNA are shown to be high in foetal liver, low in adult liver and stimulated by partial hepatectomy. That the *RYB-a* gene is regulated coordinately with tissue regeneration/cell proliferation gains further support from studies on serum-stimulated quiescent fibroblasts (NIH/3T3 cells). Ito *et al.* have shown that fol-

lowing serum stimulation, the *RYB-a* gene is rapidly activated, but blocking of progression into S phase by detaching the cells or by treating the cells with genistein, a specific inhibitor of tyrosine kinase, almost completely suppresses expression of the *RYB-a* gene[1]. Since genistein is known to inhibit the tyrosine kinase activity of EGFR and pp60^{src}, the results suggest that expression of the *RYB-a* gene is involved in a signal transduction cascade leading to cell proliferation.

Y-box target genes

Several growth-associated genes have been shown to contain Y-box sequences in their promoter or enhancer regions. Among these are the genes encoding thymidine kinase[10,11], proliferating cell nuclear antigen (PCNA/cyclin)[12], DNA polymerase α[13] and EGFR[8]. The presence of a Y-box and the absence of the usual TATA-box indicate coordinate regulation of these genes during cell proliferation. The obvious candidate *trans*-acting factor is a YB protein, although detailed protein binding/transcription studies have not yet been reported in most examples.

One further example that has been studied in detail is the human *mdr*1 gene[14]. This multidrug-resistance gene encodes an energy-dependent drug efflux pump which is overexpressed in certain cancers that have become resistant to chemotherapy. Goldsmith *et al.* have made use of a doxorubicin-resistant ovarian carcinoma cell line to investigate nuclear factors binding to the *mdr*1 promoter[14]. Footprinting analysis showed that a region between nucleotide positions -85 and -70, protected from DNaseI digestion after protein binding, contains the sequence CTGATTGGCT, which matches the Y-box consensus. Mutation to CTGATGTGCT eliminated protein binding and deletion of the Y-box region led to markedly reduced expression of *mdr* promoter/CAT constructs. Although it has not been demonstrated that *mdr*1 overexpression is associated with the increased production of YB protein mRNA as detected in the proliferating tissues described earlier, it is interesting to note that multidrug-resistant cancer cells can also develop overexpression of EGFR[14].

The involvement of YB proteins in T-cell proliferation has acquired increased significance with recent studies on transcription of the human T-cell lymphotrophic virus type I (HTLV-I). Kashanchi *et al.* have studied the downstream regulatory element 1 (DRE 1) region in the LTR of HTLV-I and the site-A region in the LTR of HIV, both of which contain Y-box sequences[15]. Cotransfection of Jurkat T-cells with a YB-1 expression vector and with wild-type and mutant viral promoter/CAT constructs demonstrated that the Y-box sequence was essential for efficient transactivation[15]. Kashanchi *et al.* did not find that YB-1 expression was induced by viral factors such as Tax_1, but suggest, instead, that IL-2 induction of YB-1 , as described earlier for T-cell proliferation[2], might lead to stimulation of latent viral gene expression and viral replication.

YB proteins can also downregulate gene expression

During cell proliferation there is not only upregulation of growth-associated genes, but also suppression of tissue-specific genes. This effect is seen in embryonic and regenerating liver and in hepatocarcinoma cells, where high levels of expression of the chkYB-1 gene are associated with low levels of expression of liver-specific genes such as the serum albumin gene and the estrogen-dependent, very-low-density apolipoprotein II (apoVLDLII) gene[3]. Furthermore, there is an inverse correlation between levels of RYB-a mRNA and aldolase B mRNA during rat liver development[1].

Evidence for direct involvement of YB1 with a Y-box promoter to elicit negative regulation comes from studies on interferon-γ (IFN-γ)-induced MHC II genes (e.g. the human DRA gene). Ting *et al.* have shown that cotransfection of IFN-γ-responsive human glioblastoma cells with a YB1 expression vector and DRA-CAT constructs results in Y-box-specific suppression of the IFN-γ-activated DRA promoter[16]. Since a positive regulator of DRA expression, NF-Y/YEBP, can bind to the DRA gene Y-box, transcription control can be interpreted as an interplay of positive (NF-Y/YEBP) and negative (YB1) regulators competing for the same Y-box site[16].

YB proteins have a wide spectrum of activities

As a general principle, the actual effects of YB proteins on transcription regulation will depend on the composite structure of the promoter that contains a Y-box element, and also on the presence of other (competing or enhancing) regulatory proteins. The YB proteins are encoded by a multigene family[9,17,18], different members of which may be expressed in different tissues or at different stages of development and recognize the promoters of different sets of genes. For instance, the *Xenopus* protein FRGY2 is expressed at high levels during early oogenesis whereas FRGY1 mRNA, although present in oocytes, is not translated until early embryogenesis[19]. In this example, there is a fairly clear demarcation of functions: FRGY2 can regulate oocyte-specific genes, but more obviously binds to and packages mRNA molecules[20,21], whereas FRGY1 may have a more conventional role in promoting cell proliferation during embryogenesis. However, even within a single tissue, a range of YB protein isoforms could be created by differential phosphorylation, as has been

described for the FRGY2 proteins[22]. The significance of variation in phosphorylation of YB proteins to influence the ability to bind specific promoters is not known.

In addition to Y-box elements, one class of promoter structure highlighted in several studies, including those of Grant and Deeley[3], is a pyrimidine-rich single strand of DNA, possibly made accessible in regions of H-form (triplex) DNA. Regions of strong purine-pyrimidine strand asymmetry that can adopt the H-form are identified as promoter elements of the human c-*myc* gene[23], c-Ki-*ras* gene[23], EGFR gene[23] and foetal γ-globin gene[24]. In each instance, a YB protein specifically binds pyrimidine-rich strands. Although the pattern of expression of these genes fits the cell proliferation model already proposed, and indeed the stabilization of H-DNA by YB and other proteins has been suggested as a mechanism for up-regulating the *c-myc* gene[25], it must be concluded that regulation of gene expression by YB proteins might be effected through more than one type of promoter element. Therefore, the versatility of YB proteins, in being able to interact with Y-box promoters, H-DNA and mRNA, marks them out as key regulators of a range of growth-associated processes. Their true significance in this role should become apparent in the next few years.

References

1 Ito, K., Tsutsumi, K., Kuzumaki, T., Gomez, P.F., Otsu, K. and Ishikawa, K. (1994). A novel growth-inducible gene that encodes a protein with a conserved cold-shock domain. *Nucl. Acids Res.* **22**, 2036-2041.

2 Sabath, D.E., Podolin, P.L., Comber, P.G. and Prystowsky, M.B. (1990). cDNA cloning and characterization of interleukin 2-induced genes in a cloned T helper lymphocyte. *J. Biol. Chem.* **265**, 12671-12678.

3 Grant, C.E. and Deeley, R.G. (1993). Cloning and characterization of chicken YB-1: regulation of expression in the liver. *Mol. Cell. Biol.* **13**, 4186-4196.

4 Wolffe, A.P., Tafuri, S., Ranjan, M. and Familiari, M. (1992). The Y-box factors: a family of nucleic acid binding proteins conserved from *Escherichia coli* to man. *New Biol.* **4**, 290-298.

5 Wolffe, A.P. (1994). Structural and functional properties of the evolutionarily ancient Y-box family of nucleic acid binding proteins. *BioEssays* **16**, 245-251.

6 Schindelin, H., Jiang, W., Inouye, M. and Heinemann, U. (1994). Crystal structure of CSPA, the major cold-shock protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **91**, 5119-5123.

7 Didier, D.K, Schiffenbauer, J., Woulfe, S.L., Zacheis, M. and Schwartz, B.D. (1988). Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. *Proc. Natl. Acad. Sci. USA* **85**, 7322-7326.

8 Sakura, H., Maekawa, T., Imamoto, F., Yasuda, K. and Ishii, S. (1988). Two human genes isolated by a novel method encode DNA-binding proteins containing a common region of homology. *Gene* **73**, 499-507.

9 Spitkovsky, D.D., Royer-Pokora, B., Delius, H., Kisseljov, F., Jenkins, N.A., Gilbert, D.J., Copeland, N.G. and Royer, H.D. (1992). Tissue restricted expression and chromosomal localization of the *YB-1* gene encoding a 42kD nuclear CCAAT binding protein. *Nucl. Acids Res.* **20**, 797-803.

10 Lipson, K.E., Chen, S.-T., Koniecki, J., Ku, D.-H. and Baserga, R. (1989). S-phase-specific regulation by deletion mutants of the human thymidine kinase promoter. *Proc. Natl. Acad. Sci. USA* **86**, 6848-6852.

11 Gai, X., Lipson, K.E. and Prystowsky, M.B. (1992). Unusual DNA-binding characteristics of an *in vitro* translation product of the CCAAT binding-protein MYB-1. *Nucl. Acids Res.* **20**, 601-606.

12 Travali, S., Ku, D.-H., Rizzo, M.G., Ottavio, L., Baserga, R. and Calabretta, B. (1989). Structure of the human gene for the proliferating cell nuclear antigen. *J. Biol. Chem.* **264**, 7466-7472.

13 Pearson, B.E., Nasheuer, H.-P. and Wang, T.S.-F. (1991). Human DNA polymerase α gene: sequences controlling expression in cycling and serum-stimulated cells. *Mol. Cell. Biol.* **11**, 2081-2095.

14 Goldsmith, M.E., Madden, M.J., Morrow, C.S. and Cowan, K.H. (1993). A Y-box consensus sequence is required for basal expression of the human multidrug resistance (*mdr1*) gene. *J. Biol. Chem.* **268**, 5856-5860.

15 Kashanchi, F., Duvall, J.F., Dittmer, J. Mireskandari, A., Reid, R.L., Gitlin, S.D. and Brady, J.N. (1994). Involvement of transcription factor YB-1 in human T-cell lymphotrophic virus type I basal gene expression. *J. Virol.* **68**, 561-565.

16 Ting, J.P.-Y., Painter, A., Zeleznik-Le, N.J., MacDonald, G., Moore, T.M., Brown, A., and Schwartz, B.D. (1994). YB-1 DNA-binding protein represses interferon γ activation of class II histocompatibility complex genes. *J. Exp. Med.* **179**, 1605-1611.

17 Ozer, J., Chalkley, R. and Sealy, L. (1993). Isolation of the *CCAAT* transcription factor subunit *EFI$_A$* cDNA and a potentially functional *EFI$_A$* processed pseudogene from *Bos taurus*: insights into the evolution of the *EFI$_A$/dbpb/YB1* gene family. *Gene* **124**, 223-230.

18 Familiari, M., Almouzni, G. and Wolffe, A. (1994). Isolation of a potentially functional Y-box protein (MSY-1) processed pseudogene fom mouse: evolutionary relationships within the *EFI$_A$ /dbpb/YB1* gene family. *Gene* **141**, 255-259.

19 Tafuri, S.R. and Wolffe, A.P. (1990). *Xenopus* Y-box transcription factors: molecular cloning, functional analysis and developmental regulation. *Proc. Natl. Acad. Sci. USA* **87**, 9028-9032.

20 Marello, K., LaRovere, J. and Sommerville, J. (1992). Binding of *Xenopus* oocyte masking proteins to mRNA sequences. *Nucl. Acids Res.* **20**, 5593-5600.

21 Bouvet, P. and Wolffe, A.P. (1994). A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. *Cell* **77**, 931-941.

22 Cummings, A., Barrett, P. and Sommerville, J. (1989). Multiple modifications in the phosphoproteins bound to stored messenger RNA in *Xenopus* oocytes. *Biochim. Biophys. Acta* **1014**, 319-326.

23 Kolluri, R. Torrey, T.A. and Kinniburgh, A.J. (1992). A CT promoter element binding protein: definition of a double-strand and a novel single-strand DNA binding motif. *Nucl. Acids Res.* **20**, 111-116.

24 Horwitz, E.M, Maloney, K.A. and Ley, T.J. (1994). A human protein containing a cold shock domain binds specifically to H-DNA upstream from the human γ-globin genes. *J. Biol. Chem.* **269**, 14130-14139.

25 Kinniburgh, A.J., Firulli, A.B. and Kolluri, R. (1994). DNA triplexes and regulation of the *c-myc* gene. *Gene* **149**, 93-100.

Michael Ladomery and **John Sommerville** are at the School of Biological and Medical Sciences, University of St Andrews, St Andrews, Fife KY16 9TS, Scotland.

EUROPEAN SCIENCE FOUNDATION

EUROPEAN DEVELOPMENTAL BIOLOGY PROGRAMME

The European Science Foundation (ESF) announces a call for applicants for its Scientific Programme on Developmental Biology for the period 1992-1996.

The Programme focusses on:

MOLECULAR MECHANISMS OF EMBRYONIC INDUCTION AND PATTERN FORMATION IN ANIMAL DEVELOPMENT

This ESF Programme has Short and Long Term Fellowships available and organizes Conferences and Study Workshops on Developmental Biology.

Short and Long Term Fellowships are available to increase the mobility of qualified young scientists (generally under 35) and to encourage collaborative projects within the scientific scope of the Programme. Short Term Fellowships are available for a period up to 3 months. Long Term Fellowships have a duration up to 12 months with a possibility of extension up to another 12 months.

Fellowships are open to applicants from all European countries (Belgium, Denmark, Germany, Finland, France, Italy, Netherlands, Norway, Sweden, Switzerland, United Kingdom) and Israel, and will be awarded for exchange between these countries provided:

(i) the applicant is a permanent resident or is a citizen of one of the countries participating in the Programme, or

(ii) the applicant is a citizen of a non-participating European country or Israel and applies for a stay in a country participating in the Programme.

Long Term Fellowships: applications can be submitted twice a year, the deadlines being April 1st and October 1st. Annually a maximum of 12 fellowships will be awarded.

Short Term Fellowships: applications can be submitted at any time, but minimally 6 weeks before the intended visit.

EDB Conferences and Study Workshops on Developmental Biology will be organized annually to provide a forum to discuss and evaluate the scientific progress within the scope of the EDB Programme. The bi-annual Study Workshops offer an opportunity to young researchers to be informed about the current views and research strategies in this field, and to discuss their own research in this perspective.

Information:

S.W. de Laat, EDB Programme Chairman,
Hubrecht Laboratory, Netherlands
Institute for Developmental Biology,
Uppsalalaan 8, 3584 CT Utrecht,
The Netherlands
Tel. (31) 30 - 51 02 11
Fax. (31) 30 - 51 64 64

J.-H. Kock, ESF Scientific Secretary,
European Science Foundation,
1 quai Lezay Marnésia,
F-67080 Strasbourg Cedex,
France
Tel. (33) 88 - 76 71 27
Fax. (33) 88 - 37 05 32

5582–5589 Nucleic Acids Research, 1994, Vol. 22, No. 25

Binding of Y-box proteins to RNA: involvement of different protein domains

Michael Ladomery and John Sommerville*
School of Biological and Medical Sciences, University of St Andrews, St Andrews, Fife KY16 9TS, UK

Received September 27, 1994; Revised and Accepted November 18, 1994

ABSTRACT

Eukaryotic Y-box proteins are reported to interact with a wide variety of nucleic acid structures to act as transcription factors and mRNA masking proteins. The modular structure of Y-box proteins includes a highly conserved N-terminal cold-shock domain (CSD, equivalent to the bacterial cold-shock proteins) plus four basic C-terminal domains containing arginine clusters and aromatic residues. In addition, the basic domains are separated by acidic regions which contain several potential sites for serine/threonine phosphorylation. The interaction of Y-box proteins, isolated from *Xenopus* oocytes (FRGY2 type), with RNA molecules has been studied by UV crosslinking and protein fragmentation. We have identified two distinct binding activities. The CSD interacts preferentially with the polypurines poly(A,G) and poly(G) but not poly(A), this activity being sensitive to 5 mM $MgCl_2$ but not to 5 mM spermidine. In the presence of 1 mM $MgCl_2$ or 1 mM spermidine, the basic domains interact preferentially with poly(C,U), this activity being sensitive to 0.5 M NaCl. Binding of the basic domains is also sensitive to low concentrations of heparin. The basic domains can be crosslinked individually to labelled RNA. These results are discussed with reference to the various specificities noted in the binding of Y-box proteins to RNA and DNA.

INTRODUCTION

The Y-box proteins are the most evolutionarily conserved nucleic acid-binding proteins yet described, found in bacteria, plants and animals (1,2). The eukaryotic Y-box proteins were originally identified through their ability to interact with DNA containing a reverse CCAAT box, the Y-box sequence CTGATTGGCCAA (3). This sequence is found in a variety of promoter regions, including those of the MHC class II genes (3,4) and genes encoding germ cell-specific functions (1) and in these contexts the Y-box proteins are considered to act as regulators of transcription. However, a range of Y-box proteins were subsequently characterized through their selective interaction with promoter sequences containing pyrimidine-rich single-stranded DNA (5–7), apurinic DNA (8,9) and even purine-rich single-stranded DNA (10). Furthermore, the most abundantly expressed Y-box proteins, the FRGY2 class from *Xenopus* oocytes (11) and the MSY1 class from mouse spermatocytes, are found bound to mRNA (12,13). Thus a universal family of highly conserved proteins has been reported to recognize a diversity of nucleic acid structures. Either the remarkable versatility in nucleic acid recognition is a special property of all Y-box proteins and the *in vitro* conditions for their diverse binding properties have not yet been properly defined, or else individual Y-box proteins have evolved different specificities. The Y-box proteins consist of a modular series of domains, each of which has the potential to bind nucleic acids. Therefore the presence of multiple binding domains and variations in their arrangement might explain the different binding specificities of different Y-box proteins.

The most highly conserved feature of eukaryotic Y-box proteins is the presence of a 70 amino acid domain which is 43% homologous to the cold-shock protein, CSPA, of *E.coli* (14). The structure of a corresponding protein, CSPB, from *B.subtilis* has been solved by crystallographic (15) and NMR (16) analysis and consists of an antiparallel five-stranded β-barrel. More recently, this structure has been confirmed for CSPA (17,18). The three N-terminal β-strands present a face with exposed aromatic and basic side chains which could interact with nucleic acids. This cold-shock domain (CSD) is sufficient to bind Y-box sequences *in vitro* and to drive the expression of bacterial cold-shock-inducible genes (19,20). Apart from the mammalian protein unr (encoded by an ORF located upstream of n-*ras*), which consists of five tandem repeats of a CSD (21), eukaryotic Y-box proteins contain a single CSD plus a series of C-terminal charged domains. FRGY2 is quite typical in containing four basic domains, rich in arginine and aromatic residues, separated by four domains rich in acidic residues and potential sites for serine/threonine phosphorylation (11,22,24). Arginine-rich domains are a common feature of many RNA-binding proteins (25) and their multiple presence in Y-box proteins increases the potential for nucleic acid binding.

In *Xenopus* oocytes, Y-box proteins are expressed at high concentration (~0.1 μg/oocyte) as an equimolar pair (here FRGY2a = pp60 and FRGY2b = pp56; cf 11). They can be readily isolated from native mRNP particles (26,27) and by

*To whom correspondence should be addressed

simple heat treatment of oocyte homogenates (28). The particular advantage of working with native rather than with recombinant forms is that they are isolated in a state of interaction with nucleic acid. Since phosphorylation of FRGY2 is relevant, and possibly crucial, for binding to mRNA (26,29,30) and until phosphorylation sites are pinpointed and manipulated *in vitro*, proteins properly modified *in vivo* are to be preferred.

In this report we examine further the RNA binding preferences of the FRGY2 proteins and discriminate between different binding domains by UV crosslinking and protein fragmentation.

MATERIALS AND METHODS

Isolation of Y-box proteins

Poly(A)$^+$ mRNP was isolated from the previtellogenic ovary of *Xenopus laevis* as described previously (30). To select the Y-box proteins, the eluate from oligo(dT) cellulose (Pharmacia) was adjusted to 20 mM NaCl, 20 mM Tris−HCl, pH 7.5 (HTB), heated to 80°C for 10 min, cooled at room temperature for 5 min, chilled on ice for 5 min and centrifuged at 10,000 r.p.m. for 10 min (28). Under these conditions the Y-box proteins remain in the supernatant and can be recovered quantitatively as a complex with mRNA. Multiple cycles of heat treatment/centrifugation improve the purity of the preparation. All other mRNP proteins are denatured and can be recovered from the pellets. Riboprobes added to mRNP prior to heat treatment are recovered complexed with the Y-box proteins.

To separate the Y-box proteins from mRNA, the heat-treatment supernatant is subjected to a further cycle of heat treatment in the presence of heparin−Sepharose CL 6B (Pharmacia). About 100 mg of RNP and 0.5 ml of resin was suspended in 1 ml of HTB and heated to 80°C. After cooling, with continuous shaking, the slurry was pipetted into a column and, after settling, was rinsed with HTB. The RNA was eluted either with 8 M urea or with 5 mM $MgCl_2$. 8 M urea also washed off contaminating proteins. Finally the Y-box proteins were eluted with 1 M NaCl, dispensed into aliquots and stored at −70°C. The purity of preparations was checked by SDS−PAGE, the Y-box proteins having apparent molecular weights of 60 and 56 kDa.

Synthesis of riboprobes

The RNA used for binding studies corresponds to the 3′ end (280 nucleotides) of an mRNA encoding an oocyte-specific β-tubulin. Over 90% of this mRNA is packaged as non-polysomal mRNP particles and is representative of mRNAs associated with Y-box proteins in *Xenopus* oocytes (N.Clark, M.Ladomery, and J.Sommerville, unpublished). The sequence is: GAAUUCACUG AGGCCGAGAG CAACAUGAAC GACCUGGUGU CUGAGUACCA ACAGUACCAG GAUGCCACGG CUGAGGAGGA GGGAGAGUUU AAGAGGAGGA AAAUGCCUAA AGCUCCUUAU CACUUGUAAA UUAUUCAUCC AUUUCUGCUC GUUUGUUCCA UUUAUUUGUC ACCUGCAUUU CUCUUCUCUC CAGUUUCAAU GUUACCAGUU GUACAGAACG UUCCGUUCAU UAAAAGCAUU UUUCAUACUG AAAAAAAAAAAAAAAAAAAA.

The translation stop codon UAA, the polyadenylation motif AUUAAA and a potential cytoplasmic polyadenylation element UUUUCAU are underlined.

The RNA was synthesised by run-off transcription from a cDNA subclone in pBlueScript (Stratagene) in the presence of [α-^{32}P]CTP as described previously (30). Double-stranded RNA was formed by annealing equimolar amounts of labelled sense strands and unlabelled antisense strands. Remaining single strands were digested with ribonuclease and the double-stranded probe was purified by phenol extraction and ethanol precipitation.

Various ribopolymers were dissolved at 10 mg/ml in distilled water (stock). The heteropolymers poly(C,U) and poly(A,G) were dephosphorylated and 5′ ends were end-labelled using [γ-^{32}P]ATP (3,000 Ci/mmol) and polynucleotide kinase as recommended on the enzyme data sheet (Pharmacia).

Protein−riboprobe binding

0.1 μg of riboprobe was added to 2 μg of Y-box proteins (or RNP) and binding was accomplished in a final volume of 50 μl by either: (i) dialysis from 1 M NaCl into HTB; or (ii) dilution from 20 to 2 mM $MgCl_2$ with HTB; or (iii) heat treatment as described above. Protein−riboprobe interaction was challenged by adding heparin, salts, urea or polynucleotides either before or after the binding step. Conditions for the crosslinking of RNA to proteins using UV light were as described previously (30). For quantitative assessment of crosslinking efficiency, 50 μl samples were adjusted to 1% SDS, 0.15 M NaCl, 5 mM EDTA, 50 mM Tris−HCl, pH 7.5, and extracted with an equal volume of phenol−chloroform (1:1 mixture). After phase separation, the radioactivity in aqueous and organic phases was measured and the percentage of initial radioactivity extracted into the organic phase was calculated. For identification of proteins and proteolytic fragments crosslinked to riboprobe, samples were digested with

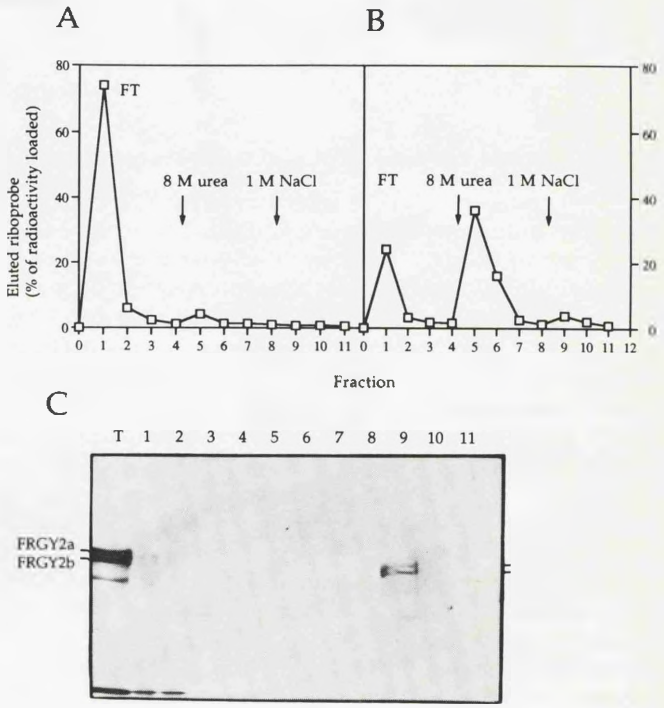


Figure 1. Elution of RNA and FRGY2 proteins from heparin−Sepharose. (**A**) Riboprobe−protein complexes applied to the column elute in the flow-through (FT) fraction. (**B**) Riboprotein complexes formed in the presence of heparin−Sepharose bind to the resin, the riboprobe being eluted with 8 M urea, the FRGY2 proteins being eluted with 1 M NaCl. (**C**) Proteins in eluted fractions as detected by SDS−PAGE. Riboprobe was eluted with 5 mM $MgCl_2$ in fractions 5 and 6; FRGY2 proteins were eluted with 1 M NaCl in fraction 9. Total protein preparation prior to chromatography is also shown (T).

ribonuclease, denatured and separated by SDS–PAGE (30). Proteins transferred to nitrocellulose or present in dried gels were then examined for bound, protected radioactivity by autoradiography. The identity of transferred Y-box proteins or fragments was subsequently confirmed by immunoblotting with antibodies raised against heat-treatment supernatant.

Binding of riboprobe to gel transfers was performed as follows. Transfers were washed twice for 10 min in 6 M urea in binding buffer (50 mM NaCl, 0.5% Tween-20, 10 mM Tris–HCl, pH 7.5). Riboprobe (0.5–1.0 μg) was added in 5–20 ml of binding buffer, then incubated for 30–60 min. Finally the transfers were washed twice for 10 min in binding buffer, either with 50 mM NaCl (low salt wash) or with 500 mM NaCl (high salt wash). The transfers were then set up for autoradiography.

Band-shift assays were used to assess the binding of FRGY2 proteins to end-labelled ribopolymers. Complexes, formed in the presence or absence of $MgCl_2$ and competitors, were electrophoresed, without crosslinking, through 1.5% agarose in 20 mM Tris–HCl, pH 7.5. The gels were fixed in 10% acetic acid, dried and set up for autoradiography.

Chemical proteolysis

Aspartate–proline bonds were hydrolysed by incubating crosslinked FRGY2–riboprobe complexes (made using 10 μg of protein and 0.5 μg of riboprobe in a final volume of 50 μl) in 70% formic acid at 40°C for 24 h (31).

Cleavage at asparagine–glycine residues was achieved by incubating FRGY2 proteins in 2% hydroxylamine at 20°C for 4 h (32). The mixture was maintained at pH 9 with NaOH. Chemically cleaved proteins were extensively dialysed against HTB before use in RNA binding.

RESULTS

FRGY2 proteins can bind RNA and heparin simultaneously

The Y-box proteins FRGY2a (pp60) and FRGY2b (pp56) can be purified from *Xenopus* oocyte mRNP particles in three steps: (i) selection of poly(A)$^+$ RNP by affinity binding to oligo(dT) cellulose (30); (ii) heat treatment of the poly(A)$^+$ RNP at 80°C followed by chilling and centrifugation (28); (iii) binding of the heat-treatment supernatant to heparin–Sepharose followed by salt elution of the Y-box proteins (described below). The RNA used in the binding studies described here represents 280 nucleotides of the 3′ end of an oocyte-specific β-tubulin mRNA (see Materials and Methods). Preliminary studies (not shown) confirmed that this sequence, but not its double-stranded form made by hybridizing sense and antisense strands, was highly efficient in binding FRGY2 proteins.

Complexes of FRGY2 proteins and RNA, formed *in vitro* but not crosslinked, do not bind to heparin–Sepharose (Fig. 1A), indicating that the arginine clusters in the protein tail domains are not accessible for binding to heparin and are neutralized through protein–RNA and/or protein–protein interaction. However, complexes formed in the presence of heparin–Sepharose are bound to the resin, the RNA component being released with 8 M urea or 5 mM $MgCl_2$, and the proteins being subsequently released with 1 M NaCl (Fig. 1B and C; see also 5,7). These results indicate that regions of positive charge (arginine clusters) in the proteins interact with the negatively charged heparin, but not at the expense of all of the protein interaction with RNA. Thus, it is possible that recognition of RNA can be achieved independently of the basic charge domains, for instance through the CSD (2).

RNA binding is blocked by a combination of heparin and Mg^{2+}

Points of contact of Y-box proteins with RNA can be fixed by crosslinking with UV irradiation (29,30). In principle, multiple RNA-binding sites within a single protein can be discriminated by differential blocking or by fragmenting the protein.

Contacts between FRGY2 proteins and RNA can be established in the presence of 10 μg/ml heparin, 3 mM $MgCl_2$, 0.5 M NaCl or 4 M urea, the percentage of riboprobe radioactivity crosslinked to protein under these conditions being similar to the control value (Fig. 2A). However, certain combinations of these agents, namely heparin plus $MgCl_2$, heparin plus urea and NaCl plus $MgCl_2$, almost completely prevent crosslinking (Fig. 2A). Even RNA–protein complexes formed first and then treated with heparin/$MgCl_2$ or heparin/urea before crosslinking are

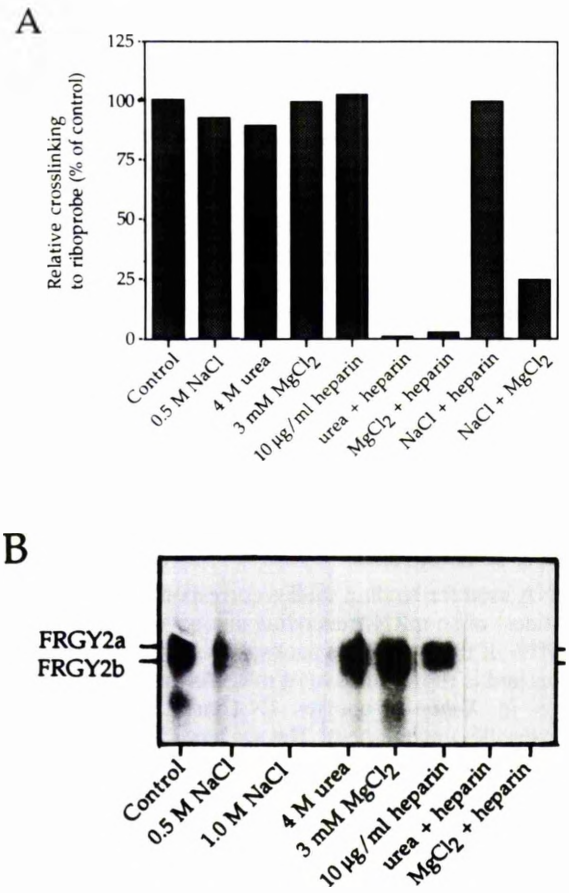


Figure 2. Effect of different agents on the UV crosslinking of riboprobe to FRGY2 proteins. (**A**) The binding buffer (20 mM NaCl, 20 mM Tris–HCl, pH 7.5) was adjusted with the agents shown prior to the binding of proteins to RNA, UV irradiation and phenol extraction. (**B**) Confirmation of crosslinking to FRGY2a and FRGY2b. In this experiment, the buffer was adjusted with the agents shown after the binding reaction. Then the complexes were UV irradiated, digested with ribonuclease and analysed by SDS–PAGE/autoradiography. Note that in (A), a minimum of one crosslinking event per protein–riboprobe complex is sufficient to be recorded as maximum binding, whereas in (B), the intensity of labelling is proportional to the number of crosslinking events.

susceptible to dissociation (Fig. 2B). That heparin and urea should cooperate to prevent binding and to disrupt complexes is hardly surprising, since protein conformation and RNA recognition are largely dependent upon charge interaction and hydrogen bonding. However, the combined effect of low concentrations of heparin and Mg^{2+} is more interesting and potentially useful.

Heparin and Mg^{2+} block different binding domains

As mentioned earlier, the effect of heparin is most likely to block the positively charged tail domains of the Y-box proteins, leaving the CSD for interaction with nucleic acid (in the absence of Mg^{2+}). The additional effect of Mg^{2+} would therefore operate mainly through the CSD. That this is, indeed, the case is demonstrated by disrupting the CSD. In all vertebrate Y-box proteins whose sequence has been determined, there exists an unique NG (asparagine−glycine) site at the beginning of the second β-strand of the β-barrel. This site (...RNGYGFINR...) shows homology with the RNP-1 site of the RRM-containing family of RNA-binding proteins (11,25). Cleavage of FRGY2 proteins at the NG site (see Fig. 5A) with hydroxylamine (HA) results in a protein preparation which still crosslinks to RNA in the presence or absence of 3 mM $MgCl_2$, this binding showing the characteristics of heparin inhibition similar to those obtained with the native proteins (Fig. 3A). These results would be explained by a Mg^{2+}-induced block operating at, or near to, the CSD and a heparin block imposed by binding to the basic tail domains. That the Mg^{2+}-induced block applies to proteins with an intact CSD, but not to the HA-cleaved proteins, is seen in Fig. 3B. It is also seen that heparin inhibition of tail domain binding (in the HA-cleaved proteins) is activated by low concentrations of $MgCl_2$ (< 1 mM). One further consequence of disrupting the CSD is that the stability of interaction with the riboprobe is substantially reduced. Thus 0.5 M NaCl appears to be sufficient to prevent binding by the tail domains of the HA-cleaved proteins, yet has little effect on crosslinking to the intact proteins (Fig. 3C).

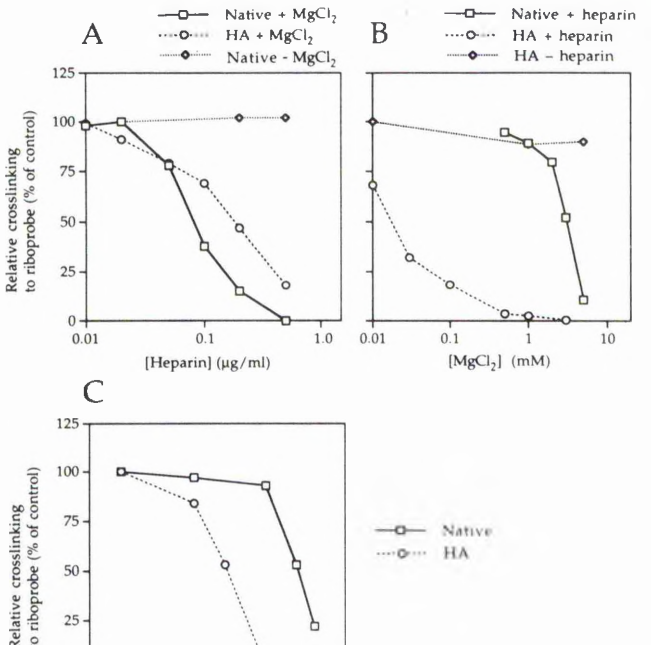


Figure 3. Effects of disrupting the CSD on the ability of FRGY2 proteins to bind riboprobe in the presence of heparin, $MgCl_2$ and NaCl. Intact FRGY2 (Native) and proteins cleaved with HA were crosslinked to riboprobe in the conditions indicated. (**A**) Heparin-sensitive binding activity in the presence and absence of $MgCl_2$ (3 mM). (**B**) Mg^{2+}-sensitive binding activity in the presence and absence of heparin (10 μg/ml). (**C**) Na^+-sensitive binding activity.

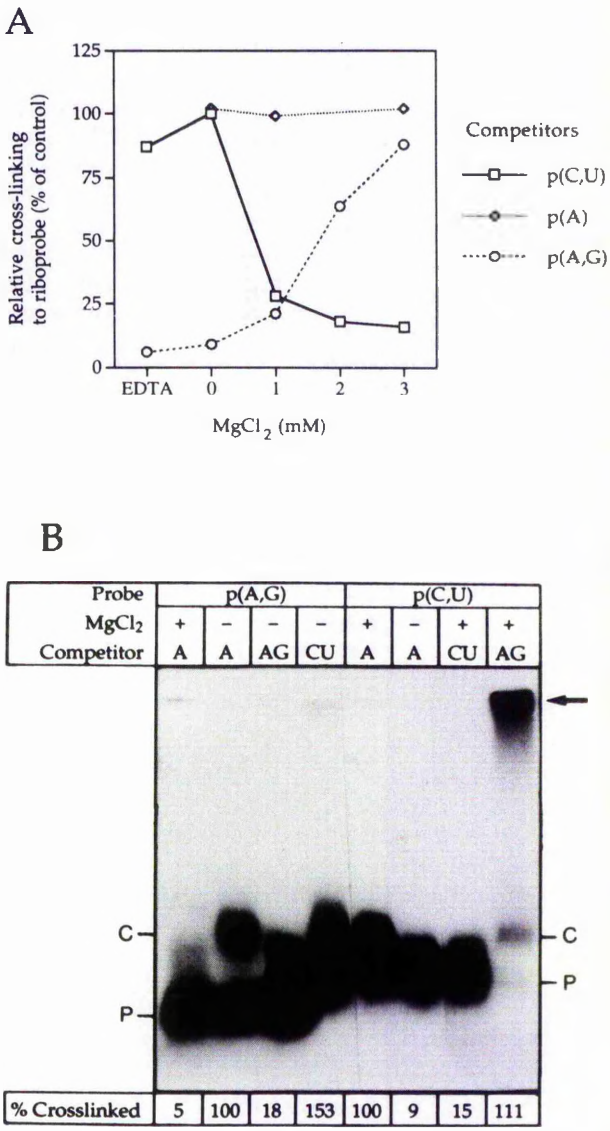


Figure 4. Effects of Mg^{2+} on the interaction of FRGY2 proteins wit ribopolymers. (**A**) The Mg^{2+}-induced switch in binding specificity. Th FRGY2−riboprobe interaction was challenged with a 100-fold excess ov riboprobe of poly(C,U), poly(A) or poly(A,G) at the concentrations of MgC shown or in the presence of 1 mM EDTA. (**B**) Band-shift assay showing dire binding of radiolabelled ribopolymers. The poly(A,G) probe (tracks 1−4) an the poly(C,U) probe (tracks 5−8) were bound to FRGY2 proteins in the absenc (tracks 2−4 and 6) and presence (tracks 1, 5, 7 and 8) of 3 mM $MgCl_2$ an in the presence of 100-fold excess of unlabelled poly(A) (tracks 1, 2, 5 and 6 poly(A,G), (tracks 3 and 8) and poly(C,U) (tracks 4 and 7). The positions unbound probe (P), protein−RNA complexes (C) and larger aggregates (arrow are indicated. Confirmation that protein−RNA complexes have been formed given by the relative crosslinking values obtained from the corresponding reaction

Sequence recognition is influenced by Mg^{2+}

We reported previously (30) that the FRGY2 proteins show a marked binding preference for polypyrimidines, the heteropolymer poly(C,U) being the best competitor of riboprobe binding tested. These experiments were performed in the presence of Mg^{2+}, diluted from 20 mM (to destabilize native mRNP complexes) to 2 mM $MgCl_2$ (to allow binding of the Y-box proteins to riboprobe). As shown here, this binding preference can now be largely attributed to the tail domains. The results (Fig. 4A) confirm that competition by poly(C,U) is actually Mg^{2+}-dependent. Of a range of polymers used to study competition in the absence of Mg^{2+}, by far the best competitors were the polypurines, poly(A,G) and poly(G), but as with Mg^{2+}-dependent binding (30), CSD binding showed no competition by poly(A) (Fig. 4A). The polypurine-binding activity of the CSD was progressively inhibited by 1–3 mM $MgCl_2$. In effect, a switch in binding affinity from poly(A,G) to poly(C,U) can be achieved simply by increasing the concentration of $MgCl_2$.

That FRGY2 proteins can interact directly with poly(A,G) or poly(C,U) depending on the binding conditions is shown by band-shift assays using labelled polymers (Fig. 4B). In the presence of 3 mM $MgCl_2$, poly(A,G) is not bound, whereas poly(C,U) is retarded as RNA–protein complexes (C). In the absence of $MgCl_2$, poly(A,G) forms complexes, whereas poly(C,U) is not bound. The binding reactions are unaffected by adding excess amounts of unlabelled poly(A), but are competed by adding the same polymer as the bound probe (Fig. 4B). Addition of excess poly(C,U) in the absence of Mg^{2+} and excess poly(A,G) in the presence of Mg^{2+} improves binding efficiency to the poly(A,G) and poly(C,U) probes, respectively. One further consequence of adding excess poly(A,G) to the reaction in which the tail domains bind the poly(C,U) probe is that the complexes are driven into larger aggregates (arrow in Fig. 4B). The molecular basis for this increased extent of interaction is not obvious.

Fragmentation of FRGY2 proteins reveals multiple binding domains

As already discussed, the CSD of the FRGY2 proteins can be disrupted with HA cleavage at a unique NG site (Fig. 5A). In addition, hydrolysis of proteins with formic acid (FA) results in preferential cleavage at aspartate–proline (DP) residues (31). A single DP site exists in FRGY2b (Fig. 5A), such that FA treatment results in an N-terminal fragment containing the CSD plus two basic domains and a C-terminal fragment containing two basic domains. Production of these fragments from isolated FRGY2b has been confirmed by immunostaining (not shown). In FRGY2a there are three further DP sites (Fig. 5A), resulting

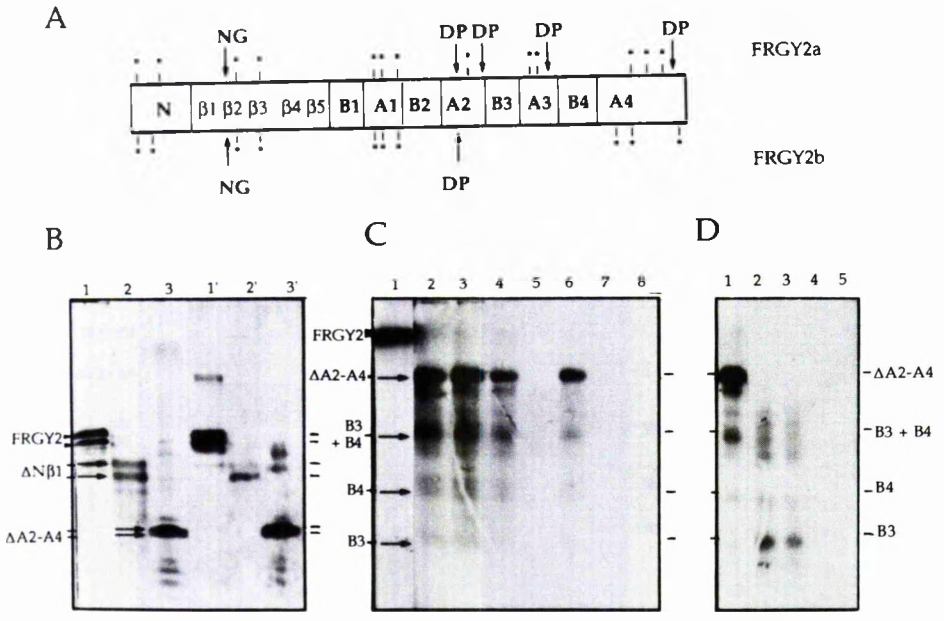


Figure 5. RNA binding to chemically cleaved FRGY2 proteins. (**A**) Diagram of FRGY2a/b showing their linear structure, consisting of: the N-terminal region (N); the CSD (β1–β5); the acidic domains (A1–A4); the basic domains (B1–B4). The position of the HA-sensitive NG site and the positions of the FA-sensitive DP sites are indicated by arrows. Also shown are the potential sites of phosphorylation by the RNP-bound casein kinase II (asterisks). (**B**) Cleaved fragments retain RNA-binding activity. FRGY2 proteins (tracks 1 and 1′) and fragments produced by HA (tracks 2 and 2′) and FA (tracks 3 and 3′) treatments were separated by SDS–PAGE, transferred to nitrocellulose and either immunostained using anti-FRGY2 (tracks 1–3) or incubated with riboprobe to produce the autoradiograph shown (tracks 1′–3′). Positions of the major fragments C-terminal to the NG site (ΔNβ1) and N-terminal to the DP sites in A2 (ΔA2–A4) are indicated by arrows on the immunoblot. (**C**) Points of crosslinking (contact) are established throughout much of the length of the protein. FRGY2–riboprobe complexes, formed under different conditions, were crosslinked, digested with ribonuclease, cleaved with FA and analysed by SDS–PAGE/autoradiography. Complexes were formed from 10 μg of protein and 0.5 μg of riboprobe in binding buffer with no addition (tracks 1 and 2) or with addition of: 3 mM $MgCl_2$ (track 3); 10 μg of heparin (track 4); $MgCl_2$ plus heparin (track 5); 20 μg of poly(C,U) in the absence (track 6) or presence (track 7) of $MgCl_2$; 20 μg of poly(A,G) in the absence of $MgCl_2$ (track 8). Indicated by arrows are the positions of: the intact proteins (FRGY2); the N-terminal fragments (ΔA2–A4); the C-terminal fragment from FRGY2b (B3 + B4); the C-terminal fragments from FRGY2a (B3 and B4). (**D**) Individual basic tail domains can bind RNA in the absence of a functional CSD. HA-cleaved proteins were bound to riboprobe and then crosslinked and treated with FA as described above. Complexes were formed in the absence (tracks 1 and 2) or presence (tracks 3–5) of 3 mM $MgCl_2$ and with the addition of 2 μg heparin (track 4) or 20 μg of poly(C,U) (track 5). Fragments could be aligned with those produced by FA treatment alone and run on the same gel (track 1). Identity of fragments as shown for (C).

in a similar N-terminal fragment plus two small fragments each containing a single basic domain.

The major HA-cleaved fragments (ΔNβ1) and FA-cleaved fragments (ΔA2-A4) can be clearly seen in the immunoblot shown (Fig. 5B, tracks 2 and 3). That these fragments retain RNA-binding activity is seen in the duplicated blot, which has been incubated with labelled riboprobe (Fig. 5B, tracks 2′ and 3′). In this particular assay, binding was accomplished in the absence of Mg^{2+}, therefore labelling of fragments with disrupted CSDs (ΔNβ1) is poor in comparison with fragments containing intact CSDs (ΔA2-A4). Furthermore, washing of the filters in higher salt (0.5 M NaCl) resulted in complete stripping of label from fragments lacking a functional CSD (not shown), again emphasizing that binding by the basic tail domains is relatively salt-sensitive. Further analysis focussed on binding of riboprobe to proteins and their fragments which had not been previously denatured.

The intact FRGY2 proteins were crosslinked to riboprobe, fragmented by acid treatment, treated with ribonuclease and then separated by SDS−PAGE. Autoradiography revealed which of the fragments were in contact with the radiolabelled RNA. As shown in Fig. 5C (track 2), the FA-cleaved fragments containing isolated basic domains were crosslinked to the riboprobe. The best competitive combinations were again heparin (track 5) and poly(C,U) (track 7) in the presence of Mg^{2+}. Note that poly(A,G) is a strong competitor when the intact protein with a functional CSD is involved (track 8). The use of HA-cleaved fragments in binding prior to crosslinking and separation confirmed that they behave as basic tail domains, i.e. riboprobe binding is competed by poly(C,U) and heparin and not by poly(A,G) in the presence of low concentrations of Mg^{2+} (not shown). Further FA digestion of the HA-cleaved material at DP sites generates individual labelled basic domains (Fig. 5D, tracks 2 and 3), which correspond to those produced by FA cleavage of FRGY2−riboprobe complexes (track 1). Again, binding is competed by heparin (track 4) and poly(C,U) (track 5) in the presence of Mg^{2+}.

To what extent sequence recognition by the different binding domains of FRGY2 proteins is Mg^{2+}-dependent rather than cation-sensitive was checked by substituting spermidine for $MgCl_2$ in binding reactions prior to crosslinking. The results are summarized in Table 1 and lead to the main conclusions that: (i) The CSD interacts preferentially with poly(A,G), to a slightly lesser degree with poly(G), (not shown), but not at all with poly(A), this activity being specifically inhibited by 5 mM $MgCl_2$; (ii) a low concentration of cations, 1 mM of either $MgCl_2$ or spermidine, is required for a specific interaction of tail domains with poly(C,U), this activity being lost in 0.5 M NaCl or competed by heparin. As shown previously (30), poly(C,U) competes for protein binding, on a mass:mass basis, almost equally with the more complex riboprobe sequence. Similarly, in conditions favouring poly(A,G) binding, the synthetic ribopolymer competes almost as well as the riboprobe (not shown). However, the optimal recognition sequence, and hence its frequency of occurrence in the different riboprobes, is still unknown for each of the binding activities.

DISCUSSION

The procedure outlined here for the purification of Y-box proteins from mRNP could in principle be applied elsewhere, for example to extract Y-box proteins from somatic tissues, where they may be much less abundant than in *Xenopus* oocytes. It is assumed that the treatment employed does not disrupt the binding specificities of the proteins and, in particular, that the cold-shock domain is either stable at 80°C or is at least able to renature correctly. Thermostability appears to be a general property of Y-box proteins; for instance, the human YB-1 Y-box−DNA complex is resistant to heating to at least 69°C (3).

The initial observation was that the combined action of heparin and Mg^{2+} completely blocks binding of FRGY2 proteins to RNA. That Mg^{2+} influences the activity of Y-box proteins is apparent in other studies, notably with YB-1 (9), where it is suggested that Mg^{2+} enhances binding to apurinic DNA, and with unr (33) where binding to both DNA and RNA targets reveals a sensitivity towards Mg^{2+} at about 1 mM. The unr protein consists of a 5-fold repeat of the CSD with no auxiliary tail domains. It may be inferred that a similar sensitivity to Mg^{2+} applies to the single CSD of Y-box proteins. The effect of Mg^{2+}, at least on FRGY2 binding to RNA, appears to be 2-fold: to interfere with binding by the CSD and to favour binding by the tail domains over the range 1−5 mM. These two effects appear to be due to different types of molecular mechanism because the first is specific to Mg^{2+}, whereas the second can be obtained with the structurally distinct cation spermidine. The contribution of heparin to the blocking of RNA binding is explained by the interaction of this polyanion with the arginine clusters contained in the tail domains, thus competing efficiently with RNA for binding sites.

To date, the ability of FRGY2 tail domains to bind RNA has been discussed (34), but not considered in detail. In the present study we show that after RNA binding, UV irradiation and protein fragmentation, individual tail domains of FRGY2 have been crosslinked to riboprobe. Furthermore, crosslinking of riboprobe

Table 1. Effect of cations on the binding specificities of FRGY2 domains*

Cations added:	Cold-shock domain poly(A,G)	poly(C,U)	Basic tail domains poly(A,G)	poly(C,U)
0.1 M Na^+	+	−	+	−
0.5 M Na^+	+	−	0	0
1 mM Mg^{2+}	+	−	(+)	+
5 mM Mg^{2+}	0	0	−	+
5 mM $Spermidine^{3+}$	+	(+)	−	+

*Efficiency of crosslinking to riboprobe was measured in the presence of poly(A,G), poly(C,U), both at 40 μg/ml and the concentrations of NaCl, $MgCl_2$ or spermidine shown. Scores of + represent competition levels of >90% of controls (minus competing polymer); scores of − represent competition levels of <20% of controls; and scores of (+) represent intermediate values. Scores of 0 indicate that no binding to riboprobe could be obtained under the conditions shown.

to HA-cleaved fragments confirms that the tail domains can bind to RNA independently of a functional CSD. Loss of function of the CSD after treatment of FRGY2 proteins with HA is inferred from the marked increase observed in sensitivity of RNA binding to heparin and NaCl. The NG cleavage site lies within the β2-strand of the CSD and cleavage would result in complete loss of the β1-strand from the main part of the protein and disruption of the β2-strand itself. Since the β1−β3-strands of the bacterial cold-shock proteins have been identified as forming the structure which interacts with nucleic acids (15−18) and since deletion of part of the β2-strand of the NSEP-1 Y-box protein results in loss of DNA-binding activity (5), it is reasonable to assume that FRGY2 would similarly lose the function of its CSD on cleavage.

The reported binding specificity of Y-box proteins to DNA is remarkably varied. In addition to interaction of Y-box proteins with Y-box promoter elements of vertebrates (3,22,35,36), HTLV, HIV and RSV (37,38) and bacteria (19,20), there appears to be an affinity for pyrimidine-rich elements in the promoters of c-*myc* (5), γ-globin (7) and the liver-specific gene *apoVLDLII* (6), and now also for purine-rich elements in the LTR of RSV (10). One possibility is that the different specificities observed are due to differences between Y-box proteins in regions less conserved than the essentially invariant CSD. Although the tail domains generally consist of alternating basic and acidic regions, their sequence is not conserved between different proteins. A second possibility is that the Y-box proteins all recognize a similar range of sequences but that different targets are recognized by different protein domains. This aspect has been investigated in the RNA-binding studies reported here, and similarities in the mode of binding of Y-box proteins to DNA and to RNA can be considered. For example, most of the studies made on the binding of Y-box proteins to DNA emphasize the preferred interaction with single-stranded targets; likewise, we note that double-stranded RNA presents a poor binding template. Furthermore, two Y-box proteins involved in developmental regulation, FRGY2 (22) and MSY1 (12,13), have important roles in masking mRNA from translation, in addition to their gene regulatory activities. Indeed, Y-box proteins in general may have a role in mRNA packaging.

In our RNA-binding assays we describe a preference for the polypurines p(A,G) and p(G), but not p(A), a binding mediated by the CSD. In addition we describe a cation-dependent preference for polypyrimidines due to the tail domains. This second type of binding has all of the characteristics that we described earlier when Mg^{2+} was an integral component of the binding reaction (30). By analogy, the various DNA binding activities so far reported may be due to the activity of the CSD, tail domains or a cooperative combination of both, depending on the reaction conditions used. In the context of binding to promoter elements, the Y-box proteins may be associating with DNA via the cooperation of both types of domain. The presence of an actual Y-box DNA element appears not to be essential for binding to certain promoters where, for example, an H-DNA structure is induced in regions of strong purine−pyrimidine strand asymmetry, exposing a pyrimidine-rich single strand from a triplex strand (30). It is interesting to note that the structure of H-DNA appears to be stabilized by Mg^{2+} (7), which is a factor determined here for activating tail domain binding to RNA. Mg^{2+} is also known to stabilize secondary structures in RNA (40), raising the possibility that its influence on FRGY2 binding is through modifying the structure of the RNA itself. An alternative possibility is that Mg^{2+} is involved directly in the interaction between amino acids and nucleotide residues, as has been suggested for the binding of HeLa transcription factor USF to the E-box (41).

Although the exact nature of the molecular interaction between Y-box proteins and nucleic acids is unknown, we have shown that *in vitro* binding to RNA is highly sensitive to ionic environment, so much so that the activity of the CSD is destroyed by 5 mM $MgCl_2$ and that the activity of the basic tail domains in binding polypyrimidines is dependent upon the presence of low concentrations of multivalent cations (Mg^{2+} and spermidine^{3+}). Since the preference in sequence interaction is different for the two binding reactions, the mixed polypurine, poly(A,G), being favoured by the CSD and the mixed polypyrimidine poly(C,U) being favoured by the tail domains, a wide range of binding specificities could be achieved. Such versatility might be advantageous in packaging a wide variety of mRNA sequences, while at the level of promoter binding, different elements such as the Y-box sequence and CT-rich strands could be recognized under different *in vivo* conditions.

REFERENCES

1. Wolffe, A.P., Tafuri, S., Ranjan, M., and Familiari, M. (1992) *New Biol.* **4**, 290−298.
2. Wolffe, A.P. (1994) *BioEssays* **16**, 245−251.
3. Didier, D.K., Schiffenbauer, J., Woulfe, S.L., Zacheis, M., and Schwartz, B.D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7322−7326.
4. Dorn, A., Durand, B., Marfing, C., Lemeur, M., Benoist, C., and Mathist, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6249−6253.
5. Kolluri, R., Torrey, T.A., and Kinniburgh, A.J. (1992) *Nucl. Acids Res.* **20**, 111−116.
6. Grant, C.E., and Deeley, R.G. (1993) *Mol. Cell. Biol.* **13**, 4186−4196.
7. Horwitz, E.M., Maloney, K.A., and Ley, T.J. (1994) *J. Biol. Chem.* **269**, 14130−14139.
8. Lenz, J., Okenguist, S.A., LoSardo, J.E., Hamilton, K.K., and Doetsch, P.W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3396−3400.
9. Hasegawa, S.L., Doetsch, P.W., Hamilton, K.K., Martin, A.M., Okenquist, S.A., Lenz, J., and Boss, J.M. (1991) *Nucleic Acids Res.* **19**, 4915−4920.
10. Kandala, J.C., and Guntaka, R.V. (1994) *Virology* **198**, 514−523.
11. Sommerville, J. (1992) *BioEssays* **14**, 337−339.
12. Kwon, Y.K., Murray, M.T., and Hecht, N.B. (1993) *Dev. Biol.* **158**, 90−100.
13. Tafuri, S.R., Familiari, M., and Wolffe, A.P. (1993) *J. Biochem.* **268**, 12213−12220.
14. Wistow, G. (1990). *Nature* **344**, 823−824.
15. Schindelin, H., Marahiel, M.A., and Heinemann, U. (1993) *Nature* **364**, 164−168.
16. Schnuchel, A., Wiltschek, R., Czisch, M., Herrier, M., Willimsky, G., Graumann, P., Marahiel, M.A., and Holak, T.A. (1993) *Nature* **364**, 169−171.
17. Schindelin, H., Jiang, W., Inouye, M., and Heinemann, U. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5119−5123.
18. Newkirk, K., Feng, W., Jiang, W., Tejero, R., Emerson, S.D., Inouye, M., and Montelione, G.T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5114−5118.
19. La Teana, A., Brandi, A., Falconi, M., Spurio, R., Pon, C.L., and Gualerzi, C.O. (1991) *Proc. Natl. Acad. Sci USA* **88**, 10907−10911.
20. Graumann, P., and Marahiel, M.A. (1994) *FEBS Lett.* **338**, 157−160.
21. Doniger, J., Landsman, D., Gonda, M., and Wistow, G. (1992). *New Biol.* **4**, 389−395.
22. Tafuri, S.R., and Wolffe, A.P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9028−9032.
23. Murray, M.T., Schiller, D.T., and Franke, W. W. (1992). *Proc. Natl. Acad. Sci. USA* **89**, 11−15.
24. Deschamps, S., Viel, A., Garrigos, M, Denis, H., and Le Maire, M. (1992). *J. Biol. Chem.* **267**, 13799−13802.
25. Burd, C.G., and Dreyfuss, G. (1994) *Science* **265**, 615−621.
26. Kick, D., Barrett, P., Cummings, A., and Sommerville, J. (1987) *Nucleic Acids Res.* **15**, 4099−4109.

27. Cummings, A., Barrett, P., and Sommerville, J. (1989) *Biochim. Biophys. Acta* **1014,** 319–326.
28. Deschamps, S., Viel, A., Denis, H., and Le Maire, M. (1991) *FEBS Lett.* **282,** 110–114.
29. Murray, M.T., Krohne, G., and Franke, W.W. (1991) *J. Cell Biol.* **112,** 1–11.
30. Marello, K., La Rovere, J., and Sommerville, J. (1992) *Nucleic Acids Res.* **20,** 5593–5600.
31. Nute, P.E., and Mahoney, W.C. (1979) *Biochemistry* **18,** 467–472.
32. Enfield, D.L., Ericsson, L.H., Fujikawa, K., Walsh, K.A., Neurath, H., and Titani, R. (1980) *Biochemistry* **19,** 659–667.
33. Jacquemin-Sablon, H., Triqueneaux, G., Deschamps, S., Le Maire, M., Doniger, J., and Dautry, F. (1994), *Nucleic Acids. Res.* **22,** 2643–2650.
34. Ranjan, M., Tafuri, S.T., and Wolffe, A.P. (1993) *Genes Dev.* **7,** 1725–1736.
35. Spitkovsky, D.D., Royer-Pokora, B., Delius, H., Kisseljov, F., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., and Royer, H. (1992) *Nucleic Acids Res.* **20,** 797–803.
36. Gai, X., Lipson, K.E., and Prystowsky, M. (1992) *Nucleic Acids Res.* **20,** 601–606.
37. Kashanchi, F., Duvall, J.F., Dittmer, J., Mireskandari, A., Reid, R.L., Gitlin, S.D., and Brady, J.N. (1994) *J. Virol.* **68,** 561–565.
38. Ozer, J., Faber, M., Chalkley, R., and Sealy, L. (1990) *J. Biol. Chem.* **265,** 22143–22152.
39. Firulli, A.B., Maibenco D.C., and Kinniburgh, A. (1992) *Biochem. Biophys. Res. Commun.* **185,** 264–270.
40. Laing, L.G. Gluick, T.C., and Draper, D.E. (1994) *J. Mol. Biol.* **237,** 577–587.
41. Bendall, A.J., and Molloy, P.L. (1994) *Nucleic Acids Res.* **22,** 2801–2810.