Studies on Two Putative G_0/G_1 Switch Genes in Human T- Lymphocytes, *RGS2/G0S8* and *G0S24*.

by

Scott Patrick Heximer

A thesis submitted to the Department of Biochemistry
in conformity with the requirements for the degree of
Doctorate of Philosophy

Queen's University
Kingston, Ontario, Canada
March, 1997

QUEEN'S UNIVERSITY AT KINGSTON

SCHOOL OF GRADUATE STUDIES AND RESEARCH

PERMISSION OF CO-AUTHOR(S)

I/we, the undersigned, hereby grant permission to microfilm any material designated as being co-authored by me/us in the thesis copyrighted to the person named below:

SCOTT PATRICK HEXTMER — Signature of copyrighted author

Name of Copyrighted Author

Name(s) of co-author(s)	Signature(s) of co-author(s)
D. R Fordyce.	
Anthony Cristillo	
LaVerne Russell	LaVerne Russell

DATE: 17/04/97

ABSTRACT

G0S24 and *RGS2/G0S8* were among the set of genes identified by differential screening of human blood mononuclear cells, which had been treated for 2 hr with Concanavalin A (Con-A) and cycloheximide. The full-length cDNA and genomic sequences of both genes were determined. Their differential expression under a variety of conditions have also been examined.

RGS2/G0S8 encodes a 24 kDa protein which is a member of a family of regulators of G protein signalling (RGS). As several of these proteins behave as GTPase activating proteins (GAPs) for the G_i subfamily of G protein alpha subunits, we anticipate a similar function for RGS2. The responses of *RGS2* and *RGS1/1R20/BL34* to Con-A and other stimuli were compared using competitive RT-PCR. On the basis of its responsiveness to Con-A and the calcium ionophore, ionomycin, we suggest that *RGS2* expression is regulated by changes in intracellular calcium levels. Recombinant RGS2 protein does not bind *in vitro* to the G_i subfamily of G protein α-subunits for which recombinant RGS1 has high affinity.

G0S24 encodes a zinc-binding nuclear 326 amino acid protein whose biological function remains poorly understood. It belongs to a family of proteins which are characterized by the presence of one or more $CX_{7-9}CX_5CX_3H$ (CCCH) amino acid repeats. Early work in this field suggested these domains were zinc-binding moieties and that proteins which contained them behaved as transcriptional regulators. Our data may suggest a different role for this protein, perhaps in RNA binding or processing. Competitive RT-PCR was used to study the kinetics of *G0S24* expression. Potential regulatory signals for 3' end processing and a CpG island in the 3' flank suggest the potential for a strong transcription

termination signal. RT-PCR characterization of *G0S24* mRNA expression identified a nuclear 3'-extended processing intermediate. Transiently high levels of this RNA species in TPA-treated and freshly explanted cells suggest its levels are dependent on the rate of *G0S24* transcription. Response of this 3' extended RNA to cycloheximide is consistent with *G0S24* transcription activation through a cycloheximide-dependent mechanism.

CO-AUTHORSHIP

Much of this work was based on the preliminary identification of a set of G_0/G_1 switch genes in peripheral blood lymphocytes (1). As the results presented in this work are based on a series of papers (2-6) on which there are more than one author, the following is a summary of the specific contributions to each paper made by each author.

The second chapter combines the results from two papers on *G0S8/RGS2*, a new member of a class of novel regulators of G protein signalling (RGS) proteins. Data from the first paper (3) includes the characterization of the *G0S8/RGS2* gene. D. Forsdyke and L. Russell cloned the original *G0S8/RGS2* 3'end cDNA fragment. D. Siderovski isolated and characterized the full-length version. My contributions included: screening of the genomic library in λ phage; preparation of phage DNA and mapping of one genomic clone containing the *G0S8* gene; and confirmation of 5' end sequences using a RACE procedure. Mr. K. Taher cloned, purified, and mapped a second overlapping genomic clone. A. Yip and L. Russell sequenced most of the genomic region containing the *G0S8* gene. This paper was written by D. Forsdyke with subsequent input from D. Siderovski and myself. For reasons of brevity the original manuscript has been significantly shortened. Within the same chapter, data from

the second paper describes the mRNA expression of *RGS1* and *RGS2/G0S8* (5). All experiments, including purification of RNA from small tissue samples and optimization of the conditions for RT-PCR analysis of G0S gene mRNAs were designed and carried out by myself. Subsequent RT-PCR analyses of *RGS2/G0S8* and *G0S24* were performed by this author while PCR analysis of *RGS1* and *FOS/FOSB* genes were carried out by A. Cristillo and L. Russell respectively. The original manuscript was written by myself, with subsequent modifications by D. Forsdyke.

The third chapter describes the characterization of the hman *G0S24* gene (2, for characterization of the mouse gene see 7). W. Sangrar screened a human genomic library, identified, and purified the DNA for the genomic region containing the *G0S24* gene. All other experiments were performed by myself. The original paper was written in first draft by D. Forsdyke with contributions to later drafts being made by myself. For presentation in this thesis, the manuscript has been modified to include more recent biological and functional data for the recombinant G0S24 protein.

The fourth chapter describes efforts to characterize the expression and post-transcriptional processing of the *G0S24* mRNA. The manuscript was written by myself with input from D. Forsdyke and is currently in preparation for submission (6). All experiments reported in this work, including recent attempts to characterize the function of the G0S24 CCCH domains and 3' flank (based on work in 8 and 9) were designed and carried out by myself. L. Russell performed RT-PCR analysis of the *F0S, F0SB,* and *MIP1α* mRNA controls.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my research supervisor, D. Forsdyke for his assistance and technical guidance throughout my studies.

Secondly, I would like to acknowledge D. Siderovski, L. Russell, and W. Sangrar for their work leading to the start of this project and A. Cristillo for valuable discussions.

Thirdly, I would like to acknowledge NHRDP/MRC and Queen's University for financial support.

In addition, I would like to thank D. Back, P. Davies, P. Greer, A. Mak, M. Petkovich, G. Cote, D. Maurice, J. Jia and R. Deeley for their kind advice at various stages of this project.

Finally, I would like to thank and dedicate this thesis to my parents, my family, and Dianne, my loving wife, for their continued support throughout my education.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

LIST OF ABBREVIATIONS

Units of Measure

bp	base pair (s)
cpm	counts per minute
°C	degree Celsius
pI	isoelectric point
min	minute
nt	nucleotide
O.D.	optical density
r.p.m.	revolutions per minute
U	unit(s) enzyme activity
w/v	weight per unit volume

Prefixes

ds	double-stranded
f	fempto
k	kilo
μ	micro
m	milli
n	nano
p	pico
ss	single-stranded

Suffixes

A	amp
Ci	Curie
D	Dalton
g	gram
hr	hour(s)
l	litre
m	metre
M	molar
s	seconds

Ribonucleoside Diphosphates

ADP	adenosine 5'-diphosphate
GDP	guanosine 5'-diphosphate

Deoxyribonucleoside Diphosphates

dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
dITP	deoxyinosine 5-triphosphate
dNTP	deoxynucleotide 5'-triphosphate
dTTP	deoxythymidine 5'-triphosphate

Chemicals

MOPS	3-[N-Morpholino] propanesulfonic acid
$CHCl_3$	chloroform
dpc	diethylpyrocarbonate
DTT	dithiothreitol
DMSO	dimethylsulfoxide
EtOH	ethanol
EtBr	ethidium bromide
EDTA	ethylenediaminetetra-acetic acid
GuHCl	guanidine hydrochloride
GuSCN	guanidine thiocyanate
IPTG	isopropyl-β-D-thiogalactopyranoside
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulphonic acid
NP-40	Nonidet 40 (Octylphenoxypolyethoxyethanol)
PMSF	phenylmethylsulphonylfluoride
PIPES	piperazine-N,N'-bis[2-ethane-sulfonic acid
PBS	phosophate buffered saline
lubrol	polyoxyethylene 12-lauryl ether
NaAc	sodium acetate
SDS	sodium dodecyl sulphate
$NaPP_i$	sodium pyrophosphate
SSC	standard saline citrate
Tris	Tris-(hydroxymethyl)aminomethane
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
x-Gal	5-bromo-4 chloro-3 indolyl-β-D-galactoside

Amino Acids

A, Ala	alanine
R, Arg	arginine
D, Asp	aspartic acid

N, Asn	asparagine
C, Cys	cysteine
Q, Gln	glutamine
E, Glu	glutamic acid
G, Gly	glycine
H, His	histidine
I, Ile	isoleucine
L, Leu	leucine
K, Lys	lysine
M, Met	methionine
F, Phe	phenylalanine
P, Pro	proline
S, Ser	serine
T, Thr	tryptophan
Y, Tyr	tyrosine
V, Val	valine

Species

C. elegans	*Caenorhabditis elegans*
E. Coli	*Escherichia coli*
S. pombe	*Schizosaccharomyces pombe*
S. cerevisae	*Saccharomyces cerevisiae*
D. melanogaster	*Drosophila melanogaster*

Ribonucleoside Triphosphates

rATP	adenosine 5'-trisphosphate
rUTP	uridine 5'-trisphosphate
rCTP	cytidine 5'-trisphosphate
rGTP	guanosine 5'-trisphosphate
rNTP	ribonucleoside triphosphates

Nucleic Acids

cDNA	complementary deoxyribonucleic acid
cRNA	control ribonucleic acid
DNA	deoxyribonucleic acid
mRNA	messenger ribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid

Miscellaneous

TPA	12-o-Tetradecanoyl phorbol-13-acetate
HAc	acetic acid
AML	acute myelogenous leukemia
ALL	acute lymphoblastic leukemia
ARE	AU-rich DNA element
CLL	chronic lymphoblastic leukemia
CML	chronic myelogenous leukemia
cAMP	cyclic adenosine monophosphate
BLAST	basic local alignment search tool
βARK	beta adrenergic receptor kinase
BSA	bovine serum albumin
Con-A	concanavalin-A
ECL	enhanced chemiluminesence
EGF	epidermal growth factor
EBNA	Epstein Barr Virus nuclear antigen
EST	expressed sequence tag
FGF	fibroblast growth factor
GAIP	G-alpha interacting protein
GH	G0S8 homology domain
GAP	GTPase activating protein
GM-CSF	granulocyte macrophage colony stimulating factor
HIS-TAG	histidine tag
IMAC	immobilized metal affinity chromatography
ITAM	immunological tyrosine-based activation motif
LB	Luria Bertani
MRE	metallothionein response element
MPD	myeloproliferative disorder
MMLV	Moloney murine leukemia virus
NHL	non-Hodgkin's lymphoma
NMR	nuclear magnetic resonance
orf	open reading frame
PBMC	peripheral blood mononuclear cell
PMSF	phenyl methyl sulphonyl fluoride
PAGE	Polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PVDF	polyvinylidene difluoride
PKC	protein kinase C
PTK	protein tyrosine kinase
RACE	rapid amplification of cDNA ends
RGS	regulator of G protein signalling
RT	reverse transcriptase
RT-PCR	reverse transcriptase polymerase chain reaction
RNAse	ribonuclease
TCR	T cell receptor
TdT	terminal deoxynucleotidyl transferase
utr	untranslated region

CHAPTER 1

GENERAL INTRODUCTION

Overview

Despite significant recent progress toward developing an understanding of the molecular mechanisms that control eukaryotic cell cycle progression, many aspects of this process remain poorly understood (10). One popular method of studying this process has been the identification of specific growth-regulated genes, whose mRNA levels are increased at critical regulatory stages of the cell cycle. (1, 11, 12). By the characterization of mRNA expression kinetics for these genes and the development of functional assays for their protein products, researchers may begin to understand their role in growth-related processes. It was speculated by our laboratory and others that genes whose mRNA levels increase at one such step, transition from the quiescent (G_0) to the growth (G_1) phase, may play an important role in cellular growth and proliferation (13). T-lymphocytes, which exist naturally in the G_0 phase during a latency stage of their cell cycle and can be readily stimulated to enter G_1 using a polyclonal T cell mitogen, provided an ideal system in which to study the G_0/G_1 switch. The genes identified on the basis of their differential expression in stimulated mononuclear cells (summarized in Table 1-1), were named *G_0/G_1 switch* (*G0S*) genes owing to their putative involvement in this process. While it should be noted that this approach has been used to identify a number of genes which are *growth-related* but not necessarily *growth-regulatory,* it is nonetheless of interest to us that leukemias and myeloproliferative disorders

result from the uncontrolled proliferation of hematopoietic cells. In some cases, such disease states may reflect the inability of these cells to normally "arrest" in the quiescent (nonproliferative) phase of the cell cycle. We anticipated that the dysregulated expression of a subset of G0S genes may lead to uncontrolled proliferation (loss of control at the G_0/G_1 switch) of lymphoid cells and thus contribute to the leukemic state in affected individuals. We further predicted that characterization of the genomic organization and normal expression kinetics of these genes in normal individuals will be a necessary first step toward defining their role in abnormally proliferating lymphoid tissues.

Specific Aims

The specific aims of this work were to first, characterize the full-length cDNA and genomic sequences of *G0S8/RGS2* and *G0S24*, two genes whose mRNA levels are increased during the G_0/G_1 switch in T-lymphocytes. Secondly, in order to gain an understanding of the timing of G0S gene expression during this critical regulatory switch, we developed a sensitive competetive RT-PCR assay for each of the G0S gene mRNAs. Finally, based on the similarity of the predicted G0S8 and G0S24 proteins to previously characterized protein species, we attempted to identify their specific roles during the activation process.

Stages of the eukaryotic cell cycle

The cell cycle is defined as the period between completion of parental cell mitosis and the completion of daughter cell mitosis. In continuously cycling cells, it is thought to be divided into 4 stages or phases: G_1, a period of cell growth marked by an increase in cytoplasmic volume; S, the DNA synthesis phase; G_2, a second short period of cell growth; and M, the mitotic (cell division) phase. Another "latent" phase, G_0, has been defined for non-dividing cells which are capable of re-entry into the cell cycle under appropriate conditions. Regulation of cell growth and proliferation has been suggested to occur at three critical steps: entry into G_1; transition from G_1 to S; and transition from G_2 to M (10). We have directed our efforts toward understanding the events occurring at G_1 entry.

Physiological significance of quiescence in cell lines and resting lymphocytes.

Several of the studies to define molecular events at the G_0/G_1 switch have used "G_0 phase-arrested" murine fibroblasts as a model system. In many of these cases, the arrest of these cells was introduced artificially, through the deprivation of serum and serum growth

factors (12, 14). Such harsh treatment has been reported to cause irreversible damage to these cells which could potentially affect subsequent cycling characteristics (15). Also, it is likely that as continuously dividing cells, they have already escaped the "normal" eukaryotic growth control mechanisms. For these reasons, the physiological significance of the artificially induced G_0 state is subject to review.

In our studies, we routinely used freshly explanted human peripheral blood mononuclear cells (1) which contain large numbers of T lymphocytes. It was our contention that since most of these lymphocytes are naturally arrested in G_0 , this system would offer a physiologically relevant model system in which to study events during the G_0/G_1 switch.

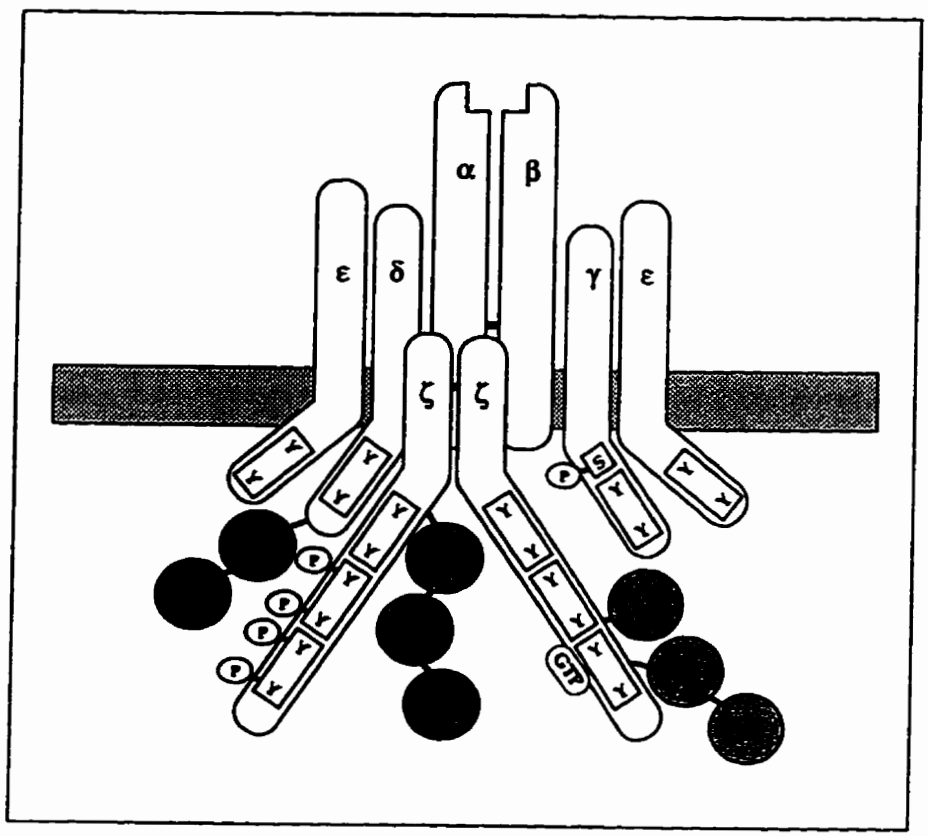
Antigenic Stimulation of T lymphocytes

The term "T-cell activation" refers to the series of events initiated by antigenic stimulation and resulting in the production of cytokines, such as IL-2 and cytokine receptors. (16, 17). The physiological trigger of T-cell activation is the specific binding of the T-cell receptor (TCR) with antigen displayed, in the context of the major histocompatability complex (MHC) class I or class II molecules, on the surface of antigen presenting cells (APCs). The signal transduction cascades which follow will, under the appropriate set of conditions, lead to cytokine production, proliferation, and clonal expansion of the subset of activated cells (18).

Signal Transduction following T-Cell Activation

The TCR/CD3 complex is responsible for mediating signal transduction following ligand binding. This complex (Figure 1-1) consists of a disulfide-linked heterodimer of the idiotypic $\alpha\beta$-chains noncovalently associated with the complex composed of the invariant

Figure 1-1: **TCR including occupancy-dependent modifications.** Tyrosine based activation motifs are shown as boxed regions, as is the nucleotide binding region. Well-characterized sites of phosphorylation (P) a serine (S), threonine (T), and tyrosine (Y) are also shown. Diagram adapted from (19).


α
β
ε
δ
γ
ε
ζ
ζ
S
GTP

CD3 γ-, δ-, ϵ- chains and the ζ- (16 kD) and η- (22 kD) chains (20, 21). Ligand binding to the TCR/CD3 complex initiates a cascade of biochemical events, the earliest of which is the activation of intracellular protein tyrosine kinases (PTKs, 22), a process which appears to be absolutely required for all subsequent T cell responses (23). Cellular substrates for TPKs include CD3 and ζ-chains of the TCR/CD3 complex and the tyrosine protein kinase ZAP-70 (24) and the (PI)-specific phospholipase C-γ1 (25). Rapid phosphorylation of these substrates results in activation of PI turnover, mobilization of intracellular calcium and stimulation of the serine/threonine-specific protein kinase C (16). These signals converge at the nucleus and result in increased mRNA expression of a number of activation-dependent and growth-related genes.

Although several studies have suggested a role for GTP binding proteins in the signal transduction cascade, the significance of such mechanisms with respect to the activation of PLC or other cellular effectors remains poorly understood (26, 27). Recent evidence has suggested an association between CD3 and $G_{\alpha 11}$ in Jurkat cell membranes (28). The physiological significance of this finding in intact cells, however, remains to be established. It is also of interest that the ζ-chain itself can bind GTP through its nucleotide binding site (Figure 1-1), and that a reduction in GTP binding may be correlated with decreases in tyrosine phosphorylation and IL-2 production (29).

Concanavalin-A stimulation of T-lymphocytes - a model for T-cell activation in vivo

Activation of resting T-lymphocytes *in vitro* may be induced by plant lectins which act as polyclonal T-cell mitogens. Lectins, such as Concanavalin-A, bind toTCR/CD3 and

other surface glycoproteins resulting in crosslinking events which appear to mimic events during antigen presentation and thus result in the appropriate T-cell activation response (30).

We have used Con-A treatment of quiescent cells as a model for studying the molecular events which occur early following T-cell activation (1, 31). As cells pass through the G_0/G_1 switch, previously silent genes are transcribed within 30 min of lectin stimulation of blood mononuclear cells, a process which was shown to be marked by increased levels of ^{3}H-uridine into their mRNAs (32). In our studies, cycloheximide was also added to mitogen-treated cultures to enhance the detection of genes whose mRNA levels are increased following stimulation. Briefly, the addition of this protein synthesis inhibitor ensures that: i) only primary response genes are isolated; ii) short-lived mRNAs are stabilized (protected from cytoplasmic degradation by polysome stalling); and iii) superinduction occurs from the promoters of genes whose transcription is normally repressed by a labile repressor (for review see ref. 9).

Isolation of previously silent cDNA clones

Cultures treated for 2 hr with Con-A and cycloheximide were used to generate an activation cDNA library using the method of Gubler and Hoffman (1983, ref. 33). Differential screening of the resulting library resulted in the identification a set of genes whose mRNA levels are specifically increased as cells enter into the G_1 growth phase (Table 1-1, 31).

Among these isolates were partial cDNAs encoding homologs of murine transcription factors (Zif268/egr-1/Krox24; c-Fos/Tis28; Fos-b) and cytokines (Mip1α). Also present were

TABLE 1-1. SUMMARY OF ALTERNATIVE NOMENCLATURES AND FUNCTIONS OF SOME G0S GENES

G0S **gene**	**Some other isolates** Rodent	Human	**Role**	**References**
2	-	-	?	(34)
3	*FOSB*	*FOSB*	transcription (AP1)	(1, 4, 222)
7	*FOS/TIS28*	*FOS*	transcription (AP1)	(77)
8	-	-	regulator of G-protein coupled receptor	(3, 39, 45, 49)
9	-	*PBEF*	cytokine	(1, 223)
19-1	*MIP-1α*	*LD78α /464.1*	cytokine inhibitor of stem cells. HIV-1 suppressor	(84, 224)
19-2	-	*LD78β /464.2*	cytokine	
CUS-	-	*LD78γ /464.3*	pseudogene?	(225)
24	*TIS11/NUP475 TTP*	-	transcription? RNA Binding zinc finger?	(2, 34)
30	*NGFI-1/TIS8 EGR-1/ZIF-268*	-	transcription/ zinc finger	(184)

two cDNA clones corresponding to the 3' ends of the *G0S24* (0.32 kb) and *G0S8* (0.8 kb) mRNAs. The characterization of the genes from which these two cDNAs were derived is the focus of this study. During the course of this study, the apparent nuclear localization of G0S24 (7) and G0S8 (K. Druey, personal communication) proteins, together with a number of amino acid sequence features consistent with their role as transcription factors (G0S24, ref. 2; G0S8, ref. 3) suggested that these proteins were required for modulation of transciptional events early in the G_1 phase of the cell cycle. Thus, it seemed a logical strategy to carry out characterization of both full-length cDNAs and subsequent cloning into bacterial expression vectors in parallel with the intent to identify the consensus DNA-binding elements to which these two proteins bound. However, recent functional evidence for these proteins, obtained in both our lab, and the laboratories of our collaborators suggest alternative roles for these proteins. This exemplifies the difficulties in predicting functional characteristics of differentially expressed genes in a mammalian system.

Characterization of G0S gene organization

As part of the characterization of the *G0S8* and *G0S24* genes, a significant amount of time and effort was spent on the characterization of the genomic sequence and organization of these genes. Although these genes are relatively small in size, such projects were not a trivial exercise as the complete sequencing of both strands was carried out manually, before the widespread availability of automated DNA sequencing technology. It is notable that to date, the *G0S8/RGS2* and *G0S24* genomic sequences reported in this work remain as the only such human gene sequences in the GenBank database (September 1996).

This is surprising since there are now known to be at least 17 human family members of the *RGS2/G0S8* gene.

We are confident that the number of important studies which these sequences will make immediately possible are justification for the time spent in their characterization. First, these sequences indicate a number of potential regulatory elements in the 5' flank and intronic regions, and will thus provide the necessary tools for future in depth promoter analysis. Also, knowledge of these sequences may provide clues to understanding the developmental expression patterns of these genes. Secondly, chromosomal aberrations leading to abnormal expression of *G0S24/HUMTTP* and *G0S8/RGS2* mRNAs have been found in several human leukemias and other malignancies (for review see general discussion). While the nature of these abnormalities remains to be established, it is possible that such abnormal expression may in some cases be the result of point mutations or small deletions in the promoter regions of these genes. In such cases, the clones and sequences generated in this work will allow the rapid development of clinical reagents to screen patients for these abnormalities. The detailed genomic restriction maps which we have generated for these regions may also be useful for the interpretation of Restriction Fragment Length Polymorphism (RFLP) analyses of genomic DNA from affected individuals. Finally, characterization of the *G0S24* gene has made possible a study of nuclear processing and transcription termination (see Chapter 4). The design of RT-PCR primers for such studies would not have been possible without genomic sequence data.

Competitive RT-PCR as a sensitive alternative to Northern blotting

Early attempts at the characterization of *G0S24* and *G0S8* mRNA expression kinetics suggested that basal levels were often below the level of sensitivity of Northern blotting. Furthermore, Con-A stimulation of mononuclear cells in the absence of cycloheximide, did not result in the marked changes in mRNA levels which typify immediate-early gene responses in other systems (1, 34). As the supply of normal human mononuclear cells was, for practical purposes, limiting, we needed a very sensitive assay to study changes in the mRNA levels of these genes. Thus, to study multiple genes in response to a variety of stimuli, we developed competitive RT-PCR technology. Although a more complete description of the assay method is provided in Heximer *et al.*,(1996, ref. 4), this paper has been excluded from this thesis for reasons of brevity. A brief summary of the development and interpretation of the competitive RT-PCR assay procedure is shown in Appendix A1.

CHAPTER 2

TITLE: Comparison of mRNA Expression of Two Regulators of G-Protein Signalling, *RGS1/BL34/1R20* and *RGS2/G0S8*, in Cultured Human Blood Mononuclear Cells.

PREFACE

The following chapter combines the work from the two papers on *G0S8/RGS2* which were published by our lab (3, 5). It describes the original characterization of the *G0S8/RGS2* gene and the subsequent comparison of mRNA kinetics of *G0S8/RGS2* and a related family member, *RGS1/BL34/1R20.* The first paper (3), describes the only published complete gene sequence for any member of the RGS family of genes and will be an invaluable reference for future clinical studies designed to characterize a role for this protein in development of human leukemias (see general discussion). In the original paper we noted a significant region of amino acid similarity between G0S8/RGS2 and a family of helix-loop-helix, DNA-binding transcription factors, thus leading to our original assumption that this protein played a role in the transcriptional regulation of events early during T cell activation (3). Perhaps more significantly, however, was our identification of a region of Sst2p (yeast pheromone desensitization protein) similarity found in the G0S8/RGS2 protein and a growing family of modulators of G protein coupled signal transduction including β adrenergic receptor kinases (βARKS). This domain (now known as the 120 amino acid "RGS core domain") has been shown to be required for the function of RGS proteins, a new class of GTPase activator proteins (GAPs) for Gα subunits of heterotrimeric G proteins (35, 36). The recent discovery of the biological function of the RGS proteins provided a timely backdrop for the mRNA

studies which compare the expression patterns of *RGS2/G0S8* and *RGS1/BL34/1R20* in mitogen-treated peripheral blood mononuclear cells. The second paper points to several differences in the RNA expression and potential biological activity of these two RGS proteins, and suggests a role for RGS2 in the control of calcium signalling, perhaps through the G_q subunit of G proteins. Subsequently, this hypothesis was confirmed by M. Linder, one of our collaborators at Washington University, St. Louis, who has, in a preliminary set of experiments, demonstrated specific interaction between G_q and recombinant RGS2 *in vitro*. For presentation as part of this thesis, a more detailed description of the genomic mapping, $(His)_{10}$-G0S8 expression and characteristics of the recombinant protein have been included.

ABSTRACT

RGS2/G0S8 was identified as a differentially expressed mRNA in human blood mononuclear cells treated with the T cell lectin Concanavalin A (Con-A) and cycloheximide (1). Comparison of a full-length cDNA sequence with the corresponding genomic sequence reveals an open reading frame of 211 amino acids, distributed across 5 exons. RGS2/G0S8, the 24 kDa predicted gene product, and RGS1/1R20/BL34, the product of a phorbol ester (TPA)- inducible gene in human B lymphocytes, belong to a new class of regulators of G protein signalling. The *RGS1* gene shows low basal mRNA expression in freshly purified blood mononuclear cells, which increases upon incubation for a day. In contrast, *RGS2* initially shows high basal levels of mRNA expression, which subsequently decreases. Expression of both genes increases in response to Con-A, with *RGS2* mRNA levels increasing briskly to a maximum between 0.5-1 hr and decreasing to baseline by 6 hr,

whereas the *RGS1* mRNA increase is delayed reaching a maximum between 1-2 hr. *RGS1* mRNA levels increase much more in response to a protein kinase C activator (TPA), than to a calcium ionophore (ionomycin), whereas the opposite is true for *RGS2*. We suggest that Con-A elevates RGS2 on the basis of its ability to increase intracellular calcium, and that RGS2 may be involved in the regulation of intracellular calcium. The distinction between RGS1 and RGS2 is further emphasized by studies indicating that recombinant RGS2 does not bind *in vitro* to two members of the G_i subfamily of G protein α-subunits for which recombinant RGS1 has high affinity.

INTRODUCTION

Heterotrimeric G proteins, containing α, β and γ subunits, play a major role in coupling receptors to effectors in eukaryotic signal transduction pathways. Continued stimulation of receptors brings about desensitization by three groups of proteins, receptor kinases which phosphorylate the receptors (37), arrestins which compete with G proteins for receptor binding (38), and the recently recognized RGS proteins (regulators of G-protein signalling), which interact with α-subunits (39, 40). Desensitization appears to provide a basis for attenuation to light, hormones, antigens and neuroregulatory molecules, and may influence the efficacy of the many therapeutic compounds which operate by way of G-protein coupled pathways.

RGS1 was identified as a phorbol ester-inducible, cycloheximide-superinducible, mRNA in lymphocytes from patients with B cell chronic lymphocytic leukemia (*1R20* ; refs. 41-42), and was independently isolated from human tonsillar B lymphocytes by Hong and coworkers

in 1993 (*BL34;* ref. 43). *RGS2* was identified on the basis of increased mRNA expression in human blood mononuclear cells cultured for 2 h with the T-cell specific lectin Con-A and cycloheximide (*G0S8;* cell cycle G_0/G_1 switch gene number 8; refs. 1, 34).

RGS2/G0S8 cDNA hybridized to multiple bands on genomic Southern blots indicating a large family of related genes (1). Both *RGS1* and *RGS2* were found at chromosome band 1q31 (44, 45). It was noted that the derived protein sequences showed similarities to the yeast Sst2p (supersensitivity to pheromone when mutated) protein, which desensitizes against pheromone-imposed arrest in the G_1-phase of the cell cycle, and to domains conserved among eukaryotic G protein-associated receptor kinases (3, 43).

Sst2p mRNA and protein levels were found to increase on exposing haploid yeast cells to the appropriate pheromone, concomitant with the imposition of G_1 arrest (46, 47, 48; W. E. Courchesne and J. Thorner, personal communication). Based on data from dominant gain-of-function *SST2* mutants, Dohlman and coworkers (1995, 1996; refs. 40, 48) suggested that Sst2p was an RGS protein mediating its adaptive effects at the level of Gpa1p, the G protein α subunit required for transduction of pheromone-induced signals. Further evidence that RGS proteins interact with α-subunits came from the characterization of new members of the family.

Using a two-hybrid assay, De Vries and coworkers (1995; ref. 49) identified a human cDNA clone from HeLa cell libraries which encoded a protein (GAIP) interacting with the α-subunit $G_{i\alpha 3}$. Similarly, EGL-10, a protein localized to neural processes in *C. elegans*, was shown to regulate G protein-mediated egg-laying behaviour (50). In humans there may be at least 15 RGS proteins, with varying specificities for different α-subunits and for

different tissues, and with varying abilities to modulate responses to different signalling agents (51). RGS proteins share a common 120 amino acid C-terminal sequence, and trans-species complementation indicates functional conservation of three GH (G0S8 homology) domains (39, 51). The brain-specific rat RGS4, and the human RGS1 and GAIP proteins, activate the intrinsic GTPase activity of G protein (G_i) α-subunits (35, 36).

We here present analysis of the *RGS2/G0S8* cDNA, its predicted protein, and the corresponding genomic region (1, 3, 34). Using RT-PCR, we compare expression of its mRNA in human peripheral blood mononuclear cells relative to that of the related gene, *RGS1*. High *RGS1* mRNA levels were found in blood cells of a child with B cell acute lymphocytic leukemia (43). High *RGS2* mRNA levels were found in cells from patients with acute leukaemias of both T and B lineages (45). These observations suggest that RGS proteins are involved in control of the size of lymphoid cell populations. Noting that protein kinase C, calcium and G proteins play important roles in lymphocyte activation pathways (28, 52, 53), we also describe here the influence of a T lymphocyte activator (Con-A), a protein kinase C activator (TPA; phorbol ester), and a calcium ionophore (ionomycin), on the expression of *RGS1* and *RGS2* mRNAs in blood mononuclear cell cultures. Our results suggest a particular involvement of RGS2 in calcium signalling pathways, and interaction with a different G protein α-subunit than RGS1.

MATERIALS AND METHODS

Methods were as described previously (1, 2, 4, 54).

Mononuclear cell purification and culture conditions

Freshly collected, heparin-treated blood from healthy human donors (90 ml) was diluted with an equal volume of pre-warmed (37°C) medium 199 or medium RPMI 1640, layered over 16 ml of Ficoll-Paque (Pharmacia, Piscataway, NJ) and centrifuged 12 min at 800 x g. The mononuclear cell layer was diluted in five volumes of 1640 medium and cells were pelleted by spinning for 4 min at 600 x g. Cells were resuspended in incubation medium (RPMI 1640, 83.3 %; autologous serum, 16.7 %; 2-mercaptoethanol, 0.05 %) to a final concentration of 0.5-5.0 x 10^6 cells/ml. Unless otherwise indicated, cells were aliquotted and incubated 24 hr prior to addition of reagents to appropriate tubes (time defined as 0 hr). All experiments were repeated at least three times with cells from different human donors. TPA (phorbol 12-myristate 13-acetate 4-0-methyl ether; Sigma Chem. Co., Mississauga) was dissolved in dimethylsulfoxide (DMSO). Ionomycin (Calbiochem, San Diego), was dissolved in DMSO. Actinomycin D-Mannitol (Sigma Chem. Co.) was dissolved in water.

Preparation of ^{32}P-labelled DNA probes

Preparation of radiolabelled probes was done by random-sequence hexanucleotide using the Multiprime DNA Labelling System Kit (Amersham PLC, Arlington Heights, IL). Labelling reactions were carried out at 22°C for 3 hr using 30 μCi of 3000 Ci/mmol [$\alpha^{32}P$]-dCTP (ICN Biomedicals Inc., Costa Mesa, CA) and 2 units (U) DNA polymerase I Klenow

fragment. Labelled DNA probes were purified from unincorporated nucleotides using gel filtration chromatography (Sephadex G-50; DNA grade). DNA was routinely labelled to a specific activity of 1 x 10^9 cpm/μg and was denatured immediately before use by treatment with 0.2 M NaOH at 37°C for 10 min.

Hybridization of ^{32}P-labelled probes

Prehybridization of nylon blots was carried out at 65°C for 2 hr in 1.0 M NaCl, 50 mM Tris-Cl, pH 8.0, and 1 % SDS. ^{32}P-labelled DNA probes were mixed with sheared salmon sperm DNA (200 μg/ml hybridization mix) and alkaline denatured as described above. Hybridization was performed at 65°C for 20 hr in 1.0 M NaCl, 50 mM Tris-Cl, pH 8.0, 1 % SDS, and 10 % dextran sulfate. Following hybridization, blots were washed according to the following steps: twice with wash A (2 x SSPE, 0.5 % SDS and 0.1 % (w/v) $NaPP_i$ at 22°C for 5 min; and twice with wash B (0.1 x SSPE, 0.5 % SDS and 0.1 % $NaPP_i$) at 65°C for 30 min. Membranes were air-dried and autoradiography performed as described in the text.

Genomic library screening

A genomic library from Strategene (La Jolla, CA) contained human placental DNA inserted into the *Xho* I site of the λ FIX II vector. Two positive clones (λG0S8-1 and λG0S8-2) were identified among 500,000 plaques screened for hybridization to a ^{32}P-labelled *G0S8* cDNA. Library screening and subsequent large-scale preparations of DNA from λ phage were performed as described by Sambrook *et al.*, (1989, ref. 55).

Mapping of Genomic DNA clones

Recombinant phage DNA (0.5 μg) was digested to completion with 5-10 units (U) of one of the following enzymes: *Eco* RI , *Sal* I , *Xba* I , *Hind* III, or *Bam* HI either alone or in pairwise combinations. Reaction products sizes were determined following separation on 0.8% agarose gels, EtBr staining, and comparison to λ *Hind* III markers. Preliminary maps were generated from these data (56). Gels were then transferred to Zetaprobe-GT nylon membrane (BioRad, Hercules, CA) using alkaline capillary transfer (57). Blots were neutralized 10 min in 0.5 M Tris-Cl, pH 7.5, 1.0 M NaCl, air-dried, baked 80°C for 30 min, and stored in sealed bags until needed. To identify genomic fragments corresponding to *RGS2/G0S8* exon sequence, blots were probed with a ^{32}P-labelled *RGS2/G0S8* cDNA probe (10^5 cpm/ml). Restriction maps were confirmed using a previously described cohesive-end mapping method (58). Briefly, using a range of concentrations (0.015- 0.5 units/ 166 ng of DNA) of each of the restriction enzymes used in the complete digestion analysis, a set of partially digested λ DNA clones were produced. Pooled partial digests were hybridized to 5'-end ^{32}P-labelled oligonucleotides complementary to the cohesive ends of the left arm (ON-L) or the right arm (ON-R) of the λ vector. Labelled complexes were separated on 0.4% agarose gels, dried onto DE81 paper, and autoradiographed, and sizes compared to similarly labelled markers (Collaborative Research Inc., Lexington, MA). The order of partially labelled fragments was then used to confirm maps generated by combination digest data (other details as in legend to Fig. 2-1).

Preparation of RNA from numerous small-sized samples

For mRNA assays total RNA from small numbers of cells was prepared by a modification of the procedure of Chomczynski and Sacchi (1987, ref. 59), using Trizol reagent (Life Technologies). At various times cells were pelleted by centrifugation (600 x g, 4 min) from 2.5 ml cultures, and pellets were shaken vigorously for 5 min in 1 ml of Trizol. After addition of 0.2 ml of $CHCl_3$, samples were shaken again (30 sec), and then centrifuged to separate the upper aqueous layer containing RNA. The latter was precipitated with isopropanol and digested with RNAase-free DNAse 1, which was then removed by phenol/$CHCl_3$ extraction. After two further ethanol precipitations, RNA pellets were dissolved in 0.4 ml of water.

Polymerase chain reaction profile with a fixed concentration of competitor

The concentrations of mRNAs were determined using the reverse transcriptase-polymerase chain reaction (RT-PCR). A particular advantage of the RT-PCR assay is that only small quantities of RNA are needed, so that levels of many different mRNAs can be determined in *one* sample. Moloney murine leukaemia virus reverse transcriptase was obtained from Life Technologies. *Thermus aquaticus* thermostable DNA polymerase (Taq) was obtained from Sangon Ltd., Scarborough, Ontario. Competitive RT-PCR assays of mRNAs were carried out as described (4), except that annealing temperatures of 56°C and 53°C were used for *RGS1/BL34/1R20* and *G0S24* cDNAs respectively. Total RNA (125-500 ng) from each sample of cultured cells was reverse transcribed using a sequence-specific primer (*RGS1/BL34/1R20*, 5'-TCTGAATGCACAAATGC-3'; *RGS2/G0S8*, 5'-

GTACTGATGATCTGTGG-3'; *G0S24*, 5'-AAGGAGCTGGAGGACAG-3'). Equal aliquots from each reverse transcriptase reaction mixture (containing cDNA corresponding to 5.2-10.4 ng of total RNA) were combined with a fixed number of molecules of cDNA controls (see below), with which they were then coamplified; the products were identified by agarose gel electrophoresis with ethidium bromide staining. This generated a 'kinetic profile' for the experiment, which is displayed in most figures.

Polymerase chain reaction with varying competitor concentration

Based on the intensity of ethidium staining in the above profile assay, a range of two-fold concentrations of control cDNA plasmid was selected for each cDNA sample, and the PCR was repeated to quantitate more precisely the level of cDNA (60, 61). In calculating mRNA concentrations, it must be noted that cDNA molecules begin as single strands, whereas the controls begin as duplexes (see calculations in Appendix A1). Since we were concerned with relative changes in mRNA levels, rather than with absolute quantities, there were no controls for the efficiency of reverse transcription (62). For each set of conditions analysed, results are representative of three independent experiments. It was assumed that this efficiency would vary randomly and to approximately the same extent with different RNA samples. The reverse transcriptase primer was quite close to (within 5 to 151 nt of) the downstream PCR primer, which was either 337 nt (*FOSB/G0S3*), 212 nt (*FOS/G0S7*), 392 nt (*G0S24*), 240 nt (*RGS1/BL34*), or 373 nt (*RGS2/G0S8*) from the upstream primer. Thus only short complementary transcripts were required of reverse transcriptase.

Preparation of cDNA control templates

The competitor controls were cDNAs in plasmid vectors, with each cDNA containing a small exogenous DNA insert in the region to be amplified; this slightly increased the mobility of the PCR product. Controls for *FOS/G0S7* and *FOSB/G0S3* mRNAs were as described in Heximer *et al.*, 1996. *RGS1/BL34/1R20* cDNA in the Bluescript II plasmid vector (provided by J. Kerhl), and *RGS2/G0S8* cDNA in plasmid pBR322, were each digested at a unique *Bgl* II sites in the 3' untranslated region. For the *RGS1* cDNA control, an approximately 150 bp fragment was inserted between appropriate primer sites (upstream, 5'- AACCCAAAGGAATTGAAAGGAACC-3'; downstream, 5'-GAAACTTCCAAAGACATTTTGACC-3'). For the *RGS2* cDNA control a 74 bp *Sau* 3A fragment was inserted between primers flanking this site (upstream, 5'-CTGTGACCTGCCATAAAGACTGAC-3'; downstream, 5'-GATGTACAACCAGCTACCTAAGGC-3'). For the *G0S24* cDNA control, the complete coding sequence in plasmid pT7-7 (63) was digested at a unique *Sma* I site in the second exon. A 211 bp fragment from *Alu* I-digested pBR322 was inserted between appropriate primers (upstream, 5'-CGCGCTACAAGACTGAGCTATGTC-3'; downstream, 5'-GAGAAGCTGATGCTCTGGCGAAGC-3').

Expression of $(HIS)_{10}$-tagged recombinant RGS2/G0S8

G0S8-specific PCR primers (upstream; 5'-CTGTGACCTGCCATAAAGACTGAC-3'. downstream; 5'-GGGAAGCTGTTCTGATA-3') were used to generate a product containing the full-length predicted coding region for the G0S8 gene product. A 5' extension of the

upstream primer included an *Nde* I restriction site such that this product could be cloned directionally into the *Nde* I site of the pET19b *E. coli* expression vector. All PCR-generated products were sequenced completely on both strands. The resulting clone expressed a RGS1/G0S8 protein with an amino terminal tag of 10 histidine residues $(His)_{10}$-RGS1/G0S8 which could be expressed in *E. coli* and purified under nondenaturing conditions by Ni^{++}-chelate chromatography essentially as described by the pET system manual (Novagen). Briefly, clones were introduced into *E. coli* BL21 (DE3 lysogen) strain containing the pLys-S plasmid. Single colonies were grown in LB media containing ampicillin and chloramphenicol until an $O.D_{600}$ of 0.6 was reached. Cells were stored at 4°C overnight and were then harvested by centrifugation (600 x g, 4 min), diluted (1/100) in fresh medium, and grown at 30°C until $O.D._{600}$= 0.4. Protein expression was induced using 1 mM (IPTG) at 30°C for 3 hr. Cells from induced cultures were harvested by centrifugation (5000 x g, 10 min), and resuspended in column-binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) containing 0.1% Triton X-100 or NP-40, 1 mM phenymethylsulfonyl fluoride (PMSF), 2 μg/ml Leupeptin and 20 μg/ml Aprotinin. After storage at -70°C, thawed samples were sonicated on ice with six 40 s pulses from a Branson sonifier (model 450), and inclusion bodies were pelleted by centrifugation (16000 x g, 10 min). Filtered supernatants were loaded onto 2 ml Ni^{++}-charged His-Bind (Novagen) columns, washed successively with binding buffer and wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9), and proteins were then eluted in binding buffer containing 500 mM imidazole. Monomeric protein was further purified by column chromatography through G75 Sephadex in binding

buffer containing 10 mM β-mercaptoethanol. Purified amino-terminal-tagged $(HIS)_6$-RGS1/BL34/1R20 was provided by K. Blumer.

Analysis of proteins by SDS/PAGE, western blotting, and mass spectrophotometry

Samples in SDS gel loading buffer (50 mM Tris-HCl, pH 6.8, 100 mM 2-mercaptoethanol, 1 % (w/v) SDS, 0.1 % Bromophenol Blue, 10 % glycerol) were loaded onto 0.75 mm, 12 % dicontinuous polyacrylamide gels containing 0.1 % SDS and run in buffer, pH 8.6, containing 50 mM Tris base, 380 mM glycine, 0.1 % SDS at 25 mA for 1-1.5 h. Protein bands were stained in Coomassie Brilliant Blue R or transferred to PVDF membranes and analysed using antisera #1766 and ECL-immunodetection as previously described (39). Proper expression of recombinant proteins was verified using a Kratos Kompact MALDI III mass spectrophotometer. Briefly, purified proteins were precipitated in low ionic strength buffers, washed several times with 80 % ethanol to remove excess salt, resuspended (2 mg/ml) in 0.1 % trifluoroacetic acid and loaded in 2 µl aliquots into a precalibrated with a trypsin standard (23 290 Da).

PKC phosphorylation

Proteins (2.5 µg) were incubated at 30°C for 30 min in 100 µl kinase buffer (20 mM Tris-Cl, pH 7.5, 5 mM $MgCl_2$, 200 µM $CaCl_2$, 1 mM rATP with and without 10 µCi [γ-^{32}P]rATP) containing 40 µg/ml Lα-phosphatidyl-L- serine (Sigma Chem. Co.), 20 µg/ml diacylglycerol (Serdary Research Labs, London, Ont.) and 1 ng (1.4×10^{-3} units) of purified Protein Kinase C (mixture of isozymes, Calbiochem). Phosphorylated ^{32}P-labelled proteins

were loaded onto SDS/PAGE for quantitation of moles phosphate/mole protein. Unlabelled PKC-treated proteins were used for *in vitro* binding assays.

In vitro binding assays with Gα subunits

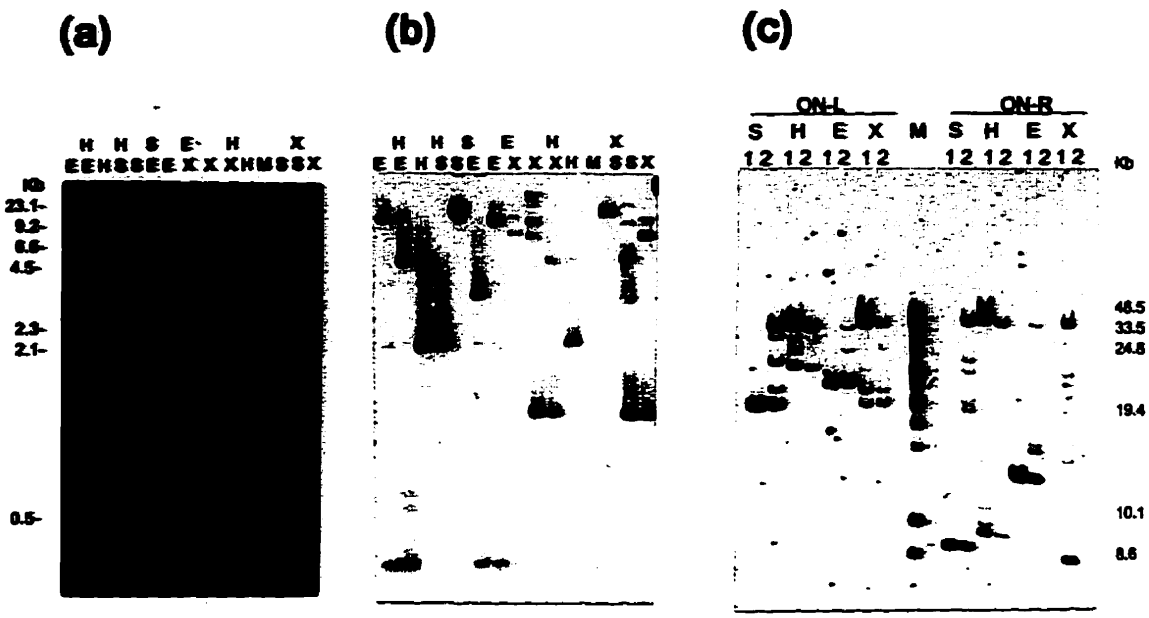
These were as described by Watson *et al.*, (1996, ref. 36). Recombinant rat myristoylated $G_{i\alpha 1}$ and $G_{o\alpha}$ proteins were provided by M. Linder (64, 65). These (3 μg) were combined with His-tagged RGS proteins (2 μg) in binding buffer containing 50 mM Tris-Cl, pH 7.9, 100 mM NaCl, 1 mM $MgSO_4$, 20 mM imidazole, 10 mM 2-mercaptoethanol, 1 mM PMSF, 0.025 % Lubrol-PX, 10 % glycerol, 10 μM GDP, 30 μM $AlCl_3$ and 10 mM NaF, and mixed by gentle agitation for 30 min at 4°C. Samples were bound to 100 μL of charged Ni^{++}-chelate resin and incubated an additional 30 min. with agitation at 4°C. Bound resin was washed three times with 500 μL binding buffer before RGS samples were eluted in 0.5 M imidazole, and assayed for bound Gα subunits by electrophoresis (12% SDS/PAGE) and Western blotting using a Gα-common antibody (antiserum P960; ref. 66) and the ECL detection system (Amersham Corp., UK).

RESULTS

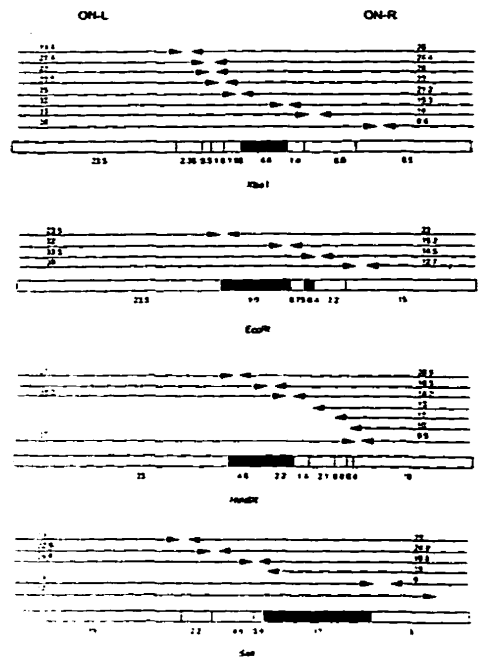
cDNA and derived protein

Our original *G0S8* cDNA clone (0.8 kb) hybridized to an approximately 1.7-kb RNA species, indicating that the cDNA was not full length (1, 34). Subsequently, this clone was used to screen genomic and size-fractionated cDNA libraries. Two overlapping genomic clones were identified from a human genomic library in bacteriophage λ. Fig. 2-1 shows how

Fig. 2-1. **Mapping of the genomic region containing the *G0S8* gene.** a) Restriction endonuclease digests of recombinant λ DNA clone G0S8-2. Recombinant phage DNA (0.5 μg) was digested to completion with 5-10 U of one of the following enzymes: *Eco* RI (E), *Sal* I (S), *Xba* I (X), *Hind* III (H), or *Bam* HI (B; data not shown) or with a combination of the two enzymes as indicated. Reaction products were analysed following separation on 0.8% agarose gels and EtBr staining. M: λ *Hind* III-digested markers the sizes for which are shown at right in kb. b) Restriction fragments from gel in A were transferred to nylon membranes under alkaline conditions (57). Blots were probed with ^{32}P-labelled *G0S8* cDNA probe (10^5 cpm/ml). c) Cohesive-end mapping analysis of the λ G0S8-2 clone. For the same set of enzymes shown in a and b, DNA was partially digested using range of enzyme concentrations (0.015- 0.5 units/ 166 ng of DNA). Pooled partial digests (high enzyme concentrations 1, low enzyme concentrations 2) were hybridized to 5'-end ^{32}P-labelled oligonucleotides complementary to the cohesive ends of the left arm (ON-L) or the right arm (ON-R) of the λ vector. Labelled complexes were separated on 0.4% agarose gels, dried onto DE81 paper and autoradiographed. M: "quick kit" λ markers the sizes for which are shown at left in kb. d) Schematic representation of Cohesive-end labelled fragments from gel in c). Derived single digest maps corresponding to fragments in a) are shown for each enzyme below the corresponding schematic. Shaded bands (in kb) represent those which hybridize with G0S8 cDNA sequence in b).

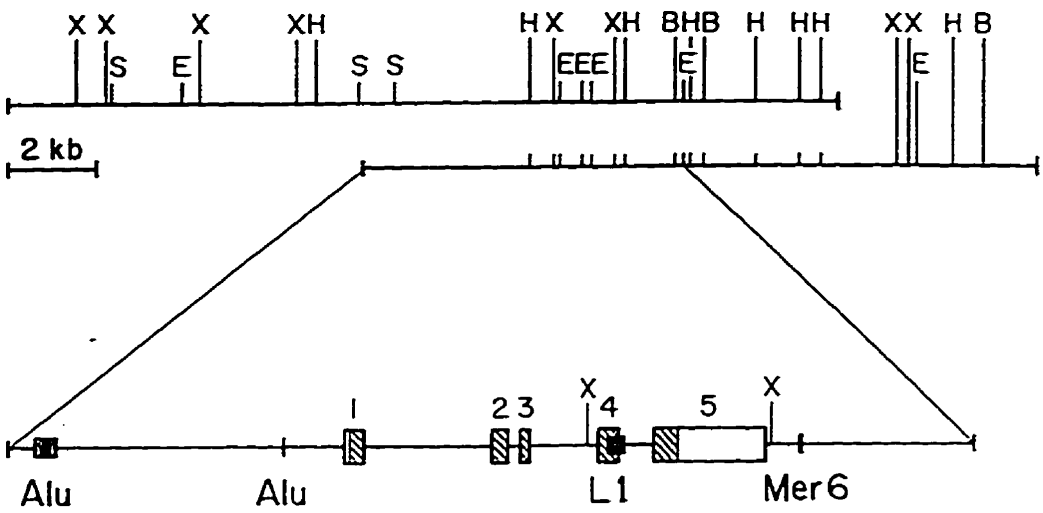


(d)



restriction analysis (Fig. 2-1a), transfer of restricted DNA to nylon blots and hybridization to a *RGS2/G0S8* cDNA probe (Fig. 2-1b), and cohesive-end labelling of partially digested λ clones (Fig. 2-1c,d), were used to characterize the genomic region corresponding to the *RGS2/G0S8* gene in one of these clones (λG0S8-02). The order, number (Fig. 2-1d), and size of cohesive-end labelled fragments for each restriction enzyme (from Fig. 2-1c), agreed well with the complete digest pattern observed in Fig. 2-1a. These data allowed construction of a map for each enzyme (Fig. 2-1d). When maps for the single enzymes were superimposed, they produced a map of this region which was confirmed by comparison to combination digest patterns for each pair of enzymes (Fig. 2-1a). The final maps generated from two independently picked human genomic clones (Fig. 2-2) overlap, confirming the order of restriction sites, and that they were derived from the same genomic region. In the overlap region, the maps are identical except for one *Sal* I site. Bands which hybridized to the *RGS2/G0S8* cDNA probe (Fig. 2-1b) were contiguous within a DNA stretch located completely within the λ clone suggesting that these clones contained the entire sequence of this gene. Overlapping restriction fragments within this region including the 4.4 (λG0S8-01) and 1.4 kb *Xba* I fragments, and the 0.85 and 1.3 kb *Hind*III/*Eco*RI fragments, were sequenced on both strands (7,345 bp, GenBank accession L13391), except for the terminal 844 3' nucleotides (lower, noncoding, strand only). The intron-exon organization of this gene as derived by comparison of genomic to cDNA sequences is shown beneath the map in Fig. 2-2. The cDNA sequence is distributed across five exons (Fig. 2-3b, upper panel). Conventional splice site sequences are found at exon-intron junctions, and the same organization can be derived independently by applying the neural network programs

Figure 2-2. **Major features of the *G0S8* gene.** Restriction maps of two overlapping phage λ human genomic clones. The 5' 7,345 bases (*Eco* RI site to end) of the lower clone (λG0S8-01) were sequenced. Exons are represented as large numbered boxes with stripes indicating protein coding regions. Small dark boxes and lines refer to sequence similarities to parts of various primate repetitive elements.

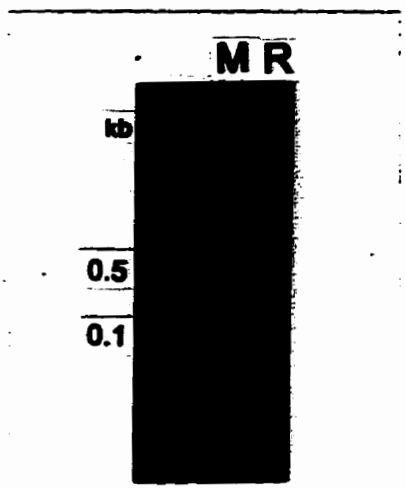

X X X XH H X XH BHB H HH XX H B
S E S S EEE E E
2 kb
1 2 3 X 4 5 X
Alu Alu L1 Mer6

NetGene and GeneID to the genomic sequence (67, 68). From size-selected cDNA libraries, a longer cDNA was obtained (1,345 bp excluding the poly(A) tail). The cDNA sequence is identical to the exon sequences except for the first nucleotide, which is a C in the genomic sequence and an A in the cDNA sequence. Sequence of an *RGS2/G0S8* 5'-end RACE product (Fig. 2-3a) obtained from another human donor matched identically those from the full-length clone with the exception that the 5'-most residue is preceded by a G in the RACE product (data not shown; *cf. G0S24*, ref. 2, Chapter 3). Although a G is found in the same position in the genomic sequence, we presume that the reverse transcription had included the 5' cap site G, which had been added post-transcriptionally to the mRNA. There is only one potential mRNA instability element (ATTTA) in the 3' noncoding region (in contrast to some other *G0S* genes; Siderovski *et al.*, 1990). A polyadenylation signal (AATAAA) begins 28 nucleotides upstream from the site of poly(A) addition.

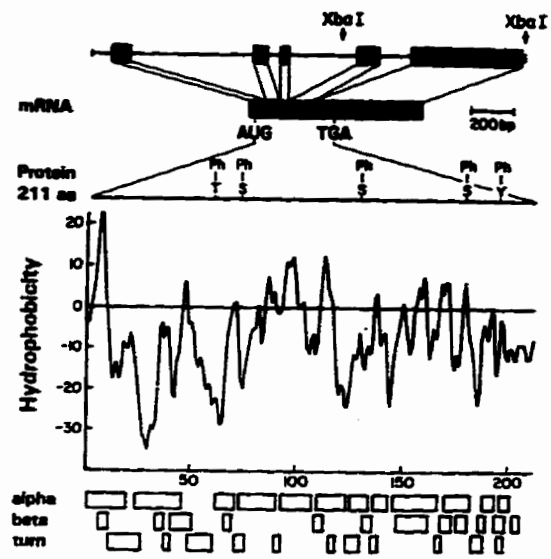
As shown in Fig. 2-3b (upper panel), the largest open reading frame encodes a hydrophilic, basic protein (pI 9.6) of 211 amino acids (molecular weight 24.4 kD). There are four in-frame start codons in the first exon; the third most amino terminal of which agrees best with the Kozak consensus (S.H. unpublished result). The protein is particularly rich in lysine, glutamate, and phenylalanine and low in glycine and valine. Secondary structure analysis by the Chou-Fasman method (69) demonstrates a propensity to form several α-helices with many intervening turns (Fig. 2-3b, lower panel).

Figure 2-3. **5' end RACE product (a) and features of *RGS2/G0S8* gene organization, mRNA and predicted protein (b).** (a) A 5'-end RACE product was amplified between the 3' (artificially polyadenylated) end of the first strand product (for details see ref. 2, Chapter 3) between an oligo$(dT)_{17}$ -adaptor primer and an antisense primer 54 nt downstream of the 5' end of the G0S8 cDNA. The RACE product (R) was electrophoresed in parallel with a 100 bp ladder marker (M) on an 0.8% agarose gel. (b) Upper. Relative positions of *Xba* I restriction sites, exons (striped boxes), and the protein coding region. The positions of certain amino acids (single letter code with "Ph" if potentially phosphorylatable) are shown on the line representing the 211-amino-acid protein. Middle. The hydropathy plot was calculated using an interval of 6 amino acids by the method of Kyte and Doolitte (see 69). Lower. Structural predictions were calculated by the method of Chou and Fasman (see 69).

(a)


M R
kb
0.5
0.1

(b)

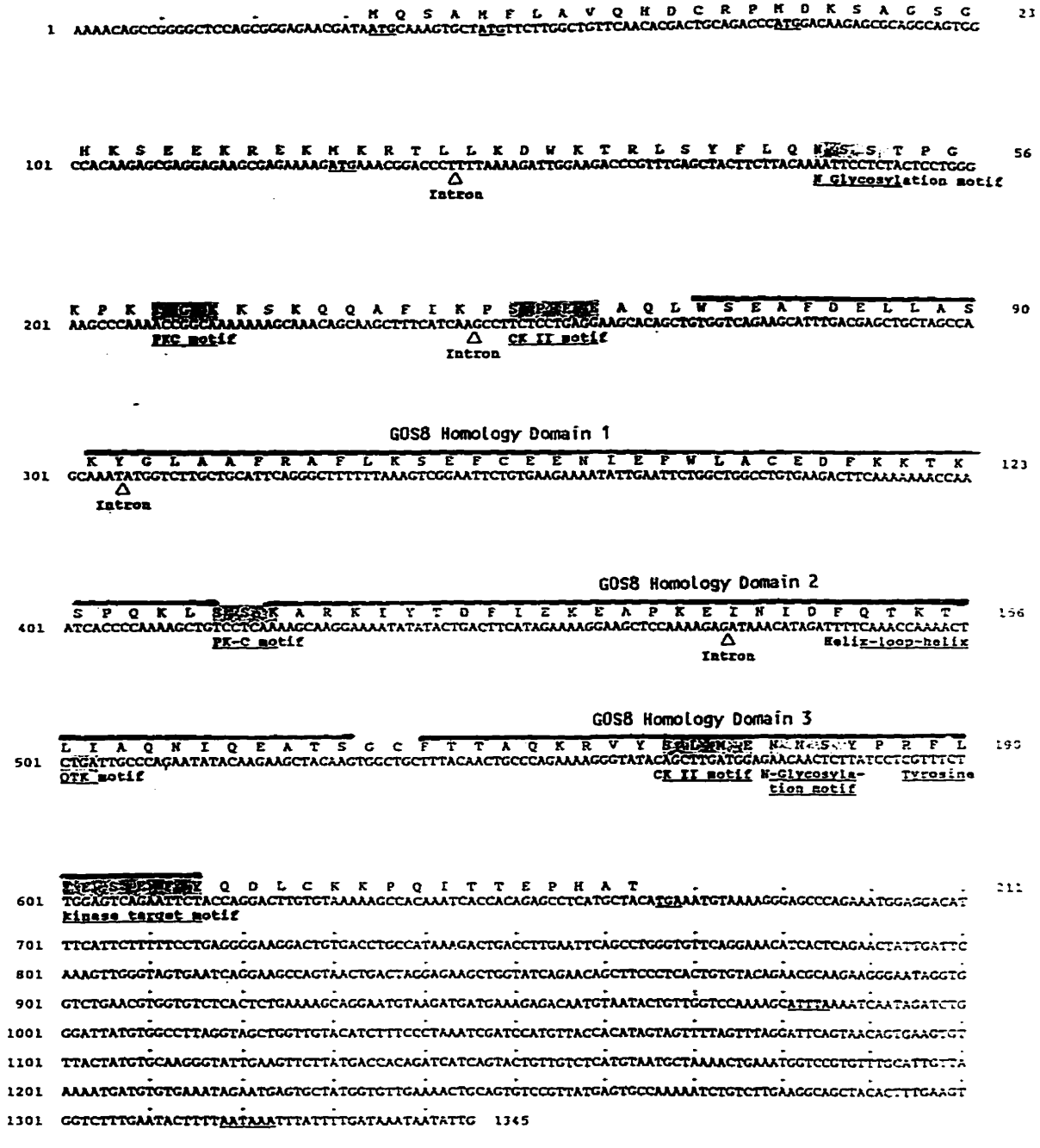

XbaI
XbaI
mRNA
200bp
AUG
TGA
Protein
211 aa
Ph T
Ph S
Ph S
Ph S
Ph Y
Hydrophobicity
20
10
0
-10
-20
-30
50
100
150
200
alpha
beta
turn

Protein database searches

As mentioned in the Chapter 1, the original paper describing the characterization of the *G0S8* gene identified an intriguing similarity (BLASTP screen of SwissProt database) at both the primary and secondary protein structure levels between G0S8 and a family of helix-loop-helix proteins (for details see ref. 3). Since this similarity occurs in the region now known as the RGS "core-domain" it is likely to reflect a conservation of structure (between two adjacent helices) rather than a functional conservation between these two protein families. For this reason, and for reasons of brevity in this thesis, a detailed discussion of the potential role of G0S8 as a transcriptional regulator has been omitted from this chapter.

Fig. 2-4 shows the cDNA sequence and predicted protein product of the RGS2/G0S8 gene. The original BLASTP database search (January 1993) also revealed a hit in the carboxyl-terminal end of this protein with a *Drosophila* G-protein-coupled receptor kinase (70). This region was later shown to correspond to the C-terminal most segment (GH3 domain) of a tripartite 120 amino acid motif found to be characteristic of the RGS family G protein regulators (39). The three domains, designated GH1, GH2, and GH3 for G0S8 homology domains, are found in a nearly contiguous stretch of amino acids in several mammalian RGS family members but are spread across a broader domain in the yeast RGS homologue, Sst2p (48). A search of the protein sequence against the ProSite database revealed two potential *N*-glycosylation motifs (N,X, S/T, ref. 71), and several potential phosphorylation motifs. These include consensus target sites for protein kinase C (S/T, X, R/K; ref. 72), and casein kinase II (S/T, XX, D/E, XX/XXX, Y; ref. 73). One of each of these sites is found within the 120 amino acid RGS-like region of similarity (Fig. 2-4). Also

Figure 2-4. **Sequences of *G0S8* cDNA and the derived protein (GenBank accession number L13463), showing potential protein functional motifs and region of similarity with the RGS family of proteins.** Numbers on the left refer to positions in the cDNA sequence. Numbers on the right refer to positions in the protein sequence. The cDNA sequence is underlined at possible alternative start codons, at the postulated stop codon, at a possible ATTTA mRNA instability element, and at the postulated polyadenylation signal. The positions of intron sequences in unspliced pre-mRNA are marked with triangles. Shaded parts of the protein indicate potential functional motifs (derived by screening version 9.1 of the ProSite database; see 69). Shown above the sequence are three components GH1, GH2, and GH3 of the tripartite 120 amino acid region of RGS similarity.

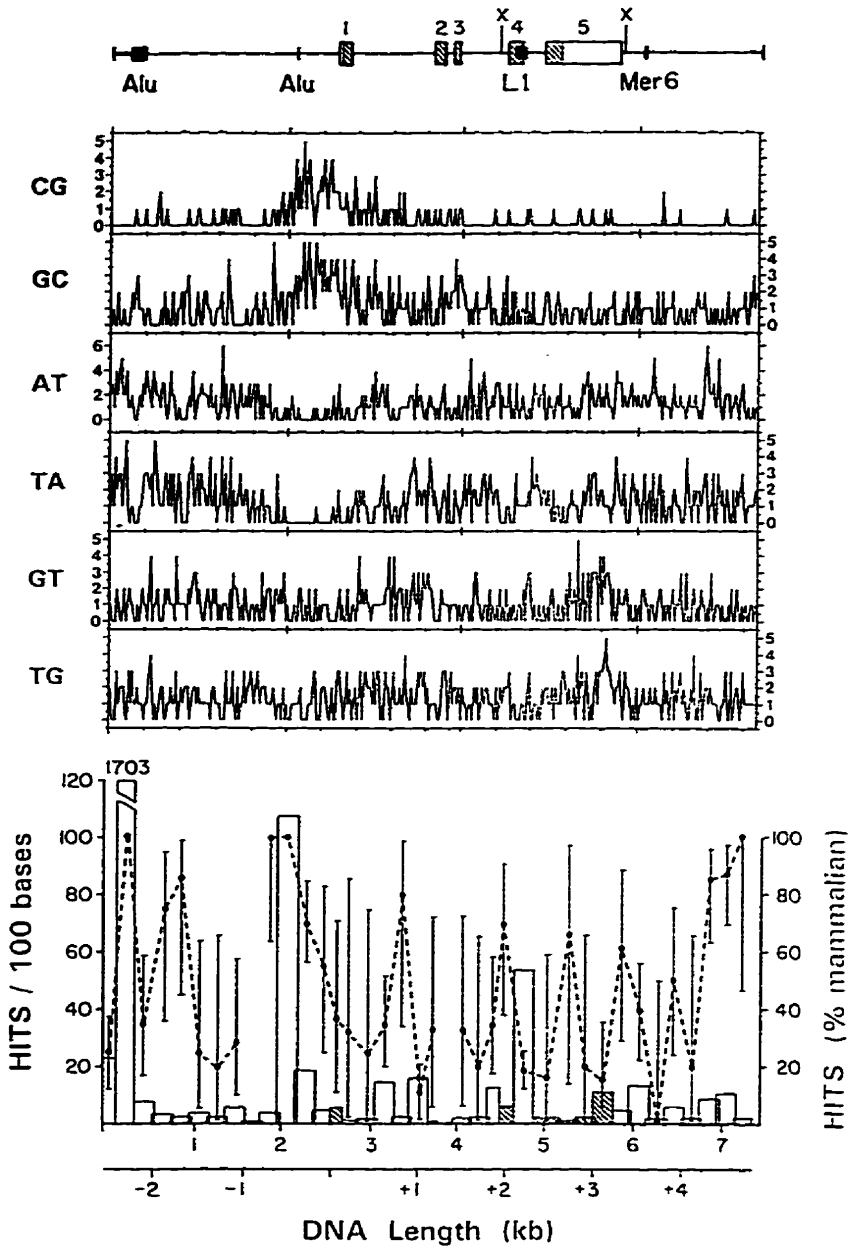

GOS8 Homology Domain 1
GOS8 Homology Domain 2
GOS8 Homology Domain 3
Intron
PKC motif
CK II motif
PK-C motif
Helix-loop-helix
OTK motif
N-Glycosylation motif
Tyrosine
Kinase target motif

interesting is a series of tyrosine residues in the carboxyl terminus, which by inspection, appear similar to the ITAM tyrosine-based activation motif (19) found in subunits of the TCR complex (G0S8, YXLX5YXXXLX4YXXL; ITAM, YXXLX$_{5-6}$YXXL).

General features of the gene

Analysis of dinucleotide frequencies demonstrates a CpG island (74) extending from approximately 500 bp upstream from the transcription start site, across the first exon and into the first intron (Fig. 2-5). The island region is also rich in GpC dinucleotides, but is depleted in ApT and TpA. The 3' noncoding region of the fifth exon shows some enrichment for GpT and TpG. The *G0S8* sequence showed no overall similarity to any sequence in the GenBank database at the time of our initial search (April 1993). Sections of the *G0S8* sequence were examined for similarities with sections of other sequences in the database (Fig. 2-5). The BLAST search program was used to screen a series of windows of approximately 220 nucleotides (both orientations). Scores exceeding a cutoff value of 100 (implying relatively high similarity) were regarded as "hits"; these are very abundant in three regions and moderately abundant in eight regions. Two of the very abundant regions correspond to the partial *Alu* elements; the third, within intron 4, corresponds to a T-rich region similar to sequences found in many invertebrate species (e.g., yeast, nematode). Invertebrate similarities are also primarily responsible for some of the moderately abundant hits. Moderately abundant hits corresponding mainly to mammalian genes occur in the first part of the CpG island and also in the extreme 3' flank.

Figure 2-5. **Characteristics of the RGS2/G0S8 gene.** Upper. Frequency of various dinucleotides in different regions displaying using the programs Windows and StatPlot (Genetics Computer Group; see 69). Readings were taken in successive 20-nucleotide windows with 1 nucleotide overlap. Lower. Frequency of "hits" (regions of sequence similarity to other genes above a cut-off BLAST score of 100) obtained from a screen of overlapping windows of *G0S8* against the GenBank database. Because window size varied (usually 220 nucleotides, with an overlap of 20 nucleotides), hits are expressed per 100 nucleotides. The value for the second window, corresponding to the *Alu* element, was off the scale (1,703 hits/100 bases). Hatched bars correspond to exons. Closed circles refer to the percentage of hits corresponding to mammalian sequences. Confidence limits (95%) were calculated as described by Blyth and Still (1983).


1
2 3
X 4
5
X
Alu
Alu
L1
Mer6
CG
GC
AT
TA
GT
TG
1703
HITS / 100 bases
HITS (% mammalian)
DNA Length (kb)

Analysis of genomic sequence for potential regulatory elements

An atypical TATA box site (CATAA) begins 28 nucleotides upstream from the transcription start site. (ref. 3, Appendix A). This and a CCAAT box beginning 90 nucleotides upstream from the transcription start site provide supporting evidence that our cDNA is full-length. Among the elements found in the 5' CpG island region are potential sites for transcription factors Sp1 (75), Zif268/NGFI-A (EGR1/G0S30), and NGFI-B (76). Potential NGFI-B sites also occur in the first intron and in the 3' flank.

Preincubation elevates RGS1 mRNA and lowers RGS2 mRNA

The *RGS2/G0S8* gene was initially identified by differential mRNA expression in freshly prepared blood mononuclear cell cultures treated with Con-A and cycloheximide for 2 hr. However, it was not possible to demonstrate a consistent response to Con-A alone by conventional RNA blotting (1). Subsequently, we found that basal levels of expression of G0S genes, and responses to inducing agents, varied depending on whether blood mononuclear cells were freshly prepared from a human donor, or had been preincubated for a day (4). Table 2-1 shows that preincubation increases the mRNA levels of RGS1/BL34/IR20 and the HIV-1-inhibitory chemokine MIP1α/G0S19, and decreases the mRNA levels of RGS2/G0S8, the transcription factors F0SB/G0S3, FOS/G0S7, and the putative zinc-finger nuclear protein G0S24/TIS11. The "spontaneously" elevated levels of the former mRNA group in freshly prepared cells was attributed to some stimulatory effect of the isolation procedure since basal levels increase transiently during the first hr of culture, a response which is less apparent in the case of preincubated cells (4).

TABLE 2-1. COMPARISON OF mRNA LEVELS[a] IN FRESHLY ISOLATED AND INCUBATED CELLS

Gene	Time		Change (%)[b]	*P*
	0 hr	24 hr		
RGS1/BL34/IR20	5.6 ± 1.8 (6)	8.3 ± 1.8 (15)	+149 ± 33.7 (6)	<0.01
RGS2/G0S8	25.0 ± 10.0 (6)	2.6 ± 0.5 (14)	-85.0 ± 2.3 (6)	<0.001
MIP1/G0S19	17.6± 5.8 (7)	64.6±15.4 (25)	+360±123 (7)	<0.05
FOS/G0S7	37.4± 8.4 (6)	4.3±1.0 (21)	-88.7±3.0 (6)	<0.001
FOSB/G0S3	41.2±21.7 (7)	1.1±0.5 (16)	<-98.0 (7)	<0.001
G0S24/TIS11	515 ± 166 (6)	107 ± 42 (14)	-73.7 ± 7.5 (6)	<0.001

[a] 10^{-6} x [Molecules /μg total RNA] with standard errors of the means. The number of experiments are in parentheses. An "experiment" involves cells from a distinct donor, cultured on a distinct occasion, for 0 hr and/or 24 hr. At these times RNA is prepared and mRNA-specific primers are added. After reverse transcription, separate aliquots are subjected to PCR amplification with mRNA-specific primer pairs.

[b] Calculated from paired values obtained from individual experiments. In the case of the *FOSB/GOS3* pairs, since values for the 24 hr preincubation were often undetectable, percentage changes approach -100%.

Responses to Con-A depend on preincubation time

"Spontaneous" transient increases in RGS1 and RGS2 mRNA levels occur in the absence of added stimulants. This is evident in freshly isolated cells (Figs.2-6a,b).

Responses to Con-A are best demonstrated with cells which have been preincubated for a day (Figs. 2-6c,d). RGS2 mRNA peaks sooner, and declines more rapidly, than RGS1 mRNA. For RGS2 mRNA, levels have decreased to baseline after 6 hr in culture. For RGS1 mRNA, levels are maintained above basal controls beyond 6 hr (Fig. 2-7). Thus, in spite of the differential expression of RGS1 mRNA in B lymphocytes (42), levels also increased in response to the T lymphocyte mitogen Con-A. However, the kinetics and magnitude of the response appear different than those for RGS2 mRNA.

Con-A does not affect RGS2 mRNA stability

Increased expression of the RGS protein Sst2p in response to pheromone in yeast is likely to reflect increased transcription rather than stabilization of rapidly turning-over mRNA (47, W. E. Courchesne and J. Thorner, unpublished data). Using a run-on transcription assay, Murphy and Norton (1990; ref. 41) directly demonstrated increased transcription as a factor in the increase in RGS1 mRNA in response to phorbol ester. That the *RGS2* response to Con-A might also reflect increased transcription rather than mRNA stabilization was suggested by studies with the transcription inhibitor Actinomycin D, in which the decay of RGS2 mRNA was compared with the decay of FOS mRNAs.

Preincubated cells were treated with Con-A and 30 minutes later, at a time when mRNA levels are still increasing in response to Con-A (Fig. 2-7), Actinomycin D (4 μM) was added.

Figure 2-6. **RT-PCR assay of the response to Con-A of RGS1 and RGS2 mRNAs levels in freshly isolated cells (a, b), or cells preincubated for 24 hr (c, d).** Small filled symbols refer to cultures containing Con-A. Large open symbols refer to cultures without Con-A. At 0 min, cultures (2.5 ml with approximately 2 x 10^6 cells/ ml) were treated with 200 µl of either Con-A solution (final concentration 200 µg/ml) or water (controls). At the indicated times total RNA was prepared, and aliquots taken for RT-PCR assay. The photographs in the upper part of the figure show ethidium bromide-stained PCR products separated by agarose gel electrophoresis. Upper bands are derived from *fixed* concentrations of control plasmids containing RGS1 or RGS2 cDNAs with small inserts to decrease electrophoretic mobility. These were coamplified with the corresponding reverse transcriptase-generated cDNAs whose concentrations vary depending on the RNA sample of origin (lower bands). A given RNA sample was used for measurement of both mRNAs. The graphs show the results of coamplification of fixed quantities of reverse transcriptase generated cDNAs with *varying* quantities of control plasmids whose concentration range was selected to include a concentration where molecules of plasmid cDNA would compete on a one-to-one basis with molecules of reverse transcriptase products. For each data point, intensities of pairs of ethidium bromide-stained bands were compared to quantitate the reverse transcriptase-generated cDNA (60). Data for RGS1 are in (a) and (c). Data for RGS2 are in (b) and (d). The ordinate scale gives molecular concentrations of mRNA assuming 100% efficiency of reverse transcription.

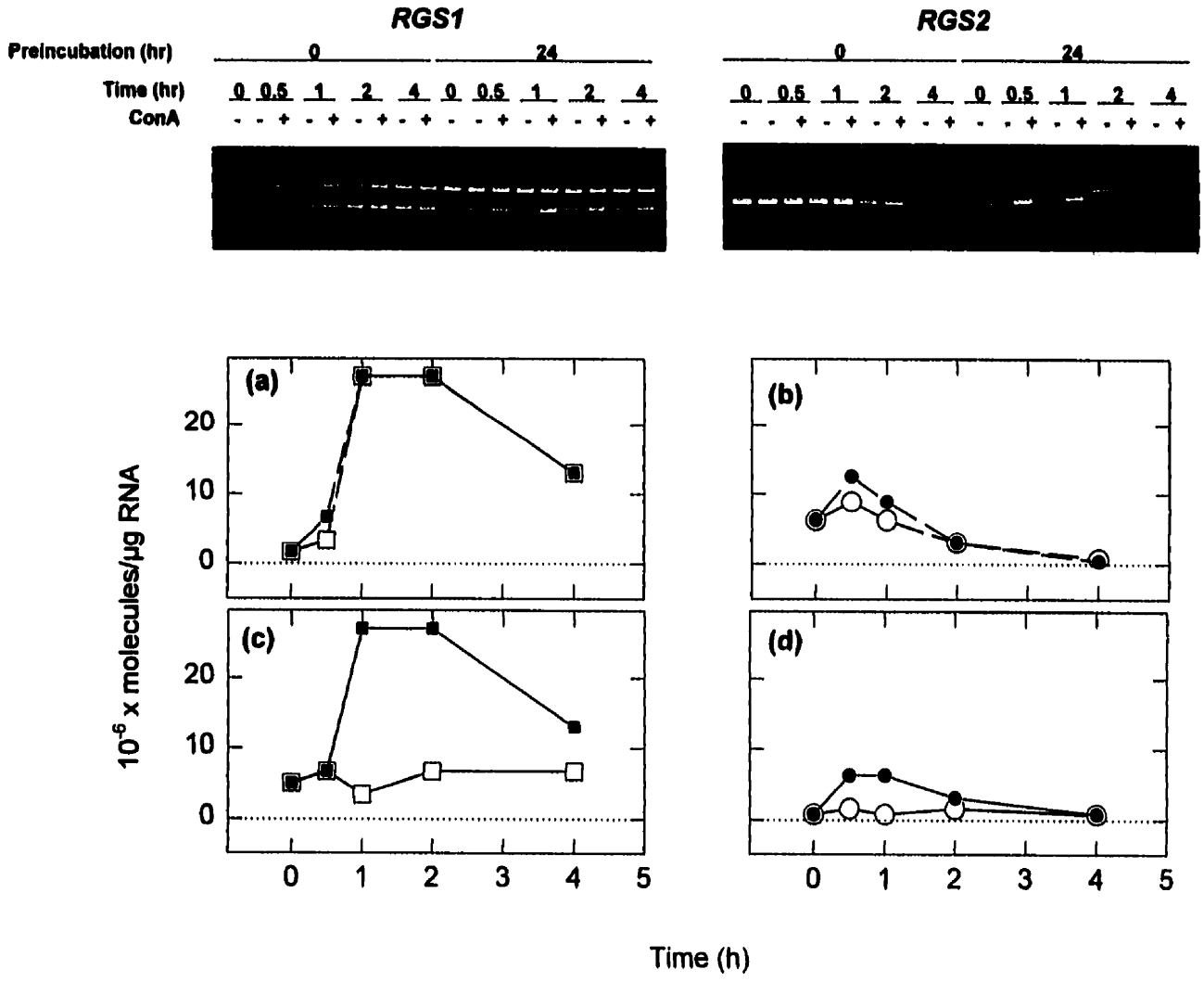
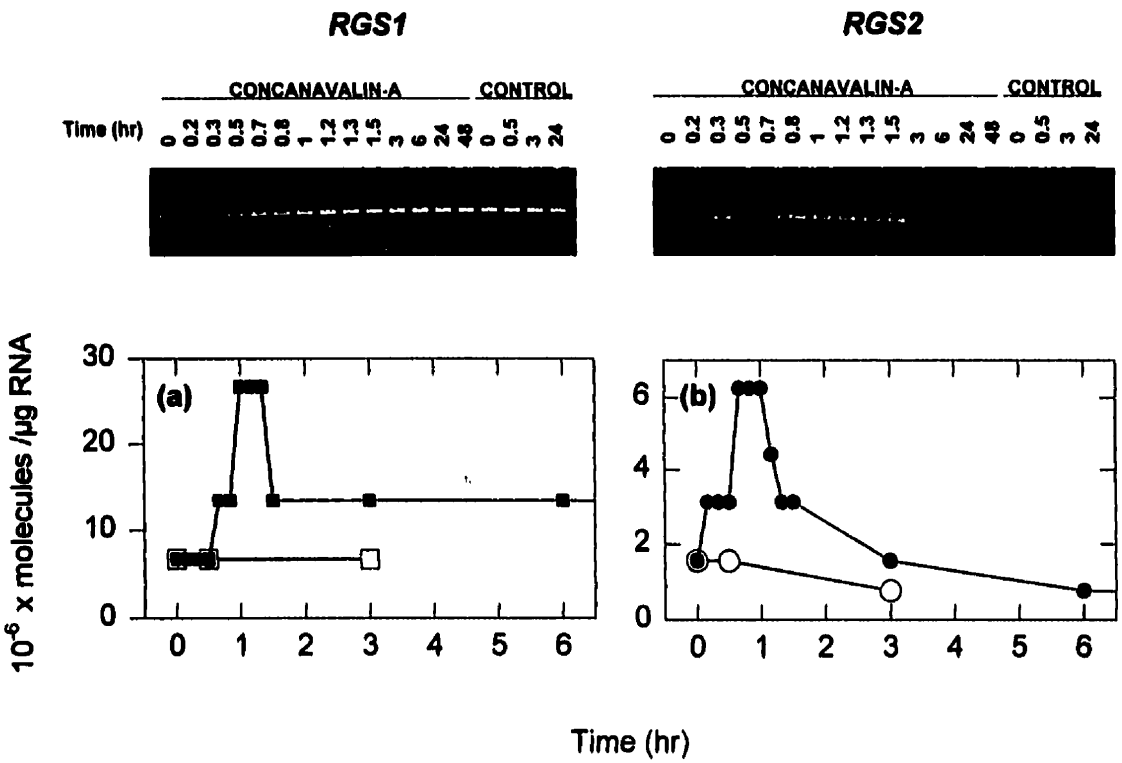

RGS1
RGS2
Preincubation (hr)
0
24
Time (hr)
0 0.5 1 2 4 0 0.5 1 2 4
ConA
(a)
(b)
(c)
(d)
20
10
0
0 1 2 3 4 5
10^{-6} x molecules/µg RNA
Time (h)

Figure 2-7. **Response to Con-A of (a) RGS1 mRNA, and (b) RGS2 mRNA levels in preincubated cells.** Details are as in Fig. 2-6.


RGS1
RGS2
CONCANAVALIN-A
CONTROL
Time (hr)
0 0.2 0.3 0.5 0.7 0.8 1 1.2 1.3 1.5 3 6 24 48 0 0.5 3 24
(a)
(b)
10^{-6} x molecules /µg RNA
Time (hr)

Messenger RNA levels were determined during the following 50 minutes (Figs. 2-8a,b,c). When expressed on a log scale, decay curves appeared to follow a linear relationship (Figs.2-8d,e,f). *RGS2* mRNA levels declined with a half-life of 29±4 min in Con-A-treated cultures and of 35±9 min in control cultures (2 experiments). Corresponding values for the *FOS/G0S7* and *FOSB/G0S3* mRNAs in Con-A-treated cultures were 14±2 min and 18±1 min, respectively (3 experiments). Pair-wise statistical analysis (t-test) of regression coefficients showed the rate of *RGS2/G0S8* decay did not differ significantly between Con-A-treated and control cultures. The decay of the *FOS* and *FOSB* mRNAs appeared to be accelerated in Con-A-treated cultures, but in some experiments levels in control cultures were so low that accurate quantitation was not possible (4).

Responses to cycloheximide depend on preincubation time

Both RGS1 and RGS2 mRNA levels increase in response to the protein synthesis inhibitor cycloheximide (1, 34, 41). In both freshly isolated cells, and cells in preincubated cultures, the levels of RGS1 and G0S24/TIS11 mRNAs increase in an approximately rectilinear fashion over the 45 min period studied, whereas the level of *RGS2* mRNA tends to increase less after the 15 min time point (Figs. 2-9a,b,c). Compared with the responses of freshly isolated cells, in preincubated cells relative responses of *RGS1* mRNA to cycloheximide (relative to levels at zero time) are decreased (Fig. 2-9d), of *RGS2* mRNA are unchanged (Fig. 2-9e), and of G0S24 mRNA are increased (Fig. 2-9e).

Figure 2-8. **Messenger RNA degradation kinetics in preincubated cultures for (a, d) *RGS2/G0S8*, (b, e) *FOS/G0S7*, and (c, f) *FOSB/G0S3*.** Con-A stimulated cultures (filled symbols) and control cultures (open symbols) were incubated at 37°C for 0.5 h prior to addition of 4 μM Actinomycin D (time defined as zero). RNA levels derived from competitive RT-PCR assays are shown either as absolute quantities (molecules/μg total RNA) on a linear scale (a, b, c), or relative to the maximum level in Con-A-treated cultures (defined as 100%) on a logarithmic scale (d, e, f). Details are as in Fig. 2-6.

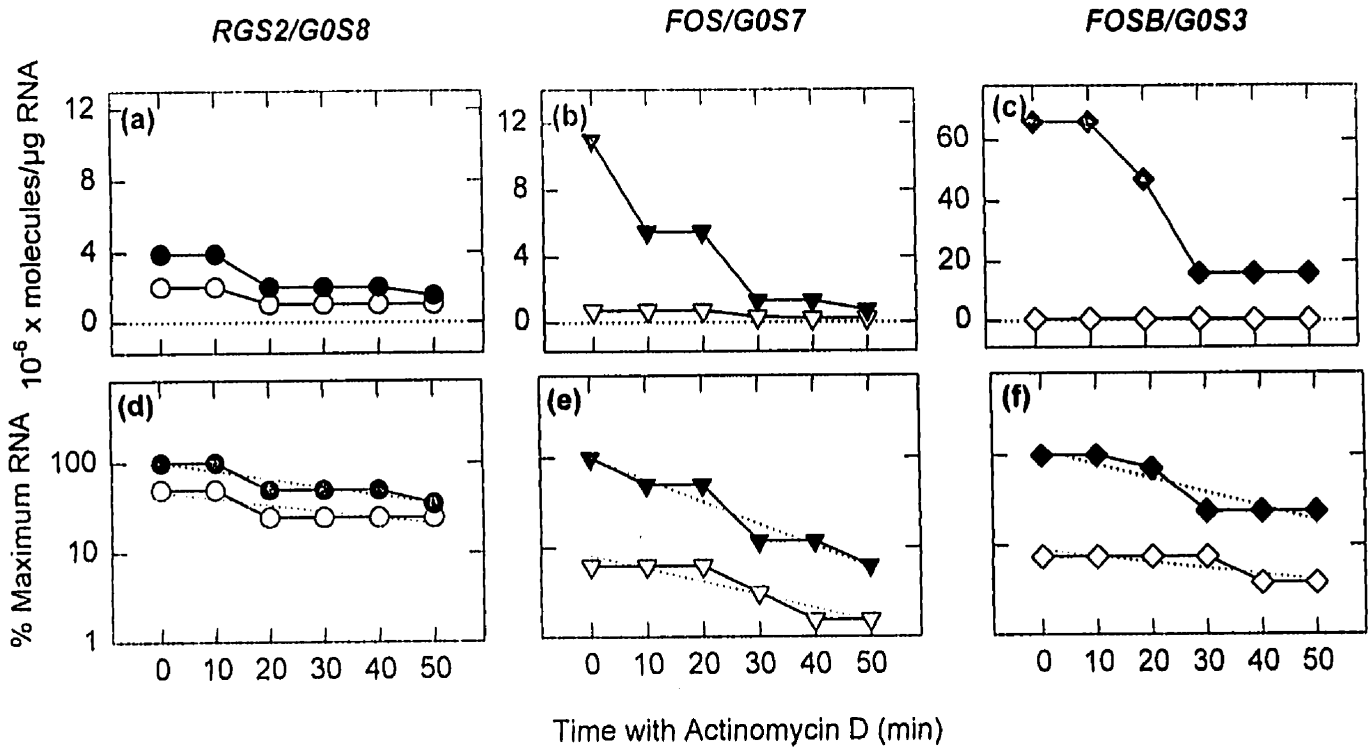
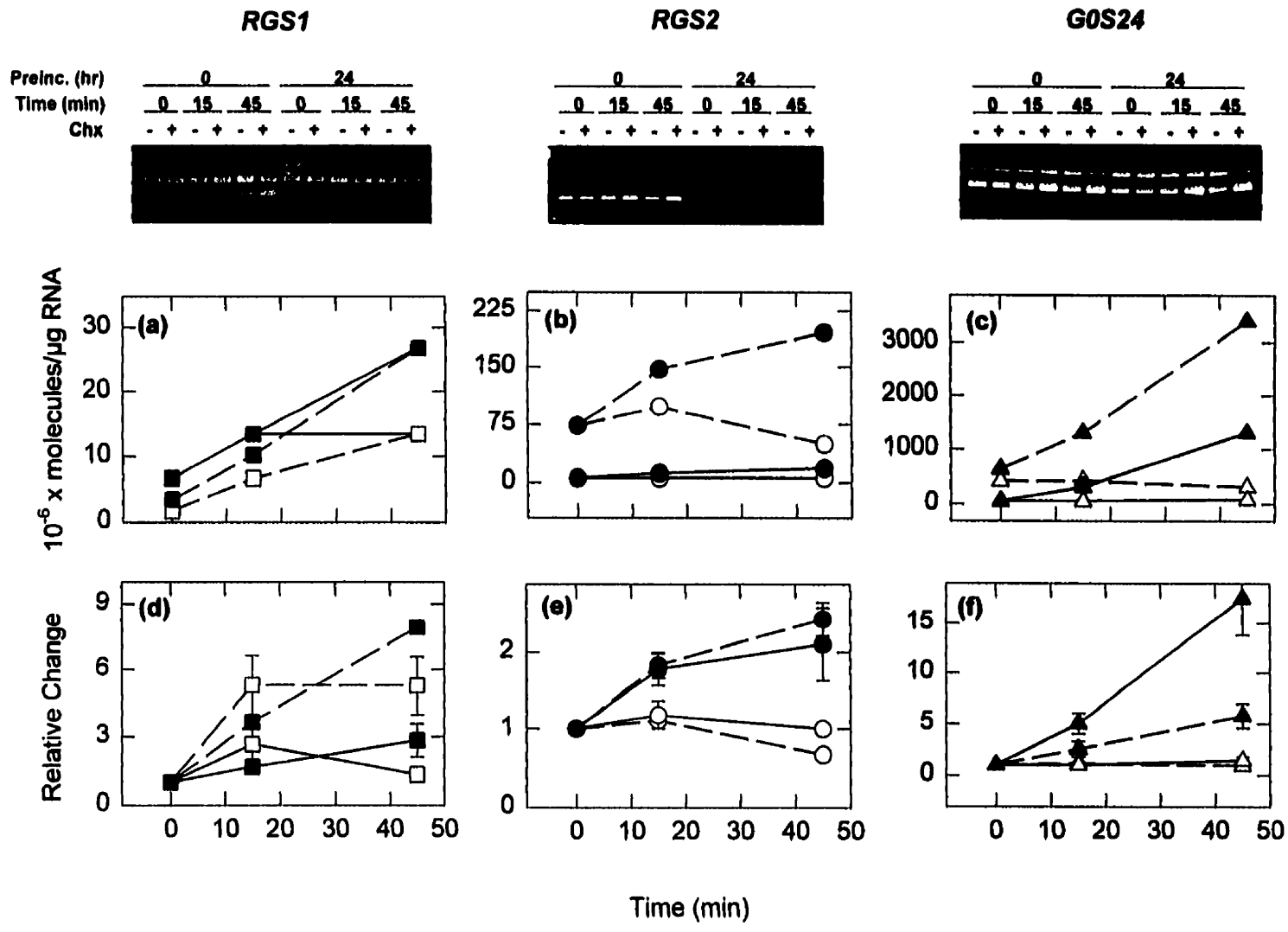

RGS2/G0S8
FOS/G0S7
FOSB/G0S3
(a)
(b)
(c)
(d)
(e)
(f)
10^{-6} x molecules/µg RNA
% Maximum RNA
Time with Actinomycin D (min)

Figure 2-9. **Effects of cycloheximide (filled symbols) on levels of (a) *RGS1* mRNA, (b) *RGS2* mRNA, and (c) *G0S24* mRNA in either freshly isolated cells (dashed lines), or preincubated cells (continuous lines).** Cells were treated with 12.5 µl of either cycloheximide solution (final concentration 0.1 mM) or water (control), and incubated for a further 0, 15,or 45 min. The upper photographs show mRNA profiles in an individual experiment. The corresponding graphs (lower) show (a, b, c) absolute mRNA levels in the same experiment, and (d, e, f) changes in mRNA levels relative to the levels in control cultures at 0 hr, summarized for three experiments. In the latter case, data are presented together with the standard errors of the mean. In the case of points shown without error-bars, the error limits fell within the area covered by the point. Other details are as in Fig. 2-6.


RGS1
RGS2
G0S24
Preinc. (hr)
Time (min)
Chx
0
24
0 15 45 0 15 45
- + - + - + - + - + - +
(a)
(b)
(c)
(d)
(e)
(f)
10^-6 x molecules/µg RNA
Relative Change
Time (min)

Unlike RGS1, RGS2 is ionomycin- but not TPA-responsive

TPA-treatment (100 nM) increases mRNA expression of a number of immediate-early genes (77). Studies with preincubated cells confirm this for *G0S24/TIS11* (Fig. 2-10c) and *RGS1* (Fig. 2-10a; ref. 41). However, mRNA levels of *RGS2* are not increased (Fig. 2-10b). On the other hand, in three experiments ionomycin treatment for 30 and 60 min resulted in transient 4.5±1.6 and 2.7±0.7-fold increases, respectively, in *RGS2* mRNA levels relative to 0 min baseline levels (Fig. 2-11b). *RGS1* mRNA levels were stimulated much less (Fig. 2-11a), consistent with the observations of Murphy and Norton (1993, ref. 42), whereas FOSB mRNA levels increased briskly (Fig. 2-11c).

Expression of recombinant RGS2/G0S8

$(HIS)_{10}$-RGS2/G0S8 protein was expressed in *E. coli* (Fig. 2-12a,b), purified by nondenaturing Ni^{++}-chelate affinity and gel filtration chromatography, and assayed as >85 % pure on Coomassie-stained gels (Fig. 2-13a). Yields of 3.5 mg/litre of *E. coli* culture were obtained. Purified protein often ran as a triplet with one major species and two more rapidly migrating minor species. Storage for short periods at 4°C under reducing conditions (10 mM 2-ME) at neutral pH resulted in increased amounts (up to 50%) of the shorter species as well as the appearance of 60 and 90 kDa bands. These could be resolved on denaturing gels as a single band only in a large excess (2 M) of reducing agent suggesting a tendency to form strong intra- and inter-molecular cysteine-cysteine bonds (data not shown). As the migration of monomeric $(HIS)_{10}$-RGS2/G0S8 was slower on SDS/PAGE than its molecular weight would predict (30.5 kD versus 27.5 kD), proper expression of RGS2/G0S8 protein was

Figure 2-10. **Effect of TPA on levels of (a) *RGS1* mRNA, (b) *RGS2* mRNA, and (c) *G0S24* mRNA in cells preincubated for 24 hr.** Photographs (upper) show expression profiles for cultures with added 100 nM TPA in dimethylsulfoxide (DMSO; lanes 3, 6, 9, and 12), or with DMSO alone (lanes 2, 5, 8, and 11), and for untreated cultures (lanes 1, 4, 7, and 10). Corresponding graphs (lower) show the mRNA levels determined using the competitive RT-PCR assay, for TPA-treated cultures (filled symbols), DMSO-treated cultures (shaded symbols), and untreated cultures (open symbols). Other details are as in Fig. 2-6.

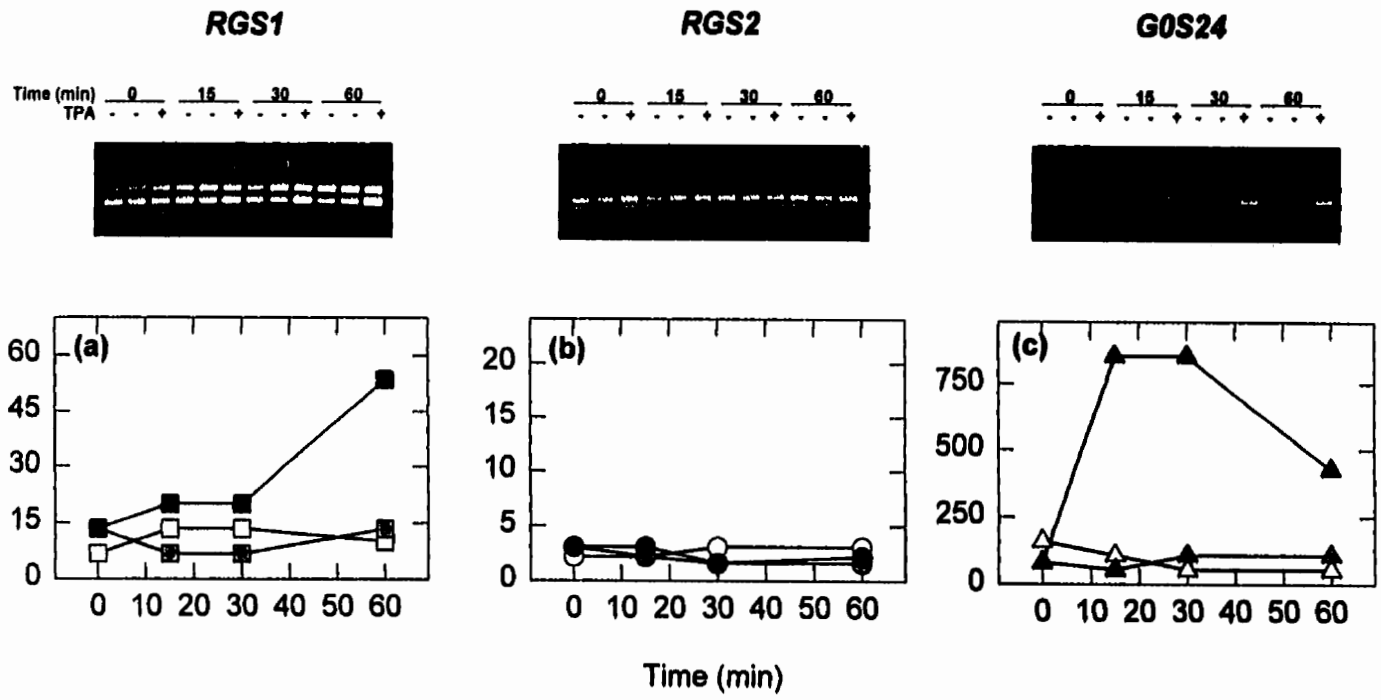
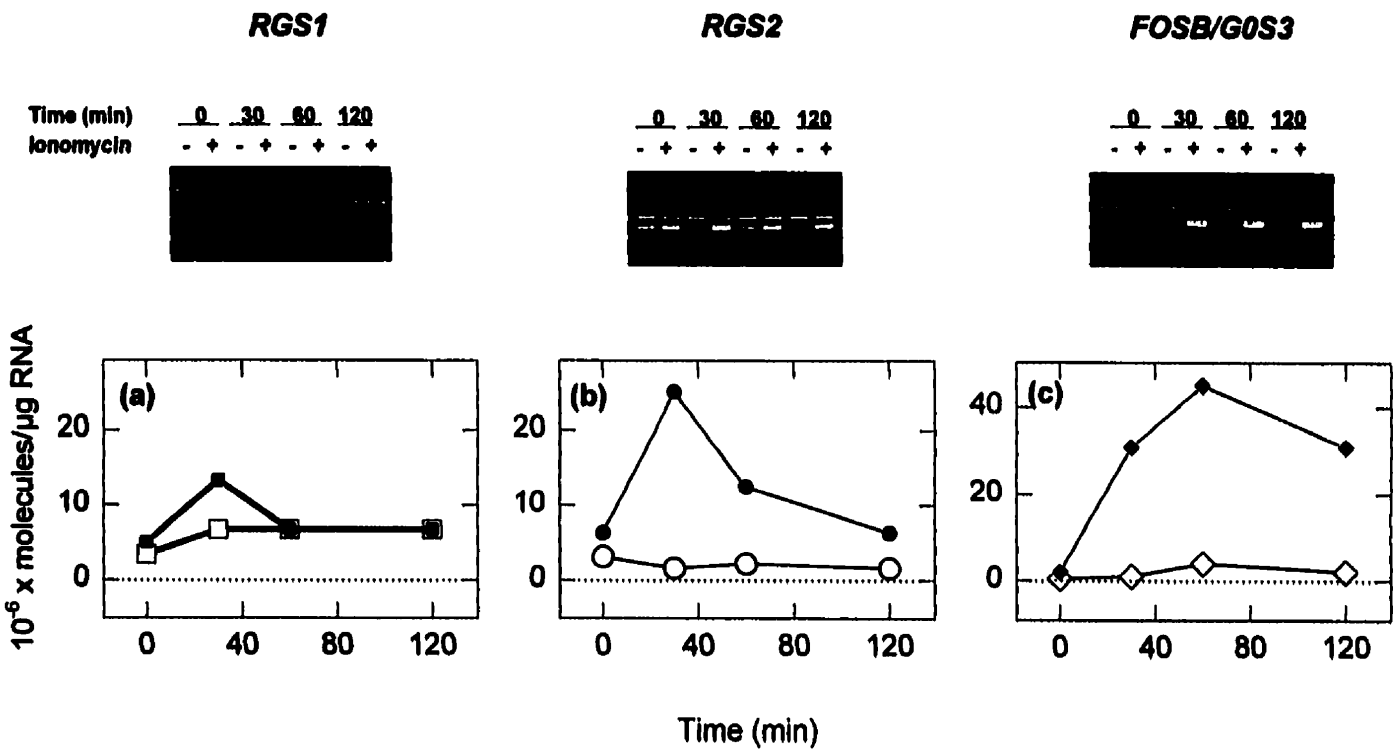

RGS1
RGS2
G0S24
Time (min) 0 15 30 60
TPA - - + - - + - - + - - +
(a)
(b)
(c)
60
45
30
15
0
20
15
10
5
0
750
500
250
0 10 20 30 40 50 60
Time (min)

Figure 2-11. **Effects of the calcium ionophore, ionomycin, on levels of (a) *RGS1* mRNA, (b) *RGS2* mRNA, and (c) *FOSB* mRNA in cells preincubated for a day.** The upper photographs show mRNA profiles for cells cultured with 2 μM ionomycin (lanes 2, 4, 6, and 8), or DMSO (lanes 1, 3, 5, and 7). Corresponding graphs (lower) show the corresponding mRNA levels determined using the competitive PCR assay, for ionomycin-treated cultures (filled symbols), and for control DMSO-treated cultures (open symbols). Other details are as in Fig. 2-6.


RGS1
RGS2
FOSB/G0S3
Time (min)
Ionomycin
0 30 60 120
- + - + - + - +
(a)
(b)
(c)
10^{-6} x molecules/µg RNA
0
10
20
40
80
120
Time (min)

verified by Western blotting (Fig. 2-12b, ref. 39), and mass spectrophotometry (Fig. 2-13b). The purified protein was most soluble in solutions containing >0.35 M NaCl. Ionic strength appeared to be the important parameter since soluble protein partially precipitated when dialysed against 0.35 M urea but remained soluble in 0.35 M GuHCl (S.H. unpublished).

$(HIS)_{10}$-RGS2/G0S8 does not bind $G_{i\alpha 1}$ or $G_{o\alpha}$ in vitro

Watson *et al.*, (1996, ref. 36) have shown that RGS proteins interact strongly *in vitro* with GDP/AlF_4^--bound Gα subunits, which are locked in the activated state conformation. Using similar binding condition we compared the binding activities of recombinant RGS1 and RGS2. In buffer containing GDP and AlF_4^-, the Gα protein subunits $G_{o\alpha}$ and myristoylated $G_{i\alpha 1}$ bind very strongly to recombinant $(HIS)_6$-RGS1, and after consecutive washes in the continued presence of AlF_4^- may still be co-eluted with this protein upon addition of imidazole to the bound Ni^{++}-charged resin (Fig. 2-14). In contrast, using identical conditions we were unable to detect association between either Gα subunit and $(HIS)_{10}$-RGS2 protein. This confirms a similar study carried out by N. Watson and K. Blumer (personal communication).

Researchers at the University of Tokyo have found that cells expressing relatively high levels of RGS2/G0S8 mRNA show Protein Kinase C-dependent modulation of α_2-adrenergic receptor-mediated inhibition of adenylate cyclase (H. Kurose, personal communication). Interestingly, no other RGS mRNAs were detected when a search using RT-PCR with degenerate oligonucleotide primers was performed. Despite being a good substrate *in vitro* for phosphorylation by purified protein kinase C (0.8 moles phosphate/mole protein), which

Figure 2-12. **Expression of $(His)_{10}$-RGS1 in total *E. coli* cell lysates.** Aliquots (100 μl) of control (-) and IPTG-induced (+) cultures started from three independent clones were pelleted , resuspended in SDS/PAGE loading buffer and separated on 12 % denaturing PAGE gels. Protein bands were analysed by Coomassie staining (a) or Western blotting (b) using antisera #1766 (39).

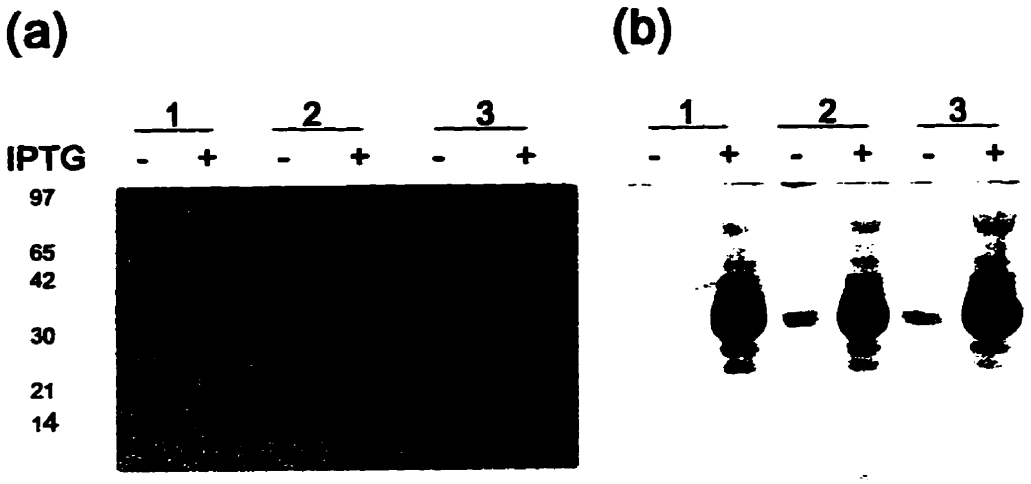

(a)
(b)
1
2
3
IPTG
- + - + - +
97
65
42
30
21
14
1
2
3
- + - + - +

Figure 2-13. **Purification of soluble $(His)_{10}$-RGS1 by nondenaturing Ni^{++}-chelate affinity chromatography.** Filtered supernatants from sonicated IPTG-induced cultures (+) were loaded onto Ni^{++}-charged immobilized metal affinity chromatography (IMAC) columns and purified as described previously (Heximer and Forsdyke, 1993, Chapter 3). (a) Protein bands at stages of the purification (pre-column +/- IPTG; column flow through, FT; 60 mM imidazole wash, WASH; 500 mM imidazole elute, ELUTE) were analysed on Coomassie-stained SDS/PAGE gels. (b) Purified $(His)_{10}$-RGS1(2 mg/ml) was analysed by mass spectrophotometry.

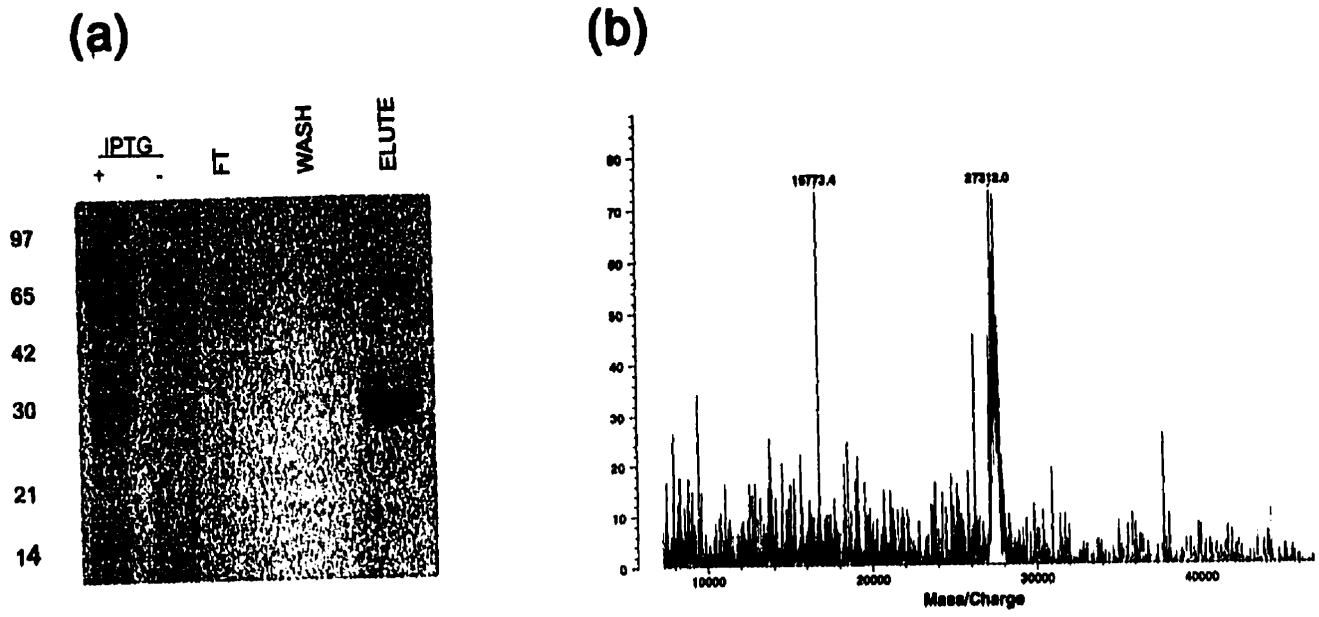

(a)
IPTG
+
-
FT
WASH
ELUTE
97
65
42
30
21
14
(b)
19773.4
27312.0
Mass/Charge

Figure 2-14. **Binding of RGS proteins to G α subunits *in vitro*.** Protein samples containing $G_{o\alpha}$ (Go) or myristoylated $G_{i\alpha 1}$ (myr-Gi) were combined with histidine-tagged recombinant RGS1 or RGS2 , incubated with charged Ni^{++}-chelate resin, washed and eluted in imidazole. Aliquots corresponding to: binding mixtures before (PRE) and after (U) Ni^{++}-chelate incubation, the final wash (W), and eluted samples (E), were assayed for G α by separation on 12% SDS/PAGE and Western blotting.

RGS1
RGS2
PRE U W E
PRE U W E
Go
myr-Gi

appeared to dramatically increase the solubility of the recombinant protein in the assay buffer (Fig. 2-15, lower panel), phosphorylation did not increase the affinity of recombinant RGS2 for either member of the G_i subfamily (Fig. 2-15, upper panel).

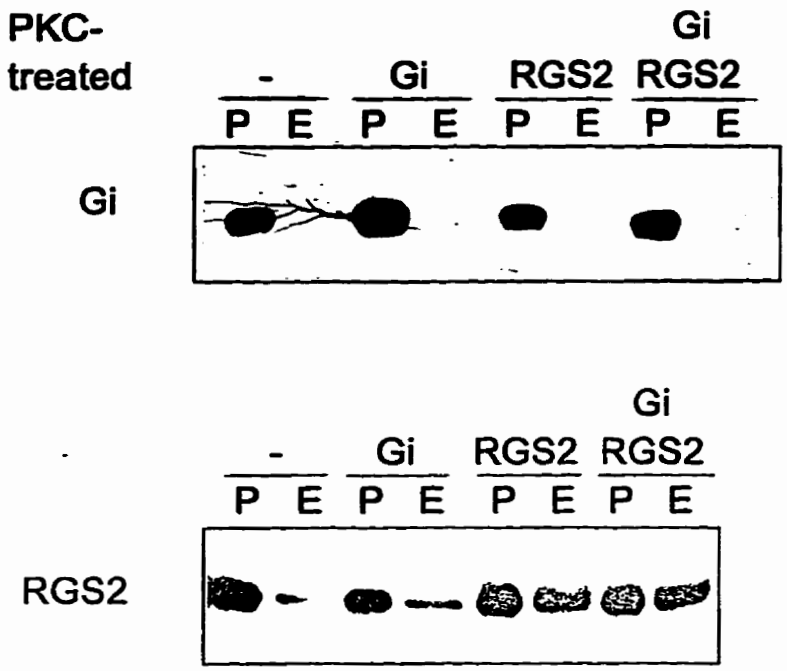
DISCUSSION

We describe here the characterization of the *RGS2/G0S8* gene and predicted protein product (Figs. 2-1 through 2-5). We point to the relationship of this protein to RGS proteins, a growing family of modulators of G protein-coupled signal transduction and study the expression of its mRNA in human mononuclear cells compared to that for RGS1/BL34, another lymphoid-specific family member.

Mammalian RGS proteins 1-4 can complement Sst2p deficiency in yeast, indicating high functional conservation throughout evolution (39, 51). Pheromone increases transcription of the yeast *SST2* gene, resulting in an increase in Sst2p. Mutational analysis indicates an important role of the protein in modulating the commitment of cells to arrest in the G_1 phase of the cell cycle, following exposure to pheromone (47, 48; W. E. Courchesne and J. Thorner, personal communication). Exposure to Con-A commits mammalian lymphocytes to switch from the G_0 to the G_1 phases of the cell cycle (78), and this may be associated with transient increases in RGS1 and RGS2 mRNAs (Figs.2-6 and 2-7), consistent with a role of the corresponding proteins in modulating Con-A-responsive pathways.

The requirement for Con-A for lymphocyte activation can be by-passed by the synergistic actions of TPA and calcium ionophore (53, 79). TPA alone increases transcription of *RGS1* (41), and our studies with actinomycin indirectly indicate that the

Figure 2-15. ***In vitro* binding assay using PKC-phosphorylated RGS2 and G proteins.** Nonphosphorylated or phosphorylated RGS2 and G protein (Gi) were combined in binding buffer in various combinations. Upper. Aliquots corresponding to binding mixtures (P) and eluted fractions (E) were analysed for *in vitro* binding as described in Fig. 2-14. Lower. Western blot analysis of RGS2 in duplicate sample series. Details as in 12b.


PKC-
treated
-
Gi
RGS2
Gi
RGS2
P E P E P E P E
Gi
Gi
RGS2
-
Gi
RGS2
Gi
RGS2
P E P E P E P E
RGS2

response to Con-A may involve increased transcription of *RGS2* (Fig. 2-8). Our failure to detect an increase in RGS2 mRNA in earlier RNA blotting studies (1), probably reflects the small magnitude of the increase compared with that of other immediate early genes (4), and the need to preincubate cells to decrease the high baseline level of mRNA in freshly isolated cells.

Approximately 85% of cells in our cultures are small lymphocytes. We do not routinely attempt further purification because various accessory cells play a role in lymphocyte activation (80), and because small lymphocytes in peripheral blood are already enriched in T lymphocytes for which Con-A is specifically mitogenic. RGS1 mRNA is considered to be B lymphocyte specific, and there are normally some B lymphocytes in blood (42, 43). Thus, it was of interest to note, paradoxically, an increase in RGS1 mRNA in response to Con-A (Fig. 2-6). However, this is delayed relative to RGS2 mRNA, which increases sooner, and declines more rapidly (Fig. 2-7). It is possible that an initial reaction of Con-A with T-lymphocytes (resulting in an increase in RGS2 mRNA), causes these cells to release chemokines which might then activate neighbouring B-lymphocytes by way of G-protein-coupled receptors (81), resulting in a subsequent increase in RGS1 mRNA.

Possible reasons for differences in mRNA levels between freshly isolated and preincubated cells have been considered previously (4). Large absolute increases in certain mRNAs in freshly isolated cells in response to cycloheximide, which stabilizes mRNA-polysome complexes, would appear to reflect the high levels of these mRNAs in freshly isolated cells (Table 2-1; Fig. 2-9). Thus, the turnover of mRNAs for RGS2 and G0S24 (Fig. 2-9), and for the two FOS oncoproteins (4) would be inhibited by cycloheximide, while

transcription would continue, or even increase if there were feedback inhibition of transcription by the corresponding gene products (82). The observation of a relatively small elevation of RGS2 mRNA levels in the presence of cycloheximide (Fig. 2-9e), suggests a low intrinsic transcription rate consistent with a requirement for transcriptional activation to generate the Con-A-induced mRNA increase (Figs. 2-6 and 2-7).

The tendency of RGS2 mRNA levels not to further increase between 15 and 45 min after addition of cycloheximide suggests mRNA stabilization, but no further transcription. Our studies with actinomycin D suggest that stabilization would be a factor in the observed elevation of the FOS oncoprotein mRNAs by cycloheximide (4), since these mRNAs have shorter half lives than RGS2 (Fig. 2-8). Consistent with this, the FOS oncoprotein mRNAs have more AUUUA-rich destabilization elements (83) in the 3' non-coding region than RGS2 mRNA.

Levels of mRNAs for RGS1 and G0S19/MIP1α increase during preincubation (Table 2-1), and in the case of G0S19/MIP1α, cycloheximide tends to cause greater absolute increases in mRNA levels in the preincubated cells, than in freshly isolated cells (4). However, this was not found with RGS1 mRNA (Fig. 2-9a), implying a relatively stable mRNA (not continually being replenished by transcription), despite the presence of two 3' non-coding AUUUA elements (44). RGS1 and G0S19/MIP1α cDNA sequences both have the 3' non-coding element TTTTGTA followed by an AT-rich sequence. This conserved regulatory element, which is involved in transcriptional activation rather than in mRNA stabilization, is found in the 3' non-coding regions of several cDNAs corresponding to "immediate early" genes (84, 85), but is not present in RGS2 mRNA.

In view of the importance of calcium and protein kinase C in T cell activation (52, 53), it was of interest to note that ionomycin, but not TPA, increased RGS2 mRNA levels (Figs. 2-10 and 2-11). This suggests that Con-A increases RGS2 mRNA expression based on its ability to increase calcium levels during lymphocyte activation, and not through activation of protein kinase C-dependent signal transduction pathways. By the same token, Con-A would appear to increase RGS1 mRNA through a protein kinase C-dependent mechanism (Fig. 2-10).

The distinction between RGS1 and RGS2 was further emphasized by the binding of RGS1, but not RGS2, to two members of the G_i subfamily of G protein α subunits (Fig. 2-14). This observation is supported by Watson and Blumer (personal communication), who showed that RGS1, but not RGS2, increased the GTPase activity of members of the same G_i subfamily. It remains to be established whether these apparent functional differences merely reflect a requirement for special *in vitro* conditions for RGS2 binding, or some difference in the recombinant forms (e.g. the numbers of added histidine residues). Although the recombinant protein is not optimally soluble in buffers required for G protein activation, the fact that PKC increases the solubility in these conditions without affecting its activity (Fig.2-15) suggests another explanation may be required. Since RGS2 mRNA induction appears to increase in response to calcium, it is possible that a biological role of the RGS2 protein in T cells is modulation of G-protein-linked mechanisms controlling calcium flux, perhaps involving G_q (28). In this respect, it is also of interest to note that a gene responsible for hyperparathyroidism jaw tumor syndrome, a disease associated with hypercalcaemia, maps

to the q25-31 interval of chromosome 1 (86, 87; M. R. Hobbs, personal communication). This region contains *RGS1* and *RGS2* (44, 45).

Mitogenic lectins, such as Con-A, bind to T lymphocyte antigen receptors (88), and thus may simulate some aspects of antigen-lymphocyte interactions. These result in either negative selection (89) or positive selection (90) of lymphocyte sub-populations. The signal transduction pathways involved in these outcomes are likely to differ, with PMA and calcium ionophore appearing to simulate signalling required for positive, but not negative, selection (79). It is possible that distinctive RGS proteins are involved, not only in modulating B or T cell specific responses, but also in modulating the positive or negative aspects of lymphocyte activation (42). Defining the basic requirements for *RGS1* and *RGS2* induction by Con-A and other agents will be a necessary first step toward an understanding of their roles in these and other processes. Currently, the polyclonal antibodies raised against RGS1 and RGS2 identify a number of nonspecific species (or perhaps related RGS isoforms) in whole cell lysates making them poor probes for specific analyses of changes in levels or localization of these proteins. The development of specific monoclonal antibody probes for RGS1 and RGS2 proteins is currently underway (Blumer lab, Washington University) and should allow us to follow changes in protein levels, phosphorylation states, and cellular localization during lymphocyte activation events.

ACKNOWLEDGEMENTS

We thank L. Russell, and Q. Ye for technical help, and J. Kerhl for providing the RGS1/BL34 cDNA clone. M. Linder and K. Blumer provided reagents for G-protein binding

assays, and with D. Siderovski, reviewed the manuscript. A. Jheon, C. Law, R. Magun, R. S. Milligan, M. Schneider and C-Y. Tso assisted in the preparation of RT-PCR control constructs as part of undergraduate projects. The work was supported by the Medical Research Council of Canada, and Queen's University. S. P. Heximer was in receipt of a studentship from the National Health Research and Development Program, Department of Health and Welfare, Canada.

CHAPTER 3

TITLE: A Human Putative Lymphocyte G_0/G_1 Switch Gene Homologous to a Rodent Gene Encoding a Zinc-Binding Nuclear Factor

PREFACE

This chapter focusses on the characterization of a second gene putatively involved in the G_0/G_1 switch, *G0S24*. This work was submitted for publication to in February 1992, and after revision, it was published in January 1993. Obtaining sequence data for this cDNA and gene by manual methods was extremely difficult, probably as a result of the GC-rich nature of the region of the genome containing this gene (see Figure 3-4). To date, this paper contains the only reported human genomic sequence for this gene (GenBank accession number M92844), including data for the intron and 3'-flanking regions. The potential significance of these regions in the regulation of *G0S24* expression is discussed more fully in Chapter 4 (6). During the course of our work, several groups working independently identified the murine homologue TIS11/Nup475/ TTP and its family members TIS11B and TIS11D (for summary see Table 3-1), as immediate-early response genes from a wide variety of systems. These data suggested the universal importance of this gene in G_1 entry. Although it was known that there were at least three G0S24-related proteins, our paper is the first to characterize domains of amino acid similarity between the different family members. For presentation as part of this thesis chapter, unpublished data on the characterization of an unusual chimeric *G0S24*-related cDNA and also on the expression and preliminary functional characterization of the recombinant G0S24 protein have been included.

ABSTRACT

G0S24 is a member of a set of genes (putative G_0/G_1 switch regulatory genes) whose mRNA levels are rapidly increased following the addition of lectin or cycloheximide to human blood mononuclear cells. Comparison of a full-length cDNA sequence with the corresponding genomic sequence reveals an open reading frame which is distributed across two exons and encodes a putative 326 amino acid polypeptide of M_r 34 kDa and pI 8.4. Comparison of the derived protein sequence with those of rodent homologs permits identification of sequence features allowing classification of mammalian G0S24-like genes into three groups. Group 1 contains *G0S24* and the rat and mouse *TIS11* genes (also known as *TTP, Nup475*, and *Zfp36*). Members of this group are distinguished by three tetraproline repeats. Groups 1 and 2 have a serine-rich region and an "arginine element" (RRLPIF) at the carboxyl terminus. In addition to a serine-phenylalanine "SFS" domain similar to part of the large subunit of eukaryotic RNA polymerase II, all groups contain two $CX_8CX_5CX_3H$ (CCCH) putative zinc finger DNA-binding domains. Protein sequences show similarities with those of a variety of proteins involved in transcription, suggesting that the *G0S24* product has a similar role. However, preliminary DNA mobility-shift assays using full-length recombinant G0S24 or zinc-finger proteins do not indicate a DNA- binding activity within these domains and are consistent with a different role for the CCCH repeats. The *G0S24* gene has potential binding sites for transcription factors in the 5' flank and intron; these include a serum response element. A CpG island in the 5' flank and first exon regions suggests expression in the germ line.

INTRODUCTION

The switch of human lymphocytes from the G_0 to the G_1 phase of the cell cycle is associated with transcriptional changes (32, 91) and a transient increase in mRNAs corresponding to various oncogenes, transcription regulators and cytokines (1, 11, 31). Similar immediate early gene products also are induced when quiescent rodent fibroblasts are activated by growth factors and other agents (12, 77). The association of the differential expression of these genes with a change in growth state suggests an active role in the process.

Among the immediate-early genes identified by differentially screening rodent fibroblast cDNA libraries is a gene encoding a basic, proline-rich nuclear protein with cysteine- and histidine-containing potential zinc-finger motifs (gene named *TIS11, TTP,* and *Nup475*; refs. 7, 14, 92). In 1990 DuBois *et al.*, reported a partial genomic sequence (5' flank and exon) for this murine gene which appears to belong to a gene family with at least three members in mammalian genomes (93, 94). Its expression has been rigorously studied in fibroblasts, epithelial cells, astrocytes, periphery and sensory neurons, and lymphocytes in response to treatment with a wide variety of mitogenic stimuli including TPA (92), epidermal growth factor (EGF; ref. 92), fibroblast growth factor (FGF; ref. 92), serum (7), granulocyte macrophage- colony stimulating factor (GM-CSF; ref. 94), insulin (14), interleukin-6 (IL-6; ref. 95), and lymphocyte inhibitory factor (LIF; ref. 96). We have studied a related gene, *G0S24*, the cDNA of which was identified by differential screening human lymphocyte cDNA libraries (1, 31, 34). We present here an analysis of the original cDNA isolates (including an anomalous G0S24-associated sequence), full-length cDNA sequence, and the

complete sequence of the human gene and 5' flank. We have deduced the exon-intron structure and have compared human and rodent protein and DNA sequences with sequences in GenBank (69) and in the Transcription Factor Database (97). A human partial cDNA sequence agreeing closely with ours, has recently been reported (98).Toward characterization of G0S24 protein function, we have assessed the putative DNA-binding activities of full-length protein and zinc finger-like (CCCH) domains.

MATERIALS AND METHODS

Methods for cell purification and culture conditions were as described previously (1, 3; see Chapter 2). When appropriate, large-scale cultures (2.5 x 10^7 cells) were treated with Concanavalin A (200 μg/ml; Pharmacia) and cycloheximide (100 μM) for 2 hr while cultured in a humidified atmosphere of air-CO_2 (95:5).

Preparation of total RNA using CsCl gradient fractionation

Total RNA was prepared by a modification of the guanidine thiocyanate/cesium chloride (GuSCN/CsCl) procedure as described previously (Forsdyke, 1984). Briefly, cultured human peripheral blood mononuclear cells PBMCs in culture (5 x 10^7) were pelleted (600 x g, 4 min), and dissolved by a 5 min vortex in lysis buffer containing 50 mM NaAc, pH 4.5, 4 M GuSCN, 7 % (v/v) 2-mercaptoethanol, and 1 mM EDTA. Following addition of 50 μL n-butanol as an anti-foaming agent, 2.5 g of CsCl were added and the solution vortexed 5 min. The resulting solution was underlayed with 1.55 ml CsCl (ρ 1.8) in a polyallomer tube and spun at 36 x K r.p.m. for 40 hr at 22°C, in a Beckman SW50.1 rotor. Tubes were pierced at

the bottom and the region containing RNA was collected (routinely drops 161-460). This was diluted with 3 volumes of dpc-water, and precipitated with EtOH. To remove excess CsCl, samples were repeatedly (at least 3 times) resuspended in dpc-treated H_2O containing 0.5 M NH_4Ac and reprecipitated with EtOH. The final RNA pellets were washed in 80% EtOH, *briefly* air-dried, resuspended in dpc-H_2O quantitated by spectrophotometry, and stored at -70°C.

Blotting of total RNA fractionated by agarose gel electrophoresis

Samples (20 μg) of total RNA were denatured in a glyoxylation cocktail (1 M glyoxal, 50 % DMSO, 0.01 M $NaPO_4$, pH 7), electrophoresed through 1 % agarose gels containing 0.01 M $NaPO_4$ and 10 mM iodoacetamide, and capillary-transferred to Zetaprobe-GT nylon membrane using 10 x SSC. Following a brief deglyoxylation in 0.05 M NaOH, moist membranes were irradiated 3 min, from a distance of 15 cm with a UVG-54 short-wave lamp to cross-link RNA. Uniformity of transfer of rRNA species was verified by inspection of nylon-fluorescence shadowing. Membranes were heated 80°C for 20 min and stored in sealed bags until use.

Blotting of genomic DNA fractionated by agarose gel electrophoresis

Lymphocyte DNA fractions from CsCl gradients (99) were used for DNA preparation. DNA (5 μg) was digested with 70 U of *Xba* I or *Sst* I in the appropriate digestion buffer (Gibco/BRL, Life Technologies, Gaithersburg, MD) at 37°C for 2 hr. Samples were

electrophoresed through 0.6% agarose in 1 x TAE at 35 V for 17 hr. Alkaline transfer and hybridization conditions were as described previously (Chapter 2).

Genomic library screening

A genomic library from Strategene (La Jolla, CA) contained human placental DNA inserted into the *Xho* I site of the λ FIX II vector. Two positive clones with inserts of 15.1 and 17.2 kb were identified among 500,000 plaques screened for hybridization to a ^{32}P-labelled *G0S24* 3'-end cDNA fragment. Other methods were as described in Siderovski *et al.*, (1994, ref. 3) and in Chapter 2.

Mapping the genomic region containing the G0S24 gene

Restriction maps of the genomic region containing the *G0S24* gene were constructed using restriction digest analysis of two overlapping λ clones containing sequences which hybridize to *G0S24* cDNA (for description of specific methods see Chapter 2). Maps were confirmed using the λ quick map kit for cohesive-end labelling and analysis of partially digested λ clones as described for the *RGS2/G0S8* gene in Chapter 2.

Sequencing of genomic and cDNA fragments

Genomic DNA fragments containing *G0S24* sequence were excised from λ clones, inserted into phage M13 (both orientations) and sequenced by the dideoxynucleotide chain termination method using [α-^{35}S] dATP (DuPont-New England Nuclear, Pointe Claire, PQ)

and Sequenase (United States Biochemical Corp., Cleveland, OH). Ambiguities within GC-rich regions of this sequence were checked using dITP in place of dGTP (63). PCR-generated cDNA and 5'-end RACE clones in pBR322 were sequenced directly on both strands using similar methods.

RACE protocol for 5'-end cDNA sequence

The protocol of Frohman *et al.* (1988, ref. 100) was modified to complete our sequencing of *G0S24* cDNA. First strand cDNA was generated from 1 μg of total RNA using a *G0S24*-specific antisense primer (5'-CAGTCACTTTGTCACTC-3') located near the predicted 5' end of the cDNA. Primer (20 pmol) and RNA were heated together at 65°C for 10 min, and cooled on ice. RNA was reverse transcribed at 42°C for 90 min using 100 U MMLV reverse transcriptase (Life Technologies Inc.) in first strand synthesis buffer containing 50 mM Tris-Cl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 1 mM of each dNTP and 15 U RNAsin ribonuclease inhibitor (Promega, Madison, WI). NaOH was added to 0.3 N and RNA hydrolysed for 1 hr at 37°C. Following neutralization, excess nucleotides and primer were removed by spin dialysis in a Centricon-100 microconcentrator (Amicon Inc., Beverly, MA). 3'-end polyadenylate tailing of the first strand was carried out at 37°C for 15 min using 15 U of terminal deoxyribonucleotidyl transferase in A-tailing buffer (0.1 M K cacodylate, pH 7.2, 2 mM $CoCl_2$, 0.2 mM DTT, 1 mM dATP). Reactions were stopped (65°C, 10 min), and tailing buffer was exchanged for 10 mM Tris-Cl, pH 8, 1 mM EDTA using spin dialysis. PCR was performed essentially as described previously using 25 pmol "anchor primer" (5'-GACTCGAGTCGACATCGT_{17}-3'), and 50 pmol each of an anchor-specific primer (5'-

GACTCGAGTCGACATCG-3') and a second internal *G0S24*-specific primer (5'-TGATGCTCTGGCGAAGC-3'). G-tailed PCR products were cloned into C-tailed pBR322 (Forsdyke, 1985) and both strands were sequenced.

Primer extension analysis

A 17 bp G0S24-specific antisense primer (5'-AGATGGCAGTCAGATCC-3') located near the 5' end of the mRNA was used for primer extension analysis. Primer (30 pmol) was suspended in kinase buffer (70 mM Tris-Cl, pH 7.4, 10 mM $MgCl_2$, 5 mM DTT) containing 20 μCi [γ^{32}P]-rATP (ICN Biochemicals Inc.) and 5'-end radiolabelled at 37°C for 60 min using 10 U T4 Polynucleotide kinase. Primer (33 fmol, 7 x 10^4 cpm) was added to 20 μg total RNA in hybridization buffer (80 % formamide, 40 mM PIPES, pH 6.4, 0.4 M NaCl, 1 mM EDTA) and incubated at 22°C for 17 hr. The mixture was brought to 0.3 M NaAc by addition of 3 M NaAc, pH 7, and precipitated with EtOH. Washed pellets were resuspended by vortexing in first strand synthesis buffer, and reverse transcription was carried out as described above. ^{32}P-labelled first strand products were loaded on a denaturing 8% polyacrylamide sequencing gel. For comparison, a lower strand genomic sequence product from the same primer was run on the same gel.

Expression of recombinant full-length G0S24 protein and zinc-finger domain peptide

A cDNA clone containing the complete *G0S24* coding sequences was constructed by inserting a 0.18 kb *Bst* EII/*Xma* III fragment from the 5' RACE cDNA clone into similarly digested double- stranded (RF) M13 phage DNA containing the complete *G0S24* genomic

sequence. A PCR fragment containing the complete coding sequences, with an *Nde* I site incorporated at the predicted site of translation initiation was cloned into *Nde* I/ *Bam* HI digested pET-19b vector (obtained from Novagen, Markham, ON). A clone for expression of a recombinant (CCCH domain-containing) polypeptide (amino acids 103-186) was constructed by PCR-mediated incorporation of an *Nde*I site (in the context of an in-frame methionine) and a stop codon (TAG) at the carboxyl-terminal boundary, and subsequent cloning into pET19b. For a brief summary of the construction of these clones, including diagrams, see Appendix A2 (Figure A2-2). All PCR-generated expression clones were completely sequenced on both strands (Core Facility, Department of Biochemistry, Queen's University). Expression in bacteria of $(His)_{10}$-tagged G0S24 protein was performed using the pET expression system (Novagen) and in Chapter 2.

Selective ^{35}S-labelling of recombinant G0S24 protein

To follow the levels of soluble expression (ratio of soluble vs. inclusion body-bound) of recombinant $(HIS)_{10}$-G0S24 protein, *E. coli* were metabolically labelled with ^{35}S-cysteine and ^{35}S-methionine. After growth to an $O.D._{600}$ of 0.4, aliquots (1-10 ml) of cells were gently pelleted by centrifugation (4 min, 600 x g). Pellets were washed in 1 volume of M9 minimal medium (0.4% glucose, 50 mM Na_2HPO_4, 22 mM KH_2PO_4, 9 mM NaCl, 18 mM NH_4Cl, and 2 mM $MgSO_4$), gently pelleted again, resuspended in 1 volume of M9 minimal medium containing 0.1% of each of 18 amino acids, excluding cysteine and methionine, and grown at 37°C for 3 hr before IPTG was added. Incubation was continued for a further 20 min when rifampicin (0.2 mg/ml), an inhibitor of *E. coli* RNA polymerase, was introduced into the

bacteria by a 10 min heat shock at 42°C. Cells were allowed to recover by incubation at 37°C for 30 min before 100 μCi/ml of [^{35}S]-Cys/Met mixture (TransLabel, DuPont-NEN) was added for a further incubation period of 15 min. Cells were then pelleted and the solubilized labelled proteins analysed by SDS/PAGE followed by Coomassie staining and autofluorography.

Purification of proteins using Ni^{+2}-chelate affinity chromatography

Soluble proteins were purified as previously described (see Methods Chapter 2). For purification of G0S24 protein from inclusion body-containing pellets, these were solubilized by sonication at 4°C in binding buffer containing 6M urea or guanidine hydrochloride (GuHCl) and then purified essentially as described for soluble proteins, except that all buffers contained the appropriate denaturant, and wash buffers contained lower concentrations of imidazole (20 mM). Denatured proteins were refolded quickly at 4°C, at low protein concentration, by dropwise addition of the protein solution into 20 volumes of binding buffer containing 10 mM 2-mercaptoethanol and 10 mM $ZnCl_2$. Soluble (refolded) protein was concentrated using nondenaturing Ni^{++}-chelate chromatography. Proteins were analysed using SDS/PAGE with Coomassie staining, or Western Blotting and ECL detection. Antisera raised against the full-length murine G0S24 homolog Nup475 was kindly provided by Dr. D. Nathans. As the zinc-finger domain could not be detected by G0S24-specific sera, its proper expression was confirmed by mass spectrophotometry as described in Chapter 2.

DNA-binding assays

Binding of proteins to a random library of double-stranded DNA fragments was performed essentially as described previously (Theisen and Bach, 1990). The sequence of the random oligonucleotide used to screen for DNA binding was 5'-CTAGGATCC$(N)_{15}$AGATCTGGG-3'. A complimentary oligonucleotide 5'-GGCCCAATTCCCAGATCT-3' was annealed and [α-^{32}P]dCTP was incorporated into the double-strand probe by Klenow extension. The reaction mixtures (25 µl) contained 1 ng (10^5 cpm) of ^{32}P-labelled DNA probe in binding buffer (20 mM HEPES, pH 7.5, 70 mM KCl, 5 mM $MgCl_2$, 100 µM $ZnCl_2$, 0.5 mM DTT, 12 % glycerol; Ragona *et al.*, 1991) with 0.1 mg/ml BSA and 2 µg poly(dI-dC). After incubation for 15 min at 30°C, protein-DNA complexes were separated on nondenaturing 8 % polyacrylamide gels, which were dried and analysed by autoradiography after 48 hr at -70°C with intensifying screens.

Sequence analysis

All DNAs were fully sequenced in both strands. Sequences were analyzed with software of the Genetics Computer Group (69) or Intelligenetics Corp (Mountainview, CA). Protein and nucleic acid sequences were aligned with the program GENALIGN using a deletion weight of 0.5 and a matching weight of 1.0. Minor modifications of the alignment were then carried out manually. Percentage identities between two aligned DNA sequences were calculated from the ratio of identical bases to the total length (bases and gaps) in the region of overlap.

Database searches

The ProSite database of functional protein motifs was searched with the QUEST program. Potential transcription factor binding sites were determined by screening the transcription factor database (97) and a personal data base using the programs FIND and FASTDB. Segments of sequences were screened using the programs FASTA and BLAST (101).

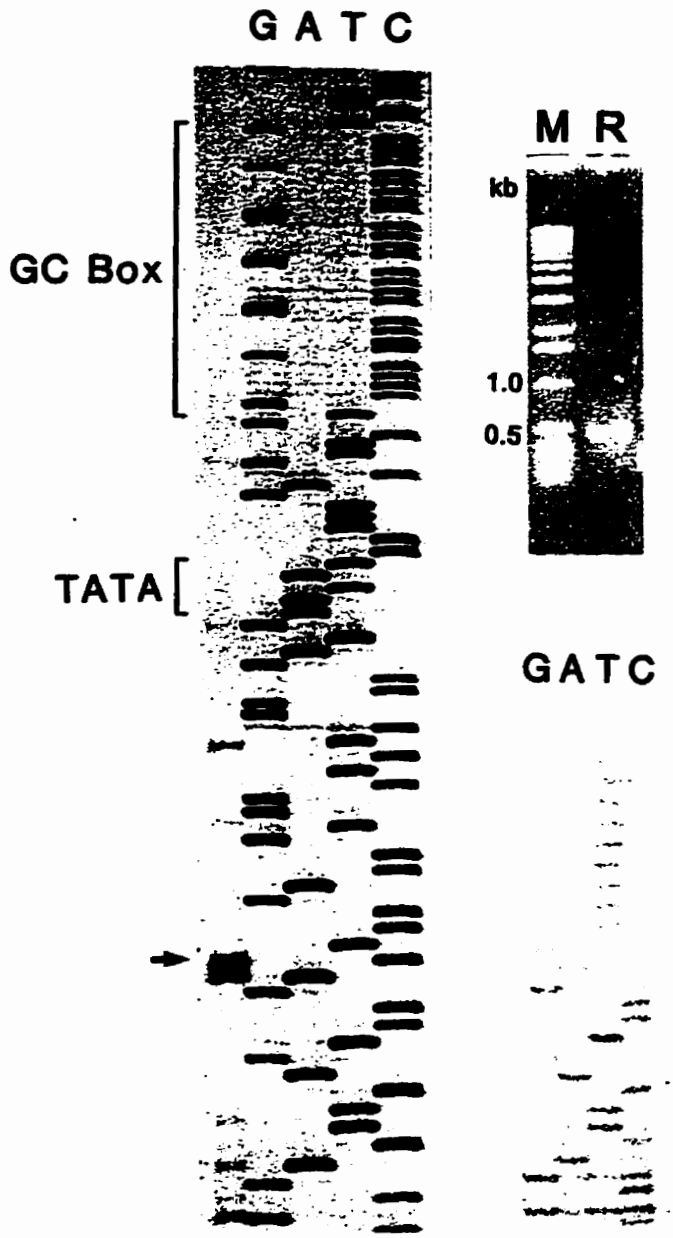
RESULTS

Identification of a full-length G0S24 cDNA

Previous studies in our laboratory showed that *G0S24* RNA is increased within 1-2 hr in human blood mononuclear cells cultured with either lectin or cycloheximide; there is then a decrease to undetectable levels by 8 hr (34). It was reported that when blots of RNA from cultures treated with lectin and cycloheximide were probed with the original small *G0S24* cDNA fragment (corresponding to 0.35 kb noncoding sequence at the 3' end of the mRNA), two bands of approximately 2.1 and 2.4 kb were noted, the upper band being cycloheximide dependent (1). Because the 3' noncoding region of the mRNA contained ATTTA repeats (associated with mRNA instability), it was suggested that cycloheximide stabilizes the mRNA and prevents shortening of the poly(A) tail. Thus, the 2.1 and 2.4 kb bands were considered likely to be the products of one gene.

By rescreening size-selected cDNA libraries with our original 3' *G0S24* cDNA probe, a number of longer cDNA clones were isolated. However, use of gene-specific RT-PCR primer extension and 5'-RACE (Fig. 3-1) were necessary to obtain the full-length sequence

Figure 3-1. **Determination of *G0S24* transcription start site by primer extension (Left) and 5' RACE (Right): Comparison with genomic and cDNA sequences.** Left. A ^{32}P-labelled 17-nucleotide primer corresponding to a sequence ending 62 bp downstream from the presumed mRNA start site was hybridized to total RNA prepared from cells cultured for 2 hr with lectin and cycloheximide. The primer was extended by reverse transcriptase in the presence of four deoxyribonucleotide triphosphates. After treatment with ribonuclease, the reaction mix was run on a standard sequencing gel (lane 1). The same primer was used for sequencing the corresponding genomic DNA (lanes 2-5) using ^{35}S-labelled dATP. (The autoradiograph is shown back-to-front to facilitate reading of the sequence.) Right. The 5'-end RACE product was amplified between the 3' (polyadenylated) end of the first strand product and a primer 450 nt downstream. Upper Panel. RACE product (R) electrophoresed and analysed compared to kilobase markers (M) on an EtBr-stained 0.8% agarose gel. Lower Panel. Sequence of the corresponding RACE product shown in alignment with genomic sequence used in primer extension analysis (Left). The cDNA begins after the poly(dT) tract, which was added by terminal deoxynucleotidyl transferase to the reverse transcriptase product.


G A T C
GC Box
TATA
M R
kb
1.0
0.5
G A T C

(deposited in GenBank; accession number M92843). The RACE product obtained in Fig. 3-1 (right, upper panel) was characteristic of this procedure, showing a distinct band beneath a long smear, consistent with repeated shortening of polyadenylated products during PCR amplification cycles. Comparison of sequence obtained from this band to the position of a characteristic doublet in the primer extension data indicated that the RACE-derived cDNA was full-length (1, 746 nucleotides). The cDNA sequence differed from that of a corresponding genomic clone at the 5' terminus (Fig. 3-1, left). A G residue was not present in this position in the genomic sequence, indicating that the cDNA sequencing had extended to the cap site (added post-transcriptionally). Further comparison with the genomic sequence (see later), indicated that the mRNA was derived by the joining of two exons. The sequence was confirmed by sequencing the exons of a genomic clone containing DNA from another individual (see later), but differed by two bases in the 3' UTR from a reported partial cDNA sequence (*HUMTTP*; Taylor *et al.*, 1991).

Three groups of G0S24*-related genes*

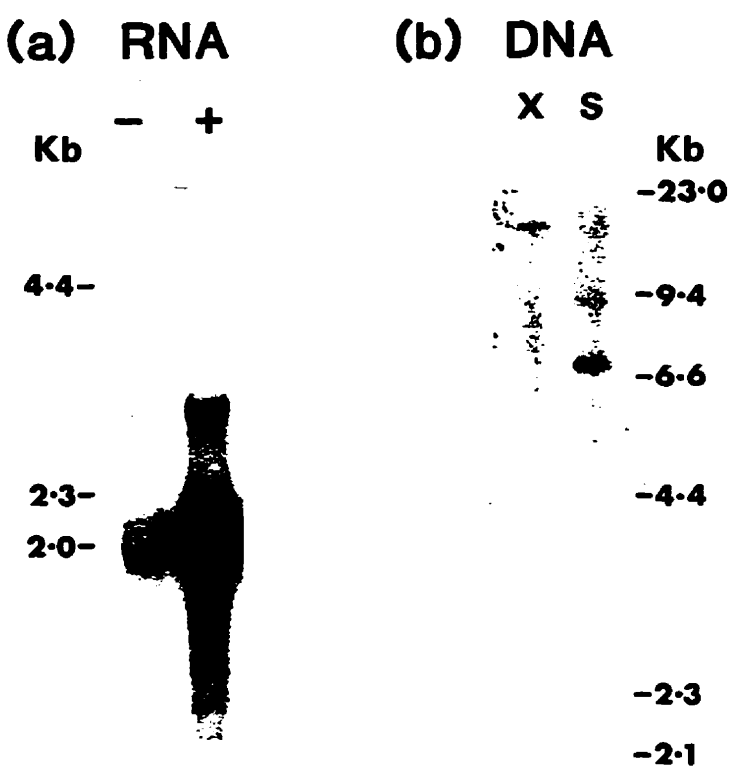
Studies in rodent systems have shown that genes related to *G0S24* fall into at least three groups (Table 3- 1). Genes in the first group, to which *G0S24* belongs, typically express an approximately 2.0 kb mRNA encoding a predicted protein of 319- to 326- amino acids. The second group corresponds to a 3.0 kb mRNA encoding a 338-amino acid protein. The third group has only a single member corresponding to a 367-amino acid protein (as of Sept 1992). To seek evidence for expression of other human family members during PBMC activation, a coding region-specific cDNA probe was hybridized to blots of total RNA. In addition to

Table 3-1. SUMMARY OF ALTERNATIVE NOMENCLATURES OF *GOS24*[a]-RELATED GENES

Group	Species	Original Name	GenBank name	RNA size[b] (kb)	Protein (amino acids)	Cell Type	Inducing Agent	Reference
1	Mouse	*TIS11*	MUSTS11[c]	2.2	319	Fibroblast	TPA/cycloheximide	(92, 94)
		TTP	MUSTTP45	2.0	319	Fibroblast	Insulin/cycloheximide	(14)
		Nup475	MUSNPGFI	1.8	319	Fibroblast	Serum/cycloheximide	(7)
	Rat	*TIS11*	RATTIS11	--	320	Pheochromocytoma	TPA/growth factors	(220)[d]
	Human	*GOS24*	HUMGOS24	2.1,2.4	326	Lymphocyte	Lectin/cycloheximide	(1, 34)
		TTP	HUMTTP	--	326	HeLa cell	--	(98)
2	Mouse	*TIS11b*	MUSTIS11B	--	338	Fibroblast	--	(94)
	Rat	*cMG1*	RATCMG1	3.0	338	Epithelial	EGF[e]/cycloheximide	(221)
3	Mouse	*TIS11d*	MSTIS11D	--	367	Fibroblast	--	(94)

[a]Assuming a probable zinc-finger structure, the Committee on Standardized Genetic Nomenclature for Mice has offered the systematic name *Zfp-36* for the mouse group 1 gene.
[b]Estimated by comparison of electrophoretic mobilities with standards.
[c]The original cDNA clone from which this sequence was derived was considered a cloning artifact (GenBank name *MUSTIS11*; (77).
[d]Kaneda *et al.* (1992). GenBank accession number X63369.
[e]EGF, Epidermal Growth Factor.

Figure 3-2. **RNA species and DNA restriction fragments that hybridize to a *G0S24* cDNA probe.** a. Human blood mononuclear cells (2 x 10^6/ml) were cultured for 2 hr in medium containing fresh autologous serum (16.7%) in the absence (-) or presence (+) of concanavalin-A (200μg/ml) and cycloheximide (100μM). Total RNA was prepared with guanidine thiocyanate. Samples (10 μg) were glyoxal-denatured, electrophoresed through a 1% agarose gel, and blotted to a nylon membrane. b. Human genomic DNA was digested with restriction enzymes X*ba* I (X) or S*st* I (S), electrophoresed through a 1% agarose gel, and blotted to a nylon membrane. The cDNA probe was prepared by PCR amplification of the sequence between nucleotide 60 and nucleotide 1,051 of the cDNA. The probe was labelled with ^{32}P to 10^6 cpm/ml.


(a) RNA
– +
Kb
4·4–
2·3–
2·0–
(b) DNA
X S
Kb
–23·0
–9·4
–6·6
–4·4
–2·3
–2·1

the main band noted above, an extra, but fainter band was noted corresponding to an RNA of approximately 3.0 kb (Fig. 3-2a). As with the main bands, the signal from this band was stronger with RNA from lectin- and cycloheximide-treated cultures. A similar-sized was found, after prolonged autoradiographic exposure, to hybridize with the original small 3' cDNA fragment as well as to an intron-specific probe (data not shown), indicating that it may be an incompletely spliced precursor of the major hybridizing (2.2 to 2.4 kb) species. Subsequent RT-PCR studies within the intron or across the intron/exon 2 boundary, however, have yet to identify a band which would support these results (see Chapter 4).

A coding region-specific probe was also hybridized to blots of restriction enzyme digests of genomic DNA from four subjects. In each case digestion with *Sst* I produced a strongly hybridizing band (6.7 kb) and a weakly hybridizing band (9.2 kb; Fig. 3-2b). Assuming complete digestion with *Sst* I, and because the 6.7-kb fragment would include both exons (see restriction map below), it is possible that the 9.2-kb band corresponds to another member of the G0S24 gene family. By contrast, the *Xba* I digest produced only one band (18 kb) and this should also contain both exons. The previously reported single bands with *Pst* I (6 kb) and *Hind* III (20 kb; Siderovski *et al.*, 1990) were confirmed. In addition fainter bands were noted (22 kb, *Pst* I; 6.6 kb, *Hind* III). These data raised the question of whether another *G0S24*-related cDNA clone identified in our size-selected library might correspond to an mRNA from another family member.

Characterization of a G0S24-associated fragment (G0S24AF)

Among the longer cDNA isolates which hybridized to the *G0S24* probe was an anomalous 1.1 kb clone containing 0.4 kb of sequence corresponding to the 3'UTR of *G0S24* and 0.7 kb of sequence which did not match that of *G0S24* cDNA or genomic clones. "Upper strand" sequence (relative to G0S24 mRNA) of the 0.7 kb region contained a 384 nt ORF corresponding to a predicted protein of 128 aa containing two in-frame methionines with poor translation initiation consensus (Figure 3-3a, ref. 102). Thus, we wondered whether this clone represented the 3' UTR of a *G0S24* family member or alternatively a *G0S24* pseudogene.

Strikingly, the "lower strand"sequence from the 0.7 kb region of *G0S24AF* contained an unfinished (no translation termination signal) 624 nt ORF which corresponded to a predicted polypeptide of at least 208 amino acids (Fig. 3-3a). Since the original library was made by annealing oligo dG-tailed cDNAs into oligo dC-tailed vector (33), the absence of a homopolymeric G/C stretch suggested the possibility that this sequence could not be explained simply as a cDNA cloning artifact. Fig. 3-3a shows the location of various primers used in the RT-PCR analysis of this novel clone. First strand products were made from a *G0S24* specific antisense primer (primer 1, Fig. 3-3a) within the 3' UTR at a sequence common to *G0S24* and *G0S24AF*. To search for cellular mRNAs corresponding to the *G0S24* cDNA, genomic (primers 1 and 2) and *G0S24AF* (primers 1 and 3) clones, specific PCR primers were used to assay the first strand products made from primer 1. A PCR band corresponding to primers based on the genomic sequence could be consistently identified in RNA from human blood mononuclear cells. The corresponding region of *G0S24AF*,

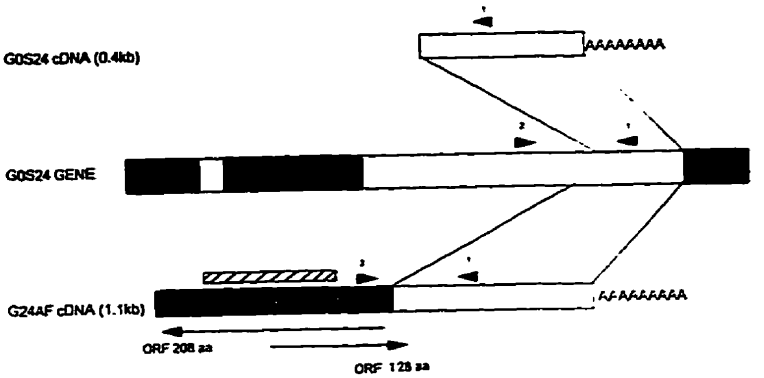
however, could not be amplified from these samples (data not shown). When a probe made from the 0.7 kb upstream region of *G0S24AF* cDNA was hybridized to blots of digested genomic DNA, it identified at approximately 17 kb *Bam*HI fragment and 6.6 , 0.6, and 0.5 kb *Xba*I fragments (Fig. 3-3b). As sequences hybridizing to *G0S24* cDNA are contained within a single approximately 18 kb *Xba*I fragment (Fig. 3-2b), the 0.7 kb sequence in the *G0S24AF* clone appeared to derive from a genomic region different from that of the *G0S24* cDNA.

G0S24-associated "lower strand" sequence corresponds to mRNA species expressed in fetal brain tissue.

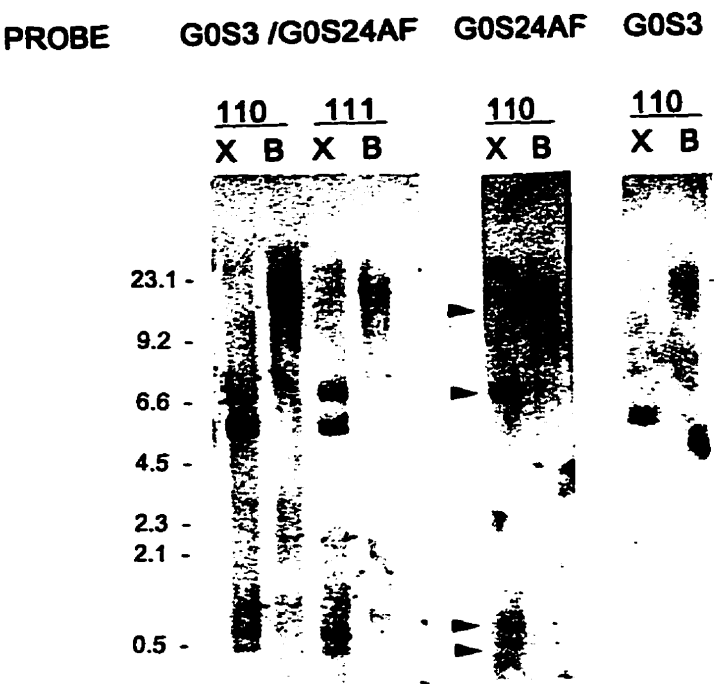
This sequence contained several in frame methionines, one of which showed a strong match to the Kozak consensus (Fig. 3-3c). When first strand RT synthesis was made from an upper strand sequence-specific primer at the 5'-end of *G0S24AF* (Fig. 3-3a, primer 3), a PCR product (made between primers 1 and 2) which hybridized to *G0S24AF* sequence was identified (data not shown). These data suggested that this anomalous clone represented a fusion of a partial *G0S24* cDNA with another opposingly oriented cDNA. This notion was confirmed by a recent screen of the GenBank EST database which identified five sequences (GenBank accession numbers R17796, R36555, R17221, R43002, R51399) from human fetal brain tissue libraries with close sequence identity to the 0.7 kb region (103). Alignment of these contiguous sequences (Fig. 3-3c) with the *G0S24AF* inverse complement (*G0S24AF'*) allows the construction of a theoretical 1.3 kb cDNA which encodes a novel polypeptide of 168 or 189 aa (potential polymorphism near the translation stop site) with no

Figure 3-3. **Characterization and sequence analysis of a G0S24-associated RNA fragment (G0S24AF).** (a) Schematic representation of the original short *G0S24* cDNA clone, *G0S24* genomic DNA, and the *G0S24AF* cDNA clone. Regions showing sequence identity to the *G0S24* cDNA are shown in stippled light grey. Filled black regions indicate intronic and genomic flanking sequences. The stippled dark grey section represents the 0.7 kb anomalous region of *G0S24AF*. The hatched box denotes sequences used as probe for *G0S24AF*. Long arrows denote the start of the 128 and 208 amino acid ORFs identified on the upper and lower strands respectively. Short arrows show positions of primers used to characterize the *G0S24AF* sequence. (b) DNA blots of *Bam* HI (B)- and *Xba* I (X)-digested human genomic DNA from two independent donors were probed either with the combination (left) or separated (right) of *G0S24AF* and FOSB/G0S3 probes. The blot from a single donor (#110) was stripped and successively hybridized with each probe separately. Details are as in legend of Fig. 3-2. (c) Multiple sequence alignment of lower strand sequence *G0S24AF* and five overlapping human EST clones (103). Alignments were done manually to bias consensus to the *G0S24AF* sequence which was rigorously sequenced on both strands. The predicted protein, in single letter amino acid code, corresponding to the longest open reading frame is shown above the sequence with the two potential translation (ATG) start site consensus (102) shown as shaded sequence. Also shaded are the alignment between the lower strand sequence (*G0S24AF'*) and the fetal brain cDNA 5'end clone (R17796) in the region of the joint between these two RNA species. Sequences corresponding to the 0.4 kb region (G0S24-identity) of the *G0S24AF'* sequence are not shown.

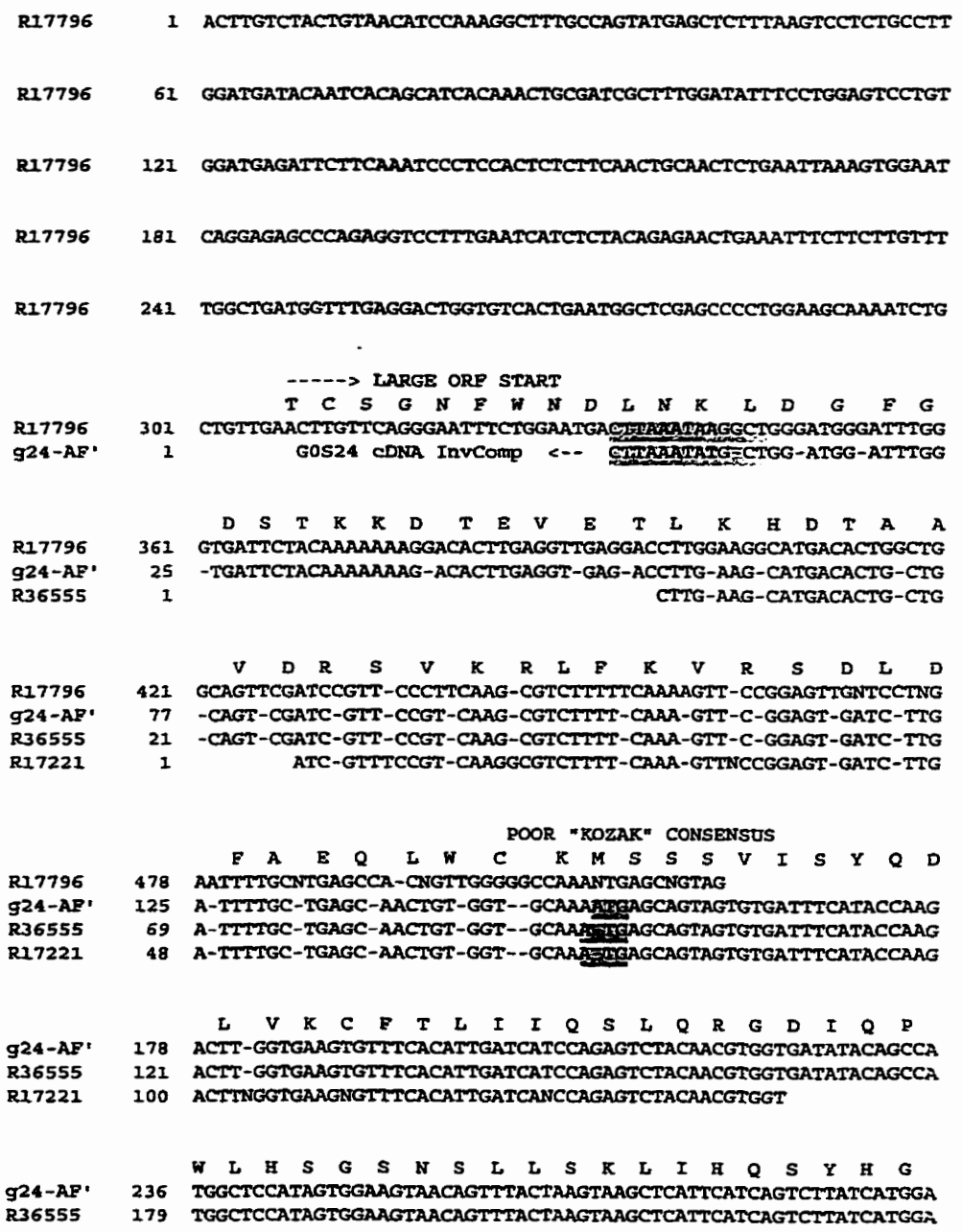
(a)


G0S24 cDNA (0.4kb)
AAAAAAAA
G0S24 GENE
G24AF cDNA (1.1kb)
AAAAAAAA
ORF 208 aa
ORF 128 aa

(b)


PROBE
G0S3 /G0S24AF
G0S24AF
G0S3
110
111
110
110
X B X B
X B
X B
23.1 -
9.2 -
6.6 -
4.5 -
2.3 -
2.1 -
0.5 -

(c)


R17796 1 ACTTGTCTACTGTAACATCCAAAGGCTTTGCCAGTATGAGCTCTTTAAGTCCTCTGCCTT
R17796 61 GGATGATACAATCACAGCATCACAAACTGCGATCGCTTTGGATATTTCCTGGAGTCCTGT
R17796 121 GGATGAGATTCTTCAAATCCCTCCACTCTCTTCAACTGCAACTCTGAATTAAAGTGGAAT
R17796 181 CAGGAGAGCCCAGAGGTCCTTTGAATCATCTCTACAGAGAACTGAAATTTCTTCTTGTTT
R17796 241 TGGCTGATGGTTTGAGGACTGGTGTCACTGAATGGCTCGAGCCCCTGGAAGCAAAATCTG
-----> LARGE ORF START
T C S G N F W N D L N K L D G F G
R17796 301 CTGTTGAACTTGTTCAGGGAATTTCTGGAATGAC
g24-AF' 1 G0S24 cDNA InvComp <-- CTGG-ATGG-ATTTGG
D S T K K D T E V E T L K H D T A A
R17796 361 GTGATTCTACAAAAAAAGGACACTTGAGGTTGAGGACCTTGGAAGGCATGACACTGGCTG
g24-AF' 25 -TGATTCTACAAAAAAAG-ACACTTGAGGT-GAG-ACCTTG-AAG-CATGACACTG-CTG
R36555 1 CTTG-AAG-CATGACACTG-CTG
V D R S V K R L F K V R S D L D
R17796 421 GCAGTTCGATCCGTT-CCCTTCAAG-CGTCTTTTTCAAAAGTT-CCGGAGTTGNTCCTNG
g24-AF' 77 -CAGT-CGATC-GTT-CCGT-CAAG-CGTCTTTT-CAAA-GTT-C-GGAGT-GATC-TTG
R36555 21 -CAGT-CGATC-GTT-CCGT-CAAG-CGTCTTTT-CAAA-GTT-C-GGAGT-GATC-TTG
R17221 1 ATC-GTTTCCGT-CAAGGCGTCTTTT-CAAA-GTTNCCGGAGT-GATC-TTG
POOR "KOZAK" CONSENSUS
F A E Q L W C K M S S S V I S Y Q D
R17796 478 AATTTTGCNTGAGCCA-CNGTTGGGGGCCAAANTGAGCNGTAG
g24-AF' 125 A-TTTTGC-TGAGC-AACTGT-GGT--GCAAAATGAGCAGTAGTGTGATTTCATACCAAG
R36555 69 A-TTTTGC-TGAGC-AACTGT-GGT--GCAAAATGAGCAGTAGTGTGATTTCATACCAAG
R17221 48 A-TTTTGC-TGAGC-AACTGT-GGT--GCAAAATGAGCAGTAGTGTGATTTCATACCAAG
L V K C F T L I I Q S L Q R G D I Q P
g24-AF' 178 ACTT-GGTGAAGTGTTTCACATTGATCATCCAGAGTCTACAACGTGGTGATATACAGCCA
R36555 121 ACTT-GGTGAAGTGTTTCACATTGATCATCCAGAGTCTACAACGTGGTGATATACAGCCA
R17221 100 ACTTNGGTGAAGNGTTTCACATTGATCANCCAGAGTCTACAACGTGGT
W L H S G S N S L L S K L I H Q S Y H G
g24-AF' 236 TGGCTCCATAGTGGAAGTAACAGTTTACTAAGTAAGCTCATTCATCAGTCTTATCATGGA
R36555 179 TGGCTCCATAGTGGAAGTAACAGTTTACTAAGTAAGCTCATTCATCAGTCTTATCATGGA

```
                GOOD "KOZAK" CONSENSUS
                T  M  D  T  V  S  L  S  G  T  I  P  V  Q  M  L  L  E  I  G
g24-AF'   296 ACCATGGACACAGTTTCTCTCAGTGGGACTATTCCAGTTCAAATGCTTTTGGAAATTGGT
R36555    239 ACCATGGACACAGTTTCTCTCAGTGGGACTATTCCAGTTCAAATGCTTTTGGAAATTGGT

                L  D  K  L  K  K  D  Y  I  S  F  F  I  G  Q  E     L  A     S  L
g24-AF'   356 TTGGACAAACTAAAGAAAGATTATATCAGTTTTTTCATAGGTCAGG-AACTTG-CATCTT
R36555    299 TTGGACAAACTAAAGAAAGATTATATCAGTTTTTTCATAGGTCAGGGAACTTGGCATCTT

                   N  H     L  E  Y  F     I  A  P  S  V  D  I  Q  E  Q  V  Y  R
g24-AF'   413 TGAATC-ATTTGGAATACTTC-ATTGCTCCATCAGTAGATATACAAGAACAGGTTTATCG
R36555    359 TGAATCCATTTGGAATACTTCCATTGCTCCATCA
R43002      1                                          TTATACAAGAAC-GGTTTATCG
R51399      1                                       AGTAGATATACAAGAACAGGTTTATCG

                V  Q  K  L  H     H  I  L     E  I  L  V  S  C  M  P     F     I  K
g24-AF'   470 TGTCCAAAAACTCC-ACCATATTC-TAGAAATATTAGTCAGTTGCATGCC-TTTC-ATTA
R43002     22 TGTCCAAAAACTCCCACCATAT-CCTAGAAATATTAGTCAGTTGCATGCCCTTTCCATTA
R51399     28 TGTCCAAAAACTCC-ACCATATTC-TAGAAATATTAGTCAGTTGCATGCC-TTTC-ATTA


                   S  Q  H  E  L  L  F  S  L  T  Q  I  C  I  K  Y  Y  K  Q  N
g24-AF'   525 AATCTCAACATGAACTCCTCTTTTCTTTAACACAGATCTGCATAAAGTATTACAAACAAA
R43002     81 AATCTCAACATGAACTCCTCTTTTCTTTAACACAGATCTGCATAAAGTATTACAAACAAA
R51399     84 AATCTCAACATGAACTCCTCTTTTCTTTAACACAGATCTGCATAAAGTATTACAAACAAA

                   P  L  D  E  Q  H  I  F  Q  L  P     V  R  P  T  A  V  K  N
g24-AF'   585 ATCCTCTTGATGAGCAACACATTTTTCAGCTGCC-AGTCAGA
R43002    141 ATCCTCTTGATGAGCAACACATTTTTCAGCTGCCCAGTCAGACCAACTGCTGTAAAGAAC
R51399    144 ATCCTCTTGATGAGCAACACATTTTTCAGCTGCC-AGTCAGACCAACTGCTGTAAAGAAC

                L  Y  Q  S  E  K  P  Q  K  W  R  V  E  I  Y  S  G  Q  K  K  I
R43002    201 TTATATCAAAGTGAGAAGCCACAGAAATGGAGAGTGGAAATATATAGTGGTCAAAAGAAGA
R51399    203 TTATATCAAAGTGAGAAGCCACAGAAATGGAGAGTGGAAATATATAGTGGTCAAAAGAAGA

                   K  T  V  W  Q  L  S  D  S  S  P  I  D  H  L  N  F  H  K  P
R43002    261 TTAAGACAGTTTGGCAACTGAGTGACAGCTCACCCATAGACCATCTGAATTTTCACAAAC
R51399    263 TTAAGACAGTTTGGCAACTGAGTGACAGCTCACCCATAGACCATCTGAATTTTCACAAAC

                         G  I  N  T  K  R
                   D  F  S  E     L  T  L  N  G  S  L  E  E  R  I  F  F  T  N
R43002    321 CTGATTTTTCGG-AATTAACACTAAACGGTAGCCTGGAAGAAAGGATATTCTTTACTAAC
R51399    323 CTGATTTTTCGGGAATTAACACTAAACGGTAG

                M  V  T  C  S  H  V  H  F  K
R43002    360 ATGGTTACCTGCAGCCAGGTGCATTTCAAGTGAAGTGTGCTGATGAAGTCCTCTATAAGC


R43002    440 ACAAGCCAAAAAAAAAAAAAAA.
```

significant similarity to previously identified proteins (BLAST search, August 1996). The 12 nt (5'-CTTAAATATGCT-3') which make up the joint region of *G0S24AF* (at the 5'-most end of the 0.4 kb *G0S24*-like sequence) show identity in 11 out of 12 nt to the corresponding region of clone R17796. There is also an extra G present in the EST within the region of overlap, however we attribute this to a stuttering effect found at numerous locations along the length of this sequence. As we could identify no precedent for creation of such a species *in vivo,* these data suggest that the two independent cDNA sequences were associated on the basis of the complementarity of the corresponding regions of the first strand products and thus were generated by recombination between two plasmids in *E. coli* or at the second strand synthesis step during preparation of the cDNA library.

High conservation of mammalian G0S24-related group 1 mRNA sequences

Of two suitably located potential protein start codons, the first was absent from the group 1 mouse sequence and the latter corresponded more closely to the consensus start site (102). Thus, the major open reading frame is considered to encode a protein of 326 amino acids. The 3' terminus of the mRNA sequence (nucleotides 1,531-1,746) was UA rich and contained three AUUUA mRNA potential instability elements (1). Group 1 sequences (human, mouse, and rat) showed high conservation in both coding and noncoding regions (alignment not shown). The conservation between the two rodent sequences was highest. Similarities between the human and mouse sequences were 67% in the 5' noncoding region, 85% in the coding region, 53% in the main part of the 3' noncoding region, and 81% in the UA-rich terminal part of the 3' noncoding region.

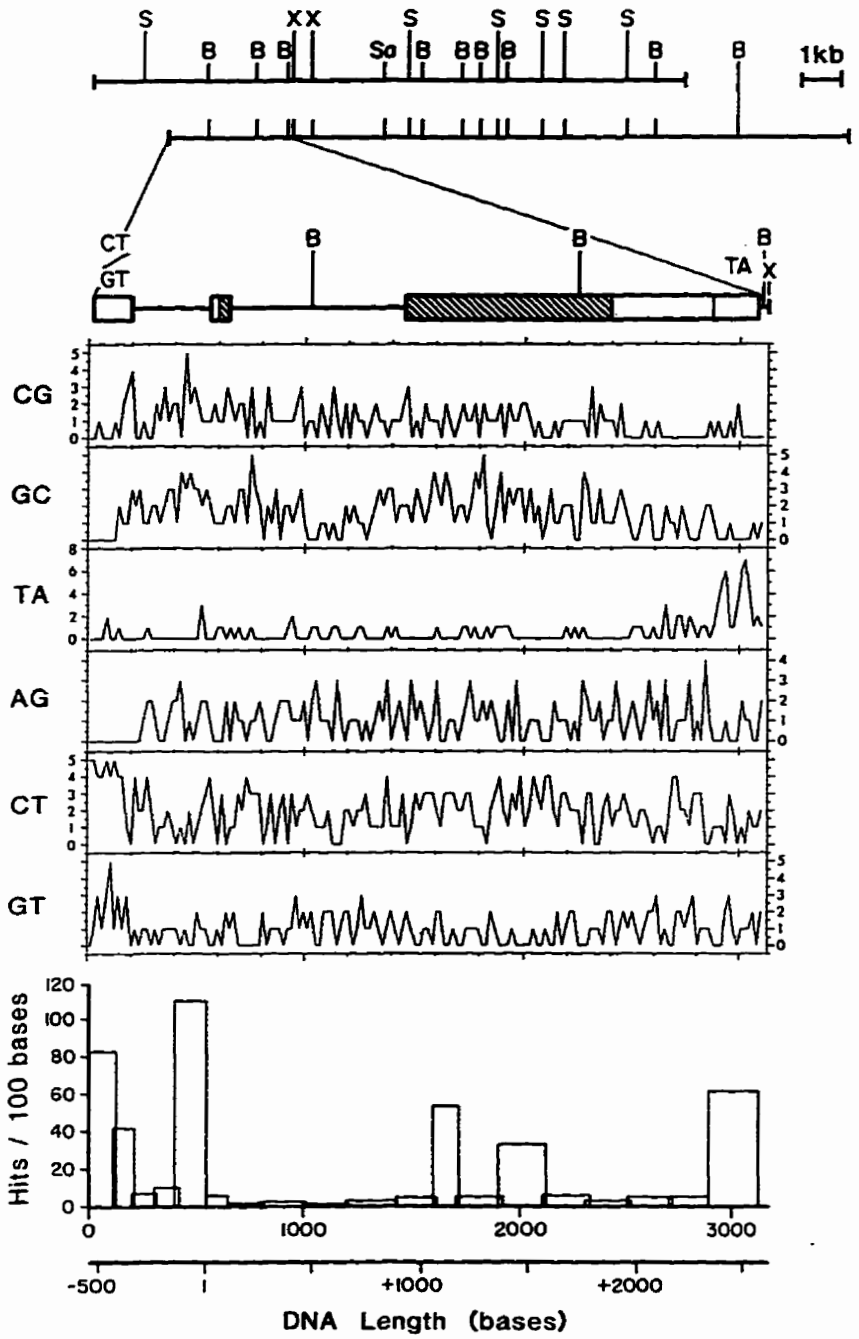
CpG island-containing gene

Fig. 3-4 shows some general features of the gene. Restriction maps were made of two overlapping genomic clones that hybridized with *G0S24* cDNA (see Chapter 2 for specific mapping methods). A 3.1-kb *Xba* I fragment generated by cleavages in the vector cloning site and an internal *Xba* I site was found to contain the entire gene and flanking regions. In the extreme 5' flank was a region enriched in the dinucleotides CpT and GpT. The rest of the 5' flank was CpG-rich; this CpG island extended across the first exon, the intron, and the coding region of the second exon. Correlating with this there was a high GpC frequency (ref. 74 ; a CpG island is also present in the mouse group 1 homolog). The entire sequence was depleted in T and A and the corresponding dinucleotides, except for the TA-rich region at the 3' end of the second exon.

General elements shared with other genes

Sections of the *G0S24* sequence were examined for similarities with other sequences in the GenBank data base. The BLAST search program (101) was used to screen a series of windows of approximately 200 nucleotides. Scores exceeding a cut-off value of 100 (implying relatively high similarities) were abundant in five regions (Fig. 3-4; lower). The first of these correlated with the 5' region rich in CpT and GpT dinucleotides; this reflected the dispersion of regions containing these dinucleotides in mammalian sequences. Thus, 91% of the "hits" were in mammalian sequences (24% of these were human). A second peak occurred in the 5' flank immediately before the first exon and again reflected predominately mammalian sequences, 51% of which were human. In this case a major element

Figure 3-4. **Major features of the *G0S24* gene.** Upper. Restriction maps of two overlapping phage λ human genomic clones (B, *Bam* HI; S, *Sal* I; S, *Sst* I; X, *Xba* I). The 5' 3,135 bases (*Xba* I site to the end) were sequenced. The 5' open box refers to a region containing repeats of the dinucleotides CpT and GpT. Other boxes refer to exons. Striped boxes refer to coding regions. The open box at the 3' end refers to the region rich in the bases T and A. Middle. Frequency of various dinucleotide sequences in different regions displayed using the programs Window and StatPlot (Genetic Computer Group; ref. 69). Readings were taken in successive windows of 20 nucleotides with a one nucleotide overlap. Lower. Frequency of "hits" (regions of sequence similarity above a cut-off score of 100) in a screen of *G0S24* against the GenBank database (consecutive windows overlapping by 20 nucleotides; (101). Since window sizes varied between 83 and 255 nucleotides, hits are expressed per 100 nucleotide bases.

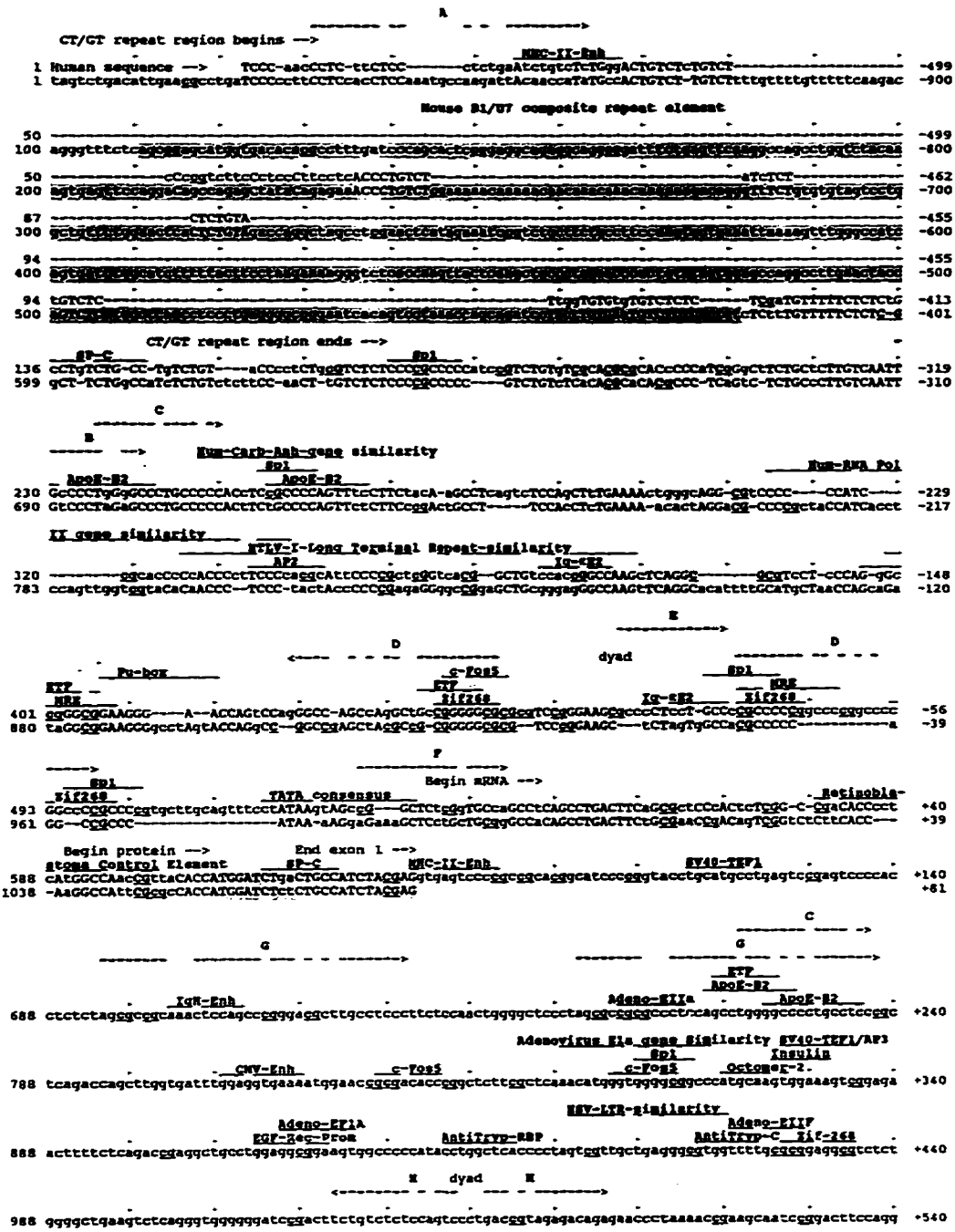

S
XX
S
S
S S
S
B B B B Sa B B B B B B 1kb
CT
GT
B
B
B
TA
X
CG
GC
TA
AG
CT
GT
Hits / 100 bases
120
100
80
60
40
20
0
0
1000
2000
3000
-500
1
+1000
+2000
DNA Length (bases)

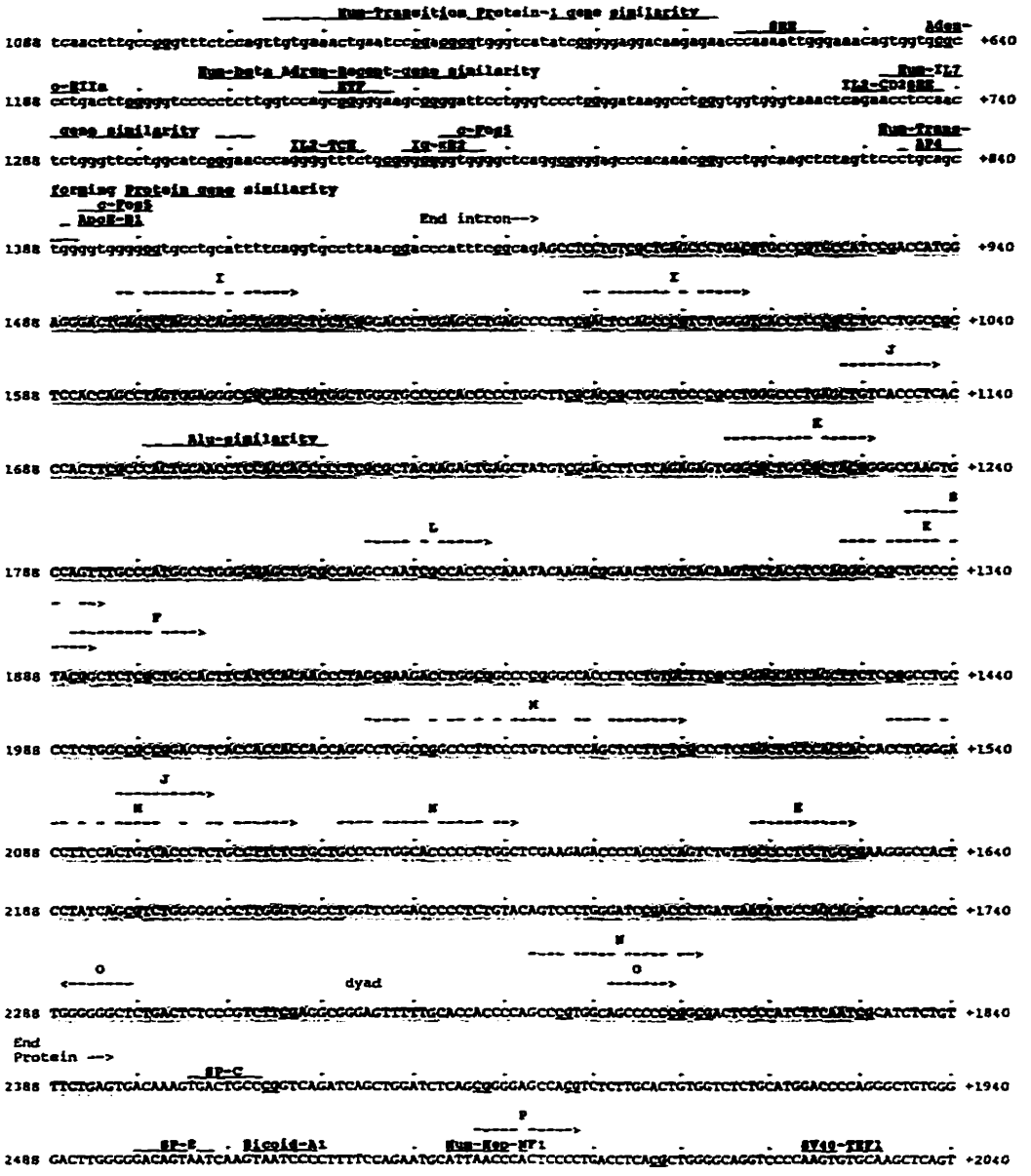
was the extensive CG-box region found in the immediate 5' flanks, 5' noncoding regions, and sometimes introns, of numerous genes, including various oncogenes and transcription factors (104). There were two peaks within the coding region. The first reflected predominantly mammalian sequences, 90% of which were human; this peak was caused by a 90% match with a 20-nucleotide section of the primate *Alu* repetative sequence. The second peak was localized in a region corresponding to the second two tetraproline repeats and reflected predominantly mammalian sequences, 33% of which were human. Some of these encoded proline-rich proteins. Other encoded various transcriptionally active proteins. The final peak corresponded to the TA-rich terminus of exon 2 and reflected mainly yeast mitochondrial TA-rich sequences.

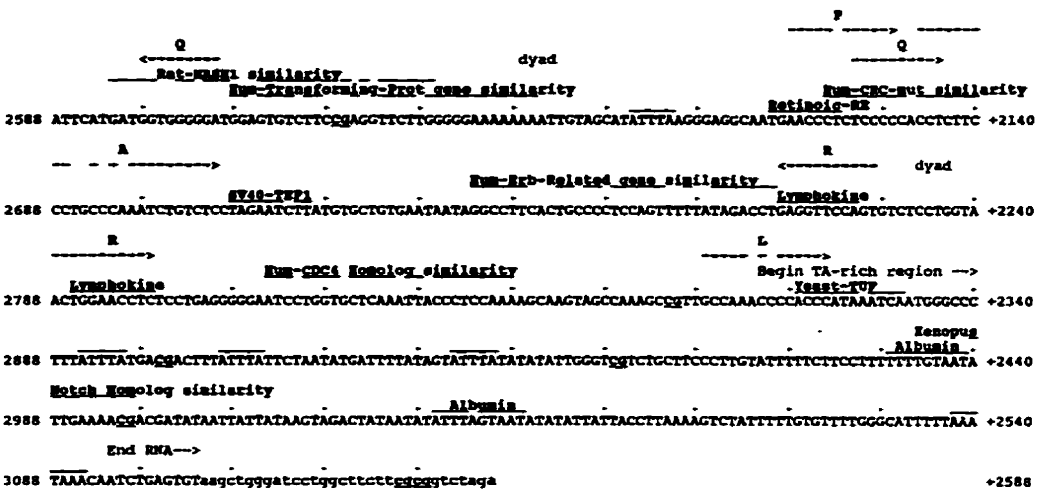
Analysis of genomic sequence

The only genomic sequence currently available from the family of *G0S24*-related genes (Table 3-1) is the 5' flank and first exon of the group 1 murine homolog, *Nup475* (accession number L42317). However, much of this sequence is occupied by a composite SINE element generated from B1 and U7 repetitive elements (105). Fig. 3-5 shows an alignment of the murine genomic sequence with the *G0S24* genomic sequence. Homology was quite high (60%) for at least 430 bases upstream from the mRNA start site and extended into the CT/GT repeat region. Although the human sequence had a TATA box sequence close to the consensus (75), this was not well conserved in the murine homolog. Just upstream from the TATA box was a region of dyad symmetry, containing Cs and Gs in combinations indicating a potential to be bound by transcription factors Sp1, ETF (106), and Zif268 (the mouse

Figure 3-5. **Sequence of the human *G0S24* gene and alignment with 5' flank and first exon of the murine homolog N*up*475 (Table 3-1).** The human sequence (GenBank accession number M92844) is aligned above the mouse sequence. Numbering is relative either to the first base (left) or to the transcription start site (right). The shaded region in the 5' flank refers to a mouse B1-U7 composite repeat element (105). Other shaded regions refer to parts of exons encoding protein. Lowercase lettering is used to indicate mismatches between the human and mouse sequences, and to indicate the sequences of the intron and the 3' flank. CpG dinucleotides are underlined. All other symbols are above the sequences to which they refer. References to regions of potential transcription factor sites (either orientation) are placed immediately above the sequence whenever possible. Regions of similarity with other genes are named as such and are separated from the sequence by one or more spaces. Repeats and dyads are represented as dashed arrows; members of pairs are identified by capital letters; mismatches are indicated by omission of dashes. Mismatches are also shown by gaps in the underlining beneath the abbreviated names of the potential transcription factor sites and of the gene similarities referred to. These are: MHC enhancer elements (MHC-II-Enh, 107), hepatitis B promoter elements (SP-C, SP-E, 108), transcription factor Sp1 (Sp1), B1 and B2 elements of the human apolipoprotein E gene (ApoE-B1, ApoE-B2, 109), 5' flank of a human carbonic anhydrase gene (Hum-Carb-AnH, 110), 3' non-coding region of human RNA polymerase II, large subunit mRNA (Hum-RNA Pol II, 111) human T-cell leukemia virus I long terminal repeat (HTL V-1), 112), transcription factor AP2 (113), enhancer element E2 of a murine immunoglobulin κ-chain gene (Ig-κ-E2, 114), element RS2 of the epidermal growth factor receptor gene (ETF, 106), human metallothionein gene response element (MRE, 115), human heat-shock purine-rich promoter element (Pu-box, 116), octomer repeat element of the C-*fos* oncogene (c-Fos5, 117), transcription factor encoded by the murine homolog of *G0S30* (Zif268, 12), TATA box consensus sequence (75), retinoblastoma control element (118), transcription factor TEF1 sites (SV40-TEF1, 119; SV40-TEF1/AP3, 120), adenovirus EIIa promoter element (Adeno-EIIa, 121), murine cytomegalovirus immediate-early enhancer (122), 5' flank of adenovirus-7 E1a gene (123), octamer B2 transcription factor (octamer-2, 124), insulin gene enhancer (insulin, 125), epidermal growth factor receptor gene promoter (EGF-Rec-Prom, 126), adenovirus Eia enhancer element (Adeno-EF1A, 127), human α1 antitrypsin gene elements (AntiTryp-RBP, AntiTryp-C, 128), herpes simplex virus long terminal repeat (HSV-LTR, 129), adenovirus EIIa enhancer element (Adeno-EIIF, 130), 5' flank of human transition protein-1 gene (131), serum response element (SRE, 12), 5' flank of human β-adrenergic receptor gene (132), interleukin-2 gene response elements (IL2-CD28RE, 133; IL2-TCE, 134), human transforming protein gene (5' flank and 3' noncoding region, 135), transcription factor AP4 (136), *Drosophila* homeodomain transcription factor (Bicoid-A1, 137), hepatic nuclear transcription factor 1 (Hum-Hep-NF1, 138), 3' noncoding region of rat homologs of *Drosophila* developmental genes (Rat-MASH1, 139), retinoic acid response element (140) 3' noncoding region of the human gene mutated in colorectal carcinoma (Hum-CRC-mut, 141), 3' noncoding region of human gene related to the *v-erb* oncogene (Hum-Erb-Related, 142), consensus sequence found in lymphokine genes (lymphokine, 143), 3' noncoding region of human homolog of yeast CDC4 gene (Hum-CDC4, 144), transcription factor for yeast ribosomal protein genes (Yeast-TUF, 145), albumin promoter element (146), and 3' noncoding region of *Xenopus* homolog of *notch* (147).


A
CT/GT repeat region begins -->
MHC-II-Enh
Human sequence -->
Mouse B1/B7 composite repeat element
CT/GT repeat region ends -->
SP-C
Sp1
B
Hum-Carb-Anh-gene similarity
ApoE-B2
Hum-RNA Pol
C
IF gene similarity
HTLV-I-Long Terminal Repeat-similarity
AP2
Ig-κE2
D
dyad
Pu-box
c-Fos5
ETF
NRE
Sif268
E
F
Begin mRNA -->
TATA consensus
Retinobla-
Begin protein -->
End exon 1 -->
gtgag Control Element
SV40-TEF1
G
IgH-Enh
Adeno-EIIa
Adenovirus Ela gene Similarity SV40-TEF1/AP3
CMV-Enh
Insulin
Octomer-2.
HBV-LTR-similarity
Adeno-EIIA
Adeno-EIIF
EGF-Reg-Prom
AntiTryp-RBP
AntiTryp-C
Sif-268
dyad


Rum-Transition Protein-1 gene similarity
SRE
Aden-
1088 tcaacttt qccgggtttctccagttgtgaaactgaatccggagggtgggtcatatcgggggaggacaagagaacccaaaattgggaaacagtgggtgggc +640
Rum-beta Adrem-Regent-gene similarity
Rum-ILZ
α-EIIa
RVZ
IL2-CD28RE
1188 cctgacttgggggtcccccctcttggtccagcgggggaagcggggattcctgggtccctgggataaggcctgggtgggtgggtaaactcagaacctccaac +740
gene similarity
c-Fos1
Rum-Trans-
IL2-TCR
Ig-HE2
AP4
1288 tctgggttcctggcatcgggaacccagggggtttctgggggggggtgggctcagggggggagcccacaaacgggcctggcaagctctagttccctgcagc +840
forming Protein gene similarity
c-Fos2
AngE-B1
End intron-->
1388 tgggggtgggggtgcctgcattttcaggtgcttaacggacccatttcggcagGCCCTCTGTGCTCTGCCCTGAGTGCCGTGCCATCCGACCATGG +940
I
I
1488 AGGGACCTGATGGCCCAGGCAGGTCTCCTGCACCCTGGAGCCTGAGCCCCTGGTCTCCATCCTGGGTCACCTCCTGGTGCCTGCCAC +1040
J
1588 TCCACCAGCCTGTGTGAGGGCGGTGCTGGTCCCCCACCCCCTGGCTTGCCGCTGGCTGCCCGCCTGGCCCTGTGTCACCCTCAC +1140
K
Alu-similarity
1688 CCACTTGCCCATGCCAACCTCTGCCTCCCTGCGCTACAGACTGAGCTATGTCGCCTTCTCAGGAGTGGCGGCTGTGGGGCCAAGTG +1240
S
L
K
1788 CCAGTTTGCCCATGCCTGGCGGCTGCCAGGCCAATCGCCACCCCAAATACAAGAGCAACTGTGTCACAAGTCACCTCCGGGCCCTGCCCC +1340
F
1888 TACGCTCTCTGCTGCCACTTCATGCCAACCTACGGAAGACCTGGCGCCCCGGCCACCCTCTGTGCTTGCCGTGTGTCTCCGCCCTGC +1440
N
1988 CCTCTGGCCGCGCCTCACCACCCCACCAGGCCTGGCGGCCCTTCCCTGTCCTCCAGCTCCTCTGCCCTCGAGTGCCACCTGGGGA +1540
J
K
K
E
2088 CCTTCCACTGTGCCCTCTGCTCTCTGCTGCCCTGGCACCCCCTGGCTCGAAGACACCCACCCCAGTCTGTTGCCCTCTGCGAAGGCCACT +1640
2188 CCTATCAGGTCTGGGGGCCCTGGGTGGCCTGGTCGGACCCCCTCTGTACAGTCCCTGGATCGACGTGATGAATATGCCGCGCGGCAGCAGCC +1740
M
O
O
dyad
2288 TGGGGGGCTCTGACTCTCCGTCTTGGAGGCGGGAGTTTTTGCACCACCCCAGCCCGTGGCAGCCCCCGGCGGACTCCCATCTTCAATGGCATCTCTGT +1840
End
Protein -->
SP-C
2388 TTCTGAGTGACAAAGTGACTGCCGGTCAGATCAGCTGGATCTCAGCGGGGAGCCAGGTCTCTTGCACTGTGGTCTCTGCATGGACCCCAGGGCTGTGGG +1940
P
SP-B
Bicoid-A1
Rum-Kap-NF1
SV40-TEF1
2488 GACTTGGGGGACAGTAATCAAGTAATCCCCTTTCCAGAATGCATTAACCCACTCCCCTGACCTCAGGCTGGGCCAGGTCCCCAAGTGTGCAAGCTCAGT +2040


Q
dyad
P
Q
Rat-MyoD1 similarity
Rvm-Transforming-Prot gene similarity
Rvm-CBC-mut similarity
Retinoic-RE
2588 ATTCATGATGGTGGGGGATGGAGTGTCTTCCGAGGTTCTTGGGGGAAAAAAAAATTGTAGCATATTTAAGGGAGGCAATGAACCCTCTCCCCCACCTCTTC +2140
A
R
dyad
Rvm-Erb-Related gene similarity
HV40-TRP1
Lymphokine
2688 CCTGCCCAAATCTGTCTCCTAGAATCTTATGTGCTGTGAATAATAGGCCTTCACTGCCCCTCCAGTTTTTATAGACCTGAGGTTCCAGTGTCTCCTGGTA +2240
R
L
Rvm-CDC4 Homolog similarity
Begin TA-rich region -->
Lymphokine
Yeast-TUP
2788 ACTGGAACCTCTCCTGAGGGGGAATCCTGGTGCTCAAATTACCCTCCAAAAGCAAGTAGCCAAAGCCGTTGCCAAACCCCACCCATAAATCAATGGGCCC +2340
Xenopus
Albumin
2888 TTTATTTATGACGACTTTATTTATTCTAATATGATTTTATAGTATTTATATATATTGGGTCGTCTGCTTCCCTTGTATTTTTCTTCCTTTTTTTGTAATA +2440
Notch Homolog similarity
Albumin
2988 TTGAAAACGACGATATAATTATTATAAGTAGACTATAATATATTTAGTAATATATATTATTACCTTAAAAGTCTATTTTTGTGTTTTGGGCATTTTTAAA +2540
End RNA-->
3088 TAAACAATCTGAGTGTaagctgggatcctgqcttctttgtgggtctaga +2588

homologue of G0S30, ref. 12). However, there was no obvious CCAAT box. Also upstream were a number of potential sites for transcription factor binding including an element putatively involved in the regulation of the metallothione in gene (MRE, ref. 115).

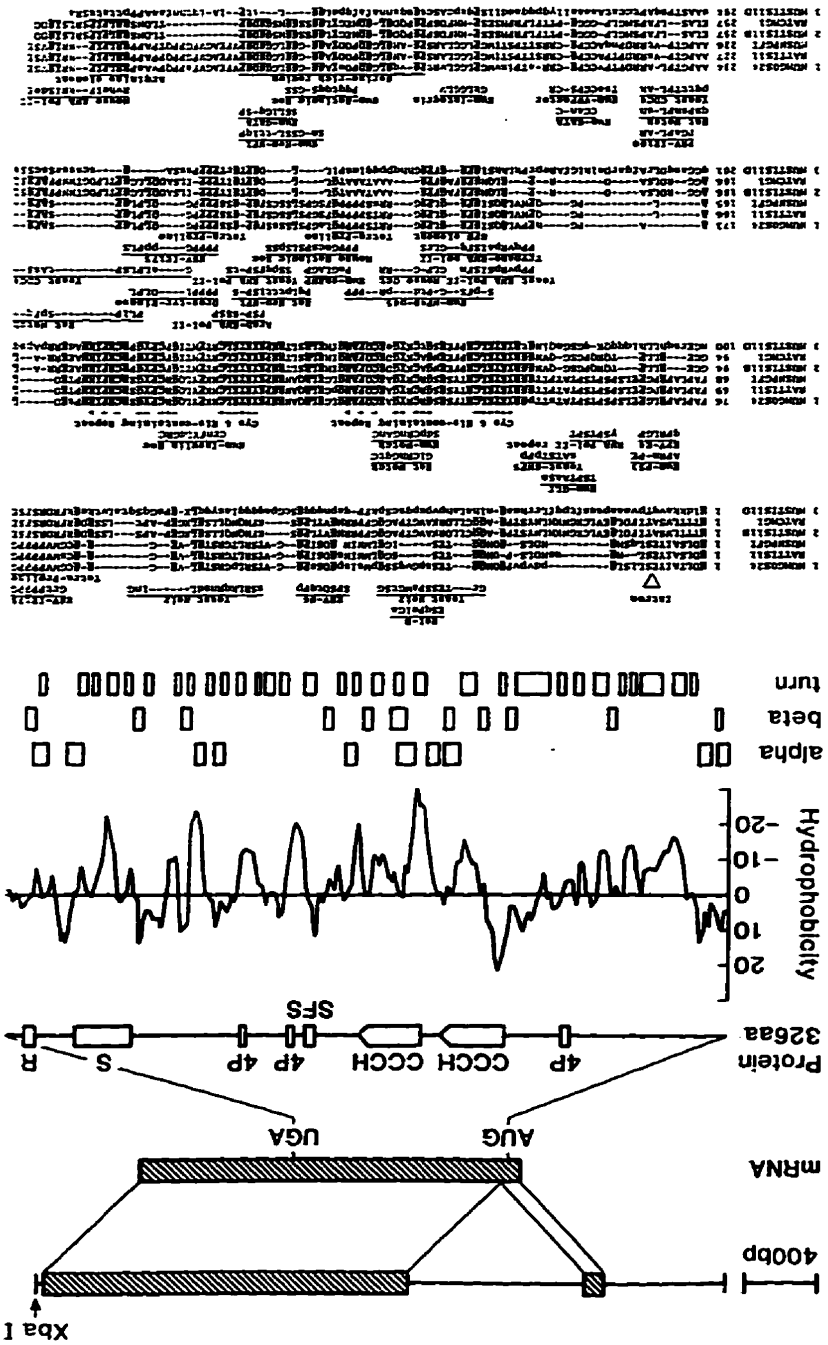
The following is a summary of some of the key features of the human sequence. Within the first exon, the 5' noncoding region contained a potential binding site for the retinoblastoma gene product (118), suggesting possible regulation at the translational level (102). The mid-part part of the intron contained a long dyad repeat sequence suggesting potential formation of strong secondary structure within the unspliced *G0S24* message. Just downstream of the dyad sequence and close to a region of extensive sequence similarity to the 5' flank of the gene for human transition protein-1 (expressed in haploid spermatozoa, ref. 131) was a perfect serum response element (SRE), known to be involved in the activation of immediate early genes (12).

The 3' noncoding region also showed various features of interest. Downstream of the three ATTTA RNA destabilization elements in the TA-rich region was located a regulatory element (TTTTTGTA) which has been identified in the 3' UTR of a number of immediate-early genes (84, 85). A conventional polyadenylation signal was identified 12 nt upstream of the site where the cleavage and polyadenylation would occur during 3' end processing of the RNA.

Distinctions between three groups at the protein level

Alignment of the *G0S24*-derived protein sequence with derived sequences of the other rodent proteins (Table 3-1) is shown in Fig. 3-6. All six proteins had cysteine- and histidine-

Figure 3-6. ***G0S24* Exon-intron organization; major motifs, hydropathy plot, and calculated structure of derived protein; comparison of G0S24 protein sequence with those of related rodent sequences and with partial sequences from other proteins.** Upper. Relative positions of *Xba* I restriction site, exons (striped boxes), and coding region. Motifs referred to are tetraproline repeats (4P), a serine-phenylalanine domain (SFS), cysteine- and histidine-containing presumed zinc-binding regions (CCCH), a serine-rich region (S), and an arginine-containing element (R). Middle. The hydropathy plot was calculated using an interval of 6 amino acids by the method of Kyte and Doolittle (see ref. 69). Structural predictions were calculated by the method of Chou and Fasman (1978, ref. 148). Lower. Alignment of derived protein sequences of members of the *G0S24*-related gene family (group members and GenBank designations shown at left. Gaps (dashes) were introduced to maximize the alignment. The initial alignment (GENALIGN program) was adjusted manually to emphasize regions of correspondence between all six proteins. Shaded regions indicate identity between 5 or more residues at a particular position. The intron position is marked with a triangle. The motifs described in the upper part of the figure are named immediately above the *G0S24* protein sequence. Dashes in the interrupted arrow indicates amino acid identities between the two cysteine and histidine-containing repeats. The results of a database similarity search with the *G0S24* sequence are presented above the sequence with amino acid identifies in uppercase lettering. The sequences correspond to: mouse rel-B (149), yeast mei2 (150), Epstein-Barr virus N6 (EBV-Ng; 151), the product of herpes simplex virus I immediate-early gene 175 (HSV-IE175; 129), human p53 oncogene product (152), human papilloma virus 5 protein (153) RNA polymerase II, large subunit,


turn
beta
alpha
Hydrophobicity
20
10
0
-10
-20
Protein
326aa
4p
CCCH
CCCH
4p
4p
S
R
SFS
mRNA
AUG
UGA
400bp
Xba I

rich ($CX_8CX_5CX_3H$) repeats corresponding to putative zinc-binding domains (7); they also had a serine- and phenylalanine-containing "SFS" domain. In the original paper (2), we reported that the sequence of the group 3 protein, *TIS11d*, differed considerably from those of the other five proteins. In fact, the murine sequence of *TIS11d*, as originally reported, has been recently suggested as being an aberrant protein with an amino-terminal truncation, which also contains a single base pair addition at the carboxyl-terminus resulting in a frame shift in this region. (167; see Table 4-1 for details). Based on these data it is now believed that TIS11b and TIS11d are very similar proteins. The analysis which follows was also based on the originally reported sequences. Group 1 proteins had three tetraproline sequences (PPPPG) in similar positions; the third tetraproline sequence was matched partially in the group 2 and 3 proteins. Near the carboxyl terminus, proteins of groups 1 and 2 had a conserved serine-rich region and an arginine-containing element.

Potential protein structural features

The putative zinc-finger repeats of the *G0S24* protein were flanked by relatively hydrophobic regions and were separated by a hydrophyllic region containing basic amino acids (Fig. 3-6). The hydrophobic amino terminus (exon 1 and the first part of exon 2) showed a propensity to form an α-helical structure (148); helical wheel analysis of the first 18 amino acids showed a tendency for hydrophobic amino acids to localize on one face of the helix. A similar amphipathic tendency was noted in the case of the potential α-helical region between the two potential finger domains. Codon preference analysis (69) showed

a clustering of least preferred codons in a region from the first tetraproline sequence to the middle of the first putative zinc-binding domain.

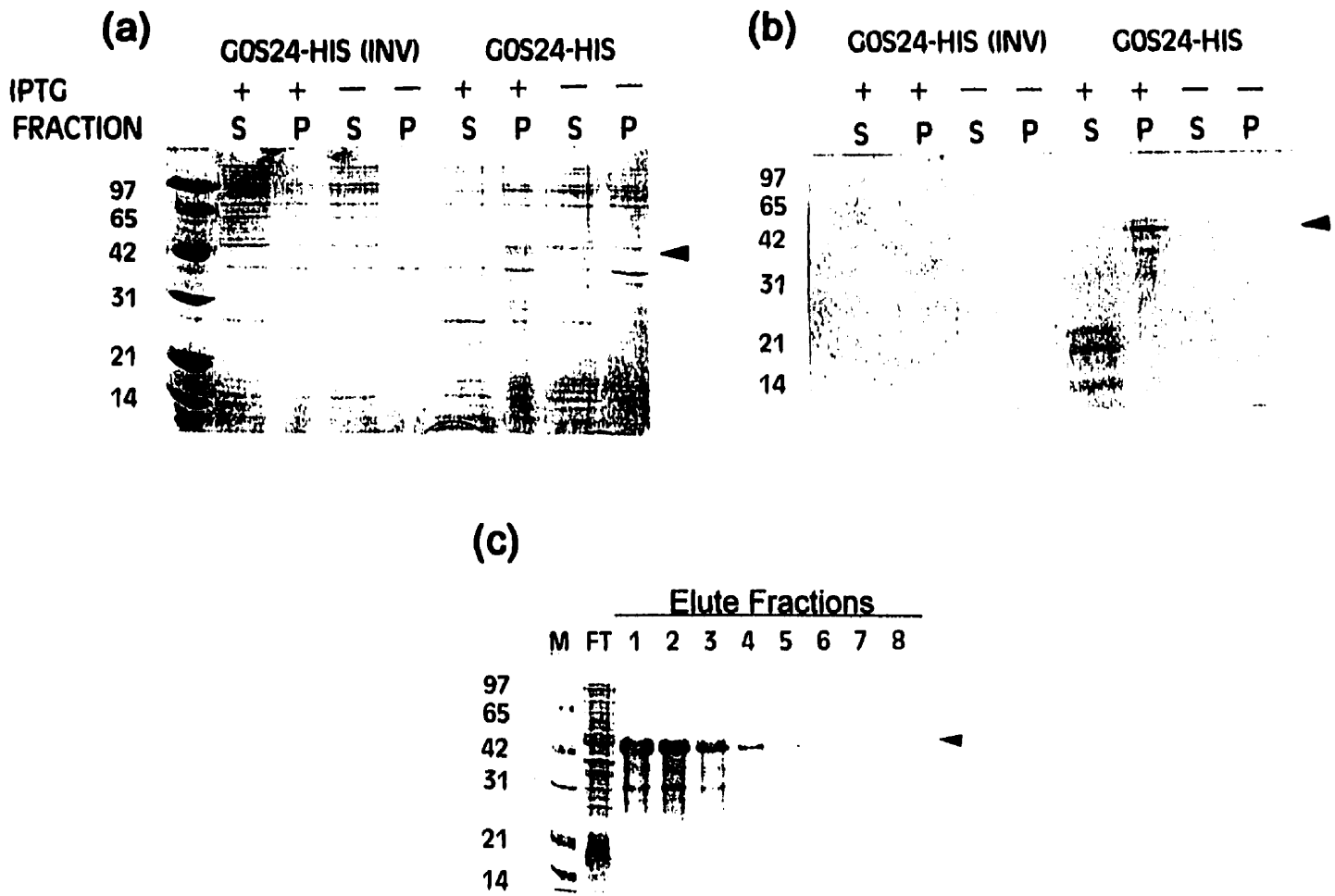
Protein database searches

By searching the ProSite database, it was found that there were several potential sites for phosphorylation by protein kinases A and C and by creatine kinase II (not shown in Fig. 3-6; 168). A search of the SwissProt and GenPept databases revealed regions of sequence similarity to a variety of proteins, many of which are localized to the nucleus and involved in transcription and RNA processing. These included various regions with similarity to parts of the large subunit of eukaryotic RNA polymerase II, which contains, in its carboxyl-terminal domain, repeats of a seven-amino-acid sequence that is a potential target of the mammalian equivalent of the yeast cell cycle control kinase cdc2 (154) and of a DNA-dependent protein kinase (169). A sequence closely resembling the cdc2 target site (YSPTSPT) was found in all three group 1 family members (PSPTSPT). Near the amino terminus of the protein were two regions showing strong similarity to a yeast nuclear RNA-binding protein that is involved in commitment to meiosis (*mei2*; ref. 150).

Expression in E. coli of recombinant histidine-tagged proteins

Specific induction of an ^{35}S-labelled recombinant histidine-tagged G0S24 protein, $(His)_{10}$-G0S24, in *E. coli* cells was indicated by the IPTG-dependent increase of a 42 kDa band when inclusion body pellets were analysed on Coomassie-stained SDS/PAGE gels (Fig. 3-7a). Detectable levels of this species were not seen in soluble cell lysates.

Figure 3-7. **Expression in *E. coli* and purification of recombinant $(His)_{10}$-G0S24 under denaturing conditions.** (a) Metabolically-labelled bacterial cells transformed with pET19b expression vector containing G0S24 full coding sequence in either the proper (G0S24-HIS) or inverse (G0S24-HIS(INV)) orientation were treated with water (-) or IPTG (+) for 3hr at 37°C, lysed by sonication and the soluble (S) and inclusion body pellet (P) fractions were analysed by SDS/PAGE and Coomassie staining. (b) Gel corresponding to upper panel was fixed, dried and analysed by autofluorography. (c) Inclusion body pellet fractions containing $(His)_{10}$-G0S24 were solubilized in 6M urea, purified using denaturing Ni^{++}-chelate chromatography and fractions eluted from the resin with 500 mM imidazole were assayed for purity on Coomassie brilliant blue-stained gels.


(a)
GOS24-HIS (INV)
GOS24-HIS
IPTG
FRACTION
+ + — — + + — —
S P S P S P S P
97
65
42
31
21
14
(b)
GOS24-HIS (INV)
GOS24-HIS
+ + — — + + — —
S P S P S P S P
97
65
42
31
21
14
(c)
Elute Fractions
M FT 1 2 3 4 5 6 7 8
97
65
42
31
21
14

Identification of this species in the pellet, but not soluble lysate fractions from metabolically-labelled cells, suggested its expression as an improperly folded species in this system (Fig. 3-7b). Decreasing the growth temperature to as low as 22°C in the presence or absence of 10 μM Zn^{++}_{aq} in the growth medium did not appreciably increase the proportion of soluble $(His)_{10}$-G0S24 protein (data not shown). Yields of <10μg purified soluble protein/litre culture were obtained from the soluble fraction after Ni^{++}-chelate chromatography, as estimated by Western blotting of SDS/PAGE gels (data not shown) using anti-Nup475 antisera JH60 (7). Recovery of inclusion body-localized protein using Ni^{++}-bound resin under denaturing buffer conditions, yielded 10-12 mg $(His)_{10}$-G0S24 protein/litre culture. This protein was judged as > 60 % pure by Coomassie brilliant blue staining of SDS/PAGE gels (Fig. 3-7c). Subsequent Western blotting studies suggested that the majority of lower molecular weight species were breakdown products rather than contaminating proteins. However, its utility in functional assays was severely compromised by the fact that it did not refold efficiently when subjected to either decreasing steps of denaturant during dialysis, or immobilization to Ni^{++}-chelate resin and slow refolding in decreasing denaturant concentration gradient, or quick refolding in a dilute solution by dropwise addition into large volume of buffer. On the assumption that a smaller polypeptide containing the putative zinc-binding features (CCCH domain) would be expressed more efficiently, yet still maintain demonstrable affinity for a target nucleic acid sequence, we expressed this domain as well. As in the case of the full-length protein, this protein, designated $(His)_{10}$-G0S24ZF, was found predominantly in the inclusion body fraction (data not shown). Denaturing Ni^{++}-chelate affinity chromatography produced a protein which was >70 % pure as judged by

Coomassie brilliant blue staining (Fig. 3-8a). Refolding of $(His)_{10}$-G0S24ZF by dropwise addition into 20 volumes of binding buffer and subsequent concentration of soluble protein on Ni^{++}-chelate resin yielded 3.5 mg/litre culture of a protein with > 85% purity (Fig. 3-8b). Since the apparent size of this protein on SDS/PAGE (15 kDa) was larger than its predicted mass (12.6 kDa), its proper expression was verified by mass spectrophotometry (Fig. 3-8c).

Recombinant $(His)_{10}$-G0S24 full-length and zinc-finger domain lack in vitro double-stranded DNA-binding activity

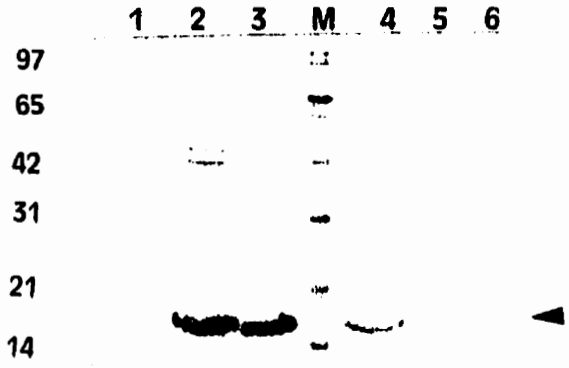
As there has been no published report of a specific DNA-binding element for G0S24 or related family members, we attempted to demonstrate its DNA-binding activity through its ability to shift the gel mobility of members of a library of randomly generated double-stranded DNA species (170). Purified soluble $(His)_{10}$-G0S24 and $(His)_{10}$-G0S24ZF were combined with a randomly-generated library of ^{32}P-labelled double-stranded DNA probes and protein-DNA complexes were separated on nondenaturing polyacrylamide gels. A representative experiment is shown in Fig. 3-9. Using binding conditions containing Zn^{++}, no detectable interaction between BSA controls or either G0S24 protein were observed. By contrast in positive control samples containing Epstein Barr Virus Nuclear Antigen (EBNA; 2 μg), we detected evidence of at least two protein-DNA complexes.

DISCUSSION

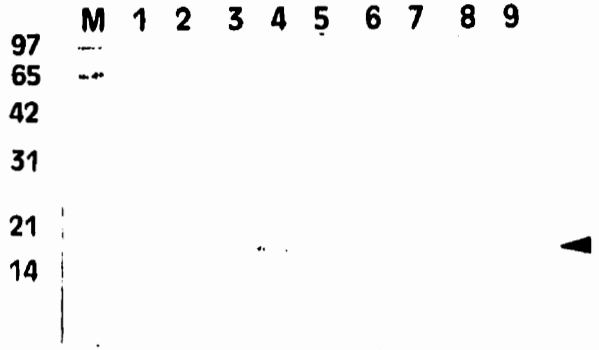
The switch from the quiescent G_0 phase to the G_1 phase of the cell cycle is an important cell cycle control point. Disordered expression of the genes involved could result in

Figure 3-8. **Purification and Mass Spectrographic analysis of recombinant $(His)_{10}$-G0S24ZF.** (a). Analysis of $(His)_{10}$-G0S24ZF purification from inclusion body pellet. Details are as in lower panel of A. (b) Fractions containing eluted $(His)_{10}$-G0S24ZF from denaturing column in (a) were pooled and refolded by dropwise addition into nondenaturing binding buffer. Soluble refolded protein was concentrated using Ni^{++}-chelate chromatography and protein fractions eluted in nondenaturing binding buffer containing 500 mM imidazole, were collected and analysed by SDS/PAGE and mass spectrophotometry (c).

(a) GOS24ZF-HIS PURIFIED USING IMAC UNDER DENATURING CONDITIONS



(b) PURIFICATION OF REFOLDED GOS24ZF-HIS USING NON-DENATURING IMAC



(c)

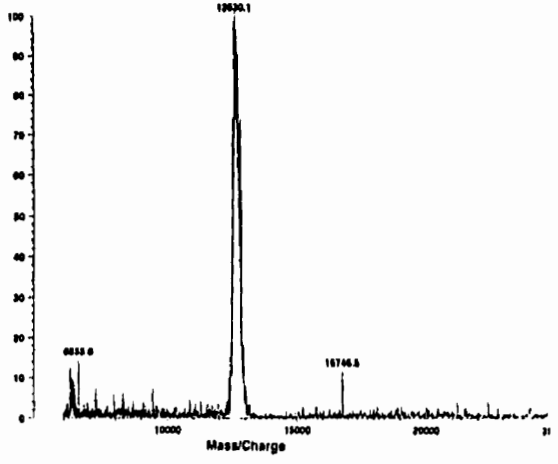
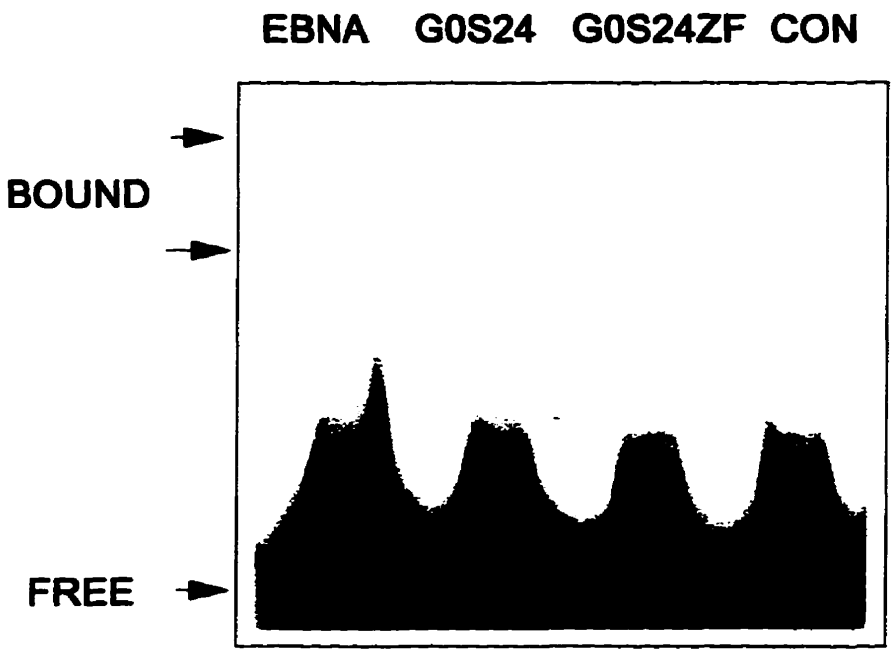


Figure 3-9. **Mobility shift assays to assess *in vitro* DNA-binding activity of recombinant $(His)_{10}$-G0S24 and $(His)_{10}$-G0S24ZF proteins.** Extracts containing no protein (-), EBNA (1 μg), or purified recombinant proteins $(His)_{10}$-G0S24 (~ 25 ng; as estimated by Western blotting) and $(His)_{10}$-G0S24ZF (1.5 μg) were combined with ^{32}P-labelled oligonucleotides (1 ng; 10^5 cpm) from a DNA binding-site selection library. Protein-DNA complexes separated on an 8 % polyacrylimide gels and analysed by autoradiography (72 hr exposure at - 70°C with intensifying screens). Protein-bound and free complexes are indicated by arrows at the appropriate band.


EBNA
G0S24
G0S24ZF
CON
BOUND
FREE

unregulated proliferation; thus, they may prove to be oncogenes. The switch has been studied both in lymphocytes, which are naturally arrested in G_0, and in various rodent cell lines that have been artificially arrested by serum deprivation. In these systems screening of cDNA libraries by differential cDNA hybridization has brought to light an intriguing cast of molecular players with a variety of roles, many of them transcriptional (1, 11, 12, 77). Thus, in our analysis of *G0S24* we were particularly interested in relationships to known transcription factors, oncogenes, and latent viruses.

Studies of the rodent group 1 homologs of *G0S24* suggest a transcriptional role. The murine protein is basic, has cysteine and histidine-rich regions with the potential to form zinc-finger structures, and appears transiently in nuclei following serum stimulation (7). While in this work we also observe a number of similarities, at both the protein and gene levels, between *G0S24* and rodent genes involved in transcription regulation (Figs. 3-5 and 3-6), the apparent lack of DNA-binding activity within the recombinant zinc-finger domains (Fig. 3-9) may suggest an alternative growth-related role for this protein. In the absence of a biological function for G0S24 and the related family of proteins, it remains to be determined whether the absence of DNA-binding of the recombinant proteins reflects a technical aspect of the *in vitro* assay (i.e. presence of N-terminal His-tag, improper folding of the proteins, or lack of accessory binding factors required for DNA-binding) or alternatively predicts a different (non DNA-binding) role for this protein.

In this regard, it is interesting that G0S24 also shows similarity, within its CCCH and proline-rich domains, to a number of RNA processing proteins (see Chapter 4 Table 4-1 and 4-2) including human U2AF35 and snRNP proteins (161, 171) and the Drosophila

suppressor of sable protein (172). In addition, we have observed recurrent similarities with the large subunit of RNA pol II, especially within proline-rich domains, which are often involved in protein-protein interactions. Such regions of sequence similarity suggest a common evolutionary origin and function. We expect, therefore, that the *recurrent* finding of sequences associated with particular genes (e.g., the gene encoding RNA polymerase II) are unlikely to be incidental.

G0S24-related genes fall into three groups in mammalian systems (Table 3-1). Criteria for distinction between the mammalian groups at the protein level readily became apparent from the alignment shown in Fig. 3-6. Thus, possession of three tetraproline repeats is a distinguishing characteristic of group 1. In addition to the human group 1 mRNA product (2.2-2.4 kb major species on Northern blots), a 3.0-kb RNA was noted (Figure 3-1a). Although, hybridization with an intron-specific probe suggested that this might be an unspliced precursor, our inability to confirm this using RT-PCR suggested that this species may correspond to a 3.0 kb group 2 mRNA. Faint bands on DNA blots (Fig. 3-1b) are consistent with at least two gene groups being represented in humans. From polymerase chain reaction (PCR) amplification studies, Varnum and co-workers (1991, ref. 94) have concluded that all three groups are represented in human DNA. Recently, this result was confirmed by the discovery of ERF-1 and ERF-2 cDNAs, whose predicted proteins encode the human group 2 and 3 family members (173, 174).

Originally, it was predicted that *G0S24AF* might represent another yet uncharacterized member of the *G0S24*-related family of genes. The genomic Southern blotting and EST search data, however, are consistent with it being a chimeric sequence, composed of two

genetically independent RNA species. Although we cannot rule out the possibility that this unusual clone was generated *in vivo* by an unusual RT priming event, a more likely explanation would be that it formed during second strand synthesis based on the complementarity of the two first strand products within the joint region.

The characterization of the *G0S24AF* 0.7kb sequence has also allowed the construction of what appear to be the complete coding sequences for a new RNA. The EST sequences were found to contain numerous ambiguities which made accurate cDNA alignment and open reading frame analysis impossible without use of the *G0S24AF* sequence which had been rigorously sequenced on both strands. The significance of the expression of this mRNA in human PBMC cultures will await the future functional characterization of the predicted gene product, a protein which bears limited similarity to any previously identified proteins (GenBank search, Fall 1996).

For some proteins, especially those with proline-rich domains, the large-scale production of soluble recombinant material can be a difficult task (175). Clearly, $(His)_{10}$-G0S24 falls into this category as its full-length form is found almost exclusively in the inclusion body-containing cell pellet fraction. Often, however, purified proteins solubilized in urea or GuSCN may be refolded by removal or dilution of the denaturant. All attempts to refold the $(His)_{10}$-G0S24 protein resulted in its precipitation, suggesting that aggregation of hydrophobic patches from partially folded species was inhibiting folding along the proper pathway(s). Evidence that regulation of folding pathways might be required *in vivo* during biosynthesis of this protein came from our identification of several least preferred human codons in the middle of the second exon. This might induce some pausing during synthesis,

perhaps to facilitate folding of the first part of the protein, a mechanism which would not be present in bacteria or refolding conditions. G0S24 contains regions rich in proline residues (up to 70% in some regions; Figure 3-6). Thus, the rate limiting step(s) of the folding pathway most likely involve the isomerization of proline residues, a process which is aided *in vivo* by a specific enzyme (cis-trans prolyl-isomerase; 176). The proper coordination of Zn^{++} within the putative CCCH finger domain may also be an important structural requirement for this protein since the alkylation of cysteine residues within this motif will result in the improper folding of the $(His)_{10}$-G0S24ZF protein (S. Heximer, unpublished data).

Many proteins involved in mitogen-dependent signal transduction are affected by kinases and phosphatases (12). Our analysis showed that the serine-rich domains in the human and murine group 1 proteins contain possible target for various kinases (168). Therefore, it was not surprising that subsequent studies of the murine protein identified Ser^{220} as a preferred site for MAP kinase phosphorylation *in vitro* (177). Conservation of this site between the mouse and human (Ser^{228}-Pro-Ser-Ala) proteins is consistent with our finding that $(His)_{10}$-G0S24 is phosphorylated *in vitro* by purified sea star $p44^{mpk}$ (S. Heximer, unpublished result). The preferred site is conserved among the group 1 proteins in mammalian species but is not found in those from group 2 or 3, suggesting a possible role for phosphorylation in the differential regulation of G0S24-related proteins; however, the biological significance of this modification is still unknown (177). Of further interest was the potential of this protein product to be regulated by the mammalian homolog of a kinase,

which plays an important role in the control of the cell cycle of *Saccharomyces pombe* (cdc2, ref. 154).

A screen of databases for known sites recognized by transcription factors revealed a variety of potential regulatory sequences, providing a guide for future investigations. These sites were found, not only in the 5' flank, but also in the intron and 3' flank. The extension of the CpG island across the intron (Fig. 3-4) would indicate a strong evolutionary pressure to conserve the intron sequence and suggests expression in the germ-line as well as in lymphoid tissue (74). This finding may be significant in light of the developmental-stage specific expression of *PIE-1*, a CCCH domain-containing factor in *C. elegans* (178), whose expression in germline cells represses expression of genes required for somatic cell differentiation (179). It is also interesting that a preliminary RT-PCR characterization of exon 1 and the 5' end of the intron shows high basal expression of these sequences which cannot be detected in the middle or 3' end of the intron (see Chapter 4).

The presence of two potential metallothionein response elements in the 5' flank (116) correlates with the observation that mRNA of the group 1 homolog (*TIS11*) accumulates in mouse fibroblasts treated with heavy metals (180, 181). However, the elements are not well conserved in the murine sequence (Fig. 3-5) suggesting that these results might represent a stabilization of the TIS11 mRNA at the level of polysome "freezing" (177). Within the intron there was a potential serum response factor-binding site that is found in a variety of other immediate-early genes, including *G0S19-1* and *G0S19-2* (12, 84). Sequences in the 5' UTR and first intron have, in the case of some early response genes, been shown to contain pausing signals for RNA polymerase II (c-fos and c-myc; ref. 182). There is a strong

dyad repeat sequence in the first intron (Fig. 3-5) which may be involved in a similar termination mechanism for the G0S24 gene. In preliminary RT-PCR studies, high levels of a nuclear RNA species containing the exon1/intron junction appear to accumulate despite the absence of RT-PCR signals from the middle or 3' end of the intron (discussed in detail in Chapter 4). Two other features (clusters of potential regulatory motifs) in the intron were of particular note. One of these, in a region showing similarity to an adenovirus-7 promoter sequence (123), contained potential sites for AP1, AP3 (120), and the lymphoid tissue specific factor involved in expression of the oncogene *c-fos* (117). In view of the ability of insulin to induce rapidly the expression of the murine group 1 homolog of *G0S24* (14), it was notable that an insulin gene enhancer motif was also present in the cluster (125). Another cluster was in a region showing similarity to part of the 5' flank of the human gene encoding interleukin-7 (183). This cluster consisted of sequences which might respond to signals generated by way of the T-cell CD28 and other receptors (133, 134). The presence of similar sequences in *G0S24* would imply an ability to respond to the same signals. Understanding the role of these and as of yet unidentified elements in the control *G0S24* mRNA expression will await a footprint analysis of the promoter region and the development of a sensitive reproducible assay (such as RT-PCR) which will allow detection of subtle changes of RNA levels in primary cultures of human mononuclear cells.

ACKNOWLEDGEMENTS

We thank Mr. W. Sangrar for screening the genomic library and Mr. D. Allan, Ms. L. Duke, and Mr. A. Hughes for assistance in restriction mapping as part of an

undergraduate project. Much valuable technical advice was received from Ms. L. Russell, Dr. D. Back, and Dr. P. Davies. The work was supported by grants from the Medical Research Council of Canada, the Leukemia Research Fund (Toronto) and the American Foundation for AIDS Research.

CHAPTER 4

TITLE: RT-PCR Characterization of *G0S24* mRNA in Human Blood Mononuclear Cells.

PREFACE

This final chapter describes the use of RT-PCR to characterize expression and post-transcriptional processing of G0S24 mRNA. This manuscript is currently in preparation for submission. Using the genomic sequence for G0S24 reported in the previous chapter, we have designed assays to study the expression and 3'-end processing of G0S24 mRNA. This is the first report of a second CpG island unusually situated downstream of this gene. The implications of this region as far as transcription termination is concerned are discussed. Furthermore, this study is unusual as it includes kinetic data for a nuclear unprocessed G0S24 RNA. Such studies have also lead to the identification of a novel cycloheximide-dependent mechanism which results in an increase in the level of G0S24 nuclear RNA.

ABSTRACT

G0S24 is a member of a set of putative G_0/G_1 *switch regulatory* genes (*G0S* genes) selected by screening libraries prepared from human blood mononuclear cells cultured for 2 hr with lectin and cycloheximide. The protein product of this gene is a member of a family of putative nuclear proteins containing two or more $CX_8CX_5CX_3H$ zinc-binding repeats. Freshly isolated cells contain high levels of *G0S24* mRNA which have declined significantly after 24 hr incubation in culture medium. Similar responses were observed for *FOS/G0S7*

and *FOSB/G0S3* mRNAs, the kinetics of which suggest that the initial high levels in freshly explanted cells are a result of a transient *spontaneous* stimulation during the purification procedure. (4, 5, Chapter 2). In cells preincubated for a day, however, *G0S24* mRNA levels remain significantly higher than those for other immediate-early genes. Con-A stimulation results in maximum *G0S24* mRNA levels between 1-2 hr which decrease to the same high baseline within 6 hr. *G0S24* mRNA levels, like those for *FOS* mRNA, increase much more in response to a protein kinase C activator (TPA), than to a calcium ionophore (ionomycin), whereas the opposite is true for *FOSB*. We suggest that Con-A elevates *G0S24* mRNA mainly on the basis of its ability to activate PKC-dependent signal transduction pathways. Several potential regulatory signals for 3' end processing and a CpG island in the immediate 3' flank of this gene suggest the potential for a strong downstream transcription termination signal. RT-PCR characterization of *G0S24* mRNA expression has led to the identification and kinetic analysis of a nuclear 3' extended processing intermediate whose 3' end maps to within 116 bp of the polyadenylation signal. Transiently high levels of this RNA species in TPA-treated and freshly explanted cells suggest its levels may be dependent on *G0S24* transcription rate. Response of this 3' extended RNA to cycloheximide is consistent with transcription activation of the *G0S24* gene either through the release of transcriptional repression normally mediated by a labile repressor (31) or through a cycloheximide-dependent signal transduction mechanism also shown to be similarly involved in FOS gene induction (9). The characterization a stable *G0S24* RNA species in Actinomycin D-treated cells suggests unusual stability of a G0S24-derived RNA in the absence of *de novo* transcription.

INTRODUCTION

Unless stimulated by specific antigens, human T-lymphocytes may persist in peripheral blood for several years in a naturally arrested state. The switch of these cells from a quiescent (G_0) state to the G_1 phase of the cell cycle is accompanied by the differential expression of various "immediate early" or *G_0/G_1 switch* ("*G0S*") genes, which are not dependent on prior protein synthesis for their upregulation (184). Among the products of such genes are cytokines, transcription factors and other proteins which may be involved in the activation process (11, 31). One such protein, G0S24, belongs to a family of proteins containing one or more $CX_8CX_5CX_3H$ (CCCH) zinc-binding motifs (Tables 4-1 and 4-2; refs. 7, 14, 185).

Depletion of serum or serum growth factors causes mammalian cell lines to enter a state analogous to that of resting lymphocytes, from which, on refeeding, they escape into the G_1-phase of the cell cycle. In this system, the differential expression of "immediate-early" murine homologs of *G0S24* was observed in G_0-arrested fibroblasts in response to serum, polypeptide growth factors, and phorbol 12-myristate 13-acetate (*Nup475,* ref. 7; *TIS11*, ref. 92; *TTP*, ref. 14).

Although the specific biological role of G0S24 and the related family of CCCH-containing proteins remains unknown, abnormal expression of these factors in lymphoid tissues may lead to dysregulated growth. A chromosomal translocation (7q35-6p21.3), characteristic of an acute human T-cell leukemia, fuses the CCCH domain-containing amino terminus of *ERF-2/HUMTIS11d* to a fragment of the T-cell receptor β-chain, suggesting a

Table 4-1. ALTERNATIVE NOMENCLATURES OF *GOS24*-RELATED MAMMALIAN GENES

Group	Species	Original Name	GenBank name	RNA size (kb)	Protein (amino acids)	Cell Type	Inducing Agent	Reference
1	Mouse	*TIS11*	MUSTS11	2.2	319	Fibroblast	TPA/cycloheximide	(92, 94)
		TTP	MUSTTP45	2.0	319	Fibroblast	Insulin/cycloheximide	(14)
		Nup475	MUSNPGFI	1.8	319	Fibroblast	Serum/cycloheximide	(7)
	Rat	*TIS11*	RATTIS11	--	320	Pheochromocytoma	TPA/growth factors	(220)
	Human	*G0S24*	HUMG0S24	2.1,2.4	326	Lymphocyte	Lectin/cycloheximide	(2)
		TTP	HUMTTP	--	326	HeLa cell	--	(98)
2	Mouse	*TIS11b*	MUSTIS11 B	--	338	Fibroblast	--	(94)
	Rat	*cMG1*	RATCMG1	3.0	338	Epithelial	EGF/cycloheximide	(221)
	Human	*ERF-1*	HSERF13	3.9	338	Epithelial	EGF/NaB/TPA	(173, 234)
		Berg36	HSBERG36	--	339	Lung		(226)
3	Mouse	*TIS11d*	MSTIS11D	--	367[a]	Fibroblast	--	(94)
	Human[b]	*ERF-2*	HSERF2	--	493	Epithelial	--	(174)
		HUMTIS11d	HSUO7802	3.7	463 ,492[c]	Pheochromocytoma	--	(167)

[a]Error in original murine nucleotide sequence (GenBank accession number M58564) resulted in frame shift and aberrant C-terminal end. It has also been suggested that original cDNA sequence was incomplete at 5' end and that the murine protein contains another 29 N-terminal amino acids similar to the human protein (167).

[b]Differences in protein size and sequences are the result of ambiguous cDNA sequences in these clones (GenBank Accession numbers are *HUMTIS11d*, U07802; *ERF-2*, X78992).

[c]Predicted protein for fusion product created when fragment from chromosome 7q35 (T-cell receptor β chain) is inserted into *TIS11d* gene at 6p 21.3 in a patient with acute T-cell leukemia.

Table 4-2. SUMMARY OF OTHER OF CCCH[a] DOMAIN-CONTAINING GENES.

Organism	Original Name	GenBank Accession	Protein (amino acids)	Number Zn^{+2} Fingers	Reference
Human[b]	U2AF35	M96982	241	2	(171)
	U2AFBPL	U51224	479	2	(227)
Drosophila melanogaster	Clipper[b]	U26549	296	5	(8)
	Dtis11/Tiscc1	U13397	437	2	(228, 233)
	Unkempt	Z11527	614, 599	5	(229)
	Suppressor (Sa)	M57889	1,322	2	(172)
Saccharomyces cerevisiae	YTIS11	S76619	285	2	(188)
Schizosaccharomyes pombe	$zfs1^{+}$	D49913	404	2	(187)
Caenorhabditis elegans	PIE-1	U62896	336	2	(178)
Respiratory syncytial virus	22K (M2) protein	M11486	195	1	(230)
		X63408	187	1	(231)
		M82816	187	1	(232)

[a]CCCH domains for this group of proteins may vary in the size of the first knuckle (usually 7-9 amino acids)
[b]homologs for this related group of human proteins have also been identified in mouse, fruit fly and yeast species.

role for this aberrant gene in leukemogenesis (167). Secondly, homozygous deletion of the murine *G0S24* homologue, *TTP/Nup475/TIS11,* results in decreased numbers of mature blood cell types, including B and T lymphocytes, and a corresponding increase in the myeloid cell population (177). These data have lead to the suggestion that this gene is required for the control of hematopoeisis at the level of myeloid cell development.

Based on its nuclear localization and zinc-binding properties it was originally predicted that this family of proteins behaved as transcriptional activators (7, 14). Subsequently, it was shown that mitogenic stimulation resulted in a nuclear to cytoplasmic translocation, suggesting a role for this protein as a repressor of immediate-early gene expression (186). However, a sequence-specific DNA- or RNA-binding activity for this class of mammalian CCCH-containing proteins has yet to be described.

Similar CCCH domains have also been identified in proteins from a wide range of lower eukaryotic species and RNA viruses suggesting an ancient origin for this motif (Table 4-2). Genes encoding CCCH-containing proteins in *Schizosaccharomyces pombe* (zfs1+, ref. 187) and *Saccharomyces cerevisiae* (YTIS11, ref. 188) show increased mRNA expression upon nitrogen starvation and may be required for entry into stationary phase, a unique developmental branch of the proliferation cycle (189). Expression of a similar CCCH domain-containing protein in *C. elegans,* PIE-1, is required by germline cells and blastomeres until the four-cell stage of embryogenesis for the specific repression of genes involved in somatic cell development (178).

Recent data from these systems suggests CCCH-containing proteins may act at the level of RNA processing. The Clipper (CLP) protein in *Drosophila melanogaster* has been shown

to act as a double-stranded RNAse capable of the nonspecific degradation of RNA hairpin structures (8) This activity was shown to be dependent on a region of this protein containing five $CX_{7-9}CX_5CX_3H$ motifs. The *S. pombe* homolog of *G0S24* ($zfs1^+$) was identified on the basis of its ability to suppress meiotic inhibition caused by the overexpression of pac1+, a ribonuclease III-like double-stranded RNAase (187).

We have developed a sensitive competitive RT-PCR assay to study the kinetics of expression and nuclear processing of *G0S24* mRNA in human blood mononuclear cells. In particular, we were interested in the site of transcription termination since genomic sequence in the immediate 3' flank of this gene contains a CpG-rich island, which is highly unusual. RT-PCR studies in the region downstream of the *G0S24* gene were used to map the 3' end of the transcription unit and to characterize the expression kinetics of a *G0S24* 3' extended RNA. The increase in levels of this unprocessed RNA species correlate strongly with the expected activity of the *G0S24* promoter and thus its increase in cycloheximide-treated cells suggest the possibility of a nuclear cycloheximide-dependent mechanism for activation of *G0S24* transcription.

MATERIALS AND METHODS

Cells and culture conditions

Cell purification and culture conditions were essentially as described previously (2, 4, 5; Chapters 2 and 3).

Purification of cytoplasmic and nuclear RNA fractions

All procedures during cytoplasmic RNA preparation (modified from 55) were carried out at 4°C. Briefly, cells in culture were washed three times in phosphate buffered saline (PBS) prior to resuspension with gentle vortexing in lysis buffer (10 mM Tris-Cl, pH 8.6, 0.14 M NaCl, 1.5 mM MgCl, 1 mM DTT, 0.5% NP40) containing 1 U/µl RNAsin ribonuclease inhibitor (Promega, Madison, WI). Samples were kept at 4 °C for 5 min before nuclei and cell debris were pelleted by centrifugation at 12000 x g for 4 min. The supernatant fraction containing cytoplasmic RNA was precipitated using 1 volume of isopropanol and the pellets were immediately solubilized by vortexing in 1 ml of Trizol Reagent (Gibco, Life Technologies, Gaithersburg, MD). Nuclei pellets were washed twice in lysis buffer and solubilized in Trizol reagent. RNA from the cytoplasmic and nuclear fractions was isolated from Trizol lysates as described previously (4). Purified nuclear and cytoplasmic fractions were made up in equal volumes of dpc-treated water for spectrophotometric quantitation. As yields of nuclear RNA were routinely < 10 % those for cytoplasmic RNA they were below the limits of spectrophotometric detection. Thus, direct comparison of RNA levels in the two compartments were carried out using equal proportions of the total nuclear and total cytoplasmic RNA.

Polymerase chain reaction profiles and mRNA analysis by competitive RT-PCR

Moloney murine leukaemia virus reverse transcriptase was obtained from Life Technologies. *Thermus aquaticus* thermostable DNA polymerase (Taq) was obtained from Sangon Ltd., Scarborough, Ontario. Competitive RT-PCR assays of mRNAs were carried

out as described (4, 5, Chapter 2). Total RNA (125-500 ng) from each sample of cultured cells was reverse transcribed using a sequence-specific primer (sequences listed in Table 4-3). Equal aliquots from each reverse transcriptase reaction mixture (containing cDNA corresponding to 5.2-10.4 ng of total RNA) were combined with a fixed number of molecules of cDNA controls (see below), and were then coamplified and the products identified by agarose gel electrophoresis with ethidium bromide. This generated a kinetic profile for the experiment, which is displayed in most figures. Based on the intensity of ethidium staining in the above profile assay, a range of two-fold concentrations of control cDNA plasmid was selected for each cDNA sample, and the PCR was repeated to quantitate more precisely the level of cDNA (4, 5, 60, Chapter 2). Since we were concerned with relative changes in mRNA levels, rather than with absolute quantities, there were no controls for the efficiency of reverse transcription (62). For each set of conditions analysed, results are representative of three independent experiments. It was assumed that this efficiency would vary randomly and to approximately the same extent with different RNA samples. In cases where it was important to compare absolute *G0S24* and *RGS-2/G0S8* mRNA levels, competitive cRNA analysis was used. Two-fold dilutions of a cRNA template were added to equal aliquots of total RNA prior to the reverse transcriptase step, and the resulting cDNA products were then coamplified as described above.

Semi-quantitative analysis of polyadenylated *G0S24* mRNA were carried out as described above except that annealing temperatures of 47°C were used (upstream primer, 5'-CAAACCCCACCCATAAATCAATGG-3'; downstream primer, 5'-$(dT)_{18}$ACACTC-3'), RNA samples were reverse transcribed using oligo $(dT)_{17}$ primer, and equal aliquots of

reverse transcriptase reaction mixture (containing cDNA corresponding to 25 ng of total RNA) were amplified in the absence of competitor. Kinetic profiles were photographed and densitometric measurements of band intensities on the negatives were made using the Molecular Dynamics Image Quant v3.3 scanning package.

Preparation of control plasmids for competitive PCR analysis of G0S24 3'-extended RNA and G0S24 mRNA

Control constructs used for competitive RT-PCR analysis of *G0S* mRNAs were made as described previously (5, Chapter 2). For the *G0S24* 3' extended RNA an 830 bp *Sma* I-*Pvu* II genomic fragment containing the region flanking the polyadenylation signal was inserted into the *Eco*RV site of pBR322. This was digested with *Xba* I (a unique site in the genomic fragment), blunt ends were generated using Klenow large fragment, and an approximately 100 bp exogenous *Alu* I-cut DNA fragment was inserted. PCR primers flanked this site. Insertion of the exogenous DNA fragment slightly decreased the mobility of the PCR product.

G0S24 and RGS-2/G0S8-specific cRNA synthesis

To prepare *G0S8* control RNA (cRNA), a 2 kb *Nco* I fragment from the *RGS2/G0S8* control cDNA sequence (5, Chapter 2) was inserted into *Nco* I-digested pET-19b plasmid vector (Novagen, Markham, Ont.), downstream of the T7 promoter. *G0S8* cRNA was generated *in vitro* from this *Bam*HI-digested plasmid vector. *G0S24* cRNA was made similarly from *Hind* III-digested pT7-7 cG0S24 (control plasmid; ref. 5, Chapter 2). Briefly,

20 U T7 RNA polymerase was added to 2 μg of linearized plasmid in 100 μL transcription buffer (40 mM Tris-Cl, pH7.5, 6 mM $MgCl_2$, 10 mM NaCl, 10 mM DTT, 2 mM spermidine, 0.5 mM each rNTP) containing 90 U RNAsin (Promega). After incubation (40°C, 90 min) samples were treated with RNAse-free DNAse I and prepared for analysis using methods described previously for total RNA samples (4, Chapter 2). RNA yields were determined spectrophotometrically ($O.D._{260}$). Integrity of *in vitro* transcripts was checked by addition of 3 volumes of sample buffer (65% formamide, 8.4% formaldehyde, 30 mM MOPS, pH 7, 6.5 mM NaAc, 0.6 mM EDTA), heating to 65°C for 10 minutes and separation by electrophoresis on 1 % agarose gels.

Sequencing and sequence analysis

Upper strand sequence corresponding to cDNA and genomic regions downstream of the *G0S24* gene were obtained by the dideoxy chain termination method using Sequenase or automated sequence (Core Facility, Queen's University). Database searches were carried out as described previously (2, Chapter 3), except that online BLAST searches were carried out through the services of the National Centre for Biotechnology Information at the National Institutes of Health, Washington (69).

In vitro translation of hTTF-I

RNA was prepared by *in vitro* transcription from *Nco* I-linearized pRSETB plasmid containing the carboxyl terminal 348 amino acids of human TTF-I, as described previously. *In vitro* translation from hTTF-I cRNA was carried out in 50 μl volume containing 4 μg

uncapped RNA, 35 µl rabbit reticulocyte lysate (Promega), 30 U RNAsin, 20 µM each amino acid (minus methionine), and 40 µCi [^{35}S]methionine. Reactions were incubated at 30 °C for 60 min, divided into 10 µl aliquots and stored at -70°C.

DNA mobility shift assays

Mobility shifts were performed essentially as described previously (190). Briefly, the reaction mixtures (25 µl) contained 2 µl of reticulocyte lysate mixture containing hTTF-I or luciferase control protein (Promega), 7.5 fmol of double-stranded ^{32}P- labelled 48 bp oligonucleotides (human Sal box,

5'-CCCGGGATTCCGCAGGGTCGACCAGCAGAATTCGCGACCCCGGGATC-3';

G0S24 Sal box-like,

5'-GGGACCAGGTGAGGTCACCAGATGGGAACCGGCAGTTAACTCTTCCTC-3') in binding buffer (12 mM Tris-HCl, pH 8.0, 100 mM KCl, 5 mM $MgCl_2$, 0.1 M EDTA, 0.5 mM DTT, 8 % glycerol), and 2 µg poly (dI-dC). After incubation for 15 min at 30°C, protein-DNA complexes were separated on nondenaturing 8 % polyacrylamide gels, which were dried and analysed by autoradiography after 48 hr, -70°C with intensifying screens.

RESULTS

High levels of G0S24 mRNA in freshly explanted and preincubated cells in culture

Mononuclear cells were prepared from human blood and cultured in autologous serum and RPMI 1640 medium. RNA levels were determined immediately, or after culture for 24

hr. On the assumption that the efficiency of reverse transcription is similar for different mRNAs (see Methods), the level of *G0S24* mRNA in freshly explanted cells (515±166 molecules/μg RNA) was much higher than that of all of the other *G0S* genes (Table 2-1). As described previously (4, 5, Chapter 2), these genes could be grouped into two classes on the basis of the change in their mRNA levels during the preincubation period. *G0S24, G0S8/RGS2* and *FOS*-related genes (*G0S3, G0S7*) show high mRNA levels in freshly isolated cells, which decrease upon incubation. In contrast, levels for the chemokine *MIP1α/G0S19* and *RGS1/BL34/1R-20* are relatively low in freshly explanted cells and increase during 24 hr culture. *G0S24* mRNA levels after 24 hr incubation (107±42 molecules/μg RNA) are significantly higher ($p<0.05$ in pairwise comparison) than the other mRNAs of the same class.

To assess if the higher levels of *G0S24* mRNA relative to other *G0S* gene mRNAs could be explained by greater RT-PCR efficiency we compared absolute *G0S24* and *G0S8* mRNA levels using a competitive cRNA technique (see Methods). Using cRNA data, relative differences between the baseline levels of *G0S8* and *G0S24* mRNA in freshly explanted cultures (17.2-fold) and in preincubated cultures (67-fold), were similar to those obtained by the standard technique (20.2- and 42.8-fold respectively). These data are consistent with a high absolute level of *G0S24* mRNA in these cells, rather than differential reverse transcriptase efficiency for different mRNAs.

G0S mRNA responses to Con-A depend on preincubation time

High levels of *G0S* mRNAs in unstimulated freshly explanted cells (Table 2-1) are interpreted as resulting from spontaneous stimulation of these cells during the isolation procedure (4, Chapter 2). Fig. 4-1 shows the effect of preincubation period on the response to Con-A. The high levels of *G0S24* and *FOS* mRNAs in freshly isolated cells make it difficult to determine a response to Con-A at this time (Figs. 4-1a,e,f; ref. 4, Chapter 2). In preincubated cells, however, levels in control cultures are constant and significant transient responses to Con-A are observed (Figs.4-1c,g,h). Fig. 4-1b and d show the slower kinetics of the *MIP1α/G0S19* mRNA response.

A detailed kinetic analysis of the responses to Con-A after 24 hr preincubation are shown in Fig. 4-2. At this time, initial *G0S24* mRNA levels are still high (Table 2-1), and increase to a maximum between 1 and 2 hr following Con-A addition. *FOS/G0S7* and *FOSB/G0S3* mRNA levels start much lower (*FOSB* is often undetectable) and increase to maximum levels within 20 and 40 min (Fig. 4-2c,d, ref. 4). Levels of all three of these mRNAs have returned to their original levels by 6 hr. In prolonged culture, *G0S24* RNA levels in control and Con-A-stimulated cultures are maintained at higher levels than those for *FOS* and *FOSB* (up to 48 hr post-stimulus, see picture insert corresponding to Figs. 4-2c,d). *MIP1α/G0S19* mRNA levels increase with slightly less rapid kinetics but are maintained at high levels until at least 6 hr following Con-A treatment (Fig. 4-2b).

Figure 4-1. **RT-PCR assay of the response to Con-A of mRNA levels of *G0S24*, *FOS*-related genes and *MIP1α/G0S19* in freshly isolated cells (a, b,e,f), or cells preincubated for 24 hr (c,d,g,h).** Small filled symbols refer to cultures containing Con-A. Large open symbols refer to cultures without Con-A. At 0 min, cultures (2.5 ml with approximately 2 x 10^6 cells/ ml) were treated with 200 µl of either Con-A solution (final concentration 200 µg/ml) or water (controls). At the indicated times total RNA was prepared, and aliquots taken for RT-PCR assay. The photographs in the upper part of the figure show ethidium bromide-stained PCR products separated by agarose gel electrophoresis. Upper bands are derived from *fixed* concentrations of control plasmids containing the appropriate cDNAs with small inserts to decrease electrophoretic mobility (4, 5, Chapter 2). These were coamplified with the corresponding reverse transcriptase-generated cDNAs whose concentrations vary depending on the RNA sample of origin (lower bands). A given RNA sample was used for measurement of all mRNAs. The graphs show the results of coamplification of fixed quantities of reverse transcriptase-generated cDNAs with *varying* quantities of control plasmids whose concentration range was selected to include a concentration where molecules of plasmid cDNA would compete on a one-to-one basis with molecules of reverse transcriptase products. For each data point, intensities of pairs of ethidium bromide-stained bands were compared to quantitate the reverse transcriptase-generated cDNA (60). Data for : *G0S24* are in (a) and (c); *MIP1α/G0S19* are in (b) and (d); *FOS/G0S7* are in (e) and (g); and *FOSB/G0S3* are in (f) and (h). Plots for mRNA levels in freshly explanted and preincubated cells are shown as dotted and solid lines respectively. The ordinate scale gives concentrations of mRNA assuming 100% efficiency of reverse transcription.

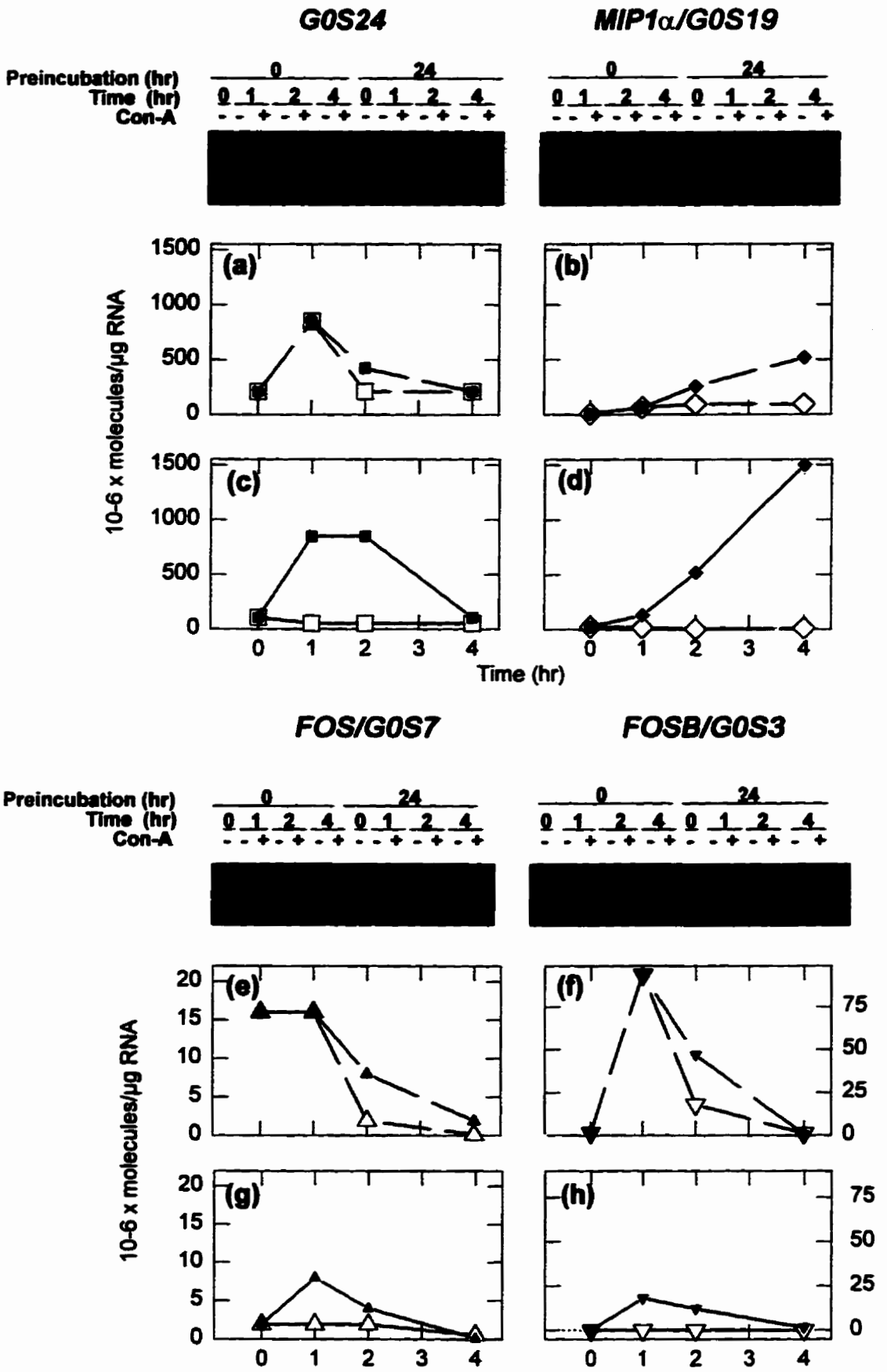
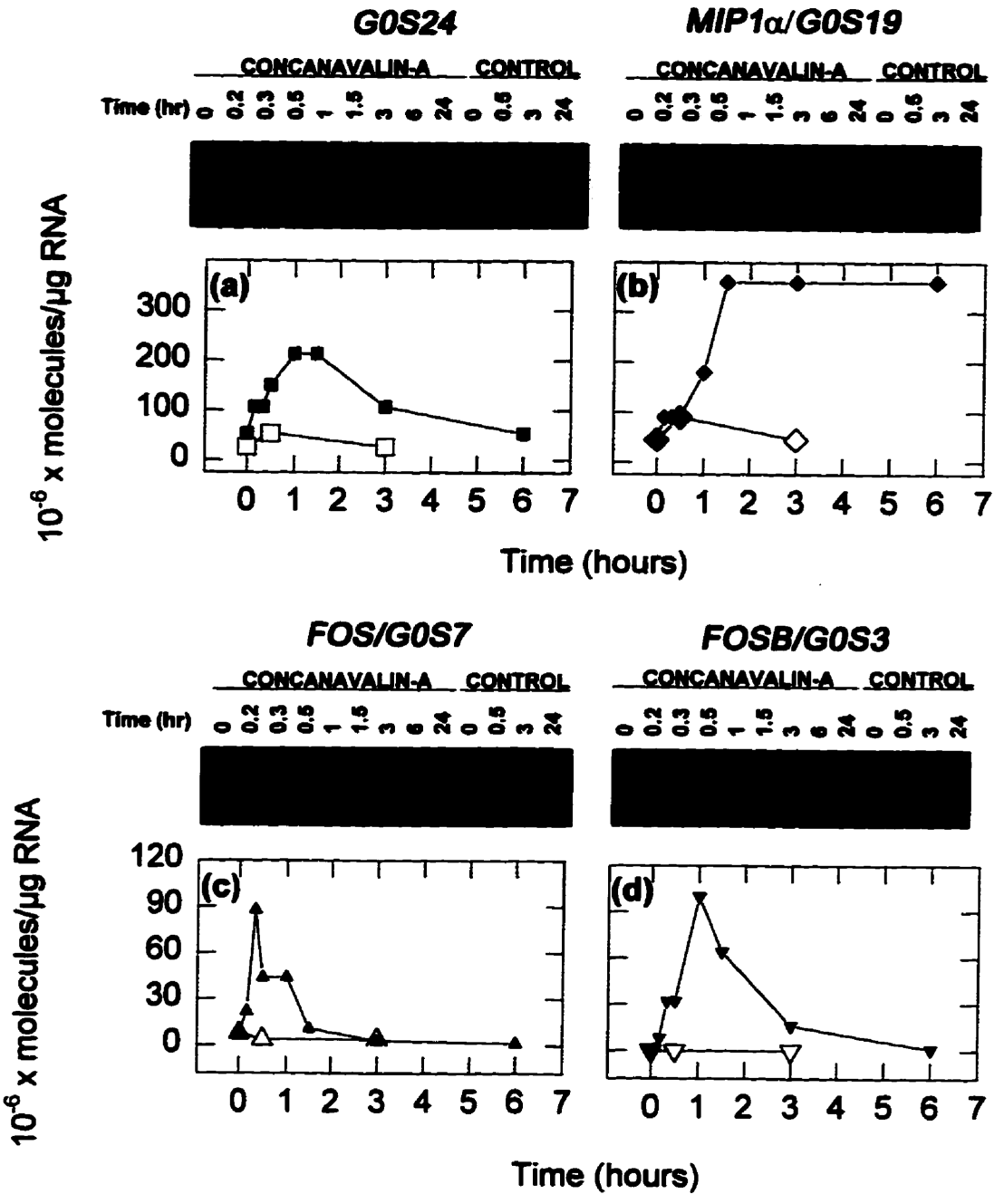

G0S24
MIP1α/G0S19
Preincubation (hr)
Time (hr)
Con-A
0
24
0 1 2 4 0 1 2 4
- - + - + - + - - + - + - +
(a)
(b)
(c)
(d)
1500
1000
500
0
10-6 x molecules/µg RNA
0 1 2 3 4
Time (hr)
FOS/G0S7
FOSB/G0S3
(e)
(f)
(g)
(h)
20
15
10
5
0
75
50
25
0
10-6 x molecules/µg RNA
Time (hr)

Figure 4-2. **Changes in mRNA levels in preincubated cells in response to Con-A for *G0S24* (a) and *MIP1α/G0S19* (b), *FOS* (c), and *FOSB* (d).** Details are as in Fig. 4-1.


G0S24
MIP1α/G0S19
CONCANAVALIN-A
CONTROL
Time (hr)
0 0.2 0.3 0.5 1 1.5 3 6 24 0 0.5 3 24
(a)
(b)
(c)
(d)
300
200
100
0
120
90
60
30
0 1 2 3 4 5 6 7
10-6 x molecules/µg RNA
Time (hours)
FOS/G0S7
FOSB/G0S3

Changes in G0S gene RNA levels in response to TPA and ionomycin

Fig. 4-3 shows the changes in *G0S24* and *FOSB/G0S3* mRNA levels in response to TPA and ionomycin, either added alone or in combination, to preincubated cultures. As expected from the name given to the rodent homolog (*"TPA inducible gene 11"; TIS11;* ref. 77)*G0S24* mRNA levels increase strongly in response to TPA treatment (Fig. 4-3a; ref. 5). Weaker responses are observed for the calcium ionophore, ionomycin. In a series of 3 experiments, cells treated 1 hr with both compounds, *G0S24* mRNA levels (354 ± 71 molecules/μg RNA) are increased to a similar degree as those in cells treated with TPA alone (425 ± 0 molecules/μg RNA). However, if an earlier time point is examined (30 min) a modest superinduction of *G0S24* mRNA levels is observed in ionomycin/TPA-treated cells (1133 ± 283 molecules/μg RNA compared to cells treated with TPA alone (625 ± 125 molecules/μg RNA). Absolute increases in *G0S24* mRNA levels in cells treated either with TPA alone or with TPA and ionomycin are comparable to those observed in Con-A-treated cultures. Similar TPA-dependent responses were observed for *FOS* mRNA (data not shown). In contrast, *FOSB/G0S3* mRNA levels appear highly sensitive to changes in intracellular calcium levels (Fig. 4-3b, ref. 5, Chapter 2) and respond weakly to TPA alone.

Unusual CpG island in 3' flank of the G0S24 gene

Consistent with expression in the germline, the promoter and coding sequences of the human *G0S24* gene contain islands rich in CpG dinucleotides (2). Fig. 4-4 shows the sequence at the 3' end of the *G0S24* gene, including ~800 bp of recently characterized sequence in the 3' flank. The sequence in this region sequence was also found to contain a

Figure 4-3. **Response to TPA and the calcium ionophore, ionomycin, of *G0S24* (a) and *F0SB* (b) mRNA levels in preincubated cells.** Photographs (upper) show expression profiles for cultures with DMSO controls (lanes 1, 5, 9, and 13), 100 nM TPA alone in DMSO (lanes 2, 6, 10, 14), 2 μM ionomycin alone in DMSO (lanes 3, 7, 11, and 15), or a combination of TPA and ionomycin in dimethylsulfoxide (DMSO; lanes 4, 8, 12, and 16). Corresponding graphs (lower) show the mRNA levels determined using the competitive RT-PCR assay, for cultures treated with DMSO controls (open symbols), TPA alone (dark shaded symbols), ionomycin alone (light shaded symbols), or TPA and ionomycin (filled symbols). Other details are as in Fig. 4-1.

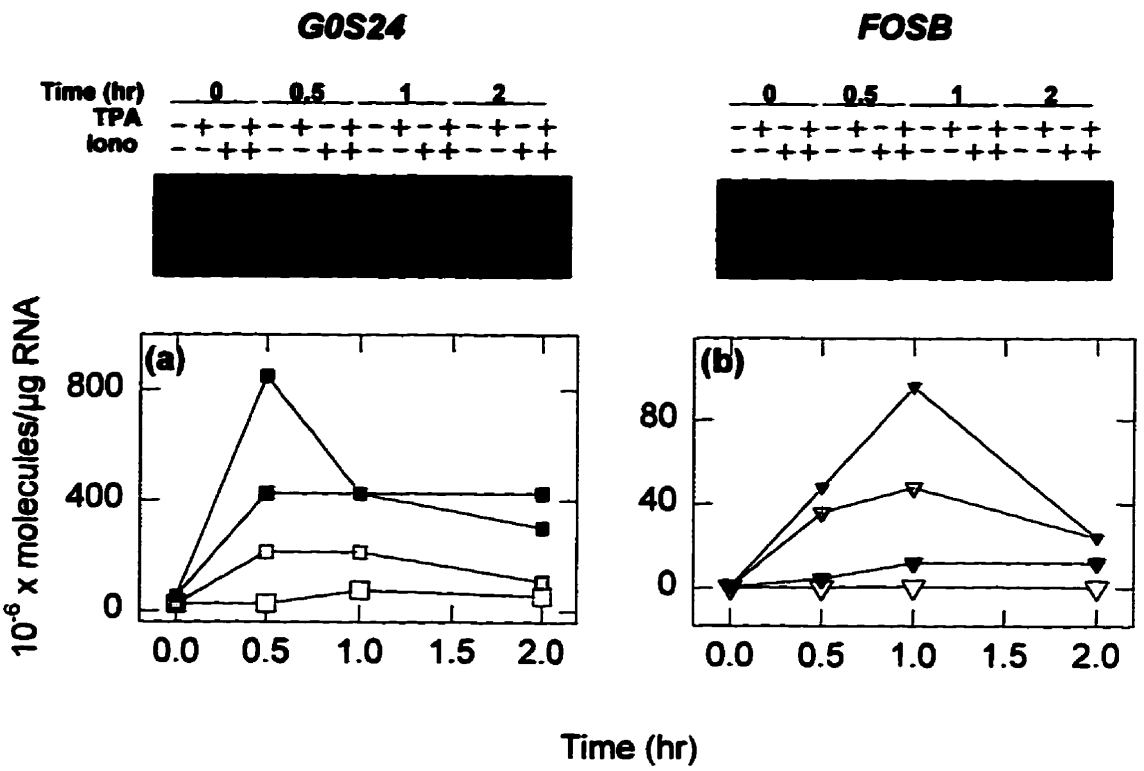
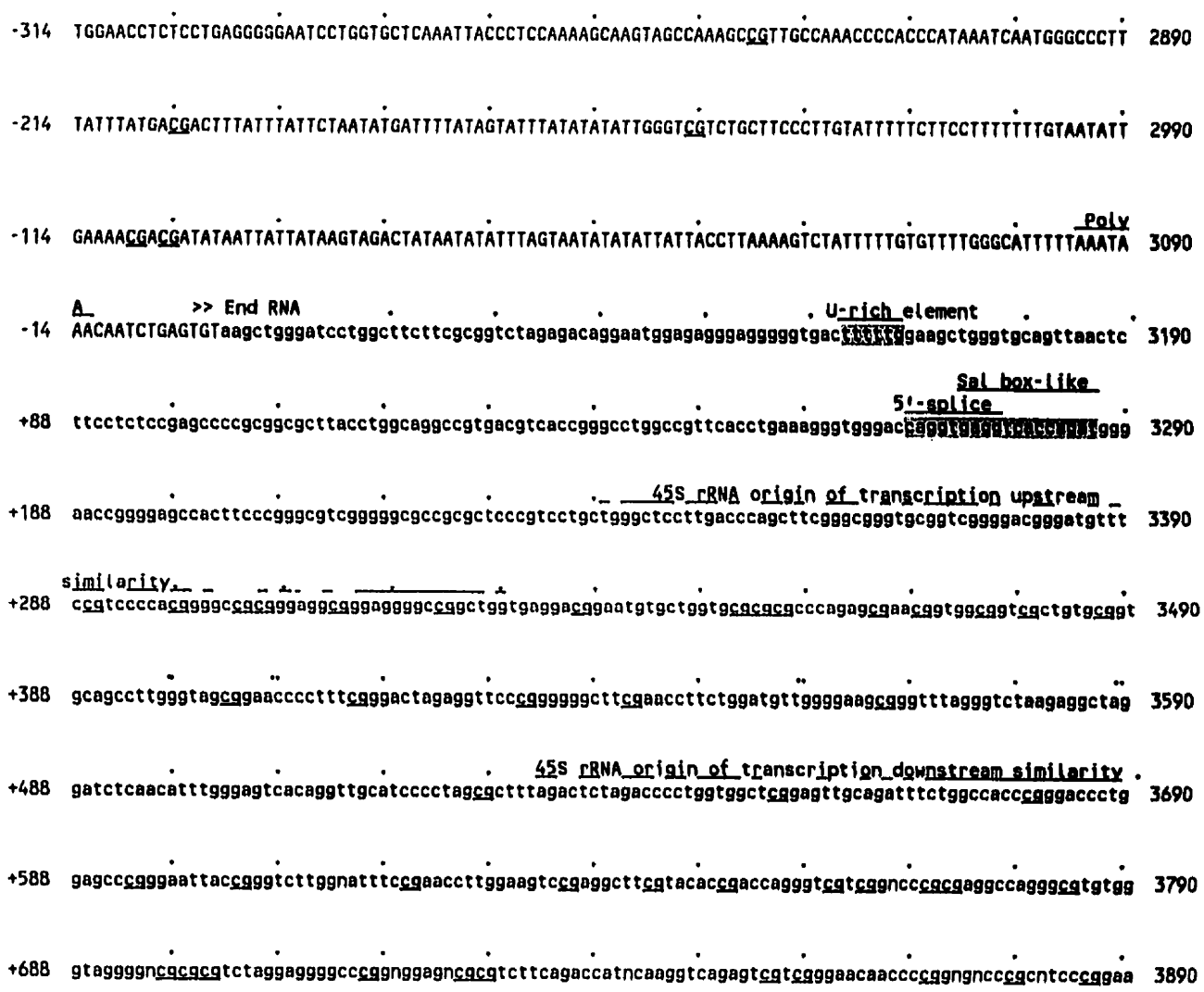

G0S24
FOSB
Time (hr)
0
0.5
1
2
TPA
Iono
(a)
(b)
800
400
0
80
40
0
0.0
0.5
1.0
1.5
2.0
10^-6 x molecules/µg RNA
Time (hr)

Figure 4-4. **Sequence of the *G0S24* 3' flank region.** Numbering is relative to either the end of exon two (left column), or the first base of the previously reported *Xba* 3.1 kb fragment containing the *G0S24* gene (GenBank accession number M92844, right column). Upper case lettering indicates the exonic sequence. CpG dinucleotides are shown in bold type and are underlined. Elements potentially involved in 3' processing/termination are shaded and named above the sequences to which they refer. These are: U-rich element (191), a Sal-box-like sequence (190), and 5' donor splice site. Regions of similarity with rRNA origin of transcription (193) are named as such. In all cases, mismatches are shown as gaps in the underlining beneath the motif name. Both strands were sequenced to nt 32 (relative to cleavage/polyadenylation site, ref. 2). For the final 787 nt only the upper (mRNA synonymous) strand was sequenced.


-314 TGGAACCTCTCCTGAGGGGGAATCCTGGTGCTCAAATTACCCTCCAAAAGCAAGTAGCCAAAGCCGTTGCCAAACCCCACCCATAAATCAATGGGCCCTT 2890
-214 TATTTATGACGACTTTATTTATTCTAATATGATTTTATAGTATTTATATATATTGGGTCGTCTGCTTCCCTTGTATTTTTCTTCCTTTTTTTGTAATATT 2990
-114 GAAAACGGACGATATAATTATTATAAGTAGACTATAATATATTTAGTAATATATATTATTACCTTAAAAGTCTATTTTTGTGTTTTGGGCATTTTTAAATA 3090
Poly
A
>> End RNA
U-rich element
-14 AACAATCTGAGTGTaagctgggatcctggcttcttcgcggtctagagacaggaatggagagggaggggtgac[redacted]gaagctgggtgcagttaactc 3190
Sal box-like
5'-splice
+88 ttcctctccgagccccgcggcgcttacctggcaggccgtgacgtcaccgggcctggccgttcacctgaaagggtgggac[redacted]ggg 3290
45S rRNA origin of transcription upstream
+188 aaccggggagccacttcccgggcgtcgggggcgccgcgctcccgtcctgctgggctccttgacccagcttcgggcgggtgcggtcggggacgggatgttt 3390
similarity
+288 ccgtccccacggggccgccgggaggccgggaggggccggctggtgaggacggaatgtgctggtgcgcgccgcccagagcgaacggtggcggtcgctgtgcggt 3490
+388 gcagccttgggtagcggaacccctttcgggactagaggttccgggggggcttcgaaccttctggatgttggggaagcgggtttagggtctaagaggctag 3590
45S rRNA origin of transcription downstream similarity
+488 gatctcaacatttgggagtcacaggttgcatcccctagccgctttagactctagacccctggtggctcggagttgcagatttctggccaccgggaccctg 3690
+588 gagccgggaattaccgggtcttggnatttccgaaccttggaagtccgaggcttcgtacaccgaccagggtcgtcggnccgccgaggccagggccgtgtgg 3790
+688 gtaggggncgcgcgctctaggaggggccggnggagncgcgtcttcagaccatncaaggtcagagtcgtcgggaacaacccgggngnccgcntccggaa 3890

CpG island (Fig. 4-5, upper panel). Since such features are usually found at the 5' end of their associated genes, this suggested either the presence of another gene, or an important regulatory domain in the immediate downstream region.

Also downstream (60 nt) of the last exon was a U-rich sequence (UUUUUG) which, in the case of many eukaryotic genes, is involved in efficient transcription termination and 3' end processing (191; for review see 192). No additional polyadenylation signals were found in this region and BLAST searches of the GenBank database (September 1996) found no similarities in this region to previously identified expressed sequence tags (ESTs). The 3' flank of the *G0S24* gene, however, contains two regions which show significant similarity to the site of transcription initiation of human ribosomal RNA genes (Fig. 4-4). The spacing between these regions is similar to that observed in a normal rRNA gene supporting the notion that these regions are derived from an rRNA gene. Also within this region was located a sequence element (5'-AGGTCACCAGAT-3') which closely resembles the Sal box transcription terminator for RNA polymerase I, which transcribes rRNA (190). This element overlaps a consensus 5' splice signal sequence which has also been implicated in the efficient 3' end processing of some RNA species (194).

Lymphocytes contain measurable levels of apparent prematurely terminated and 3'-extended nuclear G0S24-derived RNAs

To further characterize *G0S24* transcription and processing using RT-PCR, a number of primers targeted to various regions of the gene were chosen (Table 4-3, Fig. 4-5 middle and lower panels). Total RNA was reverse transcribed from primers 281 bp downstream of

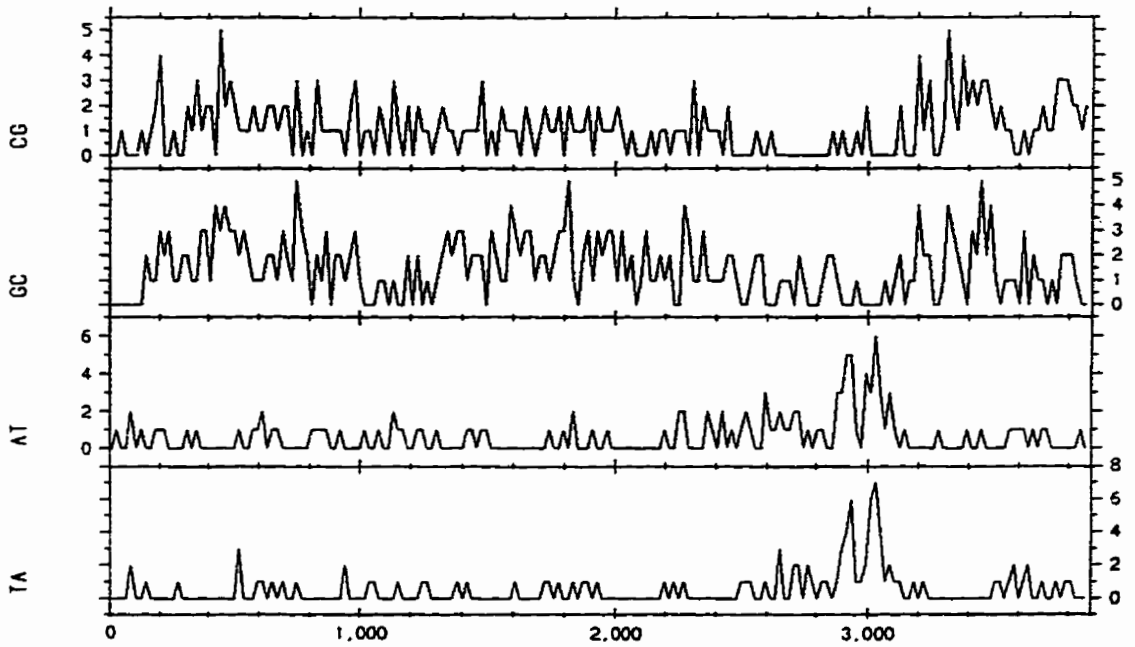
Table 4-3: Sequence of primers used in reverse transcriptase and polymerase chain reactions to characterize *G0S24* transcription and processing.

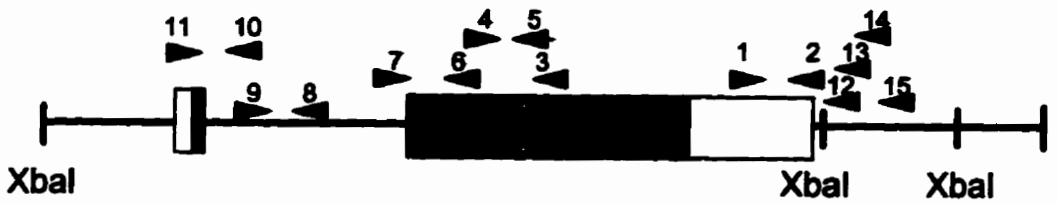
Primer Number[a]	Sequence	Strand[b]
1	5'-CAAACCCCACCCATAAATCAATGG-3'	+
2	5'-AGACCGCGAAGAAGCCA-3'	-
3	5'-AAGGAGCTGGAGGACAG-3'	-
4	5'-GAGAAGCTGATGCTCTGGCGAAGC-3'	-
5	5'-CGCGCTACAAGACTGAGCTATGTC-3'	+
6	5'-TAGCTCAGTCTTGTAGC-3'	-
7	5'-AACCGACCCATTTCCGC-3'	+
8	5'-TATGACCCACGCGTCGG-3'	-
9	5'-CCAGCTTGGTGATTTGG-3'	+
10	5'-CCAATTCACCAAGCTGG-3'	-
11	5'-ATGGATCTGACTGCCATC-3'	+
12	5'-TGTCTCTAGACCGCGAAGAAGCCA-3'	-
13	5'-ACTGCACCCAGCTTCCA-3'	-
14	5'-GAGGAAGAGTTAACTGC-3'	-
15	5'-GGTGACGTCACTTCCTG-3'	-

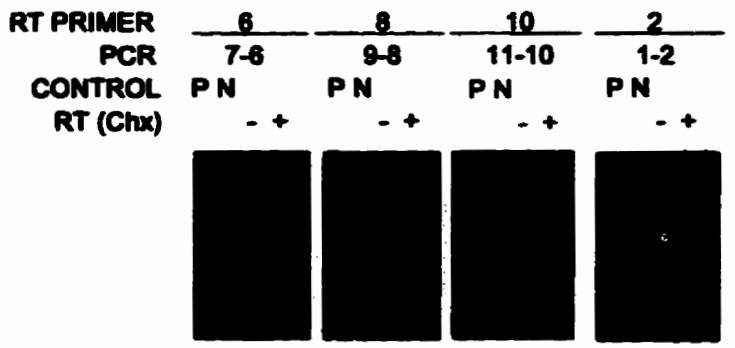
[a] Primer locations relative to *G0S24* exons are shown in Figure 4-6. Locations were chosen to avoid hybridization with nucleic acid of other members of the human gene family (*ERF1*, (173); *ERF2/TIS11d*, (174; 167).

[b] + indicates the same sequence as the "upper" *G0S24* mRNA-synonymous strand (i.e., the primer binds to the "lower" template strand).

Figure 4-5. **RT-PCR characterization *G0S24* mRNA expression.** (Upper) Profile of some dinucleotide frequencies from the genomic region containing the *G0S24* gene. (Middle) Schematic diagram of the region containing the *G0S24* gene and 3' flank. Shown above the sequence are the locations of the primers used to assay for intronic (11-10, 9- 8, 7- 6) and 3' extended (1-2, 1-12, 1-13, 1-14, 1-15) putative pre-mRNA species. A list of primer sequences is shown in Table 4-3. (Lower) *G0S24*-specific antisense oligonucleotides at various locations in the gene were used to prime first strand synthesis on RNA from control (-) and cycloheximide-treated cultures (+). Photograph shows ethidium-stained agarose gel of the RT-PCR products for each primer pair together with the corresponding positive (P; genomic DNA template) and negative (N; water) PCR controls. All RNA samples were DNAse treated prior to analysis and gave no detectable PCR signal in the absence of reverse transcriptase with any of the primer pairs shown.


CG
GC
AT
TA
0
1.000
2.000
3.000


11
10
4
5
14
7
6
3
1
2
13
9
8
12
15
XbaI
XbaI
XbaI


RT PRIMER
6
8
10
2
PCR
7-6
9-8
11-10
1-2
CONTROL
P N
P N
P N
P N
RT (Chx)
- +
- +
- +
- +

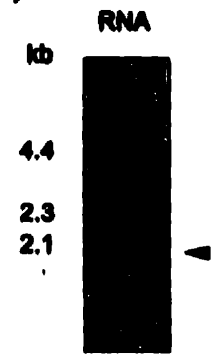
the intron/exon 2 boundary (primer 6) and 13 bp *downstream* of the end of exon 2 (i.e. in the 3' flank; primer 2). Also, intronic primers 164 bp and 497 bp downstream of exon 1 (primers 10 and 8 respectively) were used to search for stable intronic species and RNAs which had unprocessed exon 1/intron splice junctions . Fig. 4-5 (lower panel) shows the products which result when these RT products were used as templates for PCR. No significant levels of RT-PCR product could be generated corresponding to intron- or intron/exon 2 boundary-containing RNA species. However, products corresponding to the region spanning the exon1/intron splice junction and to the 3' flank were obtained both with RNA from control cultures and from cultures treated with cycloheximide (which gave stronger signals). These studies suggest that there exists a population of RNAs which terminate within the intron. We predicted it might represent a shortened transcript formed by RNA polymerase II pausing, perhaps at a potential RNA hairpin structure formed at the dyad repeat sequence in middle of the intron (2, Chapter 3). We refer to the RNA population which extends into the 3' flank a the "3' extended *G0S24* RNA" variant.

Termination of unprocessed RNA species proximal to a strong transcription terminator

Northern blots, probed with a 0.5 kb *Xba* I probe made from the immediate 3' flank of the gene (Fig. 4-6a), identified a weakly hybridizing ~1.9 kb RNA species suggesting that the 3' unprocessed G0S24 mRNA terminated within 100 bp of this gene. Using the primer combinations shown in Fig. 4-6b, the 3' end terminus of this RNA was mapped to a region between nt 76 and 116 downstream from the polyadenylation signal which was upstream of and proximal to the Sal box-like motif shown in Fig. 4-4. In band-shift studies designed to

Figure 4-6. **Mapping of the 3' end of the putative transcription termination site.** (a) Total RNA was prepared in guanidine thiocyanate from freshly explanted unstimulated cells. The RNA (20 μg) was glyoxyl-denatured, electrophoresed through 1 % agarose gel and blotted to a nylon membrane. A radioactive probe was a ^{32}P-labelled 0.5 kb *Xba* I fragment (10^6 cpm/ml) made from 3' *G0S24* flank region. The arrow points to 1.9 kb band which gives the best hybridization signal with this probe. Sizes of glyoxyl-denatured DNA markers (kb) are shown on left. (b) Summary of RT-PCR characterization of the 3' end of the 3' extended *G0S24* RNA. RT primers (1, 13, 14, and 15) at increasing distances downstream of the site of cleavage/polyadenylation (see Fig. 4-6 for locations) were used to generate first strand products from total RNA. First strand products of these reactions were analysed by PCR primer-pairs flanking the polyadenylation signal. Results in the case indicate the presence (+) or absence (-) of detectable levels of RT-PCR products in these samples.

(a)



(b)

Mapping of the 3' end of the *G0S24* nuclear 3' extended message.

RT primer	PCR primer pair	Product	Distance from poly(A) signal (bp)
1	1-2	+	13
13	1-13	+	64
14	1-14	+	76
15	1-15	-	116

predict the involvement of Sal-box binding protein, human transcription termination factor (HUMTTF-I), we were unable to detect binding of *in vitro* translated protein to a 48 bp oligonucleotide containing the *G0S24*-derived Sal box-like sequence (Fig. 4-7).

Localization of 3'-extended G0S24 RNA variant to the nucleus

The sequence of the RT-PCR product generated across the cleavage/polyadenylation site showed that this RNA contained the same sequence as the gene. Furthermore, when equal proportions of RNA from nuclear and cytoplasmic fractions were used for RT-PCR analysis of the 3'-extended *G0S24* RNA (Fig. 4-8), this species was found to be localized in the nuclear fraction. This was not seen when using exon-specific primers for the *G0S3/FOSB, G0S7/FOS, G0S8* or *G0S24* mRNAs which all show equal, if not greater, amounts of message in the cytoplasmic fraction. These data suggested that the 3' extended species represented an unprocessed RNA, rather than a differentially polyadenylated species, and were consistent with our previous failure to detect additional polyadenylation variants of the *G0S24* using 3' end RACE (S. Heximer, unpublished result).

Levels of the 3' extended *G0S24* transcript appeared to increase in response to Con-A and cycloheximide either added alone or in combination. This pattern was also observed for the normal *G0S24* mRNA product, which shows equal distribution between nuclear and cytoplasmic compartments. Such distribution is not observed for the *RGS2/G0S8*, *FOS* or *FOSB* mRNAs which are distributed mainly in the cytoplasmic fraction in control and stimulated cells (Fig. 4-8).

Figure 4-7. **DNA binding-shift analysis of hTTF-I binding to *G0S24* Sal box-like sequence.** ***In vitro*** **translation extracts containing luciferase control protein (Luc) or hTTF-I were combined with ^{32}P-labelled human Sal box (5'-**CCCGGGATTCCGCAGGGTCGACCAGCAGAATTCGCGACCCCGGGATC-3'); or *G0S24* Sal box-like oligonucleotides (5'-GGGACCAGGTGAGGTCACCAGATGGGAACCGGCAGTTAACTCTTCCTC-3') . **Protein-DNA complexes separated on an 8 % polyacrylamide gel and analysed by autoradiography. Protein-bound and free complexes are indicated by arrows at the appropriate bands.**

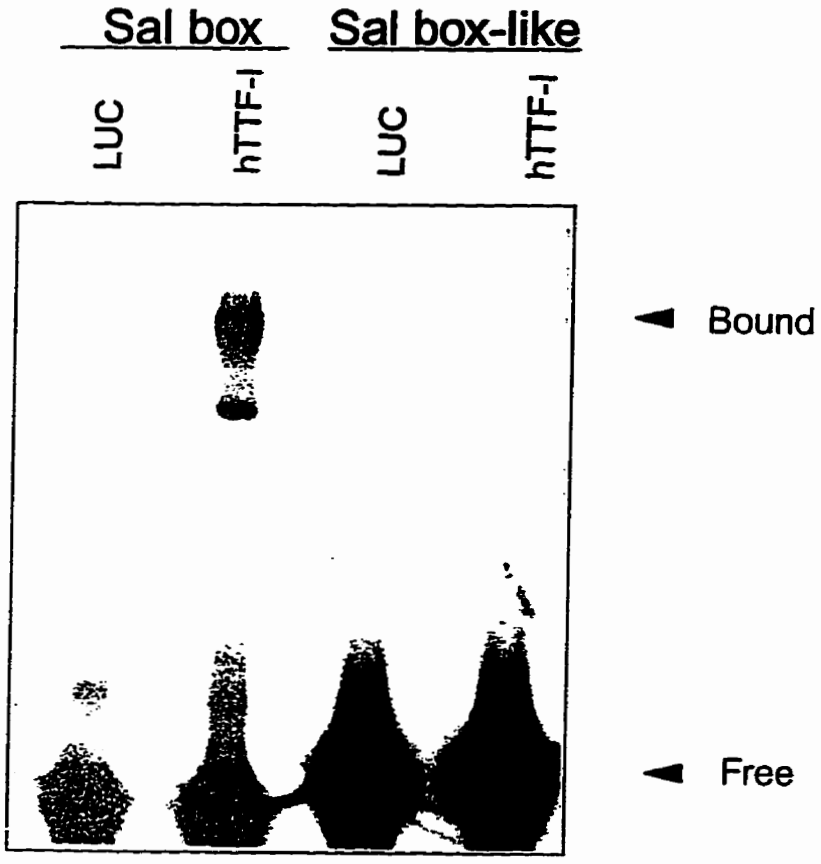
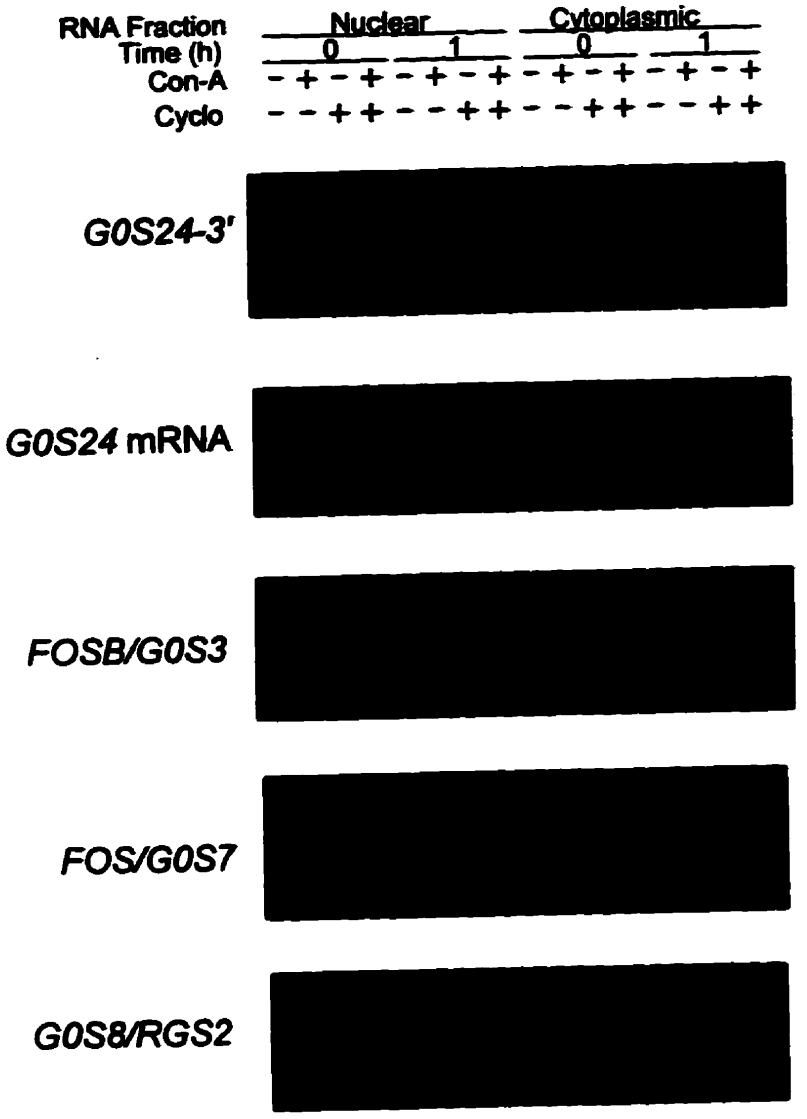

Sal box
Sal box-like
LUC
hTTF-I
LUC
hTTF-I
Bound
Free

Figure 4-8. **Cytoplasmic and nuclear distribution of 3'-extended *G0S24* RNA and *G0S* gene mRNAs.** RNA was prepared from nuclear and cytoplasmic compartments of preincubated cultures which had been treated for 0 or 1 hr with Con-A and cycloheximide, either alone or in combination. Equal proportions of the total amount of RNA purified from each compartment were assayed by RT-PCR to study the cellular distribution of *each G0S* RNA. Photographs show profiles of the cellular distribution pattern of *G0S24* 3' extended RNA, and the *G0S24*, *G0S8/RGS2*, and *FOS*-related mRNAs.


RNA Fraction
Nuclear
Cytoplasmic
Time (h)
0
1
0
1
Con-A
- + - + - + - + - + - + - +
Cyclo
- - + + - - + + - - + + - - + +
G0S24-3'
G0S24 mRNA
FOSB/G0S3
FOS/G0S7
G0S8/RGS2

3' extended nuclear RNA as a processing intermediate

Using the same competitive RT-PCR approach as was used to study levels of *G0S* mRNAs in exonic regions, we developed a specific assay for the 3' extended region of *G0S24* RNA. Baseline levels of 3' extended RNA in control cells from freshly explanted cultures (4.86±1.84 molecules/μg RNA, n=3) and preincubated cultures (1.20±0.21 molecules/μg RNA, n=6) were significantly lower than those observed for the standard *G0S24* mRNA in these cells (Table 2-1). Thus, levels of the 3' extended species were significantly higher in freshly explanted, and in spontaneously stimulated cultures under conditions when we expected a high rate of transcription from the *G0S24* promoter. The *G0S24* gene is very responsive to TPA-treatment as shown by the profiles in Fig. 4-3. In TPA-treated cultures there was a rapid transient increase in the levels of the 3' extended RNA species (Fig. 4-9b) whose levels are maximal at 15 min post-stimulus. The kinetics of this RNA response were rapid compared to the *G0S24* mRNA species, the levels for which peaked at _ _ min and remained high after 60 min (Fig. 4-9a), suggesting a precursor product relationship.

Cycloheximide induction of the 3' extended nuclear RNA species

Fig. 4-10 shows RT-PCR assays, using total RNA from freshly explanted or preincubated cultures, of the responses to cycloheximide of the 3' extended RNA and other *G0S* mRNAs. In freshly explanted cells, levels of the 3' extended RNA were maintained by cycloheximide above the levels in untreated controls, which decreased significantly within 45 min in culture (Fig. 4-10b). When baseline levels of this RNA had stabilized (in preincubated cells), absolute levels were increased by cycloheximide to those which

Figure 4-9. **Time course for response to TPA of *G0S24* mRNA (a) and *G0S24* 3'-extended RNA (b).** Photographs (upper) show expression profiles for untreated control cultures (lanes 1, 4, 7, and 10), control cultures with DMSO alone (lanes 2, 5, 8, and 11), or cultures treated with 100 nM TPA in dimethylsulfoxide (DMSO; lanes 3, 6, 9, and 12). Corresponding graphs (lower) show the mRNA levels determined using the competitive RT-PCR assay for untreated cultures (open symbols), DMSO-treated cultures (shaded symbols), and TPA-treated cultures (filled symbols). Other details are as in Fig. 4-1.

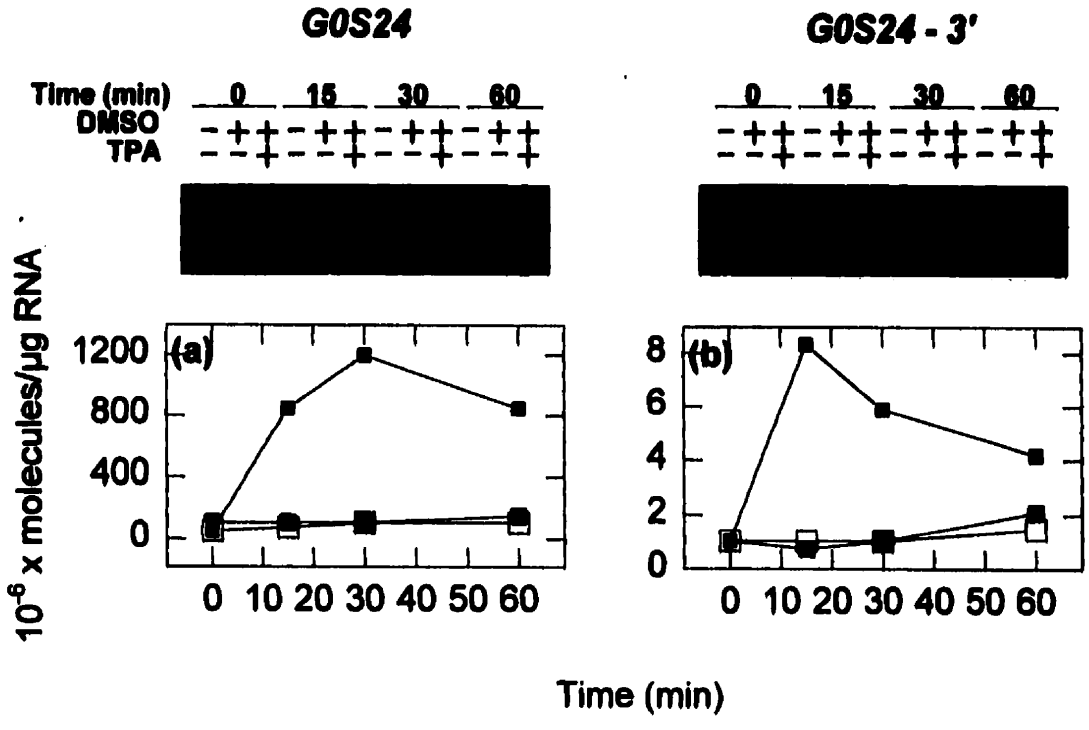
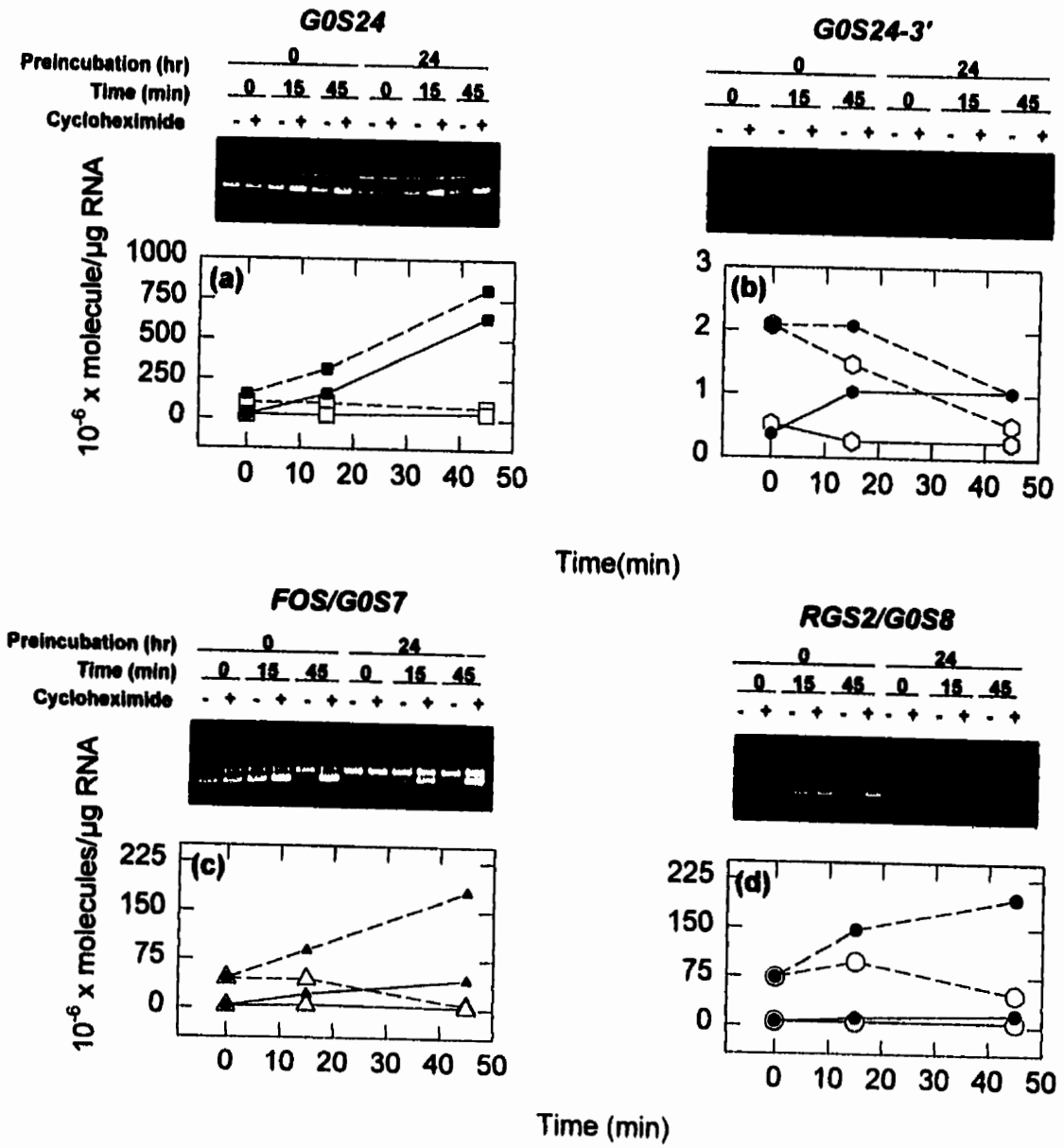

G0S24
G0S24 - 3'
Time (min)
0
15
30
60
DMSO
TPA
(a)
(b)
1200
800
400
0
8
6
4
2
0
0 10 20 30 40 50 60
10^{-6} x molecules/µg RNA
Time (min)

Figure 4-10. **Effects of cycloheximide (filled symbols) and water controls (open symbols) on levels of *G0S24* mRNA (a), *G0S24*-3' extended RNA (b), *FOS* mRNA (c) and *G0S8/RGS2* mRNA (d) in either freshly isolated cells (dashed lines), or preincubated cells (continuous lines).** Photographs (upper) show expression profiles in cells treated with 12.5 μl of either cycloheximide solution (final concentration 0.1 mM) or water (control), and incubated for a further 0, 15,or 45 min. Other details are as in Fig. 4-1.


G0S24
Preincubation (hr)
Time (min)
Cycloheximide
0
24
0 15 45 0 15 45
- + - + - + - + - +
10^-6 x molecule/µg RNA
1000
750
500
250
0
(a)
0 10 20 30 40 50
G0S24-3'
3
2
1
0
(b)
0 10 20 30 40 50
Time(min)
FOS/G0S7
10^-6 x molecules/µg RNA
225
150
75
0
(c)
0 10 20 30 40 50
RGS2/G0S8
225
150
75
0
(d)
0 10 20 30 40 50
Time (min)

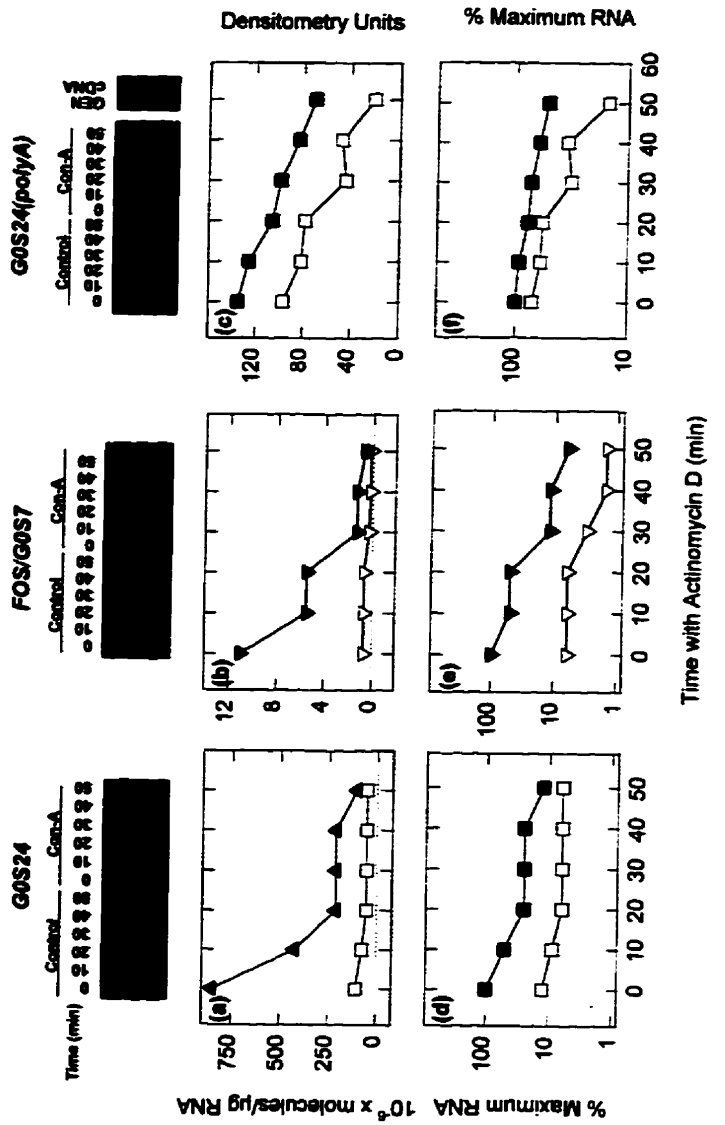
approached levels in similarly treated freshly explanted cells. The kinetics of the 3' extended RNA differ greatly from those of the *G0S24* mRNA (Fig. 4-10a). In freshly isolated cells, absolute levels of *G0S24* mRNA *increased* in a *rectilinear* fashion. After 24 hr preincubation, when control *G0S24* mRNA levels had decreased significantly (Table 2-1), the response to cycloheximide was still quite large and showed a similar rectilinear pattern. The response profile for the *G0S24* mRNA was similar to that of *FOS* mRNA (Fig. 4-10c; ref. 4), whose transcription rate has been shown to be induced by a novel cycloheximide-dependent mechanism (9, 195), but differed greatly from the response of *RGS2/G0S8* (Fig. 4-10d) whose levels appeared to increase to a lesser degree in response to cycloheximide in preincubated cells (5, Chapter 2).

RNA half-life studies

Figs. 4-11a,b shows the half-life kinetics of *G0S24* and *F0S* mRNAs following addition of the inhibitor of transcription, Actinomycin D. When plotted on an exponential scale, profiles for *G0S24* mRNA (Fig. 4-11d) could be fitted to two lines, whereas *FOS* (Fig. 4-11e) and G0S8/RGS2 (5, Chapter 2) mRNA profiles and were best represented by a single line. For *G0S24* mRNA, between 0 and 20 min following Actinomycin D treatment, a half-life value of 20.4±8.6 min (n=3) in control and 10.0±0.2 min (n=3) in stimulated cultures was observed. In the second phase, (20 to 50 min), slower rates of breakdown (half-life >>30 min) were observed in both stimulated and unstimulated cells.

To assess whether the stable species corresponded to a polyadenylated *G0S24* mRNA species the same RNA samples were reverse transcribed using oligo-dT, and amplified

Figure 4-11. **Messenger RNA degradation kinetics in preincubated cultures for *G0S24* (a,d) and *FOS/G0S7* (b,e) mRNAs or the poly(A) tail-containing fraction of *G0S24* mRNA (c,f).** Con-A stimulated cultures (filled symbols) and control cultures (open symbols) were incubated at 37°C for 0.5 hr prior to addition of 4 μM Actinomycin D (time defined as zero). In a,b,d,e, RNA levels are derived from competitive RT-PCR assays as described in Fig. 4-1 and plotted either as absolute quantities (molecules/μg total RNA) on a linear scale (upper panels), or relative to the maximum level in Con-A-treated cultures (defined as 100%) on a logarithmic scale (lower panels). In c,f *G0S24* polyadenylated mRNA-specific products (see Methods) were amplified from oligo(dT)-primed first strand products and relative levels quanitated by densitometric scanning of a negative of the picture shown above. RNA levels are plotted as densitometry units on a linear scale or as relative to maximum on logarithmic scales as described above. *G0S24* cDNA (polyadenylated) and genomic DNA ("GEN"; including 3' flank region) clones were included as positive and negative controls respectively.


Densitometry Units
% Maximum RNA
G0S24(polyA)
FOS/G0S7
G0S24
Control
Con-A
Time (min)
(a)
(b)
(c)
(d)
(e)
(f)
% Maximum RNA
10⁵ x molecules/µg RNA
Time with Actinomycin D (min)

between primer 1 (see Fig. 4-5, middle panel) and a primer specific for the 6 nt at the 3' end of the polyadenylated cDNA (Fig. 4-11c). These primers did not efficiently amplify a band from genomic DNA templates (Fig. 4-11, picture insert-lane GEN).Thus, the only product which could be amplified under the conditions used obtained must have had both a poly(A) tail and the appropriate sequence at the 3' end (Fig. 4-11, picture insert- lane cDNA). When plotted on an exponential scale, a linear profile was observed for the polyadenylated *G0S24* mRNA (Fig. 4-11e). By this criterion, it seems that the stable *G0S24*-derived RNA (detected using primers 5-4) did *not* correspond to polyadenylated *G0S24* mRNA. Relatively high levels of the polyadenylated *G0S24* mRNA were observed in unstimulated, compared to stimulated, cells consistent with the a high level of constitutive expression of a relatively unstable RNA species (Fig. 4-11c).

DISCUSSION

G0S genes were originally cDNA-cloned from freshly isolated blood mononuclear cells cultured for 2 hr with Con-A and cycloheximide in order to identify genes whose expression was specifically required for progression through the G_0/G_1 switch (31). Although work in murine fibroblast systems showed dramatic increases in *G0S24* mRNA levels following restoration of serum to quiescent cultures, in our system, only modest induction was observed in response to Con-A alone (34). We have previously shown that the cell purification procedure itself results in spontaneous transient increases of several *G0S* gene mRNAs in freshly explanted cells (4, 5, Chapter 2). Thus, in freshly explanted cells, the responses to Con-A are often masked by high levels of *G0S* mRNAs which subsequently

decrease upon incubation in culture. RT-PCR analysis of *G0S24* mRNA levels in freshly isolated and preincubated cells confirms this hypothesis and suggests that this gene responds in a similar manner to the purification procedure (Table 4-2, Fig. 4-1).

This work points to significant differences in the baseline levels of *G0S24* and other immediate-early genes in freshly explanted, compared to preincubated, mononuclear cells in culture. Although it is possible that the disparity between baseline levels of these RNAs in freshly isolated and preincubated cells indicates high *in vivo* expression levels which gradually are lost during long periods of incubation, it appears more likely that these levels represent *spontaneous* stimulation of these genes during one or more steps of the purification procedure (4). This effect is transient and does not represent lymphocyte activation *per se*, since cells in freshly explanted cultures cells, unlike Con-A stimulated cells, do not progress into G_1 and proliferate (no increased incorporation of uridine or thymidine).

Cells preincubated for 24 hr in the absence of a stimulus show high levels of *G0S24* mRNA compared to the immediate-early genes *FOS* and *FOSB* (Table 2-1, ref. 4). Such high levels are also observed in stimulated cultures after extended periods of culture. High levels of *G0S24* mRNA in mature murine blood cell types, including B- and T-lymphocytes have been previously attributed to constitutive expression of this gene (186), but the nature of this cell-type specific regulation has not been investigated.

Two signals which are required for the activation of T lymphocytes through the TCR/CD3 complex are PKC activation and an increase in the levels of intracellular calcium. These signals can be mimicked *in vitro* by the addition of TPA and the calcium ionophore, ionomycin. The weak responsiveness of G0S24 and FOS to ionomycin, compared to their

strong responses to TPA (Fig. 4-3a) suggest a similar PKC-dependent activation pathway for these two genes in Con-A-treated cells. In contrast, FOSB mRNA levels are highly sensitive to ionomycin and expression of this gene in stimulated cultures may require calcium-dependent signalling pathways (Fig. 4-3b; ref. 5, Chapter 2).

Characterization of *G0S24* mRNA expression using primers located throughout the gene (Fig. 4-5, upper and middle panels) has led to the identification of a 3' extended RNA and a species which spans the exon1/intron junction. Unsuccessful attempts to obtain RT-PCR products corresponding to the intron and the intron/exon2 junction suggested that either RT-PCR within this region was inefficient or that these cells contained very low levels of unspliced material. We favour the latter case since we were unable to generate comparable levels of RT-PCR products from two distinct downstream regions in the intron (Fig. 4-5, upper panel). We originally predicted that the 189 bp exon1/intron product corresponded to a paused transcript, similar to those observed for *c-fos* and *c-myc* (196). As previously described for prokaryotic and HIV genes, such pausing might be expected to occur proximal to, and downstream of, a strong dyad repeat sequence located 400 bp downstream of the exon1/intron junction (197). Another potential explanation for the lack of RT-PCR product across this dyad region, however, comes from the recent characterization of a double-stranded RNAse activity for a CCCH-containing protein in *Drosophila melanogaster* (8). Cleavage of the *G0S24* preRNA hairpin by such a protein would allow postranscriptional control of its expression and raises the intriguing possibility that the G0S24 CCCH-containing protein may autonomously regulate its own expression.

Methylated CpG dinucleotides are highly mutable. Thus, the presence of a DNA region rich in these sequences (CpG island) is thought to indicate local protection from methyltransferase activity. Such islands are often found in promoters for genes actively expressed in the germline where proteins constitutively bound to upstream activation sequences maintain a state of local hypomethylation. As previously reported (2, Chapter 3; Fig. 3-4), the *G0S24* promoter and intron are found within a region rich in CpG dinucleotides. Another CpG-island has now been identified immediately downstream of this gene (Fig. 4-5, lower panel). This highly unusual location suggests the presence of another nearby promoter/gene (perhaps for rRNA), in which case efficient 3' end processing of the *G0S24* transcript would be important. Alternatively, the 3' CpG island might be important for *G0S24* regulation.

Pre-mRNA processing at the 3' end has been shown to require a specific AAUAAA polyadenylation signal (for review see 192). Such processing may be enhanced by the presence of downstream sequences such as the U-rich element (UUUUU) located ~65 bp downstream of the AAUAAA signal in the SV40 virus late gene (191). The presence of a similar U-rich motif in the *G0S24* 3' flank (+ 61 nt) suggested the potential for efficient termination of this gene (Fig. 4-4).

It is of interest that the G0S24 3' flank sequence contains a donor splice site at position +167 bp relative to the polyadenylation signal (Fig. 4-4). Work on alternative processing of the calcitonin/calcitonin gene-related peptide (CT/CGRP) has identified a donor splice signal as part of a polyadenylation enhancer element 168 nt downstream of the site of polyadenylation (198). Activation of polyadenylation was shown to require expression of U1

snRNA implicating a role for U1 snRNPs in polyadenylation. Although future studies will be required to define the role of this sequence in efficient 3' end processing, these data suggested the potential involvement of several downstream elements in this process.

The low abundance 1.9 kb species on Northern blots, probed with 3' flank sequence, and subsequent RT-PCR mapping of the 3' end of an extended nuclear RNA species to a region 76-116 nt downstream of the polyadenylation signal (Figs. 4-6 and 4-7) supported this strong early termination hypothesis. It was of potential significance that this putative 3' termination site mapped to a region upstream and proximal to sequences containing an element similar to the human Sal-box motif and two regions with significant similarity to the human rRNA origin of transcription (Fig. 4-4). The Sal box sequence, through its binding to the transcription termination factor, hTTF-I, prevents read-through transcription of RNA polymerase I into the next tandemly arranged rRNA transcription unit (190). The analogous protein in yeast, Reb1p, is predicted to terminate transcription by mediating a pausing of RNA polymerase in the context of a release element (199). As the pausing of this is not specific to RNA polymerase I, we wondered whether a potential pause site in the *G0S24* 3' flank might be important for 3' end processing of the *G0S24* mRNA.

Although the process by which 3' end cleavage and termination are coupled during mRNA processing remains unknown, a growing body of evidence supports the notion that RNA polymerase II pausing plays an important role in this process (200). It has been shown that transcription termination depends on 3' end cleavage and that pausing of the RNA polymerase II complex at a site(s) downstream of the cleavage site increases the efficiency by which polyadenylation signals are used (201). That a *G0S24* Sal box-like sequence did

not bind *in vitro* to translated hTTF-I (Fig. 4-9) suggests either that hTTF-I is not involved, or that sequences in the adjacent regions contain the necessary pause/release elements or that pausing is not required for efficient *G0S24* RNA cleavage and termination. The existence of another hTTF-I-like factor with slightly different DNA-binding characteristics has not been reported, but the results in this study may warrant the search for such a factor.

To study the expression kinetics of the 3' extended RNA species, we used competitive RT-PCR with primers spanning the site of cleavage/polyadenylation. Assuming normal efficiency of the RT-PCR assay, levels of the 3' extended RNA species in freshly explanted and preincubated control cultures were much lower than those for the *G0S24* mRNA baseline, consistent with the notion that this species represents an unprocessed RNA. Secondly, if there were a relatively high level of activity of the *G0S24* promoter, and a correspondingly high density of RNA polymerase complexes in the 3' flank region then we might expect to detect relatively high levels of the 3' extended RNA. Consistent with this notion, levels of the 3' extended *G0S24* RNA were higher in TPA-treated and freshly explanted cells, when high levels of *G0S24* transcription were expected. Furthermore, peak levels of the 3' extended RNA expression were observed following 15 min TPA stimulation, the time point when maximum transcription is expected with immediate-early genes such as *c-fos* and *G0S24*. This is further supported by the fact that in preincubated cultures, when cells have recovered from the purification procedure, levels of the 3' extended species have reached a stable baseline. Thus, the specific measurement of such species may, therefore, provide a semi-quantitative measure of RNA polymerase II density in this region of the transcription unit.

Since cleavage precedes termination, and we are amplifying across the cleavage site, we do not expect any contribution to our assessed levels of the 3' extended RNA from a pool of released *G0S24* pre-mRNA species. Therefore, two possible lines of reasoning could be used to explain increases of the 3' extended RNA. In the first case, the cleavage step is rate limiting in cells actively transcribing the *G0S24* gene. In the context of the pausing/termination model, this would mean either that polymerase complexes are in queue upstream of the pausing element awaiting cleavage of the RNA at the pause site, or that there is an inhibition of polymerase II pausing (cleavage potentiation, for review see Wahle, 1995) and a corresponding increase in the length of the transcription unit in stimulated cells. To date, however, we have been unable to detect measurable levels of any products further than 116 bp from the poly(A) site by RT-PCR or Northern blotting with RNA from stimulated or control cultures.

In the second case, the cleavage rate is not a limiting factor. Increases of a 3' extended RNA under these conditions would simply indicate an increased number of unfinished RNA products which had been made from transcription complexes which had transcribed through the region containing the reverse transcriptase primer but had not yet been processed.

According to this model, on an actively transcribed gene, we would expect to detect an increase in the population of nascent unfinished RNA species and a greater number of these aborted species should by chance be detectable by reverse transcriptase primers located closer to the polyadenylation site. Support for this notion came from a number of competitive RT-PCR analyses of 3' extended *G0S24* RNA levels in samples from stimulated *and* control cultures. First, when samples were reverse transcribed from the standard primer (primer 14)

or from a position 12 bp further upstream, greater levels of this RNA were detected in cDNA templates made from the upstream primer.

Such interpretation may also provide an explanation of the unusual cycloheximide-dependent increases of *G0S24* mRNA in mononuclear cells. We have reported recently (5, Chapter 2) on the near-rectilinear increases of *G0S24* mRNA in response to cycloheximide. While *FOS* mRNA shows a similar response pattern (Fig. 4-11), RGS2/G0S8 mRNA levels in the same RNA sample begin to plateau suggesting cytoplasmic stabilization of the mRNA in the absence of renewed transcription. These differences suggest a nuclear mechanism for the increase in levels of *G0S24* and *FOS* mRNAs. In run-on transcription assays, it has been shown that cycloheximide increases the rate of transcription from the *FOS* gene promoter (9, 195). It is postulated that this response represents a novel cycloheximide-dependent signal transduction pathway dependent on the increased phosphorylation of two nuclear proteins, histone H3 and an unidentified chromatin-associated protein, pp33 (9). In our system, the cycloheximide-dependent increase in levels of the nuclear 3' extended *G0S24* RNA species in preincubated cells is consistent with a similar transcription activation hypothesis for the *G0S24* gene (Fig. 4-11). This is the first suggestion that such a mechanism may also exist for this gene.

The degradation of immediate-early gene mRNAs, including those for *FOS* (Wilson and Treisman, 1988) and murine *TTP/G0S24* (181), are mediated by AU-rich elements (AREs) in their 3'UTR regions. The first step in degradation of such species is the rapid synchronous shortening by deadenylation of the poly(A) tail to a size of 30-60 nucleotides after which time, the body of the RNA is degraded by apparent first order kinetics (203).

When we used competitive RT-PCR to study the stability of the G0S24 mRNA in Actinomycin D-treated cultures the unusual apparent biphasic decay curve for this mRNA, but not for *FOS* mRNA, suggested either particular sensitivity of the *G0S24* gene to the transcription inhibitor. Although this inhibitor is a powerful intercalating agent, capable of stabilizing certain cytoplasmic ARE-containing messages (204), the linear degradation of polyadenylated *G0S24* mRNA suggested that stabilization of *G0S24* ARE sequences did not result in the observed degradation pattern (Figs. 4-11c,e). Half-life values from the first and rapid phase agreed closely with the values obtained by workers in the murine field (14) and suggested degradation during this period by the normal ARE-dependent mechanism. The slow breakdown phase (half-life >> 30 min), however, was not expected and suggested the presence of a second, more stable *G0S24*-derived RNA species in Actinomycin D-treated cells. Such high levels of the apparent stable *G0S24* mRNA may represent incomplete nascent *G0S24* pre-mRNA chains from transcription complexes stalled at sites of Actinomycin D intercalation (205, 206).

CHAPTER 5

GENERAL DISCUSSION

G0S gene mRNA levels increase during G_0/G_1 switch in T-lymphocytes

An underlying assumption of this work is that the genes such as *G0S24* and *G0S8*, which were identified on the basis of their differential expression during the G_0/G_1 switch, may play important regulatory roles during the activation process. This further implies that the protein products of these genes may also be required during this critical regulatory period. In the absence of specific antibody probes for the G0S24 and G0S8 proteins, we have been unable to show changes in the levels of these proteins during the G_0/G_1 switch. Given the modest relative change in mRNA levels observed for some of the G0S genes studied in this work, the development of such tools will be an important step toward defining a role for these products during the activation process.

We must be careful, however, not to ignore the possibility that the functional product of genes involved in such regulatory processes *is* the polyadenylated RNA species. This is indeed the case for a gene expressing meiosis-specific RNA product (meiRNA) in *S. Pombe* (150).

Competetive-PCR analysis of G_0/G_1 switch gene mRNAs

In relatively small sized primary cultures of human mononuclear cells, the detection of baseline, and even induced, levels of mRNAs for *normally* expressed genes using the standard Northern blotting technique is often difficult (1, 34). Since the analysis of a range of culture conditions for each cell preparation requires the use of multiple small-sized

samples, a major objective of this work was to develop a sensitive assay which would allow semi-quantitative analysis of very low mRNA levels. We feel that the competitive RT-PCR assays are useful for this purpose as we can routinely detect mRNA baseline levels at or below 1 molecule/cell for several of the *G0S* genes studied (4, 5, Chapter 2). The sensitivity of this assay is further demonstrated by the fact that we can measure changes in the levels of 3' unprocessed nascent RNA product and correlate these to the expected transcriptional activity (i.e. polymerase density) on the *G0S24* gene 3' flank Figs. 4-9 and 4-10). The optimization of each step of this assay included: i) RNA purification from multiple small sized samples; ii) design and production of appropriate PCR and cRNA control constructs; and iii) matching *Taq* polymerase with optimal buffer conditions for detection of low-level mRNAs. This optimization was carried out largely by this author and was not a trivial exercise (see Appendix A1 for details). We are confident that these assays are accurate, reproducible, and representative of those which will be required in future studies to analysed *G0S*- and related-gene mRNA levels in small tissue samples from diseased individuals or transgenic animals (see RGS discussion).

RGS PROTEINS

Toward defining a specific biological role for RGS2/G0S8

We have identified a gene, *RGS2/G0S8,* whose mRNA levels increase in mononuclear cells treated for 2 h with Concanavalin-A and cycloheximide. (1, 3). The cDNA encodes a protein which belongs to a novel, rapidly growing family of regulators of G protein signalling (RGS proteins). RGS proteins have been recently shown to function as GTPase-activating proteins (GAPS) for the G_α subunits of heterotrimeric G proteins (35, 36) and thereby support the numerous physiological studies which have predicted a role for accessory proteins in G_α subunit regulation (207, 208, 209). Using the *RGS2/G0S8* cDNA clone characterized in this work, our collaborators have demonstrated the ability of this protein to modulate yeast (pheromone receptor) and mammalian (IL-8) receptor signalling pathways (39, 51).

To date, however, the G protein-coupled signalling pathways resulting from T-cell receptor activation are poorly understood, although potential evidence of a role for G_q in this activation pathway has been discussed (28). Thus, it was our contention that much could be learned about the specific biological function of RGS2/G0S8 protein by studying the conditions under which its mRNA was expressed. Indeed, the levels of the mRNA coding for the yeast RGS family prototype, Sst2p, are induced in response to pheromone treatment and suggest a potential feedback mechanism for regulation of this G protein-coupled signal transduction pathway in yeast. On the assumption that mammalian G protein signalling pathways might contain similar feedback mechanisms, and that the induction by specific lymphocyte pathways would provide information the potential G_α subunits RGS proteins

might specifically modulate, we compared the response to various mitogenic stimuli of the mRNA levels of *RGS1/BL34/1R20* (which is now known to specifically regulate G_i and G_q subtypes; 35, 210) and *RGS2/G0S8*, for which the specific biological target was previously unknown. Our studies suggested that these two RGS mRNAs are differentially expressed in mononuclear cells in response to PKC activation and changes in intracellular calcium levels, two typical lymphocyte activation signals. In particular, *RGS2/G0S8* mRNA levels were sensitive to changes intracellular calcium but relatively insensitive to PKC-dependent signalling pathways whereas the opposite was true for RGS1 (5, Chapter 2). These data led to our suggestion that RGS2/G0S8 protein might be specifically involved in the regulation of calcium signalling through modulation of G_q, a known regulator of Phospholipase C. In collaboration with M. Linder at Washington University at St. Louis, we have now shown *specific* interaction of RGS2 with the G_q subunit.

Since unlike the other RGS family members tested so far, RGS2 does not appear to regulate G_i subtypes, these data create the exciting possibility that RGS2 is specific for the G_q subfamily. Future studies will require confirmation of these data using purified recombinant RGS2 and G_q in biochemical GTPase assays to demonstrate GAP activity for this protein. The development of such assays is not an easy exercise since this particular G_α subunit has unusual GTP binding and hydrolysis characteristics which require that GAP assays be done with receptors and G proteins reconstituted in lipid vesicles.

FUTURE DIRECTIONS

Potential functions of RGS2 and other RGS proteins in normal hematopoietic cells

At present, very little is known about which of the 17 known RGS proteins are expressed in lymphoid or myeloid tissues, whether their levels are constitutive or inducible, and whether the function of these proteins may be regulated post-translationally. It will be important to develop competitive RT-PCR strategies to assay for expression of each of the known RGS mRNAs. Such tools will allow us to gain a better understanding of which of these genes are involved in the regulation of normal lymphoid function and proliferation. Furthermore, as there are currently no good antibodies for any of the RGS proteins (see Discussion ref. 5, Chapter 2), development of such tools will be essential for the definition of patterns of RGS protein expression and localization. The production of monoclonal antibodies raised against various RGS proteins will likely be required, since polyclonal antibodies (even after affinity purification) are often not specific for the RGS proteins against which they were raised.

RGS regulation of chemokine signalling

To date, there has not been developed a good physiological model in which to study RGS protein function. The regulation of chemokine signalling may provide an ideal system to investigate these issues as a role of RGS1-4 proteins in the control of IL-8-dependent signalling has been recently demonstrated (51). In fact, many of the chemokine receptors belong to the superfamily of seven transmembrane spanning receptors which exert their effects through G protein-coupled mechanisms (for review see Ben *et al.*, 1995). Stimulation

of these receptors by chemoattractant ligands such as IL-8, results in a number of measurable physiological changes including cell motility and actin assembly, changes in intracellular Ca^{++} levels, and production of cytokines. Early attempts to study RGS function will be aimed at the overexpression of RGS proteins and determining their effects on these processes. The successful development of such a model system will allow the future study of the ability of RGS proteins to control: sensitivity to agonist stimulation; specific requirements for receptor desensitization; ligand discrimination (many of these receptors bind more than one ligand); and efficiency of chemotactic migration.

RGS regulation of cell differentiation and proliferation

Another aim of studying genes which are induced during the G_0/G_1 switch is to identify genes, which when abnormally expressed (either by deletion or hyperactivation), may lead to errors in proliferation or development of hematopoietic cells. It may be that the RGS1 and RGS2 proteins are required for the modulation of specific G protein activity during critical decision-making periods of these cells. Indeed, our data suggest a role for these proteins in crossing the G_0/G_1 threshold during normal lymphoid proliferation (5, Chapter 2). Furthermore, pertussis toxin (an inhibitor of G_i subtypes) may stimulate the proliferation of monocytes and T-cells (212, 213). Other recent data suggests a role for specific G proteins during myeloid development as $G_{i\alpha 2}$-deficient mice display defects in thymocyte development and function (214). Thus, it is possible that RGS proteins could promote or inhibit proliferation of hematopoietic cells, depending on the function of the specific G proteins in the cells type being examined.

These findings may also be important to the interpretation of data which suggests that abnormally high levels *RGS1* or *RGS2* expression may promote leukemia in humans (43, 45). Abnormally high levels of *RGS2* mRNAs have been identified in various hematopoietic malignancies including acute myelogenous leukemia (AML), acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), and chronic myelogenous leukemia (CML). The human *RGS1* and *RGS2* genes may also be implicated in hematopoietic malignancies on the basis of their chromosomal localization to loci at chromosome 1q31(44, 45). Isochromosomes of the long arm of chromosome 1 and duplications of the 1q31 region frequently occur as secondary genetic aberrations in ALL, Hodgkin's disease, non-Hodgkin's lymphoma (NHL), and chronic myeloproliferative disorders (MPD). It remains to be established whether such abnormalities directly affect the expression or structure of RGS proteins. A study in which transgenic RGS overexpressor mice were generated (placing *RGS* genes under control of differentiation-specific promoters) would allow rapid identification of any leukemogenic properties of these proteins.

It should be noted that RGS1 and RGS2 overexpression is not restricted to hematopoietic malignancies. *RGS2* mRNA is also found to be highly expressed in cervical carcinoma- (HeLa), ovarian cancer-(OVICAR-3), and brain glioblastoma-derived (A-172) cell lines (45). Thus, these genes may contribute to oncogenic transformation in various human cell types. For this reason, it is also of interest that the *RGS2* gene maps directly to genetic markers for hyperthyroidism jaw tumour syndrome (86, 87; M.R. Hobbs, personal communication) and region implicated in development of prostate cancer (215). Characterization of *RGS* gene mRNAs in tissues from patients of these disorders will be a

necessary first step toward identifying a potential role for these genes in carcinogenesis in non-lymphoid tissues. Such studies will require a rapid sensitive method to analyse these levels such as the RT-PCR assay developed in this work.

CCCH CONTAINING PROTEINS

Toward defining a role for G0S24

Based on the nuclear localization and novel CCCH zinc-finger structures of the murine G0S24 protein, most of the early studies on this gene, including our own characterization of human *G0S24* gene (2, Chapter 3), suggested a role for this gene as a transcriptional regulator. Despite repeated attempts by our lab (Figure 3-9) and other groups working on CCCH-containing proteins, there has been no definitive description of double-stranded DNA binding activity which would support this original hypothesis.

Although we cannot rule out a role for these proteins in transcriptional modulation, the surge of recent data suggests that these proteins may be involved in binding to and processing of single-stranded nucleic acids. In an expression cloning screen, designed to identify genes which encode ssDNA binding proteins from an *Drosophila melanogaster* ovarian cDNA library, the Clipper (CLP) gene was first identified (216). Subsequent cloning of the full length cDNA revealed the presence of 5 CCCH domains whose presence was required for binding to and cleavage of double stranded RNA hairpins. Secondly, a nuclear CCCH-containing protein in *C. elegans*, PIE-1, is responsible for the germline-specific repression of genes whose expression is required for somatogenesis (178, 179). These authors suggest specific degradation (RNAse) of newly transcribed RNAs from these genes

as a potential mechanism for PIE-1 protein activity (179). If such a mechanism exists for repression of specific gene expression, it is likely that it has been conserved throughout evolution as similar motifs are being identified in genes in a wide range of organisms including mammals, nematodes, plants and dsRNA viruses (Tables 4-1 and 4-2). It is of interest that in the case of two of these other CCCH proteins, they been implicated in RNA processing at the level of 5' (suppressor of sable; ref. 172) or 3' (U2AF25; ref. 194) splice site selection.

It should be pointed out that the CCCH domains in these latter proteins often contain either a shorter (4-7 amino acids), or longer (9 amino acids) first knuckel compared to the immediate G0S24 family of mammalian and yeast proteins (8 amino acids). Although recent NMR structure of this region predicts a loop for this sequence, which is not likely to interact with the rest of the finger structure (185), it remains to be established how changes in the length of this knuckle would affect binding site specificity or dsRNAse activity. Thus it is possible that all of these proteins, despite the presence of very similar structural domains, carry out a wide variety biological roles. Work being carried out presently in the *C. elegans* CCCH project may offer the greatest potential for rapid advancement of this field. Since the target genes (mRNAs negatively regulated in PIE-1 overexpressors embryos) of the PIE-1 protein have already been identified using genetic studies (179), they will be able look at these specific pre-RNAs for effects of PIE-1 binding and processing.

In preliminary ssDNA binding studies we appear to have demonstrated the ability of recombinant $(HIS)_{10}$-G0S24ZF (CCCH domain) to shift a population of species within a library of randomly generated DNA oligonucleotides (A. Cristillo and S. Heximer,

unpublished data). This binding appears to be specific (shifts only a small proportion of a random library) compared to that for gene 32 protein which binds ssDNA nonspecifically. If binding is reproducible and can be "supershifted"with G0S24-specific antibodies, further rounds of site-selection will be used to identify the specific binding sequence.

High level of constitutive expression of G0S24 mRNA in mononuclear cells

Our studies suggest that *G0S24* mRNA levels may be highly expressed in resting human mononuclear cells. Under such conditions, it is also likely that the protein levels would be maintained at relatively high levels, although this proof of this will await production of good antibodies against the human protein. If we use the PIE-1 protein-mediated repression of somatogenesis in *C. elegans* as a potential model system, we can imagine a similar role for G0S24 in repression of immediate-early gene RNAs during the quiescent (G_0) phase of the cell cycle. The reported nuclear to cytoplasmic translocation of the murine G0S24 protein following mitogenic stimulation (186) might relate to rapid derepression of genes involved in G_1 entry and cell-cycle progression. If G0S24 mediates its effects through an RNA-binding activity, then two different mechanisms for such repression can be imagined based on its similarity to other CCCH domain-containing proteins: i) it acts as a double stranded RNAse resulting in degradation of target mRNAs; or ii) it acts as a regulator of splice site selection resulting in the production of nonfunctional RNAs.

Such a mechanism might also allow for autologous regulation of the G0S24 gene, as is the case U1A snRNP and other splicing proteins. Indeed, our identification of an RT-PCR product across the exon 1/intron despite our inability to identify downstream intronic

products is consistent with either unusual splice selection from a cryptic intronic site or dsRNAse digestion within the intronic region containing the dyad repeat. Although much work is needed to characterize this unusual expression pattern, preliminary RNA localization data suggests a role for alternative splicing since the exon1/intron species (which if translated would produce a truncated protein without CCCH domains) is found in both the nuclear and cytoplasmic RNA fractions.

CONCLUSIONS

This work has described the characterization of two genes whose mRNA levels are increased when resting lymphocytes are treated for 2 hr with Con-A and cycloheximide. Although a specific role for these genes in the activation process remains to be established, recent work by our lab and others suggest an important role for them in growth-related events.

G0S8/RGS2 is a regulator of G protein signalling and will be a valuable tool for the future study of G protein involvement in TCR/CD3-mediated activation of tyrosine kinase- and calcium-dependent signalling events. We were among the first of several groups to identify a member of the RGS family and note its similarity to Sst2p, a negative regulator of G protein signalling in the yeast pheromone response pathway. Our report of the human genomic sequence remains the only such sequence in the database and will be an invaluable tool for future promoter analyses and studies of the role of this gene leukemogenesis. It was previously reported that mononuclear cells contained no detectable expression of RGS2/G0S8 mRNA (45). Owing to the sensitivity of the RT-PCR assay developed in this

work, we were able to demonstrate basal expression of RGS2 mRNA in blood mononuclear cells (5, Chapter 2). The potential for the constitutive expression of the RGS2/G0S8 and other RGS proteins may have a significant impact on our understanding of the general "state-of-sensitization" of particular cell types as this relates to their profile of RGS isoform expression. Furthermore, in this work is presented the first suggestion of activation of the RGS2/G0S8 gene through an intracellular Ca^{++}-dependent pathway, an observation which may lead to our understanding of the specific G protein targets of the RGS2 protein. Although the specific biological target for this protein remains unknown, our suggestion of its role in Gq modulation has led to fruitful collaborations with leading G protein biochemistry laboratories including J. Hepler (Emory University) and M. Linder (Washington University).

G0S24 appears to belong to a very large family of RNA binding/processing proteins, the timely regulation of which may play a significant role in regulation of mRNA expression at the posttranscriptional level. Our genomic sequence data point to the presence of an unusual CpG island in the 3' flank of the *G0S24* gene, and suggest the need for further studies to uncover the role of these sequences in regulation of its expression, particularly at the level of transcription termination. Using Northern blotting and RT-PCR, we have mapped the upstream (5' RACE) and downstream boundaries of the transcription unit. The mRNA expression data suggest an unusually high basal level of activity from the *G0S24* promoter, and that this activity may be greatly increased through PKC- and a less well defined cycloheximide-dependent signal transduction cascade. These studies also suggest the presence of an unusual *G0S24*-derived RNA species containing exon1 and 5' intronic, but

lacking 3' intronic sequence. The need for further studies to uncover potential splice variants or transcriptional pause sites in the G0S24 mRNA is discussed.

BIBLIOGRAPHY

1. Siderovski, D.P., Blum, S., Forsdyke, R.E., and Forsdyke, D.R. (1990) A set of human putative lymphocyte G_0/G_1 switch genes includes genes homologous to rodent cytokine and zinc-finger protein-encoding genes. *DNA Cell Biol.* **9**, 579-587.

2. Heximer, S.P., and Forsdyke, D.R. (1993) A Human Putative Lymphocyte G_0/G_1 Switch Gene Homologous to a Rodent Gene Encoding a Zinc-Binding Potential Transcription Factor. *DNA and Cell Biol.* **12**, 73-88.

3. Siderovski, D. P., Heximer, S. P., and Forsdyke, D. R. (1994) A human gene encoding a putative basic helix-loop-helix phosphoprotein whose mRNA increases rapidly in cycloheximide-treated blood mononuclear cells. *DNA Cell Biol.* **13**, 125-147.

4. Heximer, S.P., Cristillo, A.D., Russell, L., and Forsdyke, D.R. (1996) Sequence analysis and expression in cultured lymphocytes of the human *FOSB* gene (G0S3). *DNA Cell Biol.* **15**, 1025-1038.

5. Heximer, S.P, Cristillo, A.D., and Forsdyke, D.R. (1997) Comparison of mRNA expression of two regulators of G protein signalling, RGS1/BL34/1R20 and RGS2/G0S8, in cultured human blood mononuclear cells. *DNA Cell. Biol.***16**, 589-598.

6. Heximer, S.P., Russell, L.R., and Forsdyke, D.R. (1997) RT-PCR characterization of *G0S24* in human blood mononuclear cells. (In preparation, Chapter 4)

7. Dubois, R.N., McLane, M.W., Ryder, K., Lau, L.F., and Nathans, D. (1990) A growth factor-inducible nuclear protein with a novel cysteine/histidine repetitive sequence. *J. Biol. Chem.* **265**, 19185-19191.

8. Bai, C., and Tolias, P.P. (1996) Cleavage of RNA Hairpins Mediated by a Developmentally Regulated CCCH Zinc Finger Protein. *Mol. Cell. Biol.* **16**, 6661-6667.

9. Edwards, D.R., and Mahadevan, L.C. Protein synthesis inhibitors differentially superinduce *c-fos* and *c-jun* by three distinct mechanisms: lack of evidence for labile repressors. *EMBO J.* **11**, 2415-2424.

10. Baserga, R. (1990) The Cell Cycle: Myths and Realities. *Cancer Research* **50**, 6769-6771.

11. Zipfel, P.F., Irving, S.G., Kelly, K., and Siebenlist, U. (1989) Complexity of the primary genetic response to mitogenic activation of human T-cells. *Mol. Cell. Biol.* **9**, 1041-1048.

12. Lau, L.F., and Nathans, D. (1991) Genes induced by serum growth factors. *Mol. Asp. Cell. Reg.* **6**, 257-332.

13. Baserga, R. and Surmacz, E. (1987) Oncogenes, cell cycle genes, and the control of cell proliferation. *Bio/Technology* **5**, 355-358.

14. Lai, W.S., Stumpo, D.J., and Blackshear, P.J. (1990) Rapid Insulin-stimulated Accumulation of an mRNA Encoding a Proline-rich Protein. *J. Biol. Chem.* **265**, 16556-16563.

15. Johnson, R.T., Downes, C.S., and Meyn, R.E. (1993) In: The Cell Cycle: A Practical Approach. eds. Fantes, P., and Brooks, R. Oxford University Inc., New York, USA.

16. Crabtree, G.R. (1989) Contingent genetic regulatory event in T lymphocyte activation. *Science* **243**, 355-361.

17. Janeway, C.A., and Goldstein, P. (1992) Lymphocyte activation and effector functions. *Curr. Opin. Immunol.* **4**, 265-270.

18. Janeway, C.R. (1992) The T cell receptor as a multicomponent signalling machine:CD4/CD8 coreceptors and CD45 in T cell activation. *Ann. Rev. Immunol.* **10**: 645-674.

19. Weissman, A. (1994) The T cell antigen receptor: A multisubunit signalling complex. In Lymphocyte Activation. *Chemical Immunology*: **59**, 1-18. ed. Samuelson, L.E., Basel, Karger.

20. Reth, M. (1989) Antigen receptor tail clue. *Nature* **33**, 383-384.

21. Weiss, A. (1993) Tcell antigen receptor signal transduction: A tale of tails and cytoplasmic protein-tyrosine kinases. *Cell* **73**, 209-212.

22. Klausner, R.D. and Samuelson, L.E. (1991) T cell antigen receptor activation pathways: The tyrosine kinase connection. *Cell* **64**, 875-787.

23. June, C.H., Fletcher, M.C., Ledbetter, J.A., and Samuelson, L.E. (1990) Increases in tyrosinekinase phosphorylation are detectable before phospholipase C activation after T cell receptor stimulation. *J. Immunol.* **144**, 1591-1599.

24. Chan, A.C., Iwashima, M., Turuk, C.W., and Weiss, A. (1992) ZAP-70: A 70 kD protein-tyrosine kinase that associates with the TCR ζ-chain. *Cell* **71**, 649-662.

25. Weiss, A., Koretzky, G., Schatzmann, R.C., and Kadlecec, T. (1991) Functional activation of the T cell antigen receptor induces tyrosine phosphorylation of phospholipase C gamma 1. *Proc. Natl. Acad. Sci.USA* **88**, 5484-5488.

26. Anderson, D.L., and Tsoukas, C.D. (1989) Cholera toxin inhibits resting human T cell activation via a cAMP-independent pathway. *J. Immunol.* **143**, 3647-3652.

27. Graves, J.D., and Cantrell, D.A. (1991) An analysis of the role of guanine nucleotide binding proteins in antigen receptor/CD3 antigen coupling to phospholipase C. *J. Immunol.* **146**, 2102-2107.

28. Stanners, J., Kabouridis, P. S., McGuire, K. L., and Tsoukas, C. D. (1995) Interaction between G proteins and tyrosine kinases upon T cell receptor/CD3-mediated signalling. *J. Biol. Chem.* **270**, 30635-30642.

29. Frank,S.J., Niklinska, B.B., Orloff, D.G., Mercep, M., Ashwell, J.D., and Klausner, R.D. (1990) Structural mutations reveal a novel role for the TCR ζ-chain in T cell activation. *Science* **249**, 174-177.

30. Tsien, R.Y.,Pozzan, T., and Rink, T.J. (1982) T-cell mitogens cause early-changes in cytoplasmic free Ca^{++} and membrane potential in lymphocytes. *Nature* **295**, 68-71.

31. Forsdyke, D. R. (1985) cDNA cloning of mRNAs which increase rapidly in human lymphocytes treated with concanavalin-A and cycloheximide. *Biochem. Biophys. Res. Commun.* **129**, 619-625.

32. Forsdyke, D.R. (1968) Studies of the incorporation of [5-^{3}H]-uridine during activation and transformation of lymphocytes induced by phytohaemagglutinin. *Biochem. J.* **107**, 197-205.

33. Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-269.

34. Russell, L., and Forsdyke, D. R. (1991) A human putative lymphocyte G_0/G_1 switch gene containing a CpG-rich island encodes a small basic protein with the potential to be phosphorylated. *DNA Cell. Biol.* **10**, 581-591.

35. Berman, D. M., Wilkie, T. M., and Gilman, A. G. (1996) GAIP and RGS4 are GTPase-activating proteins for the G_i subfamily of G protein α-subunits. *Cell* **86**, 445-452.

36. Watson, N., Linder, M. E., Druey, K. M., Kehrl, J. H., and Blumer, K. J. (1996) RGS family members: GTPase-activating proteins for heterotrimeric G protein alpha subunits. *Nature* (in press).

37. Lefkowitz, R. J. (1993) G protein-coupled receptor kinases. *Cell,* **74**, 409-412.

38. Wilson, C. J., and Applebury, M. L. (1993) Arresting G protein-coupled receptor activity. *Curr. Biol.* **3**, 683-686.

39. Siderovski, D. P., Hessell, A., Chung, S., Mak, T. W., and Tyers, M. (1996) A new family of regulators of G-protein-coupled receptors? *Curr. Biol.* **6**, 211-212.

40. Dohlman, H.G., Apaniesk, D., Chen, Y., Song, J., and Nusskern, D. (1995) Inhibition of G-Protein Signalling by Dominant Gain-of-Function Mutations in Sst2p, a Pheromone Desensitization Factor in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **15**, 3635-3643.

41. Murphy, J. J., and Norton, J. D. (1990) Cell-type-specific early response gene expression during plasmacytoid differentiation of human B lymphocytic leukemia cells. *Biochim. Biophys. Acta,* **1049**, 261-271.

42. Murphy, J. J., and Norton, J. D. (1993) Multiple signalling pathways mediate anti-Ig and IL-4-induced early response gene expression in human tonsillar B cells. *Eur. J. Immunol.* **23**, 2876-2881.

43. Hong, J. X., Wilson, G. L., Fox, C. H., and Kehrl, J. H. (1993) Isolation and characterization of a novel B cell activation gene. *J. Immunol.* **150**, 3895-3904.

44. Newton, J.S., Deed, R.W., Mitchell, E.L.D., Murphy, J. J., and Norton, J.D. (1993) A B cell specific immediate early human gene is located on chromosome band 1q31 and encodes an α helical basic phosphoprotein. *Biochim. Biophys. Acta,* **1216**, 314-316.

45. Wu, H-K., Heng, H.H.Q., Shi, X-M., Forsdyke, D.R., Tsui, L-C, Mak, T.W., Minden, M.D.,and Siderovski, D. P. (1995) Differential expression of a basic helix-

loop-helix phosphoprotein gene, *G0S8*, in acute leukemia and localization to human chromosome 1q31. *Leukemia* **9**, 1291-1298.

46. Chan, R.K. and Otte, C.A. (1982) Isolation and genetic analysis of *Saccharomyces cerevisiae* mutants supersensitive to G_1 arrest by a-factor and α-factor pheromones. *Mol. Cell. Biol.* **2**, 11-20.

47. Dietzel, C. and Kurjan, J. (1987) Pheromonal regulation and sequence of the *Saccharomyces cerevisiae SST2* gene: a model for desensitization to pheromone. *Mol. Cell. Biol.* **7**, 4169-4177.

48. Dohlman, H. G., Song, J., Ma, D., Courshesne, W. E., and Thorner, J. (1996) Sst2, a negative regulator of pheromone signalling in the yeast *Saccharomyces cerevisiae*: expression, localization, genetic interaction and physical association with Gpa1 (G protein α subunit). *Mol. Cell. Biol.* **16**, 5194-5209.

49. De Vries, L., Mousli, M., Wurmser, A., and Farquhar, M. G. (1995) GAIP, a protein that specifically interacts with the trimeric G protein $G\alpha_{i3}$, is a member of a protein family with a highly conserved core domain. *Proc. Natl. Acad. Sci. USA* **92**, 11916-11920.

50. Koelle, M. R., and Horvitz, H.R. (1996) EGL-10 regulates G-protein signalling in the *C. elegans* nervous system and shares a conserved domain with many mammalian proteins. *Cell*, **84**, 115-125.

51. Druey, K. M., Blumer, K.J., Kang, V.H. and Kehrl, J.H. (1996) Inhibition of G-protein-mediated MAP kinase activation by a new mammalian gene family. *Nature*, **379**-742-746.

52. Gardner, P. (1989) Calcium and T lymphocyte activation. *Cell* **59**, 15-20.

53. Crabtree, G. R., and Clipstone, N. A. (1994) Signal transduction between the plasma membrane and the nucleus of T lymphocytes. *Annu. Rev. Biochem.* **63**, 1045-1083.

54. Cristillo, A. D., Heximer, S. P., and Forsdyke, D. R. (1996) A "stealth" approach to inhibition of lymphocyte activation by oligonucleotide complementary to the putative G_0/G_1 switch regulatory gene *G0S30/EGR1/NGFI-A*. *DNA Cell Biol.* **15**, 561-570.

55. Molecular Cloning-A Laboratory Manual (2nd. edition) eds. Sambrook, J., Fritsch, E.F., and Maniatis., T. (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY.

56. Fitch, W.M., Smith, T.F., and Ralph, W.W. (1983) Mapping the order of restriction fragments. *Gene* **22**: 19-29.

57. Chomoczynski, P. and Qasba, P.K. (1984) Alkaline transfer of DNA to plastic membrane. *Biochem. Biophys. Res. Commun.* **122**, 340-344.

58. Rackwitz, H.R., Zehetner, G., Frichauf, A.M., and Lehrach, H. (1984) Rapid restriction mapping of DNA cloned into lambda phage vectors. Gene **30**, 195-200.

59. Chomoczynski, P., and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159.

60. Foley, K. P., Leonard, M. W. and Engel, J. D. (1993) Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* **9**, 380-384.

61. Siebert, P., D., and Larrick, W. (1993) PCR MIMICS: Competitive DNA fragments for use as internal standards in quantitative PCR. *BioTechniques* **14**: 244-249.

62. Duchmann, R., Strober, W., and James, S. P. (1993). Quantitative measurement of human T-cell receptor Vβ subfamilies by reverse transcription-polymerase chain reaction using synthetic internal mRNA standards. *DNA Cell Biol.* **12**, 217-225.

63. Tabor, S., and Richardson, C. C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci.* **82**, 1074-1078.

64. Linder, M. E., Middleton, P., Hepler, J. R., Taussig, R., Gilman, A. G., and Mumby, S.M. (1993) Lipid modifications of G proteins: α subunits are palmitoylated. *Proc. Natl. Acad. Sci. USA*, **90**, 3675-3679.

65. Linder, M. E., Pang, I-H., Duronio, R. J., Gordon, J. I., Sternweis, P. C., and Gilman, A. G. (1991) Lipid modification of G-proteins: myristoylation of G_{oa} increases its affinity for βγ. *J. Biol. Chem.* **266**, 4654-4659.

66. Casey, P., Fong, H., Simon, M., and Gilman, A. G. (1990) Gz, a guanine nucleotide-binding protein with unique biochemical properties. *J. Biol. Chem.* **265**, 2383-2390.

67. Brunak, S., Eegelbrecht, J., and Knudsen, S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequences. *J. Mol. Biol.* **220**, 49-65.

68. Guigo, R., Knudsen, S., Drake, N., and Smith, T. (1992) Prediction of gene structure. *J. Mol. Biol.* **226**, 141-157.

69. Grivskov, M. and Devereux, J. (1991) *Sequence Analysis Primer* (Stockton Press, New York).

70. Cassill, J.A., Whitney, M., Joazeiro, C.A., Becker, A., and Zuker, C.S. (1991) Isolation of *Drosophila* genes encoding G-protein-coupled receptor kinases. *Proc. Natl. Acad. Sci. USA* **88**, 11067-11070.

71. Marshall, R.D. (1972) Glycoproteins. *Annu. Rev. Biochem.* **41**, 673-702.

72. Woodgett, J.R., Gould, K.L. and Hunter, T. (1986) Substrate specificity of protein kinase. *C. Eur. J. Biochem.* **161**, 177-184.

73. Cooper, J.A., Esch, F.S., Taylor, S.S., and Hunter, T. (1984) Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases *in vivo* and *in vitro*. *J. Biol. Chem.* **259**, 7835-7841.

74. Bernardi, G. (1989) The isochore organization of the human genome. *Annu. Rev. Genet.* **23**, 637-661.

75. Bucher, P. (1990) Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212**, 563-578.

76. Wilson, T.E., Fahrner, T.J., Johnston, M., and Milbrandt, J. (1991) Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science* **252**, 1296-1300.

77. Herschman, H. R. (1991) Primary response genes induced by growth factors and tumor promoters. *Ann. Rev. Biochem.* **60**, 281-319.

78. Forsdyke, D. R. (1980) Lectin pulses as determinants of lymphocyte activation and inactivation during the first six hours of culture. *Can. J. Biochem.* **58**, 1387-1396.

79. Takahama, Y., and Nakauchi, H. (1996) Phorbol ester and calcium ionophore can replace TCR signals that induce positive selection of CD4 T cells. *J. Immun.* **157**, 1508-1513.

80. Kern, J. A., Reed, J. C., Daniele, R. P., and Nowell, P. C. (1986) The role of the accessory cell in mitogen-stimulated human T cell gene expression. *J. Immunol.* **137**, 764-769.

81. Murphy, P. M. (1994) The molecular biology of leukocyte chemoattractant receptors. *Annu. Rev. Immunol.* **12**, 593-633.

82. Subramamiam, M., Schmidt, L. J., Crutchfield, C. E., and Getz, M. J. (1989) Negative regulation of serum-responsive enhancer elements. *Nature* **340**, 64-66.

83. Shaw, G., and Kaman, R. (1986) A conserved AU sequence from the 3'-untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659-667.

84. Blum, S., Forsdyke, R. E., and Forsdyke, D. R. (1990) Three human homologs of a murine gene encoding an inhibitor of stem cell proliferation. *DNA Cell Biol.* **9**, 589-602.

85. Freter, R. R., Irminger, J. C., Porter, J. A., Jones, S. D., and Stiles, C. D. (1992) A novel 7-nucleotide motif located in 3' untranslated sequences of the immediate-early gene set mediates platelet-derived growth factor induction of the JE gene. *Mol. Cell. Biol.* **12**, 5288-5300.

86. Szabo, J., Heath, B., Hill, V.M., Jackson, C.E., Zarbo, R. J., Mallette, L.E., Chew, S.L., Besser, G.M., Thakker, R.V., Huff, V., Leppert, M.F., and Heath, H. (1995) Hereditary hyperparathyroidism jaw tumor syndrome. The endocrine tumor gene HRPTL maps to chromosome 1q21-q31. *Am. J. Hum. Genet.* **56**, 944-950.

87. Teh, B.T., Farnebo, F., Kristofferson, U., Sundelin, B., Cardinal, J., Axelson, R., Yap, A., Epstein, M., Heath, H., Cameron, D., and Larsson, C. (1996) Autosomal dominant primary hyperparathyroidism-jaw tumor syndrome associated with adult nephroblastoma and renal cysts: linkage to 1q21-q32 and loss of the wild-type allele in nephroblastomas. *J. Clin. Endocrin. Metab.* (in press. December issue)

88. Chilson, O. P., and Kelly-Chilson, A. E. (1989) Mitogenic lectins bind to the antigen receptor on human lymphocytes. *Eur. J. Immun.* **19**, 389-396.

89. Burnet, F. M. (1958) *The Clonal Selection Theory of Acquired Immunity.* (Cambridge University Press, Cambridge).

90. Forsdyke, D. R. (1975) Further implications of a theory of immunity. *J. Theor. Biol.* **52**, 187-198.

91. Milner, J. (1978). An inducible gene involved in commitment of lymphocytes to transform. *Nature* **275**, 660-661.

92. Lim, R.W., Varnum, B.C., and Herschman, H.R. (1987) Cloning of tetradecanoyl phorbol ester-induced primary response sequences and their expression in density-arrested Swiss 3T3 cells and a TPA non-proliferative variant. *Oncogene* **1**, 263-270.

93. Gomperts, M., Corps, A.N., Pascall, J.C., and Brown, K.D. (1992) Mitogen-induced Expression of the Primary Response Gene cMCG1 in a Rat Intestinal Epithelial Cell-line (RIE-1). *Febs Lets.* **306**, 1-4.

94. Varnum, B.C., Ma, Q., Chi, T., Fletcher, B., and Herschman, H.R. (1991) The *TIS11* Primary Response Gene is a Member of a Gene Family that Encodes Proteins with a Highly Conserved Sequence Containing an Unusual Cys-His Repeat. *Mol. and Cell. Biol.* **11**, 1754-1758.

95. Nakaijima, K. and Wall, R. (1991) Interleukin-6 signals activating jun B and TIS11 gene transcription in a B-cell hybridoma. *Mol. Cell. Biol.* **11**, 1409-1418.

96. Ip , N.Y., Nye, S.H., Boulton, T.G., Davis, S., Taga, T., Li, Y., Birren, S.J., Yasukawa, K., Kishimoto, T., Anderson, D.J., Stahl, N., and Yancopoulos, G.D. (1992) CNTF and LIF act on neuronal cells via shared pathways that involve the IL-6 signal transducing receptor component gp130. *Cell* **69**,1121-1132.

97. Ghosh, D. (1991) New developments of a transcription factor database. *Trends Biol. Sci.* **16**, 445-447.

98. Taylor, G.A., Lai, W.S., Oakey, R.J., Seldin, M.F., Shows, T.B., Eddy, R.L., and Blackshear, P.J. (1991) The human TTP protein: Sequence, alignment with related proteins, and chromosomal location of the mouse and human genes. *Nucleic Acids Res.* **19**, 3454.

99. Forsdyke, D.R. (1984) Rapid quantitative changes in mRNA populations in cultured lymphocytes: Comparison of the effects of cycloheximide and concanavalin-A. *Can. J. Biochem. Cell Biol.* **62**, 859-865.

100. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.

101. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

102. Kozak, M. (1991) An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115**, 887-896.

103. Hillier, L *et al.*, (1995) WashU-Merck EST project. Unpublished (1995)

104. Ruben, S.M., Dillon, P.J., Schreck, R., Henkel, T., Chen, C.H., Maher, M., Baeuerle, P.A., and Rosen, C.A. (1991) Isolation of a *rel*-related human cDNA that potentially encodes the 65 kD subunit of NF-κB. *Science* **251**, 1490-1493.

105. Rogers, J.H. (1985) The origin and evolution of retroposons. *Int. Rev. Cytol.* **93**, 187-279.

106. Kageyama, R., Merlino, G.T., and Pastan, I. (1988) A transcription factor active on the epidermal growth factor receptor gene. *Proc. Natl. Acad. Sci. USA* **85**, 5016-5020.

107. Benoist, C., and Mathis, D. (1990) Regulation of major histocompatibility complex class II genes. *Annu. Rev. Immunol.* **8**, 681-715.

108. De-Medina, T., Faktor, O., and Shaul, Y. (1988) The S promoter of hepatitis B virus is regulated by positive and negative elements. *Mol. Cell. Biol.* **8**, 2449-2455.

109. Smith, J.D., Melian, A., Leff, T., and Breslow, J.L. (1988) Expression of the human apolipoprotein E gene is regulated by multiple positive and negative elements. *J. Biol. Chem.* **263**, 8300-8308.

110. Wade, R., Gunning, P., Eddy, R., Shows, T., and Kedes, L. (1987) Nucleotide sequence, tissue-specific expression, and chromosomal location of human carbonic anhydrase III. *Genes & Dev.* **1**, 594-602.

111. Pati, U.K., and Weissman, S.M. (1990) The amino acid sequence of the human RNA polymerase II 33-kDa subunit hRPB 33 is highly conserved among eukaryotes. *J. Biol. Chem.* **265**, 8400-8403.

112. Josephs, S.F., Wong-Staal, F., Manzari, V., Gallo, R.C., Sodroski, J.G., Trus, M.D., Perkins, D., Patarca, R., and Haseltine, W.A. (1984) Long terminal repeat structure of an American isolate of type I human T-cell leukaemia virus. *Virol.* **139**, 340-345.

113. Imagawa, M., Chiu, R., and Karin, M. (1987) Transcription factor AP2 mediates induction by two different signal transduction pathways; protein kinase C and cyclic AMP. *Cell* **51**, 251-260.

114. Murre, C., McCaw, P.S., and Baltimore, D., (1989) A new DNA binding and dimerization motif in an immunoglobulin enhancer binding daughterless, MyoD and myc proteins. *Cell* **56**, 777-783.

115. Karin, M., Haslinger, A., Holtgreve, H., Richards, R.I., Krauter, P., Westphal, H.M., and Beato, M. (1984) Characterization of DNA sequences through which cadmium

and glucocorticoid hormones induce human metallothionein-IIA gene. *Nature* **308**, 513-519.

116. Wu, B.J., Kingston, R.E., and Morimoto, R.I. (1986) The human *HSP70* promoter contains at least two distinct regulatory domains. *Proc. Natl. Acad. Sci. USA* **83**, 629-633.

117. Fisch, T.M., Prywes, R., and Roeder, R.G. (1987) c-*FOS* sequences necessary for basal expression and induction by epidermal growth factor, 12-*O*-tetradecanoyl phorbol-13-acetate and calcium ionophore. *Mol. Cell. Biol.* **7**, 3490-3502.

118. Robbins, P.D., Horwitz, J.M., and Mulligan, R.C. (1990) Negative regulation of human *c-fos* expression by the retinoblastoma gene product. *Nature* **346**, 668-670.

118. Rudolf, U., Finegold, M.J., Rich, S.S., Harriman, G.R.,Srinivasin, Y., Brabet, P., Boulay, G., Bradley, A., and Birnbaumer, L. (1995) Ulcerative colitis and adenocarcinoma of the colon in Galpha i2-deficient mice. *Nature Genetics* **10**,143-150.

119. Davidson, I., Xiao, J.H., Rosales, R., Staub, A., and Chambon, P., (1988) The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequences. *Cell* **54**, 931-942.

120. Chiu, R., Imagawa, M., Imbra, R.J., Bockoven, J., and Karin, M. (1987) Multiple cis- and trans-acting elements mediate the transcriptional response to phorbol esters. *Nature* **329**, 648-651.

121. Boeuf, H., Zajchowski, D.A., Tamura, T., Hauss, C., and Kedinger, C. (1987) Specific cellular proteins bind to critical promoter sequences of the adenovirus early EIIa promoter. *Nucleic Acids Res.* **15**, 509-527.

122. Dorsch-Hasler, K., Keil, G.M., Weber, F., Jasin, M., Schaffner, W., and Koszinowski, U.H. (1985) A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. *Proc. Natl. Acad. Sci. USA* **82**, 8325-8329.

123. Dijkema, R., Dekker, B.M., and Van Ormondt, H. (1982) Gene organization of the transforming region of adenovirus type 7 DNA. *Gene* **18**, 143-156.

124. Rosales, R., Vigneron, M., Macchi, M., Davidson, I., Xiao, J.H., and Chambon, P. (1987) The *in vitro* binding of cell-specific and ubiquitous nuclear proteins to the octomer motif of the SV40 enhancer and related motifs present in other promoters and enhancers. *EMBO J.* **6**, 3015-3025.

125. Ohlsson, H., and Edlund, T. (1986) Sequence specific interactions of nuclear factors with the insulin gene enhancer. *Cell* **45**, 35-44.

126. Johnson, A.C., Jinno, Y., and Merlino, G.T. (1988) Modulation of epidermal growth factor receptor proto-oncogene transcription by a promoter site sensitive to SI nuclease. *Mol. Cell. Biol.* **8**, 4174-4184.

127. Bruder, J.T., and Hearing, P. (1989) Nuclear factor EF-1A binds to the adenovirus EIa core enhancer element and other transcription control regions. *Mol. Cell. Biol.* **9**, 5143-5153.

128. Simone, V., D., Ciliberto, G., Hardon, E., Paonessa, G., Palla, F., Lundberg, L., and Cortese, R. (1987) *Cis* and *trans*-acting elements responsible for the cell specific expression of the human α-1-antitrypsin gene. *EMBO J.* **6**, 2759-2766.

129. McGeoch, D.J., Dolan, A., Donald, S., and Brauer, D.H. (1986) Complete DNA sequence of the short repeat region in the genome of herpes simplex virus type 1. *Nucleic Acids Res.* **14**, 1727-1745.

130. Sivaraman, L., and Thimmappaya, B., (1987) Two promoter specific host factors interact with adjacent sequences in an EIa-inducible adenovirus promoter. *Proc. Natl. Acad. Sci. USA* **84**, 6112-6116.

131. Luerssen, H., Mattei, M.G., Schroter, M., Grzeschik, K.H., Adham, I.M., and Engel, W., (1990) Nucleotide sequence of the gene for human transition protein 1 and its chromosomal location on chromosome 2. *Genomics* **8**, 324-330.

132. Emorine, L.J., Marullo, S., Delavier-Klutchko, C., Kaveri, S.V., Durieu-Trautman, O., and Strosberg, A.D. (1987) Structure of the gene for the human ß-2-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* **84**, 6995-6999.

133. Fraser, J.D., Irving, B.A., Crabtree, G.R., and Weiss, A. (1991) Regulation of interleukin-2 gene enhancer activity by the T cell accessory molecule CD28. *Science* **251**, 313-316.

134. Serfling, E., Barthelmas, R., Pfeuffer, I., Schenk, B., Zarius, S., Swoboda, R., Mercurio, F., and Karin, M. (1989) Ubiquitous and lymphocyte-specific factors are involved in the induction of the mouse interleukin-2 gene in T-lymphocytes. *EMBO J.* **8**, 465-473.

135. Yoshida, T., Miyagawa, K., Odagiri, H., Sakamoto, H., Little, P.F., Terada, M., and Sugimura, T. (1987) Genomic sequence of *HST*, a transforming gene encoding a protein homologous to fibroblast growth factors and the *INT-2* encoded protein.

Proc. Natl. Acad. Sci. USA **84**, 7305-7309.

136. Mermod, N., Williams, T.J., and Tjian, R. (1988) Enhancer-binding factors of AP4 and Ap1act in concert to activate SV40 late transcription *in vitro*. *Nature* **332**, 557-561.

137. Hanes, S.D., and Brent, R. (1991) A genetic model for interaction of the homeodomain recognition helix with DNA. *Science* **251**, 426-430.

138. Li, Y., Shen, R.F., Tasi, S.Y., and Woo, S.L. (1988) Multiple hepatic transacting factors are required for *in vitro* transcription of the human α1-antitrypsin gene. Mol. Cell. Biol. **8**, 4362-4369.

139. Johnson, J.E., Birren, S.J., and Anderson, D.J. (1990) Two rat homologs of *Drosophila achaete-scute* specifically expressed in neuronal precursors. *Nature* **346**, 858-861.

140. Thé, H., DE., Vivanio-Ruiz, M., Tiollais, P., Stunnenberg, H., and Dejean, A. (1990) Identification of a retinoic acid response element in the retinoic acid receptor β gene. *Nature* **343**, 177-179.

141. Kinzler, K.W. *et al.* (1991) Identification of a gene located at chromosome 5q21 that is mutated in colorectal carcinomas. *Science* **251**, 1366-1370.

142. Miyajima, N., Kadowaki, Y., Fukushige, S., Shimizu, S., Semba, K., Yamanashi, Y., Matsubara, K., Toyoshima, K., and Yamamoto, T., (1988) Identification of two novel members of the erbA superfamily by molecular cloning. *Nucleic Acids Res.* **16**, 11057-11074.

143. Stanley, E., Metcalf, D., Sobieszczuk, P., Gough, N.M., and Dunn, A.R. (1985) The structure and expression of the murine gene encoding granulocyte-macrophage colony stimulating factor. *EMBO J.* **4**, 2569-2753.

144. Feuchter, A., Freeman, J., and Mager, D. (1991) Strategy for detecting cellular transcripts promoted by human endogenous long terminal repeats: Identification of a new gene with homology to yeast CDC4. *Genomics* **13**, 1237-1246.

145. Vignais, M.L., Woudt, L.P., Wassenaar, G.M., Mager, W.H., Sentenac, A., and Planta, R.J. (1987) Specific binding of TUF factor to upstream activation sites of yeast ribosomal protein genes. *EMBO J.* **6**, 1451-1457.

146. Cereghini, S., Raymondjean, M., Carranca, A.G., Herbomel, P., and Yaniv, M. (1987) Factors involved in control of tissue-specific expression of the albumin gene. *Cell* **50**, 627-638.

147. Coffman, C., Harris, W., and Kintner, C. (1990) Xotch, the *Xenopus* homolog of *Drosophila* notch. *Science* **249**, 1438-1441.

148. Chou, P.Y., and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-147.

149. Ryseck, R.P., Bull, P., Takamiya, M., Bours, V., Siebenlist, U., Dobrzanski, P., and Bravo, R. (1992) RELB, a new transcriptional activator of the REL family which interacts with P50-NF-κB. *Mol. Cell. Biol.* **12**, 674-684.

150. Watanabe, Y., Iino, Y., Furuhata, K., Shimoda, C., and Yamamoto, M. (1988) The *S. pombe mei2* gene encoding a crucial molecule for commitment to meiosis is under the regulation of cyclic AMP. *EMBO J.* **7**, 761-767.

151. Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Stachwell, S.C., Seguin, C., Tuffnell, P.S., and Barrell, B.G. (1984) DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* **310**, 207-211.

152. Lamb, P., and Crawford, L. (1986) Characterization of the human P53 gene. *Mol. Cell. Biol.* **6**, 1379-1385.

153. Laurent, B.C., Treitel, M.A., and Carlson, M. (1990) The SNF5 protein of *Saccharomyces cerevisiae* is a glutamine- and proline-rich transcriptional activator that affects expression of a broad spectrum of genes. *Mol. Cell. Biol.* **10**, 5616-5625.

154. Cisek, L.J., and Corden, J.L. (1989) Phosphorylation of RNA polymerase by the murine homolog of the cell cycle control protein cdc2. *Nature* **339**, 679-684.

155. Wainmaster, G., Roberts, V.J., and Lemke, G. (1991) A homolog of *Drosophila Notch* expressed during mammalian development. *Development* **113**, 199-205.

156. Ellisen, L.W., Bird, J., West, D.C., Sorong, A.L., Reynolds, T.C., Smith, S.D., and Sklar, J. (1991) *TAN-1*, the human homolog of the *Drosophila notch* gene is broken by chromosomal translocations in human lymphoblastic neoplasms. *Cell* **66**, 649-661.

157. Ebina, Y., Ellis, L., Jarnagin, K., Ederly, M., Graf, L., Clauser, E., Ou, J., Masiarz, F., Kan, Y.W., Goldfine, I.D., Roth, R.A., and Rutter, W.J. (1985) The human insulin

receptor cDNA: The structural basis for hormone-activated transmembrane signalling. *Cell* **40**, 747-758.

158. Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599-610.

159. Smith, J.L., Levine, J.R., Ingles, C.J., and Agabian, N. (1989) In trypanosomes the homology of the large subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain. *Cell* **56**, 815-827.

160. Wirth, T., Priess, A., Ammweiler, A., Zwilling, S., and Oeler, B. (1991) Multiple Octomer 2 isoforms are generated by alternative splicing. *Nucleic Acids Res.* **19**, 43-51.

161. Rokeach, L.A., Jannatipour, M., Haselby, J.A., and Hoch, S.O. (1989) Primary structure of a human small nuclear ribonucleoprotein polypeptide as deduced by cDNA analysis. *J. Biol. Chem.* **264**, 5024-5030.

162. Zelent, A., Krust, A., Petkovich, M., Kastner, P., and Chambon, P. (1989) Cloning of murine α- and β-retinoic acid receptors and a novel receptor γ predominantly expressed in skin. *Nature* **339**, 714-717.

163. Chouard, T., Blumenfeld, M., Bach, I., Vandekerckhove, J., Cereghini, S., and Yaniv, M. (1990) A distal dimerization domain is essential for DNA-binding by the atypical HNF1 homeodomain. *Nucleic Acids Res.* **18**, 5853-5858.

164. Yochem, J., and Byers, B. (1987) Structural comparison of the yeast cell division cycle gene *CDC4* and a related pseudogene. *J. Mol. Biol.* **195**, 233-245.

165. Cheung, A.K. (1989) DNA sequence analysis of the immediate early gene of pseudorabies virus. *Nucleic Acids Res.* **17**, 4637-4660.

166. Suzuki, S., Huang, Z.S., and Tanihara, H. (1990) Cloning of an integrin β subunit exhibiting high homology with integrin-β3 subunit. *Proc. Natl. Acad. Sci. (USA).* **87**, 5354-5358.

167. Ino, T., Yasui, H., Hirano, M., and Kurosawa, Y. (1995) Identification of a member of the *TIS11* early response gene family at the insertion point of a DNA fragment containing a gene for the T-cell receptor β chain in an acute T-cell leukemia. *Oncogene* **10**, 2705-2710.

168. Kennelly, P.J., and Krebs, E.G. (1991) Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *J. Biol. Chem.* **266**, 15555-15558.

169. Peterson, S.R., Dvir, A., Anderson, C.W., and Dynan, W.C. (1992) DNA binding provides a signal for phosphorylation of the RNA polymerase II heptapeptide repeat. *Genes & Dev.* **6**, 428-438.

170. Thiesen, H.J., and Bach, C. (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucl. Acids Res.* **18**, 3203-3209.

171. Zhang, M., Zamore, P.D., Carmo-Fonseca, M.,Lamond, A.I., and Green, M.R. (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxillary factor small subunit. *Proc. Natl. Acad. Sci. (USA)* **89**, 8769-8773.

172. Voelker, R.A., Gibson, W., Graves, J.P., Sterling, J.F., and Eisenberg, M.T. (1991) The drosophila suppressor of sable gene encodes a polypeptide with regions similar to those of RNA-binding proteins. *Mol. Cell. Biol.* **11**, 894-905.

173. Bustin, S.A., Nie, X.-F., Barnard, R.C., Kumar, V., Pascall, J.C., Brown, K.D., Leigh, I.M., Williams, N.S. and McKay, I.A. (1994) Cloning and Characterization of ERF-1, a Human Member of the Tis11 Family of Early-Response Genes. *DNA and Cell Biol.***13**, 449-459.

174. Nie, X.-F., Maclean, K.N., Kumar, V., McKay, I.A., and Bustin, S.A. (1995) ERF-2, the Human Homologue of the Murine *Tis11d* Early Response Gene. *Gene* **152**, 285-286.

175. Schein, C.H. (1989) Production of soluble recombinant proteins in bacteria. *Bio/Technology* 7:1141-1147.

176. Takahashi, N., Hayano, T., and Suzuki, M. (1989) Peptidyl-prolyl *cis-trans* isomerase is the cyclosporine A binding protein cyclophilin. *Nature* **337**, 473-475.

177. Taylor, G.A., Thompson, M.J., Lai, W.S., and Blackshear, P.J. (1995) Phosphorylation of tristetraprolin, a protential zinc-finger transcription factor, by mitogen stimulation in intact cells and by mitogen-activated protein kinase *in vitro*. *J. Biol. Chem.* **270**, 13341-13347.

178. Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Preiss, J.R. (1996) The PIE-1 protein and germline specification in *C. elegans* embryos. *Nature* **382**, 710-712.

179. Seydoux, G., Mello, C.C., Pettitt, J., Wood, W.B., Priess, J.R., and Fire, A. (1996) Repression of gene expression in the embryonic germ lineage of *C. elegans*. *Nature* **382**, 713-716.

180. Epner, D.E., and Herschman, H.R. (1991) Heavy Metals Induce Expression of the TPA-Inducible Sequence (TIS) Genes. *J. Cell. Physiol.* **148**, 68-74.

181. Taylor, G.A., and Blackshear, P.J. (1995) Zinc Inhibits Turnover of Labile mRNAs in Intact Cells. *J. Cell. Physiol.* **162**, 378-387.

182. Krumm, A., Meulia, T., Brunvand, M., and Groudine, M. (1992) The block to transcription within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev.* **6**: 2201-2213.

183. Lupton, S.D., Gimpel, S., Jerzy, R., Brunton, L.L., Hjerrild, K.A., Cosman, D., and Goodwin, R.G. (1990) Characterization of the human and mouse interleukin-7 genes. *J. Immunol.* **144**, 3592-3601.

184. McMahon, S. B. and Monroe, J. G. (1992) Role of primary response genes in generating cellular responses to growth factors. *FASEB. J.* **6**, 2707-2715.

185. Worthington, M.T., Amann, B.T., Nathans, D., and Berg, J.M. (1996) Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proc. Natl. Acad. Sci. USA* **96**, 13754-13759.

186. Taylor, G.A., Thompson, M.J., Lai, W.S., and Blackshear, P.J. (1996) Mitogens stimulate the rapid nuclear to cytosolic translocation of tristetraprolin, a potential zinc-finger transcription factor. *Molecular Endocrinology* **10**, 140-146.

187. Kanoh, J., Sugimoto, A., and Yamamoto, M. (1995) *Schizosaccharomyces pombe* zfs^+ Encoding a Zinc-Finger Protein Functions in the Mating Pheromone Recognition Pathway. *Mol. Biol. Cell.* **6**, 1185-1195.

188. Ma, Q., and Herschman, H.R. (1995) The Yeast Homologue *YTIS11*, of the Mammalian *YTIS11* Gene Family is a Non-essential, Glucose Repressible Gene. *Oncogene* **10**, 487-494.

189. Ireland, L.S., Johnston, G.R., Drebot, M.A., Dhillon, M., DeMaggio, A.J., Hoekstra, M.F., and Singer, R.A. (1994) A member of a novel family of yeast zinc-finger proteins mediates the transition from stationary phase to cell proliferation. *EMBO J.* **13**, 3812-3821.

190. Evers, R. and Grummt, I. (1995) molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. *Proc. Nat. Acad. Sci. (USA)* **92**, 5827-5831.

191. Chou, Z.-F.,Chen, F., and Wilusz, J. (1994) Sequence and positional requirements of uridylate-rich downstream elements of polyadenylation signals. *NAR* **22**, 2525-2531.

192. Wahle (1995) 3'-ends cleavage and polyadenylation of mRNA precursors. *Biochem. Biophys. Acta* **1261**, 183-194.

193. Miesfeld, R. and Arnheim, N. (1982) Identification of the *in vivo* and *in vitro* origin of transcription in human rDNA. *NAR* **10**, 3933-3949.

194. Zuo, P., and Maniatis, T. (1996) The splicing factor U2Af^{35} mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes and Dev.* **10**, 1356-1368.

195. Rosenwald, I.B., Setkov, N.A., Kazakov, V.N., Chen, J.J., Ryazanov, A.G., London, I.M., and Epifanova, O.I. (1995) Transiet inhibition of protein synthesis induces expression of proto-oncogenes and stimulates resting cells to enter the cell cycle. *Cell Prolif.* **28**, 631-644.

196. Plet, A., Eick, D., and Blanchard, J.M. (1995) Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene* **10**, 319-328.

197. Greenblatt, J., Nodwell, J.R., and Mason, S.W. (1993) *Nature* **364**: 401.

198. Lou, H., Gagel, R.F., and Berger, S.M. (1996) An intron enhancer recognized by splicing factors activates polyadenylation. *Genes and Devel.* **10**, 208-219.

199. Lang, W.H., Morrow, B.E., Ju, Q., Warner, J.R., and Reeder, R.H. (1994) A model for transcription termination by RNA polymerase I. *Cell* **79**, 527-534.

200. Connely, S. and Manley, J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.***2**, 440-452.

201. Edwalds-Gilbert, G., Precott, J., and Falck-Pederson., E. (1993) 3' RNA processing efficiency plays a primary role in generating termination-competent RNA polymerase II elongation complexes. *Mol. Cell. Biol.* **13**, 3472-3480.

202. Wilson, T. and Treisman, R. (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature* **336**, 396-399.

203. Chen, C-Y..A., and Shyu, A-B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *TIBS* **20**, 465-470.

204. Chen, C.Y., Xu, N., and Shyu, A.B. (1995) mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation. *Mol. Cell. Biol.* **15**, 5777-5788.

205. Phillips, D.R., and Crothers, D.M. (1986) Kinetics and sequence-specificity of drug-DNA interactions: an in vitro transcription assay. *Biochemistry* **25**, 7355-7362.

206. Chen, T.A., Sterner, R., Cozzolino, A., and Allfrey, V.G. (1990) Reversible and irreversible changes in nucleosome structure along the c-fos and c-myc oncogenes following inhibition of transcription. *J. Mol. Biol.* **212**, 481-493.

207. Shui, Z., Boyett, M.R., Zang, W.J., Haga, T., and Kameyama, K. (1995) Receptor kinase-dependent desensitization of the muscarinic K+ current in rat atrial cells. *J Physiol.* **487**, 359-366.

208. Vuong, T.M. and Chambre, M. (1990) Subsecond deactivation of the transducin by endogenous GTP hydrolysis. *Nature* **346**, 71-74.

209. Yatani, A., and Brown, A.M. (1989) Rapid beta-adrenergic modulation of cardiac calcium channel currents by a fast G protein pathway. *Science* **245**: 71-74.

210. Berman, D.M., Kozasa, T., and Gilman, A.G. (1996) The GTPase-activating protein RGS4 stabilizes the transition state for nucleotide hydrolysis. *J. Biol. Chem.* **271**, 27209-27212.

211. Ben, B.A., Michiel, D.F., and Oppenheim, J.J (1995) Signals and receptors involved in recruitment of inflammatory cells. *J. Biol. Chem.* **270**, 11703-11706.

212. Hendrie, P.C., and Broxmeyer, H.E. (1994) Myeloid cell proliferation stimulated by steel factor is pertussis toxin sensitive and enhanced by cholera toxin. *Int. J. Immunopharm.* **16**, 547-560.

213. Taub, D.D., Ortaldo, J.R., Turcovski, C.S., Key, M.L., Longo, D.L., and Murphy, W.J. (1996) Beta chemokines costimulate lymphocyte cytolysis, proliferation and lymphokine production. *J. Leuk. Biol.* **59**, 81-89.

214. Rudolf, U., Fingold, M.J., Rich, S.S., Harriman, G. R., Srinvasan, Y., Brabet, P., Boulet, G., Bradley, A., and Birnbaumer, L. (1995) Ulcererative colitis and adenocarcinoma of the colon in Galphai2-deficient mice. *Nature Genetics* **10**, 143-150.

215. Smith *et al.* (1996) Major susceptibility locus for Prostate Cancer on Chromosome 1 suggested by a genome-wide search. *Science* **274**, 1371-1374.

216. Stroumbakis, N.D., Li, Z., and Tolias, P.P. (1994) RNA- and single-stranded DNA-binding (SSB) proteins expressed during Drosophila melanogaster oogenesis: a homologue of bacterial and eukaryotic mitochondrial SSBs. *Gene* **143**, 171-177.

217. White, B.A., and Bancroft, F.C. (1982) Cytoplasmic dot hybridization. Simple analysis of relative mRNA levels in multiple small cell or tissue samples. *J Biol.Chem.* **257**, 8569-8572.

218. Brenner, C.A., Tam, A.W., Nelson,P.A., Engleman, E.G., Suzuki, N., Fry, K.E., and Larrick, J.W. (1989) Message amplification phenotyping (MAPPing): a technique to simultaneously measure multiple mRNAs from small numbers of cells. *Biotechniques* **7**, 1096-1103.

219. Foley, K.P., and Engel, J.D. (1992) Individual stage-selector element mutations lead to reciprocal changes in beta- vs. epsilon-globin gene transcription: genetic confirmation of promoter competition during globin gene switching. *Genes Dev.* **6**, 730-744.

220. Kaneda, N., Oshima, M., Chung, S.Y., and Guroff, G. (1992) Sequence of a Rat *TIS11* cDNA, an Immediate Early Gene Induced by Growth Factors and Phorbol Esters. *Gene* **118**, 289-291.

221. Martin-Gallardo, A.McCombie, W.R., Gocayne, J.D., Fitzerald, M.G., Wallace, S., Lee, B.M., Lamerdin, J., Trapp, S., Kelley, J.M., Liu, L-I., Dubnick, M., Johnston-Dow, L.A., Kervalege, A.R., DeJong, P., Carrano, A., Fields, S. and Ventner, J.C. (1992) Automatic DNA sequencing and analysis of human chromosome 19q13.3. *Nature Genetics* **1**: 34-39.

221. Gomperts, M., Pascall, J.C., and Brown, K.D. (1990) The Nucleotide Sequence of a cDNA Encoding an EGF-inducible Gene Indicates the Existence of a New Family of Mitogen-induced genes. *Oncogene* **5**, 1081-1083.

223. Samal, B., Sun, Y., Stearns, G., Xie, C., Suggs, S., and McNiece, I. (1994) Cloning and characterization of the cDNA encoding a novel human pre-B-cell colony-enhancing factor. *Mol. Cell. Biol.* **14**, 1431-1437.

224. Cocchi, F., DeVico, A.L., Garzino-Demo, A., Arya, S.K., Gallo, R.C., and Lusso, P. (1995) Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. *Science* **270**, 1811-1815.

225. Russell, L., and Forsdyke, D.R. (1993) The third human homolog of a murine gene encoding an inhibitor of stem cell proliferation is truncated and linked to a CpG island-containing upstream sequence. *DNA Cell. Biol.* **12**, 157-175.

226. Ning, Z.Q., Norton, J.D., and Murphy, J.J. (1996) Distinct mechanisms for rescue from apoptosis in Ramos human B cells by signalling through CD40 and interleukin-4 receptor: role for inhibition of an immediate early response gene. *Eur. J. Immunol.***26**, 2356-2363.

227. Pearsall, R.S., Shibata, H., Browsowska, A., Yoshino, K., Okuda, K., Plass, C., Chapman, V., deJong, P., Hayashizaki, Y. and Held, W.A. (1996) Absence of imprinting for U2AFBPL, a human homologue of the imprinted mouse gene U2afbp-rs (unpublished).

228. Ma, Q., Wadleigh, D., Chi, T., and Herschman, H. (1994) The Drosophila TIS11 Homologue Encodes a Developmentally Controlled Gene. *Oncogene* **9**, 3373-3377.

229. Mohler, J., Weiss, N. Murli, S., Mohammadi, S., Vani, K., Vasilakis, G., Song, C.H., Epstein, A., Kuang, T., English, J., and Cherdak, D. (1992) The embryonically active gene, *unkempt*, of drosophila encodes a Cys_3His protein. *Genetics* **131**, 377-388.

230. Collins, P.L. and Wertz, G.W. (1985) The envelope-associated 22K protein of human respiratory syncytial virus: nucleotide sequence of the mRNA and a related polytranscript. *J. Virology* **54**, 65-71.

231. Yu, Q., Davis, P.J., Brown, T.D. and Cavanaugh, D. (1992) Sequence and in vitro expression of the M2 gene of tukey rhinotracheitis virus. *J. Gen. Virol.* **73**: 1355-1363.

232. Zamora, M. and Samal, S.K. (1991) Sequence analysis of the M2 mRNA of Bovine Syncytial Virus obtained from a F-M2 dicistronic mRNA suggests structural homology with that of the Human Respiratory Syncytial Virus. (unpublished)

233. Warbrick, E., and Glover, D. (1994) A *Drosophila melanogaster* Homolog of the *crd1 cdc25* Mutant Strain of Fission Yeast. *Gene* **151**, 243-246.

234. Barnard, R.C., Pascall, J.C., Brown, K.D., McKay, I.A., Williams, N.S., and Bustin, S.A. (1993) Coding Sequence of ERF-1, the Human Homologue of Tis11b/cMG1, Members of the Tis11 Family of Early Response Genes. *Nucleic Acids Res.* **21**, 3580.

APPENDIX A1

This section discusses many of the stages involved in the development of the RT-PCR assay. This technique was originally developed for the measurement of the *G0S24* and *RGS2/G0S8* mRNAs. Subsequently, we have used it to measure kinetics of *RGS1/BL34/1R20*, *FOS/G0S7*, *FOSB/G0S3*, *MIP1α/G0S19/LD78*, *EGR-1/G0S30*, and *G0S2* mRNAs. It offers the unique advantage that mRNA levels for all of these genes can be studied using total RNA purified from < 1 x 10^6 cells. Techniques were developed according to the guidelines established by two leading laboratories in the field of RT-PCR analysis (60, 61). Other details are as described previously (4, 5, 54).

Preparation of total RNA

Several methods of RNA purification from small numbers of cells in culture, including preparation of cytoplasmic RNA (217) and pelleting of small samples of RNA by ultracentrifugation through cesium gradients (218), were attempted. In our hands, the most reproducible method was found to be Trizol reagent (Gibco/BRL, Life Technologies), a commercial preparation based on the Acid Phenol/Guanidinium thiocyanate method first described by Chomoczynski and Sacchi (1987, ref. 59). As many of the *G0S* gene RT-PCR products did not span intronic sequences, RNA samples were routinely treated with RNAse-free DNAse (RQ1; Promega) to eliminate the possibility of genomic DNA contamination.

Quantitation and storage of RNA

Purified RNA samples were resuspended in 400 µl dpc-treated water and quantitated using spectrophotometry. Aliquots of the quantitated material giving the appropriate amount of RNA (usually 250 ng) were added directly to RT reactions. RNA was stored at -70°C until needed.

Reverse transcription

Samples were reverse transcribed from the 17 bp gene-specific primers rather than random oligonucleotides to increase the sensitivity and specificity of the RT reaction. Specificity of this reaction was also increased by using a 42°C temperature during the RT reaction. We found that in the case of some RT primers, multiple band patterns were obtained if the RT reaction was carried out at lower temperatures. RT primers were chosen to be close to the downstream primer of the PCR pair to increase the chance of extending cDNA synthesis completely through the desired amplicon. The same set of RT primers was always used to assay each RNA sample to avoid the potential problems of changes in RT efficiency with different primer combinations. Indeed, we noted some decrease in RT efficiency with increasing numbers of RT primers in the reaction. Completed reverse transcriptase reactions were heated for 10 min at 94°C to remove any remaining RT activity which has been shown to inhibit efficiency of the PCR reaction (219). RT mixtures were routinely diluted in 6 volumes of TE8 buffer (10 mM Tris-Cl, pH 8, 1 mM EDTA) to dilute the concentration of RT primer prior to use in PCR reactions.

RT-PCR DNA and cRNA controls

PCR controls were prepared as described previously (4, 5, Chapter 2). Circular plasmids were linearized to facilitate uniform denaturation during early amplification cycles. Two-fold dilutions of plasmids used in competitive assays were made in TE8 and stored at 4°C prior to use. Excessive freeze thaw was found to decrease amplicon signal from controls presumably through freeze-thaw damage of the control DNA. Thus, stock samples of controls were stored frozen in small aliquots just before use.

Conditions for optimal PCR efficiency

For optimal efficiency primer pairs are routinely chosen to be approximately 200 bp apart so that with the addition of a small exogenous fragment in control constructs these amplicons do not exceed 450 bp. It was important to verify, using ^{32}P-dATP incorporation, that the control and mRNA amplicons were being produced with similar efficiency and that the reaction was stopped at a point when the two products were still being made exponentially. It was found that in some cases, under normal cycling conditions, the smaller mRNA amplicon was amplified more efficiently than the control. We found this problem could be corrected by slowing the rate of ramping between temperatures which allowed longer control species to amplify more efficiently.

Choice of reaction buffers

We have found the results were most reproducible, especially when using extremely low concentrations of control DNA, using Promega *Taq* Polymerase buffer containing $MgCl_2$,

compared to the comparable BRL and Sangon buffers. We attribute the difference to the presence of Triton X-100 in this mixture (it is not found in the other recipes) as it may prevent the nonspecific adherence of DNA to the polypropylene eppendorf tubes.

Order of addition of reagents is a critical parameter at low DNA concentrations

We believe that for similar reasons, the order of addition of reagents was critical to the reproducibility of this method. If control DNA was added first, especially at low DNA concentrations, the results were quite variable- presumably for reasons related to nonspecific adherence. However, if the PCR reaction mixture was added first, and the controls and template cDNA added directly into the liquid of the reaction, very reproducible results were obtained. Primers were routinely left out of the reaction mixture and added at hot start as described previously.

Interpretation of competitive RT-PCR results

Shown in Fig. A1-1 is a schematic representation of two competitive mRNA analyses. The equivalence point is determined as being directly on a dilution (dilution= 3.0, Fig. A1-1a) or halfway between a pair of dilutions (dilution= 3.5; Fig. A1-1b). The amount of RNA (molecules/μg total RNA) can then be calculated as shown below.

Figure A1-1. **Schematic representation equivalence point determination using competitive RT-PCR.** Schematic diagram illustrating interpretation of typical competitive analysis profiles in which the equivalence point (arrows) falls *on* a control dilution (a) or *in between* two adjacent dilutions (b). Results are as determined by visual inspection of EtBr intensities.

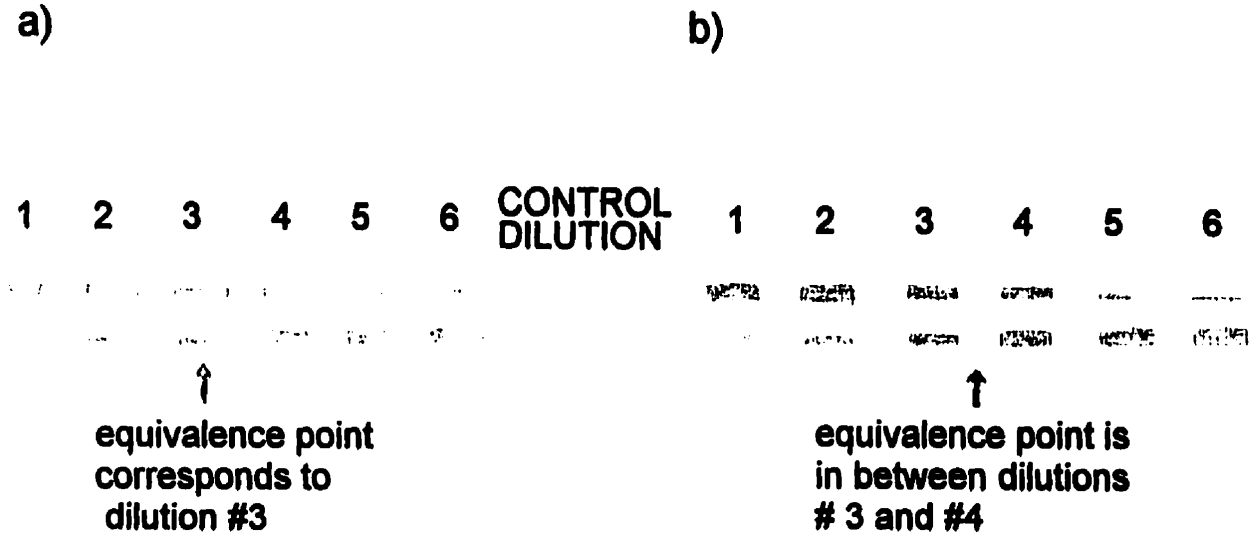
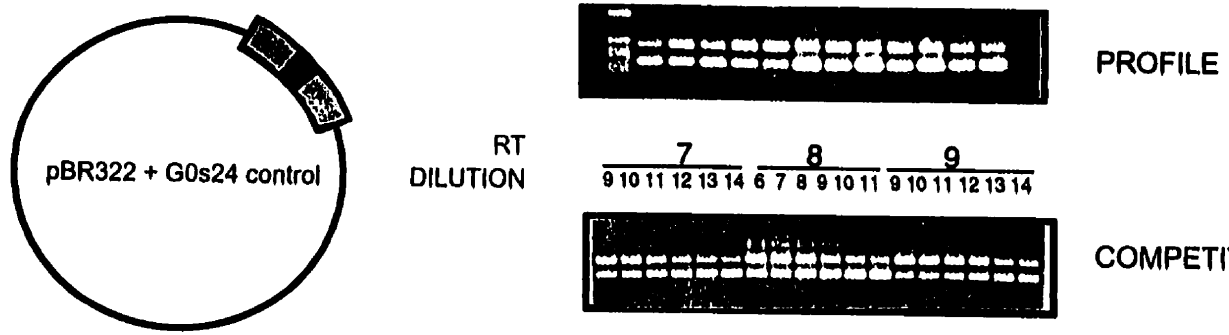
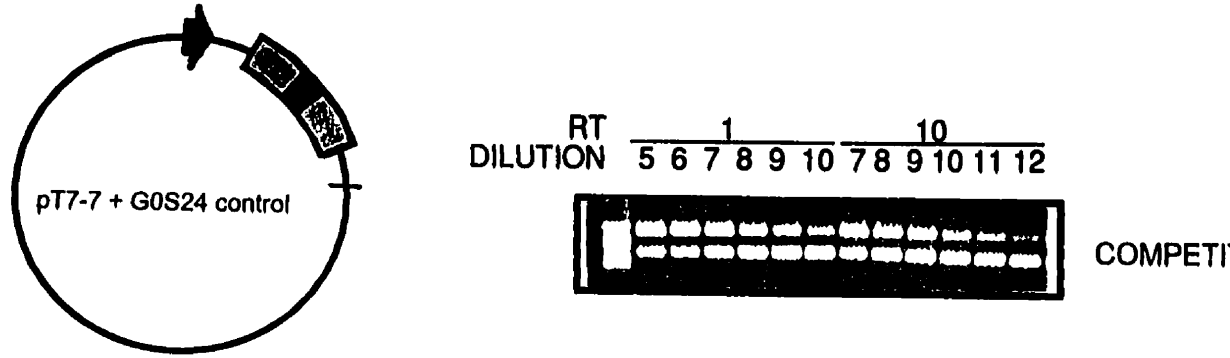

a)
b)
1 2 3 4 5 6
CONTROL DILUTION
1 2 3 4 5 6
equivalence point corresponds to dilution #3
equivalence point is in between dilutions # 3 and #4

Figure A1-2. **mRNA quantitation by competitive RT-PCR using exogenous DNA (a) and cRNA control (b) templates.** Demonstration of the control plasmids and raw data used to determine mRNA levels from samples of total RNA. (a) Upper photograph shows ethidium-stained agarose gel profile of the PCR products which resulted from coamplification of 12 RT samples (specifically transcribed from *G0S24*-specific primers) with a constant amount of control plasmid (shown at left; see also methods Chapter 2). Lower photograph shows coamplification of constant amounts of RT samples 7, 8, and 9 (from upper photograph) with an appropriate range of two-fold dilutions of control plasmid. This range can be determined intuitively from the profile data. (b) *G0S24* cRNA may be made from the linearized (hash mark represents unique restriction site) control plasmid at left (see also Methods in Chapter 4). Photograph shows ethidium-stained gels of competitive cRNA profiles made by addition of the appropriate range of cRNA dilutions to the reverse transcriptase mix prior to RT-PCR analysis. Equivalence points are determined as described in Fig. A1-1 above. mRNA levels are calculated as described in the text.

a) Competitive DNA Method



b) Competitve cRNA Analysis



Calculation of mRNA levels-competitive DNA and cRNA analyses

Fig. A1-2 shows a sample RT-PCR data for G0S24 mRNA and is useful for a description of the calculation used in the mRNA quantitation. In Fig. A1-2a the DNA assay is shown. Competitive profiles are shown for 3 RT samples (7, 8, and 9). Control DNA dilutions are a two-fold series (starting from number 1) decreasing as the numbers increase (i.e. control DNA in dilution 5 is half as much as in dilution 4). The concentration of control DNA in dilution 1 is 5.67×10^8 molecules/5 µl. In RT sample 7, which shows an equivalence point with control dilution 12 the calculation is as follows:

molecules/µg RNA= $(((5.67 \times 10^8 \text{ molecules}/5\ \mu l) \times 5\ \mu l)/2^{k-n}) \times ((0.25\ \mu g\ RNA/240\ \mu l) \times 5\ \mu l) \times a \times b$

$= 1.96 \times 10^8$ molecules/µg RNA

a=2 , correction for strand number (controls are double-stranded while cDNA are single-stranded)

b=1.84, correction for size of amplicon assuming EtBr-staining directly proportional to DNA length

k=12, control dilution at which there is an equivalence point

n=1, first control dilution in the series (concentration= $5,67 \times 10^8$ molecules/5µl)

Fig. A1-2b illustrates the use of a competitive cRNA analysis for a pair of RNA samples. Note that in this case the calculated mRNA level is simply the number of cRNA molecules at the equivalence point x b=1.84, as there is no need to correct for strand number. The concentration of cRNA in dilution 1 was 2.2×10^9 molecules/5 µl. In RNA sample 1, which shows an equivalence point with cRNA dilution 8, the calculation is as follows:

molecules/µg RNA= $(((2.2 \times 10^9 \text{ molecules}/5\ \mu l) \times 5\ \mu l)/ 2^{k-n}) \times b \times (1/0.25\ \mu g)$

$= 1.27 \times 10^8$ molecules/µg RNA

APPENDIX A2

Construction of full-length cDNAs and expression constructs

Shown in Figure A2-1 and A2-2 are the strategies used to create a full length *G0S24* cDNA and expression clones for the various recombinant G0S proteins used in this work.

Figure A2-1. **Generation of full-length *G0S24* cDNA using genomic and 5' end RACE clones.**

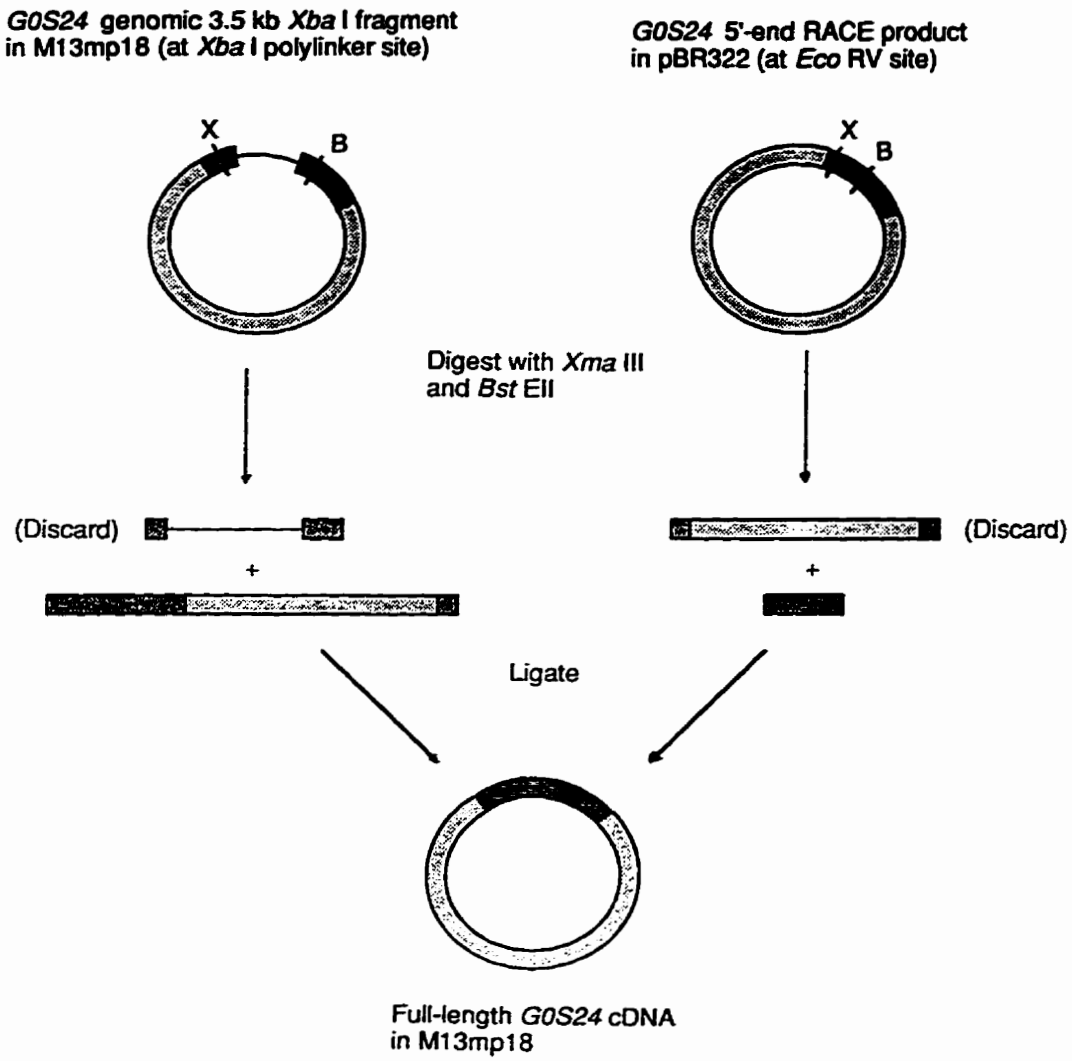

G0S24 genomic 3.5 kb *Xba* I fragment
in M13mp18 (at *Xba* I polylinker site)
G0S24 5'-end RACE product
in pBR322 (at *Eco* RV site)
X
B
X
B
Digest with *Xma* III
and *Bst* EII
(Discard)
+
+
(Discard)
Ligate
Full-length *G0S24* cDNA
in M13mp18

Figure A2-2. **PCR primers used for cloning of G0S cDNAs (a) into the pET-19b *E. coli* expression vector (b) to generate clones expressing $(His)_{10}$-tagged G0S proteins.** Indicated are the *Nde* I cloning sites (underlined) and initiation codons (bold) in the upstream (U) primers which allow G0S proteins to be cloned in frame behind a 10 Histidine residue amino terminal tag in the pET-19b multiple cloning site (b). Also in bold is the artificially created stop codon in the downstream (D) primer for G0S24ZF. In the case of the G0S8/RGS2 and full-length G0S24 proteins the endogenous stop codon is used.

a)

His-tagged protein	Amino Acids	PCR Primers
G0S24	1-326	5'-GCATGCATATGCAAAGTGCTATGTT-3' (U) 5'-CAGTCACTTTGTCACTC-3' (D)
G0S24ZF(CCCH)	103-186	5'-GCATGCATATGCGCTACAAGACTGAGCT-3' (U) 5'-GCTAGCTGATGCTCTGGCGAA-3' (D)
G0S8/RGS2	1-211	5'-GCATGCATATGCAAAGTGCTATGTT-3' (U) 5'-GGGAAGCTGTTCTGATA-3' (D)

b)