The c-*myc* coding region determinant-binding protein: a member of a family of KH domain RNA-binding proteins

Glenn A. R. Doyle, Natalie A. Betz¹, Peter F. Leeds, Ani J. Fleisig, Rebecca D. Prokipcak² and Jeff Ross*

McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, 1400 University Avenue, Madison, WI 53706, USA, ¹Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711-5399, USA and ²Department of Pharmacology, Room 4308 Medical Sciences Building, 8 Taddle Creek Road, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received September 11, 1998; Accepted September 28, 1998

DDBJ/EMBL/GenBank accession no. AF061569

ABSTRACT

The half-life of c-myc mRNA is regulated when cells change their growth rates or differentiate. Two regions within c-myc mRNA determine its short half-life. One is in the 3'-untranslated region, the other is in the coding region. A cytoplasmic protein, the coding region determinant-binding protein (CRD-BP), binds in vitro to the c-myc coding region instability determinant. We have proposed that the CRD-BP, when bound to the mRNA, shields the mRNA from endonucleolytic attack and thereby prolongs the mRNA half-life. Here we report the cloning and further characterization of the mouse CRD-BP, a 577 amino acid protein containing four hnRNP K-homology domains, two RNP domains, an RGG RNA-binding domain and nuclear import and export signals. The CRD-BP is closely related to the chicken β -actin zipcode-binding protein and is similar to three other proteins, one of which is overexpressed in some human cancers. Recombinant mouse CRD-BP binds specifically to c-myc CRD RNA in vitro and reacts with antibody against human CRD-BP. Most of the CRD-BP in the cell is cytoplasmic and co-sediments with ribosomal subunits.

INTRODUCTION

The c-Myc protein is a helix–loop–helix/leucine zipper (HLH/ LZ) type transcription factor that forms heterodimers with Max (1–3). c-Myc abundance influences cell proliferation, differentiation and neoplastic transformation (4–7) and c-Myc overexpression is implicated in tumor formation (8,9). These and other consequences of aberrant c-*myc* expression highlight the importance of understanding c-*myc* gene regulation.

c-Myc protein expression is regulated at multiple levels and is strongly influenced by changes in *c-myc* mRNA stability (10–12). The half-life of *c-myc* mRNA is usually 15–30 min but can be 2 h or more in some cells under some conditions. For example, the *c-myc* gene is active and *c-myc* mRNA is relatively stable in fetal rodent hepatocytes. As a result, fetal hepatocytes contain abundant *c-myc* mRNA. The *c-myc* gene is also active in adult hepatocys, but the mRNA is quite unstable. As a result, adult hepatocytes contain little or no c-myc mRNA (13,14). The mRNA is transiently up-regulated and stabilized in adult hepatocytes following partial hepatectomy (15,16). These findings illustrate how c-myc mRNA stability affects c-MYC protein expression.

Two cis-acting sequence elements affect the post-transcriptional regulation of c-myc mRNA: an AU-rich element (AURE) in the 3'-untranslated region (3'-UTR) and an ~250 nt coding region instability determinant (CRD). The CRD encodes part of the HLH/LZ domain and is located at the 3'-terminus of the coding region. Four observations show that this CRD functions independently of the AURE to affect c-myc mRNA expression: (i) c-myc mRNA lacking its CRD is more stable than wild-type c-myc mRNA (17-20); (ii) the CRD is required for the post-transcriptional down-regulation of c-myc mRNA that occurs when cultured myoblasts fuse to form myotubes (20,21); (iii) the c-myc CRD functions as an mRNA-destabilizing element when fused in-frame within the coding region of β -globin mRNA (22); (iv) the c-myc CRD is required for up- and down-regulating c-myc mRNA abundance post-transcriptionally in transgenic mice undergoing liver regeneration following partial hepatectomy (13,15,16,23-25).

We have investigated how the CRD affects c-myc expression using a cell-free mRNA decay system that includes polysomes from cultured cells. The polysomes contain both the substrates (mRNAs) for decay and at least some of the enzymes and co-factors that affect mRNA stability. Polysomes are incubated in an appropriate buffer and the decay rates of endogenous polysomal mRNAs such as c-myc are monitored by hybridization. This in vitro assay reflects many aspects of mRNA decay in intact cells (26–29). For example, mRNAs that are unstable in cells are also unstable in vitro, while mRNAs that are stable in cells are stable in vitro (26). In standard reactions, polysome-associated c-myc mRNA is degraded rapidly in a $3' \rightarrow 5'$ direction, perhaps by an exonuclease (29). An alternative endonucleolytic decay pathway is activated when the reactions are supplemented with a 180 nt sense strand competitor RNA corresponding to part of the c-myc CRD. This RNA induces endonucleolytic cleavage within the c-myc CRD, thereby destabilizing c-myc mRNA 8-fold (30). This effect is highly specific. Other competitor RNAs do not destabilize c-myc mRNA and c-myc CRD competitor RNA does not destabilize other mRNAs tested.

*To whom correspondence should be addressed. Tel: +1 608 262 3413; Fax: +1 608 262 9464; Email: ross@oncology.wisc.edu

Based on these observations, we suggested the following model: (i) the *c-myc* CRD is susceptible to attack by a ribosome-associated endoribonuclease; (ii) a protein can bind to the CRD and shield the CRD from the RNase; (iii) *in vitro*, when we add competitor CRD RNA to the reactions, the competitor titrates the protein off *c-myc* mRNA, leaving the CRD unprotected and open to attack by the RNase. Consistent with this model, we detected a protein that binds *in vitro* to *c-myc* CRD [³²P]RNA (30). This protein, the *c-myc* coding region determinant-binding protein (CRD-BP), was purified to homogeneity (31). We then found that the CRD-BP is developmentally regulated, being expressed in fetal and neonatal rats but not in adult animals (32).

Here we report the cloning of full-length mouse and partial human CRD-BP cDNAs. CRD-BP derived from the mouse cDNA exhibits RNA-binding properties identical to those of CRD-BP purified from cells. The CRD-BP is primarily a cytoplasmic protein that co-sediments with polysomes. These observations are consistent with a role for the CRD-BP in shielding polysomal c-myc mRNA from endonucleolytic attack. The CRD-BP is a member of a family of RNA-binding proteins containing four KH regions.

MATERIALS AND METHODS

Cell lines and preparation of subcellular fractions

Cell lines were obtained from the American Type Culture Collection (Rockville, MD). K562 human erythroleukemia cells were cultured as described (26). NIH 3T3, H4IIE and 293 cells were cultured in DMEM (4.5 g/l glucose) containing 10% calf serum and antibiotics (Gibco BRL). Polysomes and post-polysomal supernatant (S130) were prepared as described (26,30,31), but the cells were lysed in buffer A (1 mM potassium acetate, 1.5 mM magnesium acetate, 2 mM DTT, 10 mM Tris-HCl, pH 7.4) supplemented with 0.1 mM EGTA, 100 µg/ml PMSF and 2 µg/ml each of aprotinin, leupeptin and pepstatin A (Sigma). The nuclear pellet from the first low speed centrifugation was washed once in buffer A and centrifuged. Nuclear wash material in the supernatant was harvested and saved. The nuclei were then resuspended in 300 µl of buffer B (1.5 mM MgCl₂, 140 mM NaCl, 20% v/v glycerol, 10 mM Tris-HCl, pH 8.0), lysed by adding 2.7 ml of buffer C (5.0% w/v SDS, 10% v/v glycerol, 5% v/v \beta-mercaptoethanol, 62.5 mM Tris-HCl, pH 6.8), passed 10 times through an 18 gauge needle and boiled for 15 min. Ribosomal salt washes (RSW) from cells and from translation reactions were prepared as described (31). All fractions were stored at -70° C.

Protein purification and microsequencing

Two independent preparations of human c-myc CRD-BP were purified from K562 cell RSW (31). One was microsequenced at the Protein Sequence and Peptide Synthesis Facility of the University of Wisconsin Biotechnology Center (Madison, WI), the other at the Keck Laboratories, Yale University (New Haven, CT). This second sequence was distinct from the first, did not overlap and was used to design PCR primers.

Cloning of CRD-BP cDNAs

Human CRD-BP cDNA cloning. We first prepared a human CRD-BP cDNA and used it to identify mouse CRD-BP cDNA. A K562 (human) cell cDNA λ library (Clontech) was screened by degenerate PCR to amplify a 45 bp DNA sequence based on the

Mouse CRD-BP cDNA cloning. The partial human CRD-BP cDNA was used to identify two mouse CRD-BP cDNAs in the EST database. The larger one, AA073514, was obtained from Genome Systems and was sequenced. Its amino acid sequence was 99% identical to that of our human CRD-BP, indicating that it corresponded to mouse CRD-BP. It contained the entire 3'-UTR and most of the coding region. To extend the 5' sequence, 5'-RACE was performed on a 17 day mouse embryo Marathon-Ready cDNA Library (Clontech). 'Touchdown PCR' was performed with oligomers AP1 (Clontech) and CRD-BP2 (5'-AGGTTCCGTC-CTTCCTTGCCAATG-3') as 5' and 3' primers, respectively. Secondary PCR was then performed using nested 5' and 3' primers [AP2 (Clontech) and CRD-BP3 (5'-AACTTCATCTGCCGTTT-TGG-5'), respectively]. Since the resulting clone lacked the translation start site and 5'-UTR, a mouse BAC library (Genome Systems) was screened by PCR with primers CRD-BP4 (5'-CATC-AACTGGAGAACCATG-3') and CRD-BP5 (5'-GACTGCGTC-TGTTTTGTGATG-3'). A BAC clone containing the mouse CRD-BP gene was obtained from Genome Systems. The remainder of the coding region and at least part of the 5'-UTR were sequenced from this BAC clone using CRD-BP6 (5'-CTGTAGGA-GATCTTGTGCTC-3') as primer.

In vitro translation of mouse CRD-BP

The mouse CRD-BP cDNA ORF was subcloned via PCR into pSPUTK (Stratagene) to create the translation clone pSPUTK-CRD-BP. A single base mutation (underlined) in the 5' primer (5'-CGCACCGCCACCATGGACAAGCTTTACATCGG-3') changed an asparagine to an aspartic acid at amino acid position 2. The 3' primer (5'-ACTGGGATCTGACCCATCCT-3') was from the CRD-BP 3'-UTR. To express mouse CRD-BP with a C-terminal 6-histidine (His₆) affinity tag, mouse CRD-BP was subloned into pET28b(+) (Novagen) cut with *NcoI* and *XhoI*. His₆-CRD-BP was purified from *Escherichia coli* using standard Ni²⁺–agarose affinity chromatography (Qiagen). CRD-BP eluted from the column was dialyzed against 200 mM KCl, 1 mM EDTA, 10% (v/v) glycerol, 1 mM reduced glutathione, 0.1 mM oxidized glutathione, 20 mM triethanolamine, pH 7.6. cDNAs were transcribed and translated in the TnT[®] Coupled Reticulocyte Lysate System (Promega).

Immunoblotting and gel retardation assays

Immunoblotting for CRD-BP was performed using chicken anti-CRD-BP IgY against purified human protein (31,32). To detect heat shock protein 90 (HSP90), the primary antibody was

rabbit anti-mouse HSP90 polyclonal IgG (from Dr Alan Poland) and the secondary antibody was HRP-conjugated goat anti-rabbit IgG (Sigma). Reactive bands were visualized using enhanced chemiluminescence (ECL). Distinct bands were not detected with preimmune serum, normal human serum or secondary antibodies alone (data not shown). Where noted, blots were stripped for 30 min at 50 °C in 2% SDS, 100 mM β-mercaptoethanol, 50 mM K₂HPO₄, pH 6.8, and were then washed extensively in buffer containing 5% non-fat dry milk. Gel retardation assays were performed as described (31,32). In competition assays, the CRD-BP was pre-incubated with or without unlabeled competitor RNA for 10 min at 30°C and then [³²P]RNA probe was added.

Sucrose gradient centrifugation and ribosomal RNA analysis

All procedures were performed at 4°C. A 50 μ l aliquot of K562 cell polysomes was brought to 20 mM EDTA, mixed gently, left on ice for 20 min, layered over a 10 ml linear 5–30% sucrose gradient in buffer D (100 mM KCl, 10 mM potassium acetate, 5 mM EDTA, 1 mM DTT, 5 mM HEPES, pH 7.3; 34) and centrifuged in a Beckman SW41.1 rotor for 4 h at 38 000 r.p.m. (178 000 g). Fractions (500 μ l) were pipetted sequentially from the top. Pelleted material was resuspended in 500 μ l of buffer D containing 5% sucrose. Proteins were precipitated with methanol and chloroform prior to immunoblotting. RNA was isolated using TRIzol reagent (Gibco BRL), electrophoresed in a 1% agarose gel and visualized with ethidium bromide.

Transfections and immunofluorescence

To construct a mouse CRD-BP cDNA encoding an N-terminal FLAG tag, oligomers encoding the FLAG sequence (Top, 5'-CAT-GGACTACAAGGACGACGATGACAAGGT-3'; Bottom, 5'-CA-TGACCTTGTCATCGTCGTCCTTGTAGTC-3') were kinased, annealed and ligated into NcoI-digested, dephosphorylated pSPUTK-CRD-BP to form pSPUTK-FLAG-CRD-BP. FLAG-CRD-BP and untagged CRD-BP cDNAs were then subcloned into pcDNA3 (Invitrogen) for mammalian expression. Each DNA was transfected by calcium phosphate precipitation into 293 or H4IIE cells grown on coverslips. Forty eight hours after transfection, cells were fixed in ice-cold methanol (H4IIE cells) or phosphate-buffered saline (PBS) containing 4% (v/v) formaldehyde (293 cells). Cells were then washed in PBS, permeabilized in PBS containing 0.1% (v/v) Triton X-100 and stained for FLAG-CRD-BP expression by immunofluorescence using anti-FLAG M2 monoclonal (Sigma) and FITC-conjugated goat anti-mouse IgG polyclonal (Sigma) as the secondary antibody (35). Fluorescently labeled cells were visualized under dark field optics at 63× magnification. Images were captured and digitized directly to a computer for storage and printing.

RESULTS

Cloning of CRD-BP cDNAs and identification of related genes

Two preparations of highly purified human CRD-BP were isolated and microsequenced. Each gave a different, non-overlapping sequence: P-A-Q-A-V-G-A-I-Q/I-G-k/r-I/K-Y/G-Q-X-i/l-k and V-N-E-L-Q-N-L-T-A-A-E-V-V-P (lower case letters indicate residues of less confidence than upper case letters). A 1069 bp

1	60	ст	~~~	-CT		TAG		-	raci	2001	rccc					CAC		та	TCO	GC	ATA	G
61	GA	GG	AT)	TC	CGC	ccc	CGC	CCCC	cca	GAT	rccc	CAT	TGA	ATC	GAN	CAC	TG	rcc	TTO	cc	ccc	c
121	сA	cc	GCC	:AC	CAT	GAA	cyv,	<u>ec</u> r:	T.M	- MT (Geo	:VVc	CTC	AAC	GAG	IAG1	GTO	GAC	ccc	co	CYC	x
1					м	N	Щ.	<u>.</u>	<u>.</u>	<u>_</u>	<u> </u>	<u>-</u>	L	N	E	s	v	T	r	^	U	
191	ст	TG	GAC		AGT	ATT	CGC	GGA	3C N	-	ATC	TCC	TAC	AGO	GGG	:CA	TT	стт	GG	rca	AAT	с
18	L		E	ĸ	v	F	A	E	H	ĸ	I	S	Y	s	G	Q	F	L	v	X	S	
																					~	_
241	50	GÇ	τ _ν ς	ŝ	CIT	CGT V	CCV	TS	200	GAC	CGAG F	H H	TGG	CCC	M	K	200	T	CG E	~~~	CTT F	Ŧ
	۳,	-		-	÷	÷	-	<u> </u>	•		•			^				•		•		
301	CT	CG	GGC		AGT.	AGA	ACT	GCA	AGG		ACG1	TCT.	GAG	AT	ſGN	CN	TC.	AG1	CCO	CCN	***	x
58	8		G	ĸ	v	E	L	Q	G	ĸ	R	L	E	I	E	н	s	v	P	×	K	_
361			hace		TCG			ACM		cca			icc)	.cci	CAC	сто	200	ATG	GG	A A G	TGC	т
78	Q		R	s	R	ĸ	ī	0	I	R	N	I	P	₽	Q	L	R	W	E.	v	I	
								_	_													~
421	AG	AT	AGO	<u>70:</u>	<u>cc</u> t	GGC	TCA	GTA	266 6	TACI	NGTO	GAG	N	CTG	TGAC T	CAU 0	NGT V	GAJ	CA T	CTG F		نان ا
50				-	-	~	*	·	•	•	•	-		-	Ξ.	-			-			
481	ŤG	AG	ACĮ	<u>مو</u> د	ççt	6 <u>6</u> 7	CVV	CGT	C A O	CT A	CTCI	•	cee	GA	C A C	AC	CAG	GC)	UN G	CTA	TCA	т
118	E		Τ.	<u>.</u> .	<u></u>	<u> </u>	<u>N</u>	<u>×</u>	Ξ	¥	S	N	R	E	Q	Ţ	R	Q	•	-		L
541	G	AG	сти		TGG	cc.	TCA	ACT	GGA	GAA	CCAI	rcco	CTO		GT	TC	CTA	CAT	AC	ста	ATG	ia.
138	K		L	N	G	н	Q	L	E	N	H	λ	L	ĸ	v	s	Y	I	P	0	2	
601	GC	AG	AT/	UAC	ACA	AGG	TCC	TGA	GAA	TGG	<u>ece</u>	rcc:	i <u>cci</u>	<u>\GG(</u>	CTT'	rcc.	GTC	TCO	- <u>99</u> ;	<u>600</u>	AGC	ñ
128		2	1	T	Q	G	r	E	N	<u> </u>	~	<u>~</u>			-		-		-		-	
661	сс	GG	CN	٩GG	GTC	GCC	CGT	GGC.	AGC	AGG	GGC:	rccJ	AGCO		GCA	GC A	GCC	AG	rGG.	NC)	TCC	c
178	ľ	ł	0	G	S	P	v	x	λ	G	x	P	λ	ĸ	Q	Q	P	v	D	1	. 1	•
221	-		~~~			cc.,		73 C			TGT	ACCO		-	-	TGG		ഹ	166	GTO	ccu	LC.
198	10		R	L	L	V	P	T	0	Y	v	G	A	ī	I	G	ĸ	E	G		1	
				-		-	-		-	-	-									_		:
781	c	TC	CG	~~~	CAT	CVC	:***	YCY	GYC	GCN	GTC	cw	NAT.	NGA	CGT	GCA	TAG	GN	VCC	YCI	ATC.	C.
210	-		R	N	1	Т	K	Q	Ŧ	ų.	3	K	+	0	×	n	~	<u>^</u>	-	_		
841	GC	GC	GC	rcc	GGA	GN	GGC	CAT	CAG	CGT	GCA	TTC	AAC	ccc	TGA.	AGG	CTG	CT	сст	cco	CG:	rG
238	9	;	A	λ	E	ĸ	λ	I	\$	۷	н	s	T	₽	E	G	с	S	S	- 1	• •	2
														~~ .	~~~	~		~~~	~ * ~			37
901	C.	LAG r	AT .	GAT	CTT L	GGJ	T AGA	TAT: M	H H	K	E	ليانياني الأ	×~~	ם ב	T	K	T	۸	D		5 1	v.
		•		•	-	-	-				-			-								
961	T	ccc	CT	GN	GAT	cc	rGGC	TCA	TAA	CAA	CTT	CGT	ccc	GCG	ACT	CAT	TGO	сA	AGG		GC	CG
278	I	9	L	к	I	L	<u> </u>	н	N	N	F	V	G	R	L	1	G	ĸ	E		3	Ē
1021	G		ст	Gλ	GN	GC,	IGGJ		GGJ	CAC	AGA	GAG	Gλλ	GAT	cac	CAT	сто	TAC	cGC	TC	CAG	GA
		N	τ.	к	v		F	~	D	T	E	т	к	Ŧ		T	S	s	T		0	Ď
298					n.		<u> </u>	¥			-	<u> </u>		<u> </u>		_	-					
298	-				R R			¥		~	~		-								7	- TG
298 1081 318	ç	CTO		GC1		NTA.		TGJ	IGAC R	GAC T	CAT I	слс т	TGT	GAA K		ic go	CA!	TTG E	yc)		TGT C	TG C
298 1081 318	c	CTC L	CAC T	GC1 L		NTA.	ACCO	TGJ E	IGAC R	GAC T	CAT I	слс т	TGT V	GAA K		ic go	CA: I	TTG E	YC)		TGT C	TG C
298 1081 318 1141	с с	CTO L	T T	GC1 L CGI		NTA.	ACCO P	E TGJ E	IGAC R	T T	I I	T	TGT	GAA K	GGG		CA!	ITG E	AGJ	AC TG	TGT C GCC	TG C GC
298 1081 318 1141 338	с с	CTO L AGO	T T SGC	GC1 L CGI		NTA. N NGG. E	ACCO P AGA	TGJ E ICA1	IGAC R IGAJ K	I T KGAJ K	I I VAGT	T T T CAC T	TGT V AGA	GAA K GGAA K	GGG G TTA Y	ю сос х .c.с.н Е	I I I I I I I N	NCG	AGJ	LAC I STG	TGT C GCC A	TG C GC A
298 1081 318 1141 338		CTO L AGO R	T T SGC A	GC1 L CGI E		NTA NGG E		TGJ E ICA1	IGAG R IGAJ K	GAC T KGAJ K	I I VAGT	TCAC T TCG R	TGT V AGA E	GAA K GAA GGC A GGC A		icgo A icgu E	I I NGAL N	TTG E ACG D TAG	AG	AC TG	TGT C GCC A TTC	TG C GC A
298 1081 318 1141 338 1201 358	0 " 0 0	CTC L AGC R ATC	GGC A SAG	GC1 L CGI E		NTA NGG E NGT S	ACCO P AGA I CCCC	TGJ E ICA1 M ACC1 L	IGAC R IGAJ K ICA:	AGAC T AGAJ K TCCCC	I I VAGT V CTGG	TCAC T TCG R KGCT	TGT V AGA E TAJ			CGC A CGJ E TGC	CAT I GAL N TG' V	TTG E ACG D TAG	AG	AC TG	TGT C GCC A TTC F	TG C GC A CC P
298 1081 318 1141 338 1201 358	с с	CTC L AGC R ATC	GC T GC A GAG S			NTA. NGG. E NGT S	ACCO P AGA: I CCCL H	TGJ E TCA1 M ACC1 L	IGAG R IGAJ K ICA: I	GAC T K K F CCC P	I I VAGT V CTGG G	TCAC TCG R GCT L	TGT V AGA E TAJ	GAN K GGC A CC1 L		CGC A CGJ E TGC A	I I I I I I I I I I I I I I I I I I I		AG	TG	TGT C GCC A TTC F	TG C GC A CC P
298 1081 318 1141 338 1201 359 1261		CTCL L AGC R ATC	EAC T GGC A SAG S TTC	GCT L CGJ E CT L		AGG. E AGT S SCG	ACCO P AGA? I CCCJ H CTG'	TCA1 M ACC1 L	IGAU R IGAU K ICA: I I CTCO		LAGT VAGT V CTGG CTCC	TCAC T TCG R GCT L CAG	TGT V AGA E TAJ	GAA K GGC A GGC A L CC1 L TG1 V			I I I I I I I I I I I I I I I I I I I				TGT C GCC A TTC F TAT	TG C GC A CC P
298 1081 318 1141 338 1201 358 1261 378		CTC L AGC R ATC M GC	GGC T GGC A SAG S TTC S	GCT L CGJ E CT L S		AGG E NGT S GCG A	AGA? AGA? I CCCJ H CTG' V		IGAGAC R IGAJ K ICAT CTCC P	AGAO T K R TCCO P	I I VAGT V CTGG G CTCC	TCAC TCG R GCT L SCAG	TGT V AGA E TAJ N SCAC	GAA K GGC A CCC1 L CCC1 L TG7 V		CGC A CGJ E TGC A CTGC G	I I I I I I I I I I I I I I I I I I I			AC TG TG CTC	TGT C GCCC A TTCC F TAT Y	TG C GC A CC P S
298 1081 318 1141 338 1201 358 1261 378 1321		CTCL AGC AGC ACC ACC ACC	CAC T GGC A SAG S SAG S CT			AGG AGG AGCG AGG	AGA? I CCCL H CTG	E TCAT H NCCT L TCCC P CCC	IGAG R IGAJ K ICA I CTCC P AGC	AGAO T AGAJ K FCCCO P CTCCO P	LAGT UAGT V CTGG CTCC P	TCAC TCG R GCT L CCAG S	TAJ	GAN K GGC A CC1 L CC1 L TG7 V AAG	G G G TTA Y NGGC A TCAC T	LCGU A LCGJ E TTGK A TTGK	I AGAI N CTG ¹ V GGG ² A TCC				TGT C GCCC A TTCC F TAT Y	TG C GC A CC P AG S
298 1081 318 1141 338 1201 358 1261 378 1321 398		CTC L AGC R ATC M GC' A S	CAC T GGC A SAG S TTC S CTT F	GCT L CGJ E CCT L L CCT L S CCT L L CCT L L CCJ CCJ		AGG AGG S GCG A AGG A	AGAN AGAN I CCCC H CCCC H CCCC V V CCCC V	CTGJ E ICAI N ACCI L TCCC P CGG E	IGAC R R R R R R R R R R R R R R R R R R R	AGAJ K R CTCC P CTCC P AGG. E	CCAT I VAGT V CTGG G CTCC P AGA	TCAC TCG R GCCT L CCAG S TGG? V	TAGA E TAGA E TAGA N SCAC S TACA Q	GAA K GGC A CC1 L TG7 V AAG	GGG G CTTA Y CGGC A TCAC T TGT F	CGC A CGJ E CTGC A CTGC C CTGC C TCA	I I I I I I I I I I I I I I I I I I I				TGT C GCCC A TTCC F TAT Y KGC1 A	
298 1081 318 1141 338 1201 358 1261 378 1321 394		CTC L AGC R ATC M SCC'S	CAC T GGC A GGC A GAG S TTC S CTT F CGC	GCT L CGJ E CCJ L CCJ C CCJ C CCJ C C CCJ C CCJ C C C C		AGG B CG AGT CG AGC AGC AGC ACG ACG ACG ACG	AGA AGA I CCC H CTG V CTG V	CTGJ E ICAI M ACCT L ICCO P CGG E AGA	IGAJ RGAJ K IGAJ K ICA: I CTCO P AGC Q AGG	AGAJ K TCCCC P CTCC P AGG E GCC.	L CAT I V AGT G CTCC P AGA M AGC	TCAC TCG R GGCT L CCAG S TGG? V ACA?	TAGA AGA E TAGA E TAGA S CAC S TACA	GAA K GGC A CC1 L TG1 V NAG	LGGG G CTTA Y Y CGGC A TCAC T TGT' F	E CGC A CGJ E TGC C TGC C TGC C TCA	I I I I I I I I I I I I I I I I I I I				TGT C GCCC λ TTCC F TλT Y KGC1 λ CAGC	TG C GC A C C P A G S T <u>V</u> C C
298 1081 318 1141 338 1201 358 1261 376 1321 396 1381 410		CTC L AGC R ATC A C C S G G G G G	TTC SGC A SAG S TTC S CT F CGC A	GCT L CGI E CT L L S CT S CA S CA S CA S CA		AGG AGG AGT S SCG A AGG A TTG	AGA AGA I CCCCI H CCTC V CCTC P CCTC C CCCCI	CTGJ E TCAT H ACC1 L TCCC0 P CCGC E ACGA K	IGAC R IGAJ K ICAN I CTCO P AGC Q AGC G	AGAO P CTCCC P AGG E GCCC	CTGG G CTGG G CTCC P AGA: H AGC:	TCAC TCG R GCT L CCAG S TGG? V ACA?	TGT V AGA E TAJ N SCAC S TACA S TACA C S TACA C S	GAA K GGC A GGC A CC1 L STG1 V V V NAG V NAC	LGGG G G TTTM Y NGGC A TCAC T TGT F A A C	CGC A CGC E CTGC C CTGC C C C C C C C C C C C C C C	I I I I I I I I I I I I I I I I I I I				TGT C GCCC A TTCC F TAT Y KGC1 A CAGC S	TG GC A CC P AG S TG V GC A
298 1081 318 1141 338 1201 358 1261 378 1321 396 1383 410		CTC L AGC R ATC S GG G G C TC	CAC T CAC T CAC S CAC S CAC CAC CAC	GCT L CGI E CGI E CT L S AT S S A M CA		MGG E NGT S GCG A AGG A TTG	ACCCC P AGA I CCCCC H CCCCCC H CCCCCCC C CCCCCCCCCC		AGAA R RGAA K CACA CAGC CAGC		L CCAT I I V CTGG G CTCC P AGA2 M AGC2 H CAC	TCAC TCG R GGCT L CCAG S TGGT V ACA: I	TGT V JAGA E TAJA N SCAC S TACA O TCAL	GAA K GGC A GGC A CC1 L GGC A CC1 L GGC A CC1 L CC1 CC1 CC1 CC1 CC1 CC1	L GGG G G TTTM Y NGGC A TCAC T TGT F AAC	E CGC A E CGC E CTGC A CTGC G TCA TCT S	I I I I I I I I I I I I I I I I I I I				TGT C GCCC A TTCC F TAT Y SGC1 A CAGC S	
298 1081 318 1141 338 1201 350 1261 370 1321 390 1383 410		CTCLAGE R ATC S GG C TC S	CAC T CAC T CAC S CT S CT S CT S C CAC C CAC T C CAC T S C C C C C C C C C C C C C C C C C	GCT L CGI E CT L L S CA S CA H CA		AGG AGG SGCG A AGG A TTG C TTG	AGA P AGA I CCCC H CCCC H CCCC V V CCCC V V CCCC P CCCC A P C CCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCC H C CCCCCC	TCAT E TCAT H ACCT L TCCCO P CCCC E AGA K CAC	IGAA R IGAA K ICA I CTCO P AGGC G CAGG CAG	AGAO T AGAJ K FCCCC P CTCCC P CTCCC P AGG C C C C C C C C C C C C C C C C C C	ZCAT I V CTGG G CTCC P AGA: H AGC: H CACC	TCAC TCG R GGCT L CCAG S V ACAC TGGC TGGC D	TGT V IAGA E TAJ N CAC S TACA S TACA S	CCA.	L L L L L L L L L	CGC A CGJ E CTGC CTGC C CTGC C CTGC C CTGC C C C C	I AGAJ N CTG' V GGG A TCCC R CCCC R GAA				TGT C GCCC A TTCC F TAT Y GC1 A CATC I	
298 1081 318 1141 338 1201 358 1261 378 1321 396 1383 4144 431		CTCLAGC R ATC S GG CTS	EAC T SGC A SAG S TTC S CTT F CGC A I CAT	GCT L CGI E CT L CAT S C C C C C C C C C C C C C C C C C C		ATAL N AGG E AGT S GCG A AGG A TTG TTG	AGAN AGAN I CCCCH CCCCH CCCCH CCCCCC	TCAN E TCAN H ACCO F CCC F CCC E AGA K CAC	IGAG R IGAJ K ICA: I CTCC P AGC G CAG CAG	AGAC T AGAJ K FCCCC P CTCCC P AGG CTCCC P AGG C CTCCC P	CTGG G CTGG G CTCC P AGA: H CACC P	TCAC TCG R GCT L CCAG S TGGT V ACA: L CTGJ	TAJA N CACAC S TACAC S TACAC S TACAC S	CCA K K K K K K K K K K K K K	L GGG TTTA Y GGGC T TGT TGT L L NAG	E CGC A E CCGC E CTGC A CTGC A CTGC A TCA I TCT S TTCC R	CCA'I I AGAIN CTG'V CCCC R GAA				TGT C GCCC A TTCC F TAT Y SGC1 A CAGC S	
298 1081 318 1141 338 1201 356 1261 376 1321 394 1381 414 1441 434		CTL AG R ATH GC A TC S GG TC S GG		GCT L CGI E CT L S CT L S CA H CA H CA H CA K CA		ATAL N AGG E AGT S GCG A AGG A TTG J AGG	AGA AGA I CCCC H CCCC H CCCC C CCCC C CCCC C CCCC C CCCC C C CCCC	CTGJJ E TCA1 M ACCT P CGG E AGA K CAC P	IGAU R IGAU K ICA: I CAC CAC CAC CAC CAC		CAT I V CTGG G CTCC P AGA' H CACC	TCAC TCG R R GCT L CCAG S TGG V ACA: D ACA:	TAGA AGAA E TAJA N S CAC S TACA S TACA S TACA S CAC	GAN K GGC A GGC A CC1 L TG1 V A CC1 L TG1 V A CC1 CC1 CC1 CC1 CC1 CC1 CC1	LGGG G TTTA Y NGGC T TGT L L NAG V TTT.	CGC A CGJ E TCA TCA TCA TCA TCA TCA TCC R ATG	CAN I AGAI N CTG' V CTG' V CCC P CCCC R GAA GCA				TGT C GCCC A TTCC F TAT Y SGC1 A CAGC I	
298 1081 318 1141 338 1201 358 1201 378 1321 394 1381 414 1442 434 1505 455		CTCL AGE ATO A TO S GG TO S GG	EAC T CAC S CAC S CAC S CAC CAC I CAC P	GCT L CGI E CCI L CCI CCI C CCI CCI CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C CCI C C CCI C C C CCI C C CCI C C C CCI C C CCI C C C CCI C C CCI C C C CCI C C C C CCI C C C C CCI C C C C CCI C C C C CCI C C C C CCI C C C C CCI C		AGG AGG AGG AGG AGG AGG AGG AGG AGG	AGA AGA I CCC H CCC CCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCC H CCCCCC H CCCCCC H CCCCCC H CCCCCC H CCCCCCCC	CTGJJ E TCAT H ACCT P CGG E CAC P CAC P CAC F	IGAO R IGAJ K ICA: I CAC Q ACC Q ACC Q CAC CAC CAC CAC CAC C	AGAJ K AGAJ K P CTCCC P AGG E GCCC Q AAA T AGG	CAT I I V CTGG G CTCC P AGA: H AGC: H CACC P CTCC Q	TCAC T TCG R GCT L CCAG S TCGC S TCGC V V ACA: L CTGL C TGL D AGGG	TGT V HAGA E TAJA N SCAC S TACA S TACA S GAAA R	GAA GGAC L GGC L GGC L CCA K GAA I	L GGG G TTTA Y NGGC A T TGT' F MAC' L AAG' V TTTT. Y	CGCA A CGJ E TGCA TCA TCA TCA TCA TCA TCA TCA TCA TCA C A TCA C A TCA C G	CCA'I I AGAI N CTG' V SGGG A TCC R CCC R GAA M GCA				TGT C GCCC A TTCC F TAT Y GCT A CAGC S CATC I L AGAL	TG GA CP AS TV CA A T AE
298 1081 318 1141 338 1201 358 1261 378 1321 398 1381 410 1441 431 1501 456		TIL AGE AT GA TOS GG TOS GG A	EAC T SGC A SAG S TTC S CT F CGC A CA T T CA T T	GCI CGI E CCI CCI		ATAN NGG E NGT S SCG A AGG A TTG C TTG C C C C	AGA: I CCCC H CCCC H CCCC C C CCCC C C C C C	CTGJJ E TCAT H ACCT P CGG E AGA K CAC P CGG E AGA K CAC	IGAG R IGAJ K ICA: I CTCO P AGC G CAG CAG CAG CAG CAG CAG CAG CAG C	AGAO T AGAJ K P CTCCC P CTCC P AGG C C C C C C C C C C C C C C C C C C	CTCC CTCCC CTCC CTCCC CTCC CTCC CTCC CTCC CTCCC CTCC CTCC CTC	TCAC T TCG R R GCT L CCAG S TGG V ACA: L CTGL C TGG G AGGC	TAJI N CAC S TACA S TACA S TACA S CAC S S ACT S S ACT S S ACT S S A ACT S	GAA K LGGCA L STGT V MAG CCA K GAA I AGA	LGGG G TTTA Y NGGC A TCAC T TGT F AAC' L AAG' V TTTT Y CCC	CGCA A CGCA E TGCA CTGC C C C C C C C C C C C C C C C C C	CAN I I I I I I I I I I I I I I I I I I I				TGT C GCC A TTC F TAT Y SGC1 A CAGC S CATC I C AGAI E SGC2	
298 1081 318 1141 338 1201 359 1261 379 1322 394 1382 410 1442 433 1502 459 1562 470		CTCL AG R ATH GCA TCS GG TCS GG AN	EAC T T SGC A SAG S TTC S CT F CGC A TTC F CGC A T T T F	GCIL CGIL CGIL CGIL CGIL CGIL CGIL CGIL		ATAN NGG E SGCG A AGG A AGG A GTC E	AGA: I CCCJ H CTG' V CTC C CCCA P CTC C CCCA	CAT FCAT M CCAT M CCAT P CCGC E CCCC P CCGC CAT CAT CAT CAT A CAT A CAT A CAT A CAT A A C C A A C C A A A C C A A A C C A A A C C A A A C C A A A C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C C A A A C C A A A A C C A A A A A A A A A A A A A	IGAU R IGAU K I I CTCO P AGC Q AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGC C AGA C A C	AGAC T CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC C P CCCC P CCCC C P CCCC P CCCC C P CCCC C P CCCC C P CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C P C CCCC C C C C C C C C C C C C C C C C	CACT I VAGT V CTGG G CTCC P AGA: H CACC P CTCC P CTCC C TAA.	TCAC TTCG R GCT L CCAG S TGC V ACA: L CTGL G G ACCC L	TAJI E TAJI N CAC S TACIO S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S TAC S C CAC S C CAC S C CAC S C CAC S C CAC S C C CAC S C C CAC S C C CAC S C C C C	GAA GGC A CC1 L TG1 V AAG ^V CCA K GAA T	AGGG G TTTA Y SGGCA T TGT' F AAC T TTT. Y CCCC H	CGC A E TGC TGC TGC TGC TGC TGC TGC TGC TGC TGC	CCA I I I I I I I I I I I I I I I I I I				TIGT C GCCCA TTCCF TAT Y GC1 X GC1 X CAGC Z A GCC Z A C A GCC Z A	
298 298 1081 318 1141 338 1201 359 1261 379 1321 394 1381 414 1442 434 1500 455 1562 477 162		CTCL AGCR ATC S GG TC S GG A A C S GG A A A A A A A A A A A A A A A A A	$\frac{D}{T}$ \frac{D}	GCL GCL CGL CGL CGL CGL CGL CGL CGL CGL		ATAN NGG E SGCG A AGG A TTG TTG AGG C SGTC SGTC	ACCCC P AGA I CCCC H CCCC CCCC CCCC CCCCC CCCCC CCCCC CCCCC CCCC	CTGJ E TCA1 M CCC P CCG E CCC P CCG E CAC P CCG E CAC P CCG E CAC P C CGG E CAC P C CGG E TCCO P C CGG E TCCO P C CA1 S C C C C S C S C S C S C S C S C S C	IGAO R IGAO K ICAN I CTOO P AGC Q AGC G CAG CAG CAG CAG CAG CAG CAG CAG CAG CA	AGAC T T C T C C C C C C C C C C C C C C C	CAT I V CTGG G CTCC P AGA: H AGC: H CACC P CTCC C C C C C C C C C C C C C C C	TCAC TTCG R GCT L CCAG S TGGT V ACA: L CTGJ GCA	TAJ E TAJ E TAJ E TAJ E TAJ E TAJ E TAJ E TAJ E AAA	GAA GAA GGCA L GGCCA L GAA V CCA GAA T CCGG	L GGG G G TTTA Y GGC A TGT TGT TGT TGT TTT. Y CCCC H TGA	CGC A CGC E CGC E CGC A C CGC A C A C C C C C C C C C C C	CCA I I N CTG' V CCC R GAA GCA M GCA N TAC				TIGT C GCCC A TTCC F TAT S GC1 A S CATC I C A GC1 A S C A C A C C T T C C C C C C C C C C C C	
298 298 1081 318 1141 338 1201 358 1201 378 1201 378 1201 378 1383 414 1383 414 1383 414 1503 451 1563 471 162		TILL AG A TO S GG TO S G A N GA	TTO SGO A SAG S TTO S TT	GCL CGL CGL CGL CGL CGL CGL CGL CGL CGL		ATAL NAGG E AGT S GCG A AGG TTG J AGG C C C C C C	ACCCC P AGA I CCCCC H CCCCCC CCCCCCCCCCCCCCCCCCCCC	CGG E CGG E AGA K CAC P CGG E AGA K CAC P CGG E AGA CAC P CGG E CAC P CGG E C CGG E C C C C C C C C C C C C C	IGAU R IGAU R I CAN I CAN I CAN I CAN C AGG C AGG C AGG C AGG C AGG C AGG C AGG C AGG C AGG C AGG C AGA C A C	AGAC T AGAJ K FCCCC P CTCCC P CTCCC P AGG C C C C C C C C C C C C C C C C C C	CAT I V CTGG G CTCC P AGA: H AGC: H CAC P CTCC C C C C C C C C C C C C C C C C	TCGCT TCGCT R GCT L CCAG S TGCT V ACA: L CTGD G AGCC L GCA K	TAJ TAJ TAJ TAJ TAJ TAJ TAJ TAJ	GAA GAA GGC L GGC A CC1 L T T CC2 CCA K GAA T CCG V V	LIGGO G G TTTA Y GGCC A CAC T TGT F CAC T TGT T CAC C T TTT Y C C C C H T TGA	CGC A CGC A CGC CGC A CGC CGC A CA CGC A CA CGC A CA CGC A CA CGC CGC	I AGAL N CTG' V SGGG A TCC P CCC R GAA CCC R CCCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCCC R CCCC R CCC R CCC R CCC R CCCC R CCC R CCC R CCCC R CCC R CCC R CCCC R CCCC R CCCC CCCC R CCCC CCCC R CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC				TGT C GCCC A TTC F TAT Y GCC1 A C A GC2 A C C C C C C C C C C C C C C C C C C	
298 208 318 318 338 1201 358 1201 378 1322 394 1383 414 1442 434 1502 459 1566 477 1622 49		TIL GR ATH GA TS GG TS GG AN GA	CAC T CAC T CAC T CAC T CAC T CAC T CAC T T CAC T T CAC T C CAC T C C C C C C C C C C C C C	GCTL CGL CGL CGL CGL CGL CGL CGL CGL CGL CG		ATA N AGG E AGG A AGG A AGG A C A C C C C C C C C C	ACCCC P AGA I CCCC H CCCC P CCCC CCC CCC CCC CCC CCC	CAT FICAT M ACCT FICCO FICCO FICCO COG COG COG FICO COG FICO COG FICO COG FICO COG FICO COG FICO FI	AGAG R RGAJ K I CTCO P AGC Q AGG G CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C CAG C C C C	AGAC T XGAU T CCCC P CCCC C CCCC P CCCC C CCCC P CCCC C CCCC P CCCC C CCCC P CCCC C CCCC C CCCC C CCCC C CCCC C CCCC	CAT I V CTGG G CTCC P AGA: H AGC: H CTCC Q TAAA CTCC Q TAAA CTCC G G G G G G G G CTCC C P	TCGCT TCGCT R GCT L CCAG S TGCT V ACA: L CTGD G AGCC L GCA K	TAGA N CAC S TACA S CAC S TACA S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S CAC S TACA S CAC S TACA S S TACA S TACA S TACA S TACA S TA S T	GAA GGCA LGGCA LGGCA LGGCA LGGCA CCA K GAA LGGCA		CGCA E TGA E TCA I TCT S TCCA I TCT S TCCA I TCT S TCCA I I TCT S TCCA I I TCT S TCCA I I TCT S	I AGAL N AGAL N V SGGA A TCCC R GAA GGAA K TACC F I AGC				TGT C GCCCA TTCC F TAT S GCCTA S GCCTA S GCCTA S GCCTA L C TTC L	
298 298 1081 318 1141 338 1201 358 1201 358 1201 358 1201 358 1321 378 1321 378 1321 378 1321 378 1321 1381 411 1441 431 1505 457 1565 477 162 497 162		TILL AR ATH GA TS GG TS GG AN GA GA	ACT SGC A SAG S CT F CA ACP TT F AGA AGA	GCL GCL CGL CGL CGL CGL CGC CGC CGC CGC		ATAL N NGG E SGC A AGG A TTG TTG GTC TTG	ACCCC P AGA: I CCCCI H CCCCC P CCCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCC	TGM E CCA E CCA E CCA E CCA E CCA E CCA E CCA E CCA E E CCA E E CCA E E CCA E E CCA E E CCA E E CCA E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E CCA E E E E	AGAG R RGAJ K I CTCO P AGC Q AGC G CAG CAG CAG CAG CAG CAG CAG CAG C	AGAC T CCCC P CTCCC P AGG C CCC C C C C C C C C C C C C C C C	CAT I V CTGG G CTCC P AGA: H AGC/ H CTCC Q CTCC Q CTCC Q CTCC Q CTCC C P AGA: H CTCC Q C C C C C C C C C C C C C C C C	TCAC TCG R GCT L CCAG S TGG V ACA: D ACA: D AGG G AGC L CTG L CTG L CTG AGG C A CAC	TAGA TAGA N CAC S TACA S S TACA TACA	GAA GAA GAA GAA GAA CC1 L GAA CC1 L GAA CC1 CC2 CCA CCA CCA CCA CCA CC2 CCA CC2 CC2	L GGG G G TTTA Y NGGCA TTTA TGAC T TGT TGT TGT TGT TGA N C C C C T TTTA	CGC A E TGA E TCA I TCT R TCC R TCC A CA A CA A CA A CA A C	I COLORED COLO				TGT C GCCCA TTCC F TAT S GCC1 A C C TAT S S C C TAT S S C C C C C C C C C C C C C C C C C	
298 298 1081 318 1141 338 1201 358 1261 370 1321 394 1382 414 1382 414 1442 436 1566 456 1566 477 162 499 168 51		TILLAR AT GATS GG TS GG AN GA GA	T SGC A SAG S STC S CF CG A CA A CA P TT SGC A A A A A A A A A A A A A A	GCT L CG L CG L CG C C C C C C C C C C C C	Y AGCI Y IGCI S TGC S TGC TTCA I AGA I I CAG	ATAL NGG E AGG AGG AGG AGG AGG AGG AGG AGG AGG	AGA AGA I CTG CTG CTG CTG CTG CTG CTG CTG	TGU E TCAN M ACCT P CGG E CAC P CGG E CAC P CGG E CAC P CGG E CAC P CGG E CAC P CGG E CAC P C CGG E CAC P C CGG E CAC TCCC P CAC TCCCC P C CAC TCCCC P C CAC TCCCC P C CAC TCCCC P C CGG E C CAC TCCCC P C CGG E C C C C C C C C C C C C C C C C	AGACAR IGALAGA IGAL	AGAC T CCCC P CTCCC P AGG C CCC C C C C C C C C C C C C C C C	CTCC CTCC G CTCCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCC CTCCCC CTCCCCCCCC	TCAC TCAC TCG R GCT L CCAG S TGG V ACA: D ACA: D ACA: D ACA: CTG D ACA: T CTG C ACA: T C CAC C T C C C C C C C C C C C C C C	TAGA V IAGAA E TADA N CAC S TACA S TACA CAC S TACA S S TACA TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA S TACA TACA S T	GAA K GGA LGGCC LGGCC LGGC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC LGGCC		CGC A CGC A CGC A CGC A C CGC A C A C C C C	CCA I I I I I I I I I I I I I I I I I I				TGT C GCCC A TTCC F TAT Y GCT A CATCO L CATCO L CATCO L CATCO L	
298 298 1081 318 1141 338 1201 359 1321 378 1321 399 1383 410 1443 430 1563 450 1563 470 162 490 168 51		TILL AR AT GA TS GG TS GG AN GA GA GA	T GOA SAGS TTO S CT F COA A P TT F AGA A A A A A A A A A A A A A A A A A	GCIL COL CAL CAL CAL CAL CAL CAL CAL CAL CAL CA		AGG E AGG A	AGA: I CTG' CTG	TGC TGJ E TCAN M L TCCC P CGG E CAC P CGG E CAC P CGG E TCCC TCCC TCCC TCCC TCCC TCCC TCC	AGAGA R R R R R R R R R R R R R R R R R	AGAC T CCCC P AGGA C CCC AAAA T AGGA C AAAG AAAG C AAAG C AAAG C AAAG C AAAG C C C C	CTCC CTCC G CTCCC CTCCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC C	TCAC TCAC R CAC TCAC R CTCG S TCCAC S CTCG CTCG CTCG CTCG CTCG CTCG CTCG C	TAGA TAGA	GAA GAA GGA LGCL IGGC LGCC IGGCC IGGC IGGC IGGC IGGC IGGCC IGGCC IGGCC IGGCC IGGCC IGGCC IGGCC IGGCC IGGC	L GGG G G TTA Y GGC A TCAC T TGT TGT L CCC H TTTT Y CCC H TGA A TGA C H TGA C H TGA C C C C C C C C C C C C C C C C C C C	CGC A CGC A CGC C A C C C C A C C C A C C C C	I CCC I CCCC I CCC I CCCC I CCCCC I CCCCC I CCCCC I CCCCCCCCCC				TGT C GCCC A TTCC F TAT Y CAT C CAT C L CAT C	
298 298 1081 318 1141 338 1201 358 1261 378 1321 396 1321 396 1321 396 1321 396 1321 396 1321 396 1321 396 1355 1567 470 162 497 168 51 174 53		TILLAR AT GA TS GG TS GG AN GA GA AN A	T GC A GA GA GA I			AGG E S GCG A AGG	ACCC P AGA I CCC C CCC C CCC C CCC C CCC C CCC C CCC C	TCA1 E FCA1 H ACC1 E FCA1 H CCC P CCC E C C C C C C C C C C C C C C	AGAGA R R AGGAG CAGG CAGG CAGG CAGG CAGG	AGAC AGAD K CTCCC P CTCCC CTCCC C C CTCCC C C CTCCC C C CTCCC C C C C C C C C C C C C	CAT I V CTGG G CTCC P AGA: AGC: CTCC Q CTCC CTCC CTCC CTCC CTCCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCC CTCCC CTCC CTCC CTCCC CTCC CTCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCCCCC	TCAC TCAC TCAC R CTCG CTCG CTCG CTCG CTCG CTCG CTCG C	TAGA TAGA	GGAA K GGAA K GGAA K GGAA CC1 CCA K GAA CC1 CCA K GAA CC1 CCA CCA CCA CCA CCA CC1 CCA CC1 CC1	L GGG G G TTA Y GGC A TCAC T TGT TGT T TGT L C C C C C C C C C C C C C C C C C C	CGC A CGU E TGC CGC A CGU E TGC CGC A CGU E TGC CGC A CGU E TGC CGC A CGU E TGC A TGC CGU E TGC A TGC CGU E TGC A TGC CGU E TGC A TGC CGU E TGC A TGC CGU E TGC A TGC CGU E TGC A TGC TGC A TGC A TGC A TGC A TGC TGC TGC TGC TGC TGC TGC TGC	I AGAN I AGAN I TG' V I GGA A TCC P CCCC R A TCCC R A TCCCC R A TCCC R A TCCCC R A TCCCCC R A TCCCC R A TCCCCC R A TCCCCCCC A TCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC				TGT C GCCC A TTCC F TAT Y GC1 A C C T C C T C C T C C C C C C C C C C	
298 298 1081 318 1141 338 1201 358 1261 379 1322 1388 410 1442 430 1563 456 456 1565 477 162 49 168 51		TILLAR AT GA TS GG TS GG AN GALGALAK	T GOA SAG S CT F COA A CA I A CA A A A A A A A A A A A A A	GCTL CGI CCI CCI CCI CCI CCI CCI CCI CCI CCI	AGCI Y AGCI G G G CCA CCA	ATAL NAGG E AGT S S C G C G C C C C C C C C C C C C C C	ACCC P AGA I CCC C CCC C CCC C CCC C CCC C CCC C CCC C	TCA1 E FCA1 H ACC1 H CCC P CCC E AGA K CAC P F AGG CAC F F AGG TCC G G G TCC TCC TCC TCC TCC TCC TCC	AGAG R RGAJ K I CAG G CAG G CAG G CAG CAG CAG CAG CAG C	AGAC T CTCCC P CTCCC CTCCC C C CTCCC C C CTCCC C C CTCCC C C C CTCCC C C C C C C C C C C C C	CAT I V CTGG G CTCC P AGA: M AGC: CTCC Q CTCC CTCC Q CTCCC CTCC CTCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCCCCC	TCAC TCAC TCG R GCT L CTG S TGG S TGG S TGG C TGG C TGG C TGG C T C TGG S T C C TGG S T C C AC AC C T L C C AC C S S C T L C C AC S S C T L C C AC S S C T L C C AC S S C T L C C AC S S C T L C C AC S S C T C S S C T C S S C T C S S C T C S S C T C S S S C T C S S C T C S S S C T S S S C T S S S C T S S S C T S S S S	TAGA TAGA TAGA TAGA TAGA TAGA TAGA TAGA TAGA TAGA TAGA	GAN K GAN K GAN K CC1 L GAN CC2 CCA K GAN CC2 CCA CCA CCA CCA CCA CCA CCA	L GGG G G TTTA Y SGGC T T T T T T T T T T T T T T T T T T	CGC A CGZ CGC CGC CGC CGC CGC CGC CGC	GCA: I I I I I I I I I I I I I I I I I I I		AGU AGU AGU AGU AGU AGU AGU AGU AGU AGU		TGT C GCC A TTC F TAT Y GCT A C C TTC L C A T C C TTC L C A T C C T C T C T C C C C C C C C C C	
298 1081 318 1141 338 1201 358 1261 379 132 138 410 144: 430 156: 470 162 49 168 51 174 53 180		I TLI GR AT GA TS GGI TS GGI AN GAI GAI AKI GA	T GGA GA	GCL CGL CGL CGL CGL CGL CGL CGL CGL CGL	AGCI Y AGCI G G G CCA G TGC G CCA G TGC G CCA G TGC G CCA G TGC G CCA CCA	AGG	AGA: AGA: I CCC: H CCC:	CAC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC P CCGC CCC CC	AGAG R RGAJ K I CAG G CAG G CAG G CAG G CAG CAG CAG CAG	AGAG AGAJ K CTCCC P CTCCC P AGG C C C C C C P C C C C C P C C C C C P C C C C C P C C C C C P C C C C C P C C C C C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C P C P C C C C C C P C P C	LAGT G G CTCC P AGA: H AGC: CTCC P AGA: H CACC C CTCC Q CTCC Q CTCC Q CTCC Q CTCC Q CTCC Q CTCC Q CTCC Q C CTCC Q C CTCC C P AGA: H AGT C C C C C C C C C C C C C C C C C C C	TCAC TCAC TCAC TCCAC R R GCT L CTGD CTGD CTGD CTGD CTGD CTGD CTGD CTGD	TAGA TAGA TAGA TACA	GAA K GAA CC1 L GAA CC2 A CC2 A CC2 A CC2 CCA CCA	L GGG G G TTTA Y GGC TGT TGT TGT TGT TGT TGT TGT	CGC A CGC A CGC A CGC A C CGC A C C CGC A C C C C	GAA GAA GAA GCA I I GAA GCA I C C C C C C C C C C C C C C C C C C				TGT GCC A TTC F TAT C A C A C A C A C A C A C A C A C C A C C C A C C C A C C C A C C C A C C C A C C C C A C C C C C A C C C C C C C C C C C C C	
298 298 1081 318 1141 338 1201 359 1261 370 1322 399 1382 414 1321 399 1382 414 144, 431 1566 457 162 499 168 51 174 53 1800 55		TILL AR AT GATS GG TS GG AN GA GA AK GA	T GGA GA	GCT L CAL STATES	A GCLI Y A GCLI CCAG S TGCCC CCAG TGCCC TGCC TGCC TGCC TGCC TGCC TGCC TGCC TGCC TGCC	ATAL NAGG	AGAN H CCCCH CCCCH CCCCH CCCCH CCCCCH CCCCCH CCCCCH CCCCCH CCCCCH CCCCH CCCCH CCCCH CCCCH CCCCCH CCCCH CCCCH CCCCH CCCCH CCCCH CCCCCH CCCCH CCCCH CCCCCH CCCCCH CCCCCH CCCCCH CCCCCH CCCCCH CCC	CAC P CCGG C CCC P CCGG C C C C C C C C C	AGAGA RAJACA I GAJA K I CA I CTOP AGC G AGC G CAG CAG CAG CAG CAG CAG CAG	AGAG AGAD CTCCC P CTCCCC P CTCCC CTCCC P CTCCC CTCCC P CTCCC CTCCC P CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCCC CTCCCC CTCCCC CTCCCCCC	LAGC CAT I I V CTGG G CTCC P AGA H H AGC H CACC P CTCC Q TAAA CACC C C C C C C C C C C C C C C C	TCAC T TCG R GCT L CTGI GCA G AGAC CTGI GCA K AGAC C TGC C TGC C TGC C TGC C TGC C TGC C TGC C T C C C C	TAJ AGA SCAC S TAJ N CAC S TAJ N CAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S TAJ SCAC S S TAC S S S S S S S S S S S S S	GAA K GAA CC1 L GGA L CC1 CCA K CCA K CCA CCA CCA CCA CC	L GGG G G TTTA Y GGC A TTTA TGT TGT TGT TGT TGT TGT TGT TGT	CGC A CGC A CGC C C C C C C C C C C C C	GCA I I I I I I I I I I I I I			ACC P CAC O GCC A GV AAA N GT' CC C A C O C O	TGT C GCC A TTC F TAT S GC A C TTC I C A C TTC I C A C TTC I C A C C C C C C C C C C C C C C C C	
298 298 1081 318 1141 358 1201 378 1321 378 1361 410 1442 431 1505 450 1567 457 162 49 168 51 174 53 180 55		TIL AR AN GA TS GG TS GG AN GA GA AN GA GA AN GA ST	ACTING A SAGE TTO STORE TO SAGE A SAG	GCTL GCL CGL CGL CGL CGL CGL CGL CGL CGL CG	A GCI Y A GCI I GCI S T GC T C A A GA S T GC C C A C A GA S C C A C C C A C C C C	ATAL NAGG	AGA: AGA: I CTG· V CTG· V CTG· V CTG· V CTG· V V CTG· V V CTG· V CTG· V V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· V CTG· CTG· V CTG· V CTG· V CTG· CTG· V CTG· CTG· V CTG·	CAC P CCGG C CCC P CCGG C C C C C C C C C	AGGAC R RGAJ RGAJ RGAJ RGAJ RGAJ RGAJ RGAJ RG	AGAG AGAD CTCCC P CTCCCC P CTCCC C CTCCC P CTCCC C CTCCC C CTCCC P CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCC C CTCCCC C CTCCCC C CTCCCC C CTCCCCC C CTCCCCCC	LAGC	TCAC T TCG R CAC T CAC T CAC S C C S C C S C C S C C S C C S C C S C C S C C S C S C S C S C S C S C S C S C S C S C S C S C S C S S C S	TAGA TAGA TAJ N CAC S TACA S S TACA S S TACA S S TACA S S TACA S S TACA S S TACA S S TACA S S TACA S S S TACA S S S S A A A T S S S A A A S S S A A A S S S A A A S S S A A A S S S A A A S S S A A A S S S A A A A A A S S S A A A A A A A A S S S S A A A A A A A A A A A A A	GAA K GGA L GGA C GGA GGA C GGA GGA L GGA C GGA C GGA GGA C GGA GGA	L GGG G G TTTA Y GGC A TTTA T TTT Y CCC H A A TTTA Y CCC H A A TTTA Y CCC H A A T G C C C C C C C C C C C C C C C C C C	CGC A CGC CGC A CGC CGC A CGC CGC	GAA GAA GAA TCCC R GAA TCCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R GAA TCCC R TCCC R GAA TCCC R GAA TCCC R TCCC R TCCC R TCCC R TCCC R TCCC R TCCC R TCCC R TCCC R TCCC R TCCC TCCC R TCCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCC TCCCC TCCC TCCCC TCCC TCCCC TCCC TCCCC TCCCC TCCCC TCCCC TCCCCC TCCCCCC		AGJ AGJ AGJ AGJ AGJ AGJ AGJ AGJ	ACC P CAC Q GCC A GTC V AA C P AA C P GA C R C	TTC C GCCCA TTC F TAT Y GC C A C TTC F TAT Y C A C C T C A C C C C C C C C C C C C C	
298 298 1081 318 1141 358 1201 358 1261 370 1322 138 138 1370 1322 138 138 138 1370 1322 138 138 1370 1322 138 138 1370 1322 138 1370 1322 1370 1370 1370 1370 1370 1370 1370 1370		L CTCL L AGO R ATC S GGG G TTC S GGG G A TTC S GGG G A TTC S GGG A TTC S GGG A TTC S GGG A TTC S GGG A TTC S GGG A TTC S S A TTC S S A TTC S S A TTC S S A TTC S S S A TTC S S S S S S S S S S S S S S S S S S S	ACTING A SAGE TTO STORE TO SAGE A SAG	GCL CGL CCL CCL CCL CCL CCL CCL CCL CCL	RECTINE Y AGCIN Q CCAAS TGCC S TGCC CCAS TGCCC CCAS TGCC	ATAN NGGE SCGA SCGA SCGA AGG TTG TTG SCCC TTG SCCC SCCC SCCC	AGA: AGA:	TCAT E TCAT M ACCT P CCG C C C C C C C C C C C C C	AGGAC R RGAJ RGAJ RGAJ RGAJ RGAJ RGAJ RGAJ RG	AGAG AGAG CTCCC P CTCCC P AGG C CCC P CTCCCC P CTCCCCCC P CTCCCCCC P CTCCCCCCCC	LAGC	TCAC T TCG R GCT L CCAG S CCAG S CCAG S C CCAG S C CCAG S C C C C C C C C C C C C C C C C C C	TADA N CACS TADA TACACS TACAC S TACAC	GAA K GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGA L GGC GGC	L GGGG G G TTTA Y GGGC A A T T T T T T T T T T T T T	CGC A CGC CGC	GCA I I I I I I I I I I I I I		AGJ AGJ GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC	ACCP CAC OF CAC	TTC C GCCCA TTC F TAT Y GCC1 A CATCA I CATCA I GAG R AAC	
298 298 1081 318 1141 358 1261 358 1261 359 1261 359 1261 359 1261 359 136 410 144: 450 156: 450 156: 51 168 51 174 53 180 55 7		CTCL L CTCL L CTCL L CTCL S GG G CTCC S CTCC S CTCC S CTCC S CTCC S CTCC S CTCC CTCL C CTCC CTC	ACTING A SAGE TO CELLING A SAG	GCIL GCIL CCIL CCIL CCIL CCIL CCIL CCIL	TTGC	ATAN NGGE SGCGA AGGT TTG GGT GGT GGA GGA GGA GGA GGA		TCAT E TCAT M ACCT P CGG E AGA CAC P CGG CAC P CGG CAC CAC CAC CAC CAC CAC CAC	AGGAC RAL TGAL TGAL TCA TCA TCA TCA TCA TCA TCA TCA TCA TCA	AGAG CTCCC P AGGAG CTCCC P AGGAG CC AAA CC AAA CC AAA CC AAA CC AAA CC AAA CC AAA CC AAA CC CC	LAGE CACITOR CAC	TCAC T TCG R GCT L CCAG S S CCAG S S CCAG S C CCAG S C C CCAC S C C C C C C C C C C C C C	TAJA E TAJA E TAJA E TAJA E TAJA C CCA C C C C C C C C C C C C C	CCAA K GAA CCA	L GGGG G G TTTA Y GGCC A A T T T T T T T T T T T T T	CGC A C C C C	GCA I I I I I I I I I I I I I		AGJ AGJ GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC	ACC P CAC Q GCC A GTC V AAA N GTC V AAAA N GTC V AAAA N GTC V AAAA N GTC V AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TTC GCCC A TTC F TAT Y GCC A TTC F TAT Y GCC A TTC F TAT Y GCC A TTC F TAT Y GCC A TTC F TAT S C C A A C C C C C C C A C C C C C C C C C C C C C	
298 298 318 318 318 338 1201 358 1201 378 138 1201 378 138 138 138 138 138 138 138 138 138 13		CTTCL L CTTCL AGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CGG CTTCS CTTCS CTTCS CTTCCS CTTCCS CTTCS CTTCCS CTTCCS CTTCS CTTC	ACTING A SAGE TTE CALL ACTING A A A A A A A A A A A A A A A A A A A	GCL GCL CGL CGL CGL CGL CGL CGL	TTTG CCAG			CAC FCA1 H ACCT FCA1 H ACCT F CAC F	AGGAC RACATOR ICAN ICAN ICAN ICAN CAGG CAGG CAGG CAGG CAGG CAGG CAGG C	AGAG CTCCC P CTCCC P AGG CTCCC P AGG CTCCC C CTCCC P AGG C CTCCCC P AGG C CTCCC P AGG C CTCCC P AGG C CTCCC P AGG C C C C C C C C C C C C C	LAGE CACITICS CAT CACITICS CAC	TCAC TTCG R SGCT L CAG S SGCT L CTGG S CTGG C CTGG C CCCC C CCCCC C CCCCC C CCCCCCCC	TAGA TAGA	CCAA K GGAA GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGAA L GGCAA L GG	AGGG G G TTA Y GGCC T TGA CAC T TGA CAC TTTA AAG CAC TTTA	CGC A CGCC A CGC A CGC A CGC A CGC A CGC CGC	CCCC R CCCC CCCC R CCCC CCCC R CCCCC CCCC CCCC CCCC CCCCC CCCCC CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC			ACC P CAC C P	TTCC GCCCA TTCC F TAT Y GC1 A CAT CAT I GAG R AAAT CCA	
298 298 318 318 1141 338 1201 358 1201 378 132 1378 132 1378 1378 1378 1378 1378 1378 1378 1378		CTCLL AGG	ACT GA SAGS TTS TT GA CALLACP TT F GA GAL GAL TO GA GATA	GCL CGE CCE CCE CCE CCE CCE CCE CCE CCE CCE	AGCJ Y AGCJ Q IGCJ G G G CAG C C CAG C C C C C C C C C C C C C	ATAL NAGGE AGE AGE AGE AGE AGE AGE AGE AGE AGE	AGA I CCCC H CCCC H CCCCC CCCCCC	CAC FCA1 H ACCT FCCA F CAC CAC	AGGA AGGG CAGG CAGG CAGG CAGG CAGG CAGG	AAAG CTCCC P CTCCC P AGGA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCAA CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	CAC CAT G G CTCC G CTCC G CTCC C P AGA CAC C CTCC C C C C C C C C C C C C C C	TCAC T TCG R GCT L CTG S TGG S CTG GCAC C TGC C CCCC TGCC C TGCC	TAJ	GAA GAA GGA GGA GGA CC1 CC2 CCA CC4 CC4 CC4 CC4 CC4 CC4 CC4	LGGG GGGGG YTTA YGGCA TGTT TGTT TGTT TGTT TGTT TGTT TGTT	CGC A CGC E TGC CGC A CGC C CGC A C CGC A C C C C C C	CCAT	CTTG CTTT CTTTA CG CTTTA CTTTA CCTTC CTTC	AGJ AGJ AGJ AGJ AGT AGT AGT AGT AGT AGT AGT AGT AGT AGT	GTC CACCONCENT OF CONCENT OF CONC	TGT GCCCA TTCC F TATTY GC1 XGA1 E CATC I GAG R AAC AAATA CGA	
298 298 318 318 338 1201 358 1201 378 1322 399 138; 410 144; 431 150; 450 168 51 168 51 174 53 186 57 192 198 204 219		$\frac{1}{2} \frac{1}{2} \frac{1}$	D CACCACA T T CCCCCACA S S S T T CCCCCACA S S C T T C CCCCCACA S S C T T T T T T T T T T T T T T C CCCCCACA S S S S S S S S S S S S S S	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	AGCI Y AGCI Q IGCI S TCA I IGCI S TCA I I CAG C C C C C C C C C C C C C C C C C C	NTAL NGG E SGCG A SGCG A SGCGC	AGA I CCCC H CCCC H CCCC H CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	CAC CAC CAC CAC CAC CAC CAC CAC	AGGA AGGG CAGG CAGG CAGG CAGG CAGG CAGG	AGAG AGAG CTCCC P CTCCC CTCCC P CTCCC CTCCC P CTCCCC CTCCC CTCCCC CTCCCC CTCCCC CTCCCCC CTCCCCCCCC	CACI G G CTCC G CTCC G CTCC G CTCC G CTCC G CTCC C C C	TCG CAG R GCAT L CAG S TGG V ACA L CTGG CCG CCG CCG CCG CCG CCG CC	TAJ	GAA GAA GGA CC1 GGA CC1 CC1 CC2 CCA CC2 CCA CC2 CCA CC2 CCA CC2 CCA CC2 CCA CC2 CCA CC2 CC2	AAGG GGG CTTA Y GGC AAG TTTTY CCCC H AAGG CTTTA AAGG CCTT CCCC CCTT CCCC CCTT CCCC CCTTA CCCC CCTTA CCCCC CCCCC CCCC CCCCC CCCC CCCCC CCCC CCCC CCCC CCCCC CCCC CCCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCC CCCC CCCC CCCC CCCCC CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	CGC A CGC A CG	CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	E CAGE		CAGG CCAG	TGT GCCCA TTCC F TATTY GC1 XGA1 E CATC I GAG R AAC AAATA CGGACCT	TTG C GC A C P AG S S S S CAC T T S CAC T T S C C C C C C C C P AG S S C C P AG S S C C P AG S S C C C C C C P AG S S C C C C C C C C C C C C C
298 298 318 318 318 318 320 358 1201 378 358 1201 378 1383 414 1383 414 1383 414 1383 414 1505 451 1565 451 168 51 174 53 1800 55 51 98 204 4210 2198 204 2198 205 2198 204 2198 205 2198 204 2198 205 2198 205 2198 205 2198 205 2198 205 2198 2198 2198 2198 2198 2198 2198 2198		CTCLL CTCLL AAGCCTCL AAGCCTCS GGGGC CTCCS GGGGC ATTCCS AGCGGGCC ATTCL AGCCGGGCC AGGGGGCC AGGGGGCC AGGGGGCCC AGGGGGCCC AGGGGGCCCCCCCCCC	2 CACCAC T T CACCAC S S C T T C S C T T C C C T T C C C T T C C C T T C C C T T C C C T T C C C T T C C C T T C C C T T C C C T T C C C C T T C C C C T T C C C C T T C C C C C T T C C C C C T T C C C C C C C C C C C C C	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TTCA AGCI Q CCAA CCA	A C C C C C C C C C C C C C C C C C C C	AGA: I CCC: H CCC: H CCC: H CCC:	TCA1 H ACC1 FCA1 H ACC1 F CAC F CAC F F CAC F F CAC F F CAC F F CAC F F CAC F F CAC F F CAC F F CAC F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F F F CAC F CAC F C C C C	AGG CAG CAG CAG CAG CAG CAG CAG CAG CAG	AAAA CCCA CCCC CCCC CCCC CCCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCA CCCCCCA CCCCA CCCCA CCCCA CCCCA CCCCCC	LAGC CAT LAGT V CTGG G CTCC P AGA: H AGC: CTCCC CTCCC CTCC CTCC CTCC CTCC CTCC CTCCC CTCC CTCC CTCC CTCCC CTCC CTCC CTCCC CTCC CTCC CTCC CTCCC CTCCC CTCC CTCC CTCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCC CTCCC CTCCC CTCCC CTCCCC CTCCCCC CTCCCCC CTCCCCCCCC	TCG R GCAC R GCAC S TGG V ACA: I CTGG	TADA TADA	GAK GGC A GGC A GGC A GGC A GGC A GGC A GGC A GGC A GGC A C C C C C C C C C C C C C	AGG GG GG TTTA Y GGC A TTTTY CCCC H TTTTY CCCC H CCCC CCC CCC CCCCCC	CGC A CGC E TGC CGC A E TGC CGC C C CGC C C C C C C C C C C C C	CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	E E E E E E E E E E E E E E E E E E E	AGJ AGJ AGJ AGJ AGJ AGG AGG AGG AGG AGG	CAG G CAG CAG CAG CAG CAG CAG CA	TTC GCCC A TTCC F TAT C TTC F TAT C A C TTC C A C A C C A C C C A C C C C C C C C C C C C C	TTG C C C C C C C C C C C C C C

Figure 1. Mouse CRD-BP cDNA and predicted protein sequence. Sequences resembling nuclear localization and nuclear export signals, single underline and overlines, respectively. Sequences denoting the RGG box and KH domains, solid box and double underlines, respectively. The RNP1 and RNP2 sequences of each RRM, dashed boxes (Fig. 2). An asterisk indicates the translation termination site and the polyadenylation signal is single underlined. We have not demonstrated conclusively that the indicated translation start site is the correct or only start site. The 5'-UTR might be incomplete, since the transcription start site has not been mapped.

A		
mCRDBP	1	MNKLYIGNLNESVTPADLEKVFAEH.KISYSGQFLVKSGYAFVDCPDEHWAMKAIETFSG.KVELQGKRLEIEHSVP
chZBP1	1	MNKLYIGNLNESVTPADLEKVFNDH.KISFSGQFLVKSGYAFVDCPDEQWAMKAIETFSG.KVELHGKQLEIEHSVP
hKOC	1	MNKLYIGNLSENAAPSDLESIFKDA.KIPVSGPFLVKTGYAFVDCPDESWALKAIEALSG.KIELHGKPIEVEHSVP
xTFB3	1	MNKLYIGNLSENVSPPDLESLFKES.KIPFTGQFLVKSGYAFVDCPDETWAMKAIDTLSG.KVELHGKVIEVEHSVP
Vg1RBP/Vera	1	MNKLYIGNLSENVSPTDLESLFKES.KIPFTGQFLVKSGYAFVDCPDETWAMKAIDTLSG.KVELHGKVIEVEHSVP
mCRDBP	2	SR KIQIRN IPPQLRWEVLDSLLAQYGTVENCEQVNTESET AVVNVTY SNREQTRQAIMKLNGHQLENHALKVSY.IP
chZBP1	2	SR KIQIRN IPPQLRWEVLDGLLAQYGTVENCEQVNTDSET AVVNVTY TNREQTRQAIMKLNGHQLENHVLKVSY.IP
hKOC	2	IRKLQIRNIPPHLQWEVLDSLLVQYGVVESCEQVNTDSETAVVNVTYSSKDQARQALDKLNGFQLENFTLKVAY.IP
xTFB3	2	SR KLQIRN IPPHLQWEVLDSLLAQYGTVENCEQVNTDSET AVVNVTY ANKEHARQGLEKLNGYQLENYSLKVTY.IP
Vg1RBP/Vera	2	SRKLQIRNIPPHLQWEVLDSLLAQYGTVENCEQVNTESETAVVNVTYANKEHARQGLEKLNGYQLENYSLKVTY.IP
		*** * * * * * * * * * * * * * * * * * *
в		
hCRDBP	1	P lrl lvptqyvga iigkeg atirnitkqtqskid v hrkenagaaekais vhs
mCRDBP	1	PLRLLVPTQYVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKAISVHS
chZBP1	1	P L RLLVPTQYVGAIIGKEGATIRNITKQTQSKID V HRKENAGAAEKAISIHS
hKOC	1	PLRLLVPTQFVGAIIGKEGATIRNITKQTQSKIDVHRKENAGAAEKSITILS
xTFB3	1	P lrm lvptqfvga iigkeg atirnitkqtqskidihrkenagaaekpitih s
Vg1RBP/Vera	1	P lrm lvptqfvga iigkeg atirnitkqtqskidihrkenagaaekpitih s
hCRDBP	2	P LKILAHNNFVGRLIGKEGRNL KK V EQDTETKITISSLQDLTLYNPERTIT V KG
mCRDBP	2	P LKI LAHNNFVGR LIGKEGRNL KK V EQDTETKITISSLQDLTLYNPERTIT V K G
chZBP1	2	P LKI LAHNNFVGR LIGKEG RN L KK V EQDTETKITISSLQDLTLYNPERTIT V KG
hKOC	2	P LKI LAHNNFVGR LIGKEG RN L KKIEQDTDTKITISPLQELTLYNPERTIT V KG
xTFB3	2	PLKILAHNNFVGRLIGKEGRNLKKIEQDTDTKITISPLQDLTLYNPERTITVKG
Vg1RBP/Vera	2	P LKILAHNNFVGRLIGKEGRNL KKIEQDTDTKITISPLQDLTLYNPERTIT VKG
hCRDBP	3	MVQVFIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRMVIITG
mCRDBP	3	MYQV FIPAQAVGAIIGKKGQHIKQLSRFASASIKIAPPETPDSKVRM VVITG
chZBP1	3	T V H V FIPAQAVGA IIGR K G QHIKQLSRFASASIKIAPPETPDSKVRM V VIT G
hKOC	3	T VHQ FIPALSVGAIIG K QGQHIKQLSRFAGASIKIAPAEAPDAKVRM VIITG
xTFB3	3	T VHLF IPALAVGA IIGK Q G QHIKQLSRFAGASIKIAPAEGPDAKLRM VIITG
Vg1RBP/Vera	3	T VHLF IPALAVGAIIGKQGQHIKQLSRFAGASIKIAPAEGPDAKLRM V IITG
hCRDBP	4	HIRVPASAAGRVIGKGGKTVNELQNLTAAEVVVPRDQTPDENDQVI.VKIIG
mCRDBP	4	HIR V PASAAGR VIGK GGKT V NELQNLTAAE V V V PRDQTPDENDQVI.VKIIG
chZBP1	4	HIR V PASAAGR VIGR GGKT V NELQNLTAAEVVVPRDQTPDENEQVI.VKIIG
hKOC	4	HIR V PSFAAGR VIGK G G KT V NELQNLSSAE V V V PRDQTPDENDQVV. V KITG
xTFB3	4	HIK V PSYAAGR VIGK G G KT V NELQNLTSAE V V V PRDQTPDENDQVVKITG
Vg1RBP/Vera	4	HIK V PSYAAGR VIGK GGKT V NELQNLTSAE V VVPRDQTPDENDEVV.VKITG

Figure 2. Alignments of RNA-binding domains of CRD-BP and related proteins. Human CRD-BP sequences are from EST AA196976. (A) RRM domains. The RNP-1 and RNP-2 domains of each RRM are in bold. Asterisks indicate conserved residues proposed to be important in forming the hydrophobic RRM core structure. Crosses (†) denote positions at which 'typical' RNP domains contain either Y or F (37 and references therein). hCRDBP RRM sequences are not shown, since this region is not present in the EST. (B) KH domains. KH domain consensus sequences (38) are shown in bold and by asterisks below the alignment. GenBank accession nos are as follows: mCRDBP, AF061569; hCRDBP EST, AA196976/AA196977; chZBP1, AF026527; hKOC, U97188; xTFB3, AF042353; Vg1 RBP/Vera, AF064633 and AF055923, respectively.

human cDNA was generated. It contained an ORF that included both peptide sequences obtained by microsequencing.

To continue our analysis of CRD-BP in rodents, we used human cDNA to isolate mouse CRD-BP cDNA, which contained at least a portion of the 5'-UTR, a complete coding region and a complete 3'-UTR (Fig. 1). Two in-frame AUG start codons are present near the 5'-terminus. We have tentatively designated the downstream AUG as the translation start site for two reasons: it is embedded within a preferred translation start signal (36) and is similar to the translation start of CRD-BP homologs (Fig. 2).

The predicted CRD-BP sequence includes three characteristic RNA-binding protein motifs (37): RNA recognition motifs (RRMs), an RGG box and four hnRNP K-homology (KH) domains. There are two RRMs in the N-terminal third of the protein (Fig 2A). The first RRM (dashed boxes in Fig. 1) contains typical RNP-1 and RNP-2 consensus sequences (38) as identified in ZBP1, the chicken β -actin zipcode-binding protein (accession no. AF026527; 33). The second putative RRM was not identified in ZBP1. Like the RRMs of polypyrimidine tract-binding protein (pPTB; accession no. X60648; 39) and hnRNP L (accession no. X16135; 40), the RNP sequences in this putative RRM diverge from the consensus sequences. Yet, the second RRM contains conserved hydrophobic core (Fig. 2). Therefore, this region of the

mouse CRD-BP fits the broader definition of an RRM. A putative RGG box that follows the RRMs (Fig. 1, boxed) is similar to RGG boxes in fibrillarin (accession no. X56597; 41), nucleolin (accession no. M60858; 42) and fragile-X mental retardation protein (FMRP; accession no. X69962; 43–45). Four KH domains are arranged in pairs (Fig. 1, double underlines). Each pair is separated by ~30 amino acids and the two pairs are separated by 78 amino acids. There are two putative nuclear export signals (NES, overlined) and a putative nuclear localization signal (NLS, underlined), which is also a feature of ZBP1 (33). One of the two NESs (ITISSLQDLTLY) is similar to that in HIV Rev protein (accession no. X58781; 46) and is located within the second KH domain. The second NES (QLRWEVLDSLL) is similar to one in FMRP (45).

In the rat, the CRD-BP is expressed abundantly during fetal development and neonatal life but is undetectable in adult tissues (32). It is also expressed in transformed tissue culture cell lines derived from adult organisms. These findings imply that the CRD-BP is an oncofetal protein. Consistent with this notion, all known CRD-BP Expressed Sequence Tags (ESTs) are derived from fetal or embryonic tissue (accession nos AA619650, AA399833, AA073173, AA073514, D76662, D76781, AA196976/AA196977 and AA196759/AA196774).

Table 1. Comparison of sequences of CRD-BP and related proteins

	hCRD-BP ^a	cZBP1	hKOC	xTFB3	xVg1 RBP/Vera
mCRD-BP	88.8 99.1 <i>99.3</i>	82.5 94.6 95.7	69.5 73.9 79.1	71.0 78.1 <i>84.1</i>	70.2 77.9 83.8
hCRD-BP		82.0 95.1 <i>96.0</i>	69.5 74.6 78.6	69.5 78.1 83.5	70.4 77.9 83.9
cZBP1			70.0 75.6 80.1	72.0 78.7 84.5	72.2 78.4 <i>84.1</i>
hKOC				79.0 84.1 87.9	78.5 83.4 87.4
xTFB3					96.1 97.6 <i>98.1</i>

Values in bold indicate nucleic acid sequence identity, plain type indicates protein sequence identity and italic indicates protein sequence similarity. Only coding sequences were analyzed.

^aThe human CRD-BP cDNA is a partial sequence.

Mouse and human CRD-BP are similar to four other proteins (Fig. 2 and Table 1): ZBP1; hKOC, which is overexpressed in human pancreatic cancer (accession no. U97188; 47); xTFB3, a transcription factor from *Xenopus* that binds the B3-element of the TFIIIA promoter (accession no. AF042353; 48); Vg1 RBP or Vera, which binds to the localization sequence in *Xenopus* Vg1 mRNA (accession nos AF064633 and AF055923; 49,50). Of this group, mouse and human CRD-BP are most homologous to chicken ZBP1 and clearly are not alternative forms of hKOC, xTFB3 or Vg1 RBP/Vera. The common motif organization and considerable homology among these proteins suggest that they constitute a new subfamily of RNA-binding proteins (see below).

Comparison of *in vitro* synthesized CRD-BP with cell-derived CRD-BP

To confirm that our murine cDNA clone encodes a c-myc mRNA-binding protein, we synthesized the protein in vitro and analyzed it by immunoblotting and gel retardation. Reticulocyte transcription/translation reactions were programmed with pSPUTK-CRD-BP DNA. The CRD-BP sequences in this DNA lack the upstream in-frame AUG and translation begins with the AUG noted in Figure 1. By immunoblotting, pSPUTK-CRD-BP generated an ~68 kDa protein that migrated at or close to authentic human and mouse CRD-BP (Fig. 3, lanes 1-3). An immunoreactive band was not observed in extract programmed with pSPUTK vector (Fig. 3, lane 4) or luciferase cDNA (data not shown). Therefore, the antibody detected CRD-BP and not an endogenous reticulocyte protein and the cDNA encodes CRD-BP. The cross-reacting band (p85) in the K562 and NIH 3T3 lanes is a protein observed previously (32). It is not related functionally to CRD-BP (see below).

Gel retardation assays were performed to determine if *in vitro* translated CRD-BP binds specifically to c-*myc* CRD RNA. RSW from K562 cell polysomes or from reticulocyte polysomes was incubated with c-*myc* CRD [³²P]RNA. RNA/protein complexes were visualized following non-denaturing gel electrophoresis. A complex was observed with RSW from K562 cells and from reticulocyte extract programmed with CRD-BP cDNA (Fig. 4A, lanes 1 and 2, respectively). These complexes migrated to similar or identical positions. An RNA/protein complex was not observed with RSW from luciferase (Luc), vector or no mRNA (None) control reactions (Fig. 4A, lanes 3–5). Therefore, *in vitro* translated CRD-BP, like its cell-derived counterpart, associates with c-*myc* CRD RNA *in vitro*.

Previous work had shown that cell-derived CRD-BP did not bind to other RNAs we tested, suggesting considerable specificity



Figure 3. Immunoblotting assay showing co-migration of recombinant and cell-derived CRD-BP. RSW was prepared from polysomes from K562 or NIH 3T3 cells and from reticulocyte transcription/translation reactions programmed with CRD-BP DNA or with vector DNA. Approximately 7.5×10^5 cell equivalents of K562 or NIH 3T3 RSW or 3% of the RSW from a 50 µl translation reaction were electrophoresed and immunoblotted. Immunoreactive proteins were visualized using Supersignal chemiluminescent reagents (Pierce). The locations of the CRD-BP and a cross-reacting protein (p85) are indicated on the left. Prestained molecular mass markers are shown on the right in kDa.

for c-myc CRD RNA (30,31). A competition assay was performed to determine if *in vitro* translated CRD-BP exhibited similar specificity. Reactions contained c-myc CRD [32 P]RNA, CRD-BP and various unlabeled competitor RNAs at a 200-fold molar excess. Complex formation with CRD-BP was competed by c-myc CRD RNA but not by β -globin RNA (Fig. 4B, lanes 4–6), confirming that the cDNA encodes functional c-myc CRD-BP.

In view of the similarity between mouse CRD-BP and chicken ZBP1 (Table 1), we asked whether excess unlabeled, full-length β -actin zipcode RNA would block complex formation between recombinant CRD-BP and c-*myc* CRD [³²P]RNA. As expected, unlabeled c-*myc* CRD RNA did compete but globin RNA did not compete (Fig. 5, lanes 1–6). Unlabeled zipcode RNA containing the full-length 54 nt β -actin zipcode competed more than globin RNA but considerably less than c-*myc* CRD RNA (Fig. 5, lanes 7 and 8). Moreover, a 27 nt proximal zipcode [³²P]RNA probe that was used to purify ZBP1 (33) did not gel shift with recombinant mouse CRD-BP (data not shown). Therefore, although mouse



Figure 4. Gel retardation assay showing binding of recombinant CRD-BP to c-*myc* CRD RNA. (A) Gel retardation assay. RSW was prepared from K562 cell polysomes and from reticulocyte transcription/translation reactions programmed with mCRD-BP cDNA, luciferase cDNA (Luc) or vector DNA. Equivalent volumes (2 μ l) of each RSW were incubated with 50 000 c.p.m. of c-*myc* CRD [³²P]RNA. None indicates no added protein. The positions of mCRD-BP/CRD complexes (Bound) and unbound RNA (Free) are indicated on the left. (B) Competition gel retardation assay. The indicated RSW was incubated with c-*myc* CRD [³²P]RNA in the presence or absence of buffer (None) or a 200-fold molar excess of unlabeled c-*myc* CRD RNA or β -globin RNA.

CRD-BP and chicken ZBP1 are similar, they might have somewhat different RNA-binding specificities.

Characterization of the binding region for CRD-BP in c-myc CRD RNA

 $[^{32}P]$ RNA probes from parts of the coding region and 3'-UTR of c-*myc* mRNA were analyzed by gel shifts to map the major binding site for the CRD-BP. As expected, avid binding was observed with nt 1705–1886 of c-*myc* mRNA (Fig. 6). Binding was also observed with a 1705–1792 probe, which is consistent with footprinting data showing that 1705–1792 is the core binding region (data not shown). All other probes bound less avidly or not at all.

Localization of CRD-BP to the cytoplasm and co-fractionation of CRD-BP with polyribosomes

Consistent with its role as a putative mRNA-binding protein, much of the CRD-BP in cytoplasmic lysates fractionates with polysomes (31). However, the CRD-BP, like FMRP and ZBP1 (33,45), contains nuclear localization and nuclear export signals (Fig. 1). To assess whether any CRD-BP co-fractionated with nuclei or with other subcellular fractions, exponentially growing K562 cells were lysed and separated into six fractions. Equal cell equivalents of each fraction were analyzed by immunoblotting. At least 95% of the total cell CRD-BP was in polysomes and >90% of this CRD-BP was eluted in the 1 M salt wash (Fig. 7A, RSW). Little or no CRD-BP was detected in post-polysomal supernatant (S130) or in nuclei (Fig. 7A). This result cannot be explained by indiscriminate proteolysis during sample preparation, because intact HSP90 was detected in all fractions (Fig. 7B). We conclude that, at steady-state, most of the CRD-BP is cytoplasmic and polysomal. In contrast, the cross-reacting p85 is predominantly nuclear (Fig. 7A).



Figure 5. Gel retardation competition assays with c-*myc* CRD RNA β -globin RNA or β -actin zipcode RNA. Recombinant, His₆-tagged mCRD-BP was purified and 0.2 μ g was incubated with c-*myc* CRD [³²P]RNA (20 fmol) with or without the indicated molar excess of unlabeled competitor RNAs. WT Zip refers to the 54 nt chicken β -actin zipcode sequence that is required for localizing β -actin mRNA in fibroblasts (33). Gel retardation was performed as per Figure 4. Bound indicates RNA/mCRD-BP complexes. The relative percentages of protein/RNA complex formation are noted at the bottom.

To confirm the cytoplasmic localization of the CRD-BP, we used immunofluorescence to localize transiently expressed, FLAG-tagged CRD-BP. H4IIE and 293 cells expressing FLAG-CRD-BP exhibited strong staining that was predominantly cytoplasmic (Fig. 8). Little or no FLAG–CRD-BP was detected within nuclei. Neither untransfected cells nor cells transfected



Figure 6. Mapping the binding site of CRD-BP on c-myc CRD RNA. Portions of the c-myc mRNA coding region and 3'-UTR were amplified by PCR with the 5' primer including an SP6 promoter. [³²P]RNA was prepared from each template and analyzed by gel retardation for CRD-BP binding. (++++), (+++), (+++) and (+) indicate 100, ~70, ~40 and ~10% binding. (+/-) indicates that a weak protein/RNA complex was visualized after prolonged exposure of the blot to the PhosphorImager screen.

with untagged CRD-BP were stained with the anti-FLAG antibody and no cells were stained when incubated only with the FITC-conjugated anti-mouse IgG secondary antibody (data not shown).

We have suggested that some of the polysome-associated CRD-BP in the cell is bound to c-myc mRNA (30). To investigate whether some of it might also co-fractionate with ribosomal subunits, polysomes were dissociated into subunits plus mRNP by EDTA treatment and fractionated in a sucrose gradient. Small and large ribosomal subunits sedimented primarily in fractions 6-11 and 10-14, respectively (Fig. 9A and B). CRD-BP was detected over a large portion of the gradient and appeared to co-sediment with both large and small subunits (Fig. 9C). This pattern is similar to that of another putative mRNA-binding protein, FMRP (34,43–45). Some CRD-BP also pelleted to the bottom of the gradient. This material probably contains undissociated polysomes and monosomes. In summary, the CRD-BP co-sediments with ribosomal subunits, although the data do not address whether the CRD-BP is actually associated with the subunits. p85 was found at and near the top of the gradient and did not co-sediment with subunits (Fig. 9C), confirming that the CRD-BP and p85 are immunologically related but are functionally distinct.

DISCUSSION

Properties of recombinant CRD-BP

We have described the cloning and further characterization of the CRD-BP, a protein that binds *in vitro* to the CRD of c-*myc* mRNA. In cells, the c-*myc* CRD is a mRNA-destabilizing element whose function is somehow connected to translation (18–22). *In vitro*, the CRD binds to the CRD-BP and appears to shield c-*myc* mRNA from an endoribonuclease (22,30,31). Recombinant CRD-BP, like authentic CRD-BP from cells, migrates in gels at ~68 kDa (Fig. 3), binds with considerable specificity to CRD RNA (Figs 4 and 5) and remains bound to the RNA at physiological salt in high concentrations of heparin (Fig. 4).



Figure 7. Co-fractionation of endogenous CRD-BP with K562 cell polysomes and absence of CRD-BP in nuclei. Subcellular fractions were prepared from exponentially growing K562 cells. Equal cell equivalents (6×10^5) of each fraction were immunoblotted and the membrane was probed with (**A**) anti-CRD-BP IgY or (**B**) anti-HSP90 IgG. The positions of molecular mass markers are indicated in kDa on the left.

Comparison of the CRD-BP with related proteins

The CRD-BP is homologous to ZBP1, hKOC, xTFB3 and Vg1 RBP/Vera (Table 1), all of which appear to represent a unique subfamily of RNA-binding proteins with two RRMs, an RGG box and four KH domains. Other putative RNA-binding proteins also contain four KH domains (51), but these domains are neither organized nor spaced in the same way as in the CRD-BP family. The CRD-BP also contains one putative nuclear localization sequence and two putative nuclear export sequences (Fig. 1). We do not know if the CRD-BP actually shuttles between the nucleus and the cytoplasm, as do some other RNA-binding proteins (52). If it does shuttle, however, it appears to spend most of its time in the cytoplasm (below).

Various functions have been ascribed to members of this protein family, each of which was identified based on different features and nucleic acid-binding specificities. For example, ZBP1 was identified based on its affinity for a 27 nt RNA containing the proximal half-site of the 54 nt β -actin mRNA zipcode, an mRNA-localizing signal (33). Aside from stretches of A/C-richness, the RNA-binding sites in c-myc and β-actin mRNAs have no obvious similarities and a gel shift competition experiment indicates that the CRD-BP does not bind avidly to the 54 nt zipcode RNA (Fig. 5). Perhaps each family member has a unique function in various cytoplasmic and nuclear compartments. The CRD-BP is thought to protect the c-myc mRNA coding region from an endoribonuclease (30). In this respect, it is similar to the iron response protein (IRP), which is thought to protect the 3'-UTR of transferrin receptor mRNA from RNase attack (reviewed in 53).

CRD-BP in the cytoplasm

It was essential for us to determine where the CRD-BP is within the cell, because the CRD-BP sequence includes nuclear localization and export motifs (Fig. 1). We find that CRD-BP is cytoplasmic and co-fractionates with polysomes and ribosomal subunits (Figs 7–9). We do not know if the CRD-BP is actually associated with the subunits or merely co-sediments with them.



Figure 8. Immunofluorescent localization of FLAG-tagged CRD-BP in the cytoplasm of 293 and H4IIE cells. FLAG-tagged CRD-BP was transiently expressed in 293 (A) or H4IIE (B) cells. Cells were then stained and visualized via indirect immunofluorescence microscopy (Materials and Methods).



Figure 9. Co-sedimentation of the CRD-BP with ribosomal subunits. EDTA-dissociated K562 cell ribosomal subunits and mRNP were fractionated in a linear 5–30% sucrose gradient containing EDTA. Fraction 1 is the top of the gradient, fraction 18 is the last gradient fraction and fraction 19 is material resuspended from the bottom of the centrifuge tube. (A) Absorbance at 260 nm. (B) RNA from an aliquot of each fraction electrophoresed in a 1% agarose ged and visualized with ethidium bromide. The 28S and 18S rRNAs are noted on the left. (C) An aliquot of each fraction analyzed by immunoblotting using anti-CRD-BP IgY. Molecular mass markers are indicated in kDa on the left. CRD-BP and the cross-reacting p85 are noted on the right.

In either case, these findings are important for two reasons: (i) they show that the CRD-BP spends little time, if any, in the nucleus; (ii) they suggest that the CRD-BP might bind not only to mRNAs but also to ribosomes. Consistent with this possibility, xTFB3 protein in frog eggs is localized in the 'heavy ooplasmic fraction', which is enriched in ribosomes (48,54). Perhaps ribosome-bound CRD-BP molecules constitute a reservoir of protein that is recruited when needed to bind to mRNAs such as c-myc. In these respects, the CRD-BP is similar to FMRP (34,43–45). Both proteins co-sediment with ribosomes/polysomes (Fig. 9; 34,43–45) and FLAG–CRD-BP exhibits a punctate, cytoplasmic immunofluorescence staining pattern similar to that of FLAG–FMRP (Fig. 8; 34,53,55). We find no evidence for CRD-BP localization to a single region of the cytoplasm.

CRD-BP as an oncofetal protein

The c-myc CRD-BP is developmentally regulated, being expressed abundantly in fetal and neonatal life but not in adults (32). It is unknown whether xTFB3 is expressed in adults, but it is certainly expressed during early *Xenopus* development (48). Apparently, the CRD-BP is not required for cell viability, since perfectly normal adult animals do not express the CRD-BP at levels detectable by immunoblotting and/or gel retardation assays (32). Perhaps it has a special role in embryonic/fetal development.

The CRD-BP is also expressed in many cell lines, all of which are neoplastic or pre-neoplastic (Figs 3 and 4 and data not shown) and it is similar to the hKOC protein, which is overexpressed in human cancers (Fig. 2 and Table 1; 47). These findings suggest that the CRD-BP might be an oncofetal protein. If so, it would join a growing list of RNA-binding proteins that influence early development and/or carcinogenesis (56–60). In view of these potential links between the CRD-BP, early development and neoplasia, it should be interesting to learn if the CRD-BP is induced in tumor tissues of adult organisms and whether its overor underexpression affects embryonic/fetal development.

ACKNOWLEDGEMENTS

We thank Dr Alan Poland for anti-HSP90 antibody, Yuri Oleynikov and Rob Singer for the DNA template for preparing the 54 nt chicken β -actin zipcode RNA and Linda Robinson and Rich Stewart for thoughtful comments on the manuscript. The GenBank accession no. for mouse CRD-BP cDNA is AF061569. This work was supported by Public Health Service grants CA63676, CA23076 and CA09230. P.L. was the recipient of an American Cancer Society post-doctoral grant (PF-3870).

REFERENCES

- 1 Ayer, D.E. and Eisenman, R.N. (1993) Genes Dev., 7, 2110–2119.
- 2 Ayer, D.E., Kretzner, L. and Eisenman, R.N. (1993) Cell, 72, 211-222.
- 3 Zervos, A.S., Gyuris, J. and Brent, R. (1993) Cell, 72, 223–232.
- 4 Spencer, C.A. and Groudine, M. (1991) Adv. Cancer Res., 56, 1-48.
- 5 Coppola, J.A. and Cole, M.D. (1986) Nature, 320, 760–763.
- 6 Freytag,S.O., Dang,C.V. and Lee,W.M.F. (1990) Cell Growth Differentiat., 1, 339–343.
- 7 Evan,G.I., Wyllie,A.H., Gilbert,C.S., Littlewood,T.D., Land,H., Brooks,M., Waters,C.M., Penn,L.Z. and Hancock,D.C. (1992) *Cell*, **69**, 119–128.
- Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D. and Brinster, R.L. (1985) *Nature*, **318**, 533–538.
 Klein G. (1989) *Gauge Chromosomes Cancer*, **1**, 3–8.
- 9 Klein,G. (1989) Genes Chromosomes Cancer, 1, 3–8.
- 10 Lüscher, B. and Eisenman, R.N. (1990) *Genes Dev.*, **4**, 2025–2035.
- 11 Spotts,G.D. and Hann,S.R. (1990) Mol. Cell. Biol., 10, 3952–3964.
- Lutterbach,B. and Hann,S.R. (1994) *Mol. Cell. Biol.*, 14, 5510–5522.
 Morello,D., Asselin,C., Lavenu,A., Marcu,K.B. and Babinet,C. (1989)
- Oncogene, 4, 955–961.
 Gruppuso, P.A., FitzGerald, M.J. and Fausto, N. (1993) Pediat. Res., 33, 49A.
- 15 Morello, D., Lavenu, A. and Babinet, C. (1990) Oncogene, 5, 1511–1519.
- 16 Steer, C.J. (1996) FASEB J., 10, 559–573.
- 17 Jones, T.R. and Cole, M.D. (1987) Mol. Cell. Biol., 7, 4513-4521.
- 18 Wisdom, R. and Lee, W. (1990) J. Biol. Chem., 265, 19015–19021.
- 19 Wisdom, R. and Lee, W. (1991) Genes Dev., 5, 232-243.
- 20 Yeilding, N.M., Rehman, M.T. and Lee, W.M.F. (1996) Mol. Cell. Biol., 16, 3511–3522.
- 21 Yeilding, N.M. and Lee, W.M.F. (1997) Mol. Cell. Biol., 17, 2698–2707.
- 22 Herrick, D.J. and Ross, J. (1994) Mol. Cell. Biol., 14, 2119–2128.
- 23 Lavenu, A., Pistoi, S., Pournin, S., Babinet, C. and Morello, D. (1995) *Mol. Cell. Biol.*, **15**, 4410–4419.
- 24 Morello, D., Lavenu, A., Pournin, S. and Babinet, C. (1993) Oncogene, 8, 1921–1929.
- 25 Pistoi, S., Roland, J., Babinet, C. and Morello, D. (1996) Mol. Cell. Biol., 16, 5107–5116.
- 26 Ross, J. and Kobs, G. (1986) J. Mol. Biol., 188, 579-593.
- 27 Ross, J., Peltz, S.W., Kobs, G. and Brewer, G. (1986) Mol. Cell. Biol., 6, 4362–4371.
- 28 Peltz, S.W. and Ross, J. (1987) Mol. Cell. Biol., 7, 4345-4356.
- 29 Brewer, G. and Ross, J. (1988) Mol. Cell. Biol., 8, 1697–1708.
- 30 Bernstein, P.L., Herrick, D.J., Prokipcak, R.D. and Ross, J. (1992) Genes Dev., 6, 642–654.
- 31 Prokipcak, R.D., Herrick, D.J. and Ross, J. (1994) J. Biol. Chem., 269, 9261–9269.
- 32 Leeds, P., Kren, B.T., Boylan, J.M., Betz, N.A., Steer, C.J., Gruppuso, P.A. and Ross, J. (1997) *Oncogene*, **14**, 1279–1286.
- 33 Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L. and Singer, R.H. (1997) *Mol. Cell. Biol.*, **17**, 2158–2165.

- 34 Siomi,M.C., Zhang,Y., Siomi,H. and Dreyfuss,G. (1996) Mol. Cell. Biol., 16, 3825–3832.
- 35 Harlowe, E. and Lane, D. (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 359–420.
- 36 Kozak, M. (1989) J. Cell. Biol., 108, 229-241.
- 37 Kenan, D.J., Query, C.C. and Keene, J.D. (1991) Trends Biochem. Sci., 25, 214–220.
- 38 Burd,C.G. and Dreyfuss,G. (1994) *Science*, **265**, 615–621.
- 39 Gil,A., Sharp,P.A., Jamison,S.F. and Garcia-Blanco,M.A. (1991) *Genes Dev.*, 5, 1224–1236.
- 40 Piñol-Roma,S., Swanson,M.S, Gall,J.G. and Dreyfuss,G. (1989) J. Cell Biol., 109, 2575–2587.
- 41 Jansen, R.P., Hurt, E.C., Kern, H., Lehtonen, H., Carmo-Fonseca, M., Lapeyre, B. and Tollervey, D. (1991) J. Cell Biol., 113, 715–729.
- 42 Srivastava, M., McBride, O.W., Fleming, P.J., Pollard, H.B. and Burns, A.L. (1990) J. Biol. Chem., 265, 14922–14931.
- 43 Ashley,C.T., Wilkinson,K.D., Reines,D. and Warren,S.T. (1993) Science, 262, 563–565.
- 44 Siomi,H., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1993) Cell, 74, 291–298.
- 45 Eberhart, D.E., Malter, H.E., Feng, Y. and Warren, S.T. (1996) *Hum. Mol. Genet.*, 5, 1083–1091.
- 46 Ciccarelli, R.B., Gunyuzlu, P., Huang, J., Scott, C. and Oakes, F.T. (1991) Nucleic Acids Res., 19, 6007–6113.
- 47 Müeller-Pillasch, F., Lacher, U., Wallrapp, C., Micha, A., Zimmerhackl, F., Hameister, H., Varga, G., Friess, H., Büchler, M., Beger, H.G., Vila, M.R., Adler, G. and Gress, T.M. (1997) Oncogene, 14, 2729–2733.
- 48 Pfaff,S.L. and Taylor,W.L. (1992) Dev. Biol., 151, 306–316.
- 49 Havin,L., Git,A., Elisha,Z., Oberman,F., Yaniv,K., Scwartz,S.P., Standart,N. and Yisraeli,J.K. (1998) *Genes Dev.*, **12**, 1593–1598.
- 50 Deshler, J.O., Highett, M.I., Abramson, T. and Schnapp, B.J. (1998) *Curr. Biol.*, 8, 489–496.
- 51 Musco,G., Stier,G., Joseph,H., Antonietta,M., Morelli,C., Nilges,M., Gibson,T.J. and Pastore,A. (1996) *Cell*, 85, 237–245.
- 52 Piñol-Roma, S. and Dreyfuss, G. (1992) Nature, 355, 730–732.
- 53 Harford, J.B., Rouault, T.A. and Klausner, R.D. (1994) In Brock, J.H., Halliday, J.W., Pippard, M.J. and Powell, L.W. (eds) *Iron Metabolism in Health and Disease*. W.B.Saunders, Philadelphia, PA, pp. 123–149.
- 54 Lohka, M.J. and Masui, Y. (1984) J. Cell Biol., 98, 1222–1230.
- 55 Feng,Y., Absher,D., Eberhart,D.E., Brown,V., Malter,H.E. and Warren,S.T. (1997) *Mol. Cell*, **1**, 109–118.
- 56 Bandziulis, R.J., Swanson, M.S. and Dreyfuss, G. (1989) Genes Dev., 3, 431–437.
- 57 Herschlag, D. (1995) J. Biol. Chem., 270, 20871-20874.
- 58 Shamoo, Y., Abdul-Manan, N. and Williams, K.R. (1995) Nucleic Acids Res., 23, 725–728.
- 59 Gao, F.-B. and Keene, J.D. (1996) J. Cell Sci., 109, 579-589.
- 60 Cooke, H.J. and Elliott, D.J. (1997) Trends Genet., 13, 87-89.