Volume 250, number 1, 99-105

FEB 07280

June 1989

Cloning and sequencing of the human nucleolin cDNA

Meera Srivastava⁺, Patrick J. Fleming⁺, Harvey B. Pollard and A. Lee Burns

Laboratory of Cell Biology and Genetics, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892 and ⁺Department of Biochemistry, Georgetown University School of Medicine, Washington, DC 20007, USA

Received 5 May 1989

A cDNA containing the entire coding region for human nucleolin has been isolated from a λ gt10 human retinal library using a bovine cDNA probe. The cDNA hybridized to a transcript of 3000 bases from fast-dividing cells, as well as terminally differentiated tissues of several species. Translation of the nucleotide sequence revealed a long open reading frame which predicts a 707 amino acid protein with several distinct domains. These include repeating elements, four conserved RNA-binding regions, a glycine-rich carboxy-terminal domain and sites for phosphorylation, glycosylation and dibasic cleavage. Human and bovine nucleolin exhibited more additions and/or substitutions of aspartate, glutamate and serine residues in the chromatin-binding domains by comparison with the hamster and mouse nucleolins. These differences may be related to species-specific functions in transcription.

Nucleolin; Sequence; Retina; Homology

1. INTRODUCTION

Nucleolin, also known as C23, is an abundantly expressed acidic phosphoprotein of exponentially growing cells and has been shown to be mainly localized in dense fibrillar regions of the nucleolus [1,2]. This multifunctional protein [3] was shown to be involved in the control of transcription of ribosomal RNA genes by RNA polymerase I [4,5], in ribosome maturation and assembly [6,7], and in the nucleocytoplasmic transportation of ribosomal components [8]. The above activities are regulated not only by phosphorylation and proteolysis [9], but also by basic fibroblast growth factor [10] and by other thus far unidentified cytoplasmic regulators of nuclear activities [8].

While complete protein sequences for nucleolin have been isolated from hamster and mouse [11,12], the human nucleolin gene has thus far not been reported. This deficiency has precluded not only analysis of species-specific functions but also

Correspondence address: M. Srivastava, Department of Biochemistry, Georgetown University School of Medicine, Washington, DC 20007, USA analysis of this protein in viral or chemically transformed human cells, in which one might anticipate this protein to have important regulatory functions. In the present paper, we describe experiments leading to our discovery of the human nucleolin gene.

2. MATERIALS AND METHODS

2.1. Construction of cDNA library from adrenal medulla

Poly A⁺ containing RNA from bovine adrenal medulla was purified by oligo(dT) column chromatography [13] of the guanidine thiocyanate isolated RNA [14]. The cDNA library was constructed using 2 μ g poly A⁺ according to a previously described technique [15]. The cDNA was treated with *Eco*RI methylase and the ends were blunted by the Klenow fragment of DNA polymerase I. *Eco*RI linkers were then added (0.5 μ g linkers/ μ g cDNA) using T4 DNA ligase. After treatment with *Eco*RI, the free linkers were separated from cDNA by chromatography on a Bio-Gel A-50m column. 25 ng cDNA were ligated to 1 μ g *Eco*RI-cleaved λ gt11. The phages were packaged in vitro and were amplified using *E. coli* strain Y1088.

2.2. Isolation of bovine and human nucleolin cDNA

Recombinant proteins expressed by the bovine cDNA library were probed with anti-cytochrome b-561 antibody as described [16,17], and a 610 bp partial cDNA clone coding for nucleolin was isolated. Other libraries (bovine brain library, Clonetech;

Published by Elsevier Science Publishers B.V. (Biomedical Division) 00145793/89/\$3.50 © 1989 Federation of European Biochemical Societies bovine retinal library, Dr H.G. Khorana; and human retinal library, Dr Jeremy Nathans) were screened by plaque hybridization [18] with the bovine nucleolin cDNA, leading to the isolation of two human clones (CH2, 2.1 kb and CH4, 1.8 kb) and a 770 bp bovine clone. The human retinal library was rescreened using a 36 bp oligonucleotide (5'-AGCCACACCAG-GCAAAGCATTGGTAGCAACTCCT-3') from the 5'-end of CH2 clone. Ten positive phages were isolated from 1×10^5 plaques and five contained the same full-length cDNA.

2.3. Nucleotide sequence analysis

*Eco*RI cDNA inserts or fragments generated by restriction with *Hae*III, *Rsa*I or *Alu*I were subcloned into M13mp18 for sequencing using the dideoxy chain termination method [19] with Sequenase. Sequence analysis was performed using the Microgenie Program of Beckman and secondary structure was predicted by the method of Garnier [20] using a computer program obtained from Dr Janet Finer-Moore [21].

2.4. RNA blot hybridization

Northern blot analysis was done by separating 5 μ g poly A⁺ RNA by electrophoresis on a 1.0% agarose gel in the presence of 2.2 M formaldehyde [22] and by transferring the RNA to a Nytran membrane (Schleicher & Schuell). The blots were then hybridized and washed according to the manufacturer's recommendations. Hybridization probes were prepared by the random-primer method [23].

2.5. Other reagents

Restriction endonucleases, T4 DNA ligase, T4 polynucleotide kinase, MLV reverse transcriptase, RNase H and DNA polymerase I were purchased from either BRL or New England Biolabs. Sequenase kits were obtained from United States Biochemicals.

3. RESULTS

3.1. Determination of the human nucleolin sequence

While screening a bovine adrenal medulla cDNA library using antiserum raised against cytochrome b-561, a major membrane protein of chromaffin cells [24], a clone was isolated which had 90% homology to hamster nucleolin [11]. Several other libraries were screened for full-length clones using bovine nucleolin cDNA. The largest insert (2570 bp) was obtained from a λ gt10 human retinal library, which when sequenced turned out to be full-length. The complete nucleotide sequence of the cDNA was obtained as illustrated in fig.1A and is shown in fig.1B. The clone contains a 114 bp 5'-untranslated region and a 2121 bp coding region. The initial ATG has a well conserved Kozak [25] consensus nucleotide sequence needed for translation initiation. The 332 bp untranslated region at the 3'-end includes the unusual polyadenylation signal TATAAA at nucleotide position 2424.

The open reading frame for human nucleolin cDNA codes for a 707 amino acid long protein of 77 kDa. The amino acid sequence in the coding region is 84% and 81% identical to that of hamster and mouse nucleolin cDNA, respectively. In addition, the sequence has distinct structural features such as: (i) highly charged acidic stretches at the amino-terminus with characteristic repeats; (ii) phosphorylation and glycosylation sites; (iii) ribonucleoprotein consensus sequences and (iv) a glycine-rich carboxy-terminal sequence. These data strongly support our identification of the clone as that of human nucleolin. Comparison of the amino acid sequences of human nucleolin with those of bovine, hamster, mouse and chicken is shown in fig.2 and is presented in more detail in section 4.

3.2. Northern analysis of human nucleolin

Poly A⁺ RNA was prepared from various types of tissues and analyzed for expression of nucleolin mRNA by Northern blot. A major band of 3000 bp was detected in human retina, adrenal gland, liver, and a neuroblastoma cell line SK-N-SH. Similar results were obtained with bovine retina, adrenal medulla, liver, kidney and brain, plus rat brain and retina (fig.3). The two lower faint hybridizing bands seen in bovine liver and kidney could be alternative forms of nucleolin mRNA or degradation products. Expression of mRNA of the same size in different tissues is consistent with the notion that nucleolin is a ubiquitous protein involved with the critical processes of ribosomal RNA synthesis and subsequent ribosome maturation. As previously reported, nucleolin was found to be most abundant in ex-

Fig.1. (A) Restriction endonuclease map of nucleolin cDNAs and sequencing strategy. The thick line represents the coding region which is flanked by 5'- and 3'-noncoding regions indicated by thin lines. The direction of the sequence obtained from CH1, CH2, CH3, CH4 or restriction fragments using either specific primers (----→) or the universal primers (----→) are indicated by arrows. (B) Nucleotide and deduced amino acid sequence of the human nucleolin cDNA insert. Numbers for the nucleotides and amino acids are placed above or below their respective sequences. The presumed polyadenylation signal near the 3'-end is underlined.

B. B. B. B. B. B. B. B. B. B.	33
B. B. B. B. B. B. B. B. B. B.	1330 1330 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1390 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1300 1000 1000 1000 1000 1000
CH1 5- 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4	33
B. 30 40 413 414 412 414 414 414 414 414 414	1320 337770 402 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 13800 138000 138000 138000 10000000000
B. 200 30 60 30 60 30 60 30 60 30 60 30 60 30 60 30 60 30 60 30 60 30 60 300 50 301 300 302 300 303 300 304 300 305 300 306 300 307 300 308 300 309 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 <t< th=""><td>1320 2007204 400 1380 1380 1380 1380 1380 1380 1380 13</td></t<>	1320 2007204 400 1380 1380 1380 1380 1380 1380 1380 13
B. 30 60 0000000070000000000000000000000000000	1330 200700 200700 402 1380 1380 1480 1480 1480 1480 1480 1480 1480 14
CH2 CH3 CH3 CH4 CH4	1320 1320 1320 1300 1380 1380 1380 1380 1480 1480 1480 1480 1480 1480 1480 14
CH2 CH3 CH4 CH4 CH4 CH4 CH4 CH4 CH4 CH4	1320 1320 1320 1380 1380 1380 1380 1480 1480 1480 1480 1480 1480 1480 14
CH2 CH3 CH3 CH4 CH4 CH4 CH4 CH4 CH4 CH4 CH4	1320 300713AC 1380 1380 1380 1380 1380 1380 1380 1480 1480 1480 1480 1480 1480 1480 14
CH2	1320 3001797 402 1380 177907 1422 1440 1422 1440 1422 1440 1422 1440 1422 1440 1422 1440 1422 1440 1422
CH3	1320 301790 1380 1380 1380 1279070 1440 1480 1380 1480 1480 1480 1480 1480 1480 1480 14
CH3	1320 3007174C 3ProTyr 402 13800 4779471 422 1440 NSRC90CA NSRC90CA NSRC90CA
CH4 CH4 CH4 CH4 CH4 CH4 CH4 CH4	1320 3UUTTAC aProftyr 422 1380 KTTM7U 122 1480 KRTGCA KARTGCA KARTGCA KARTGCA
B. 30 60 30 60 60 30 60 60 30 60 60 30 60 60 30 60 60 30 60 60 30 60 60 30 60 60 30 60 300 30 60 60 30 60 60 30 60 300 30 60 300 30 60 300 300 60 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300 300	1320 3UUTMC 13F0Tyr 402 1380 NTPLATVA1 422 1440 NGAUGCA UMSPALa 442
Сн4	1320 307797 402 1380 8779970 7149970 1422 1440 NRC902A 148992A 148992A 148992A
B. B. B. B. B. B. B. B. B. B.	1320 200712AC aProfyr 402 1380 ATTPATC 7LatVal 422 1440 NRCGCA 1480 RACA A42 442
B. 10 10 10 10 10 10 10 10 10 10	1320 3327196 402 1380 KTTMTC 1380 KTTMTC 1380 KTTMTC 1440 KANGCA 1440 KANGCA 442
B. 3 3 3 3 3 3 3 3 3 3 3 3 3	1320 30011AC 402 1380 NTTATC 1380 NTTATC 1380 NTTATC 1380 NTTATC 1440 NSMIGCA 1440 NSMIGCA 442
B. 100 (3)/(3)/(3)/(3)/(3)/(3)/(3)/(3)/(3)/(3)/	1320 CCCTTMC aProTyr 402 1380 NTTMTC 1380 NTTMTC 1422 1440 NGRIGCA LAmpALa 442
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20	1320 CCTINC UPTOTYT 402 1380 MTTMFTC JIANVAI 422 1440 ISATGCA EASPALS EASPALS 442
10 ناب المحكمة المحكم	1380 MTRATC 1380 MTRATC JIANAI 422 1440 KENTGCA EASPALS 442
	402 1380 KTIMITC JIMIVA1 422 1440 KRATGCA KASPALa 442
90 130 1350 937107107056/10400000000000000000000000000000000000	1380 MTIMIC JimiCal 422 1440 RECICA MERICA MERICA
Hertral LynWal Throlinkopol utenityes: Ureal Presi Presi Ureal Presi Presi Ureal Presi Ureal Presi Ureal Presi Ureal Presi Ureal Presi	1440 IGRIGCA IANDALA 442
150 130 MacTraccanaccies/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anamatical/anaamatical/anamatical/anamatical/anamatical/anaamat	422 1440 NGANGCA NAMPAla 442
140 141 MACTROCTADICATION AND AND AND AND AND AND AND AND AND AN	1440 NGANGCA BAmpAla 442
LysLauki klyski klyklyski klyklyski klyski klyski klyski klyski klyski klyski klyski	aAspAla 442
20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20 20	442
രംഗത്തെ പ്രത്യാനമായ പ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമാന പ്രത്യാനമായ പ്രത്യം പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യാനമായ പ്രത്യം പ്രത്യം പ്രത്യം പ	
GUWalduGiuGiuMapderdiuGiuMapdiudiuMapdapderder 32 42 452 452 452 452 452 452 452 452 452	1500 3010136
270 300 1510 GRAMMANTOTICHTCHTCHTANAMANACCHTCHTCHTCHTCHTCHTCHTCHTCHTCHTCHTCHTCHT	462
GRADING CONTROLLY AND AND CONTROLLY CONTROLLY CONTROL OF THE CONTR	1540
	TIGGAGT
	482
330 360 1350	1620
MODIVATION CONTRACTOR AND A CONTRACTOR A	ACTOT
72 B2 492	502
380 420 1650	1680
archively and a second and a second and a second and a second a se	JANNET VLVBBar
92 102 512	522
	1740
Alayalini incense inconsection and an analyzing and an an an and an	Assistan
112 122 532	542
	1800
GlyfyslysGlyfalaliafiefrealarysGlyfalandafysfannifalysglys Cysannyafyrgoluliaeindiyfyryfalailefrealarysglwilaeindiyfry	ArgGly
112 142 552	562
570 600 1830 GANGKATGATGAGAGGATGAGGAGGATGAGGAGGAGGAGGAGGAG	1860 NBOGAT
GlukapSerKapGluGluGlukapkapSerGluGlukapAggGlukap 10	rGlukap
KTRENKEREDEDTTEDDTEDDTEDDTEDDTEDDTEDDTEDDTEDDT	APTINCT
ouwapu wapu wapu wi nesuwana anawenyaa taata kakaka pao mining walayasi wapu ya kakaya iyo kakaya iyo kakaya iy 172 182	WalThr 602
690 720 1950	1980
	XROAN
192 202 202	willing
750 780 2010	u01uAasp 622
In the second se	1031113Aanp 622 2040
212 222 632	иЗІцімар 622 2040 ТТОЮЮС ЛТТОЮСС ЛТТОЙОСС
810 840 2070	adlukap 622 2040 Traacc Traacc TrpAla 642
	431134ap 622 2040 7703633 777pAla 642 2100
LywYal Yal ProVal Lyakat Lyakan Yala ladi ukagi ukagi ukagi ukagi ku	2040 2040 2040 2040 2040 2040 2040 200 20
Lyawal Ya Prova Lyawal Layawal Lay	2040 2040 709900 709900 709940 642 2100 709940 81901y 662
ມສາຍໄປຊີມັນ ແລະ ແລະ ແລະ ເປັນເປັນເປັນເປັນເປັນເປັນເປັນເປັນເປັນເປັນ	2040 2040 2040 2040 2000 2100 1028038 1019019 662 2160 2160
لي المركز الم	40114Asp 622 2040 77060000 77079Ala 642 2100 7059600A 4013/01y 662 2160 777700A 777700A
LymWal Val Proval Lyakani val Al adj ukepći lukepći lukepči luk	401134ap 622 2040 27036600 77779Ala 642 2100 1053463A 8019019 662 2160 777700A 777700A 777700A
1/97%1/%1/%2/1/97%1/%1/%2/%2/%2/%2/%2/%2/%2/%2/%2/%2/%2/%2/%2/	401144ap 622 2040 CTGGGCC 7/TrPALa 642 2100 rokadaa 4013/01y 662 2160 7/TrTGLy 682 2220 4000AAG
 	401144ap 622 2040 CTC4640CC 7017PALa 642 2100 rotandch 662 2100 rotandch 662 2100 rotandch 662 2100 rotandch 662 2100 rotandch 7010 662 2220 (s0304Jy 662 2220 (s0304Jy 662 2220 (s0304Jy 662 2220 (s0304J) 662 220 (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s0304) (s
1/97%21%21%20%21%24.12/92%21%24.12/92%21%21 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2	401434p 622 2040 (TOKOKOC) (TOKOKOC) (TOKOKOC) 2100 (COKOKOC) 662 2160 707TOGA (SI)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)
LywWil Vii Provi Liyaka Liyaken vii A A ladi ukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäilukepäil	401434p 622 2040 (TVGGGC) 2077pAla 642 2100 (VGGAGA 801y01y 662 2160 77770GA (VGGAGGA 682 2220 VGGAAG 682 2220 VGGAAG 813/USA 702 2280 30A/GGA
	101134ap 622 2040 CTGGGCC 77779A1a 642 2100 rokadaa 662 2160 77770GA 77770GA 77770GA 77770GA 77770GA 77770GA 77770GA 772 2280 2280 233AGGA
LynWil Vil Proval LynAl Lynaen val Ri karl ukspräl uksp	101124ap 622 2040 CTGGGCC 7773PA1a 642 21000 rokanan 662 2160 7777GA 762 2160 7777GA 762 2160 7777GA 702 2280 30AGZA
LynWil Vil Provil LynXi Lafysben vil Lyn ei) uber fil uber f	101134ap 622 2040 CTGGGCC 777791a 622 2100 7658468A 8011901y 662 2160 7776801y 662 2160 2160 2160 2160 2160 2160 2160
LynWil Vial Provial Lynak Lafyadem val Al acti ukepsil ukepsil usepsil	101134ap 622 2040 CTGGGGC 2100 TGGGGGG 2100 TGGGGGG 2100 TGGGGG 2160 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 77770GB 777770GB 777770GB 777770GB 777770GB 777770GB 777770GB 777770GB 7777770GB 7777770GB 7777770GB 777777770GB 7777777777
LywNitWitPervel LywSerwal Riadiulwegdi ulwegdi	1011034p 622 2040 CTGRAGCC pTTPDALa 642 2100 CTGRAGCA 662 2100 2100 2100 2100 2100 2220 VGGAAG 652 2220 VGGAAG 7702 2280 33AGCA 23AGCA 23AGCA 2400 ATTACCG
International internatinterel international international international inter	101134p 622 2040 CTORGCC 7/TPDALs 642 2100 TCSARGA 1017014 662 2160 77TTOGA 77TTOGA 77TTOGA 77TTOGA 77TTOGA 7702 2280 8314077025 2280 2280 2280 2280 2280 2280 2280 2
International internatinterenational international international internationa	изіцімра 522 522 2040 Стазака: 542 2100 542 2100 542 2100 542 2100 542 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 2100 200 2
Lynwil Val Proval Lynal alysaem val Al adj ukepil	التالناهي (22 2040 ۲۲۵۵۵۵۵۲ 2100 2100 2100 2100 2100 2100 2100 210
Lywwitwi I provi I (ywi I wi I provi I provi I (ywi I wi I provi I provi I (ywi I provi I (ywi I provi I provi I provi I (ywi I provi provi I provi I provi I provi I provi I provi I provi	للتالناهي التالناهي التالناهي التالناهي التالناهي التالي الت
199%/1%12/000000000000000000000000000000000	22400 77823527 2040 77823527 2140 2140 2140 2140 2140 2140 22400 702 22400 702 22400 702 22400 22400 22400 22400 22400 22400 22400 22400 22400 22400 7022 7022
Lynwil Vel Provel Lyne 212 213 213 213 213 213 213 213 214 213 213 215 213 213 216 213 213 217 213 213 218 213 213 217 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 218 213 213 219 2130	12110690 522 2046 57020127 642 2100 7070242 2100 707026 2100 707026 22100 707026 22100 707027 21000 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22100 22200 22100 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 22200 2200 2200 2200 22200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2200 2000 2200 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 200000 2000 2000 2000 2000 2000 20000 2000 200000 2000 2000

101

FEBS LETTERS

Lbse	**	20		40	** **	60 **	*
nu:	MVKLAKAGKNOGDPKK	APPPKEVEED	SEDERMSEDER	DDSSGEE^	VVIPOKKGKK	AAATSAKKVVVSPT	x
Ha:	TH-EA	********	^			-TPO-	-
Mo:	TH-EA		^-	E		-TT-PQ-	_
	**					-	
	* 80 *	* **	100	**	120 **	140	•
Rus	KVAVATPAKKAAVTOS	KARATORKET	UTDA KAUTTER		ALVATOCKKO	A T DA KYA KWATA KA	v
Hat	P						-
Mo:	-AP	VN	IVTP		D	-VT	_
		• n		nyn	•	•1	
	* 1	.60		180		200	_
Hu:	KEDEDEEEDDDSEEDEE	DDEDEDEDED	EIEPAAM	`^^KAAAAA	PASEDEDD^^	^^^EDDEDDEDDDDD	D
Ha:	DDDDB	;-EDEE	-FPVVKG^1	QG-V	E^^	^^^-EEE-EE-E	E
MO:	DE.EDDE	;- <u>EDED</u> E	-FPIVKGVI	(PA-A-P			E
BO:						DDDDEE-EEDEE-E	Ľ
		**					
		220		240		260	
Hu:	· EEDDSEEE ^ AMETTPAK	GKKA^AKVVP	VKAKNVAEDEI	DEEEDDEDE	DDDDDEDDED		Е
Ha:	EI	P		`^DI	ND-EERIDE	<u> 3512</u> <u>E</u>	-
Mo:	E-DSEVI-T	TP)	MS^^^	`^E-DE-I	E-EED-	EEED-EDDD	-
Bo:	DE^PIV	PV-A	ST	-^D-D-	EEED-	ED	-
	** 280**		** 300		320		
Hu:	E^^^^PVKEAPGKRKK	EMAKOKAAPE	AKKOKVEGTEE	TTAFNI FV	NINFNKSAP	ELKIGISOVFARND	L
Ha:	-^^^^P		8-8-8	3T-	PVA	VAEP	_
Mo:	-^^^^		S-I		PVN	FAEL	_
Bo:	-EEEE					L	-
				-		_	
	340	360		380	**	400	
Hh1:	AVUTURI (MTRKREVUT	FESAEDLEKA		ETREERPK	TKDSKKERDA	RTLIAKNLPYKYTO	Ð
Ha:					-RV-A-	SFNI-E	
Mo:					-RV-A-	SFNI-E	-
Bo:	VS				D		_
	420		440		460	47	9
Hut	ELKEVFEDAAEIRLVSK	DGKSKGIAYI	EFRTEADAERI	TEEROGIE	DGRSISLYY	TGEKGONODYRGGK	N
Ha:	I/O)		аA	v	R-ERT^	_
Mo:	Õ	, 	SN	1LA	V	R-ERT^	т
Bo:	V						
Ch:		M	REP	L1	VAMVID-	S-^-E^S^Q-	G
						-	
		500		520		540 **	
Hu:	STWSGESKTLVLSNLSY	500 SATEETLQEVI	TEKATFIKVPC	520 NONGKSKG	AFIEFASFE	540 ** DAKEALNSCNKREI	E
Hu: Ha:	STWSGESKTLVLSNLSY	500 SATEETLQEV	FEKATFIKVPC	520 NONGKSKG 	AFIEFASFE	540 ** DAKEALNSCNKREI M	E
Hu: Ha: Mo:	STWSGESKTLVLSNLSY	500 SATEETLQEVI KE	FEKATFIKVPÇ	520 NQNGKSKG Q PHP	AFIEFASFE	540 ** DAREALNSCNRREI M	E
Hu: Ha: Mo: Bo:	STWSGESKTLVLSNLSY	500 SATEETLQEVI KE	FEKATFIKVPÇ	520 NONGKSKG3 	/AFIEFASFE	540 ** DAKEALNSCNKRET M	E - -
Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY	500 SATEETLQEVI KE A-SL	FEKATFIKVPÇ 	520 NQNGKSKG3 Q PHP NQ-RP	/AFIEFASFE 	540 ** DAREALASCNKREII M	E
Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY	500 SATEETIQEVI KE A-SIr	FEKATFIKVPG 	520 MQNGKSKGY 	(AFIEFASFE 	540 ** DAREALNSCNRREII M 	E
Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY GGER^IV-N 560	500 SATEETLQEV. KE A-SI	FEKATFIKVPÇ 	520 MQNGKSKGY 	/AFIEFASFE 	540 ** DAREALNSCNRREI M 	E
Hu: Ha: Mo: Bo: Ch: Hu:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS	500 SATEETLQEV. KE A-SL PNARS^^QP	FEKATFI KVPQ SM 580 3KTLFVKGLSE	520 MQNGKSKGY 	AFIEFASFE 	540 ** DAREALNSCNRRETI M 	E F
Hu: Ha: Mo: Bo: Ch: Hu: Ha:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS T^^^-	500 SATEFTIQEV 	FEKATFI KVPQ SM 580 3KTLFVKGLSF	520 NQNGKSKGY 	AFIEFASFE VPTA- 60 ESFDGSVRAR E	540 ** DAKEALNSCNREI M M 	E F -
Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS T^^^	500 SATEETIQEV 	FEKATFIKVPQ SM 580 3KTIFVKGLSE	520 NQNGRSKG? 	24FIEFASFE VPTA- 60 25FDGSVRAR E	540 ** DAKEALNSCNKREI M TT 0 IVTDRETG88KGFG	E - F
Hu: Ha: Bo: Ch: Hu: Ha: Mo: Bo:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS T^^^	500 SATEETLQEV. KE A-SIr PNARS^^QPI S^^^	FEKATFI KVPQ SM 580 3KTLFVKGLSE	520 NQNGKSKG 	VAFIEFASFE	540 ** DAKEALNSCNKREI 	E - F
Hu: Ha: Bo: Ch: Hu: Ha: Bo: Ch:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS TFSS-SWQK-N	500 SATEETLQEV. KE A-SL/ PNARS^^^QPI 	FEKATFI KVPQ -KSM 580 3KTLFVKGLSF R	520 NQNGKSKG 	2751EFASFE 	540 ** DAREALNSCNRREIJ M 	E - F
Hu: Ha: Bo: Ch: Hu: Ha: Bo: Ch:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. A-SL PNARS^^QPI GGFN-Q	FEKATFI KVPQ -KSM 580 SKTLFVKGLSF R	520 NQIRGRSKG Q	VAFIEFASFE 	540 ** DAKEALNSCNKREI M 	E - F
Hu: Ha: Mo: Do: Ch: Hu: Ha: Bo: Ch:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. A-SL PNARS^^^QPI 	FEKATFIKVPQ -KSM 580 3KTIFVKGLSE R	520 NQNGRSKG 	2AFIEFASFE 	540 ** DAKEALNSCNRREI M M TT 0 IVTDRETGSSRGPG	E - F
Hu: Ha: Bo: Ch: Hu: Ha: Bo: Ch: Hu:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. 	FEKATFIKVPC -KSM 580 3KTLFVKGLSF R 10 DWAKPKGEGG	520 NQNGRSKG 	24FIEFASFE 	540 ** DAKEALNSCNREI 	
Hu: Ha: Bo: Ch: Hu: Ha: Bo: Ch: Hu: Ha:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^ARGS TFSS-SWQK-N 620 VDFNSEEDAK^^EAME AAK	500 SATEETLQEV. KE A-SL/ PNARS^^QPI GGPN-QP 	FEKATFI KVPQ -XSM 580 3KTLFVKGLSE R 10 JWAKPKGEGG	520 MQNGKSKG -Q -PHP DTTEETLKI 	2AFIEFASFE 	540 ** DAKEALNSCNKREI M M M M M M M-	
Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Hu: Ha: Mo:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. 	FEKATFI KVPQ -KSM 580 SKTLFVKGLSE R 10 LDWAKPKGEGG	520 MQNGRSKG 	ZAFIEFASFE 	540 ** DAREALNSCNREI M 	
Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Eo:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS TFSS-SWQK-N 620 VDFNSEEDAK^^^EAME 	500 SATEETLQEV. A-SL PNARS^^QPI 	FEKATFI KVPQ -KSM 580 3KTLFVKGLSE R 10 10WAKPKGEGG	520 MQNGRSKG 	2771EFASFE	540 ** DAKEALNSCNREI M M 	
Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. A-S	FEKATFI KVPQ -KSM 580 3KTLFVKGLSE 	520 NQNGRSKG 	2AFIEFASFE 	540 ** DAKEALNSCNREI M 	
Hu: Ha: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS TFSS-SWQK-N 620 VDFNSEEDAK^^^EAME 	500 SATEETLQEV. A-SL PNARS^^QPH 	FEKATFI KVPQ -KSM 580 3KTLFVKGLSE R 10 LJWAKPKGEOG 	520 MQNGRSKG 	2AFIEFASFE 	540 ** DAREALNSCNRREI 	
Hu: Ha: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch:	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS T	500 SATEETLQEV. A-SL PNARS^^QPA 	FEKATFI KVPQ -KSM 580 SKTLFVKGLSF 	520 NQIRGRSKG7 PHP DTTEETLK0 NQ-RP DTTEETLK0 R- F^GGRGGGG 	ZAFIEFASFE	540 ** DAREALNSCNREI M	
Hu: Ha: Mo: Bo: Ch: Hu: Ha: Bo: Ch: Hu: Ha: Hu: Hu:	STWSGESKTLVLSNLSY 	500 SATEETLQEV. A-S	FEKATFI KVPQ -KSM 580 SKTLFVKGLSE 	520 MQNGRSKG PHP NQ-RP DTTEETLK R- E^GGRGGGG R- 	24F1EFASFE 	540 ** DAREALNSCNREI M	
Hu: Ha: Mo: Eo: Ch: Hu: Ha: Bo: Ch: Hu: Ha: Hu: Ha: Mo: Eo: Ch: Hu: Ha: Ka: Ka: Ka: Ka: Ka: Ka: Ka: Ka: Ka: K	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS TFSS-SWQK-N 620 VDFNSEEDAK^^^EAME 	500 SATEETLQEV. A-SL PNARS^^QPH 	FEKATFI KVPQ -KSM 580 3KTLFVKGLSE 	520 MQNGRSKG 	2AFIEFASFE	540 ** DAKEALNSCNREI M M M 	
Hu: Ha: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Ha: Mo: Ha: Mo: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Hu: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS TFSS-SWQK-N 620 VDFNSEEDAK^^^EAME 	500 SATEETLQEV. 	FEKATFI KVPQ -KSM 580 3KTLFVKGLGE 	520 MQNGRSKG Q PHP DTTEETLR 	2AFIEFASFE 	540 ** DAKEALNSCNRET	
Hu: Ha: Bo: Ch: Hu: Ha: Bo: Ch: Hu: Ha: Mo: Bo: Ch: Ha: Bo: Ch: Eb: Ch: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha: Ha	STWSGESKTLVLSNLSY GGER^IV-N 560 GRAIRLELQGP^^^RGS 	500 SATEETLQEV. A-SL/ PNARS^^QPM 	FEKATFI KVPQ -KSM 580 3KTLFVKGLSE 	520 NQNGRSKG Q NQ-RP DTTEETLKI 	ZAFIEFASFE	540 ** DAREALNSCNREI M	

Fig.2. Comparison of the amino acid sequences of human and bovine nucleolin with hamster, mouse and chicken. Hu, human nucleolin; Ha, hamster nucleolin; Mo, mouse nucleolin; Bo, bovine nucleolin; Ch, chicken nucleolin. The numbering of the amino acids of human nucleolin is indicated above the single letter code (top sequence). Amino acids in the other protein sequences which are identical to human nucleolin are indicated by dashes and the amino acids which are different are denoted by single letters. Circumflexes indicate spaces introduced into the sequences for better alignment. Asterisks above the human sequence indicate potential dibasic cleavage sites.



Fig.3. Northern blot analysis of nucleolin mRNA from different tissues. 5 μg from various tissues were analyzed. (SK-N-SH) Human neuroblastoma cell line; (BL) bovine liver; (BK) bovine kidney; (BB) bovine brain; (RB) rat brain; (RR) rat retina; (HA) human adrenal; (HR) human retina; (HL) human liver; (BR) bovine retina; (BA) bovine adrenal.

ponentially growing cells such as the neuroblastoma SK-N-SH. However, nucleolin synthesis was considerable in tissues containing nondividing cells such as retina, adrenal medulla and brain.

4. DISCUSSION

4.1. Analysis of human nucleolin sequence

A full-length human nucleolin cDNA from a retinal library was isolated using a bovine nucleolin partial cDNA. The 5'-untranslated region of 114 bp in human is approximately 50% different from hamster and mouse sequences. The first 105 nucleotides in the 3'-untranslated region of human nucleolin have a striking similarity (85%) to hamster and mouse, while the remainder of the nucleotides show considerable variation by 38% and 50%, respectively. As expected, the coding sequences for nucleolin in several species are highly conserved.

4.2. Nucleolin structure and function

We have observed that nucleolin has seventeen very conserved dibasic cleavage sites (15, 51, 54, 62, 70, 79, 87, 95, 109, 124, 141, 219, 279, 281, 294, 387, and 702). These could be the processing signals giving rise to different nucleolar proteins and possibly to other, as yet unidentified, biologically active peptides. Furthermore, human nucleolin has an additional site at position 545. The presence of potential dibasic cleavage sites for proteolytic processing has been observed in the case of chromogranin A [26] and some prohormones [27]. Nucleolin is highly susceptible to proteolytic degradation [28]. Polypeptides with masses of 95, 76, 70, 60 and 50 kDa have been identified with specific antisera to the 100 kDa protein, and it was suggested that these presumed nucleolin fragments resulted from thiol protease cleavage of the 100 kDa molecule [29]. It would be interesting to investigate which dibasic cleavage sites actually are utilized in vivo.

Recent reports suggest that karyophilic signals present in the proteins may be responsible for transporting cytoplasmic ribosomal components into the nucleolus [8,30]. The conserved, repeated sequence [TP(GAV)KK(GAV)2] at positions 59, 76, 84, 92, 106 and 121 could be putative nucleolar localization signals. These sequences contain a hydroxy amino acid, a proline, and paired dibasic residues, all of which are found in nuclear localization signals [31,32].

Other regions rich in aspartate and glutamate (positions 143–169, 184–209, and 234–272) are thought to be involved in binding to histones [33] and show variation between species. In the first segment, a potential phosphorylation site is displaced 5 residues toward the N-terminus of the human gene. In the second acidic segment, the phosphorylation site is abolished only in the bovine, which also has five additional glutamate residues relative to the rodent and human proteins. In the third acidic segment, the number of acidic amino acids is greatest in bovine (44), followed by human (39), hamster (34), mouse (34) and Xenopus (20). Most notably, in all three segments the substitutions and additions are only with aspartate or glutamate and occasionally with serine. In between the first two acidic segments, the human sequence exhibits both a deletion and high variability when compared to the hamster and mouse sequences. Conserved phosphorylation sites are located at position 28, 34, 145, 498, 580, 584 and 619. The variation in the number of acidic amino acids and in the location of phosphorylation

The four RNA-binding domains (positions 324-379, 406-461, 498-553 and 588-643) containing the central consensus sequences are present in human nucleolin. The fourth domain has the greatest conservation of sequence between species as recently reported in a comparison of the rodent and Xenopus sequences [36]. A glycine-rich carboxy-terminal domain (position 648-703) is also conserved and may be involved in proteinprotein and/or protein-nucleic acid interactions [37]. The three additional glycine residues after position 695 seem specific for hamster nucleolin and 2 additional glycines after position 677 seem specific for chicken. Two conserved glycosylation sites are observed at position 317 and 492. The glycosylation sites at position 399 and 403, previously observed in hamster and mouse, are not detected in either human or bovine, and the glycosylation site at 478 is present in human and hamster, but not in mouse and chicken. We report an additional glycosylation site at position 541.

We also compared the predicted secondary structures for the human, hamster, and mouse nucleolin amino acid sequences. All three predicted structures were almost identical with more than 60% helix predicted for each protein. Even where the human sequence differed with the mouse or hamster, such as residues 401-405 and 470-474, the predicted secondary structures were essentially identical. Significant lengths of predicted α -helix were found in all regions of the protein except the glycine-rich carboxy-terminal region. Very large probabilities of α -helix formation were observed immediately on either side of the four consensus sequences (positions 344-354, 426-436, 520-530 and 607-617) of RNA binding. These predicted helix structures were approximately 10-15 residues in length. The consensus sequences themselves are predicted to have β -structures with low probability. Thus, it appears that a variation of the helix-turn-helix motif may be present in each of the four RNA-binding segments. A number of DNA-binding proteins with known structure exhibit a helix-turn-helix segment which interacts with the DNA [38]. In the case of nucleolin the turn segments (containing the consensus sequences) are between 10 and 20 residues in length which is larger than in the DNAbinding proteins. Although the mode of interaction in this region of nucleolin with RNA may be different than the interaction of DNA-binding proteins with double-stranded DNA, it is possible that the structural motif of a helix-turn-helix in various proteins has been adapted for binding to both types of polynucleotide.

Acknowledgements: We are grateful to Drs Karin Magendzo and Anat Shirvan for their valuable advice and cooperation throughout this work; Drs Frank Gonzales, Ettore Appella and Anil K. Jaiswal for their helpful discussions; Barbara Chung and Conny Cultraro for technical assistance; Dr Vicente Notario for helping in Northern blot analysis; Katie Sidman and Lois Hunt for helping in computer analysis and Delma Tyler for typing the manuscript. This work was supported by grants from USPHS (GM 27695) and the Cystic Fibrosis Foundation.

REFERENCES

- Escande, M.L., Gas, N. and Stevens, B.J. (1985) Biol. Cell 53, 99-110.
- [2] Smetana, K., Ochs, R., Lischwe, M.A., Gyorkey, F., Freireich, E., Chudomel, V. and Busch, H. (1984) Exp. Cell Res. 1521, 195-203.
- [3] Jordon, G. (1987) Nature 329, 489-490.
- [4] Bousche, G., Caizergues-Ferrer, M., Bugler, B. and Amalric, F. (1984) Nucleic Acids Res. 12, 3025-3035.
- [5] Olson, M.O.J., Rivers, Z.M., Thompson, B.A., Kao, W.Y. and Case, S.T. (1983) Biochemistry 22, 3345–3351.
- [6] Herrara, A.H. and Olson, M.O.J. (1986) Biochemistry 25, 6258-6264.
- Bugler, B., Bourbon, H.M., Lapeyre, B., Wallace, M., Chang, J.H., Amalric, F. and Olson, M.O.J. (1987) J. Biol. Chem. 262, 10922-10925.
- [8] Borer, R.A., Lehner, C.F., Eppenberger, H.M. and Nigg, E.A. (1989) Cell 56, 379–390.
- [9] Bourbon, H.M., Bugler, B., Caizergues-Ferrer, M. and Amalric, F. (1983) FEBS Lett. 155, 218-222.
- [10] Bousche, G., Gas, N., Prats, H., Baldin, V., Tauber, J.-P., Teissie, J. and Amalric, F. (1987) Proc. Natl. Acad. Sci. USA 84, 6770-6774.
- [11] Lapeyre, B., Bourbon, H.M. and Amalric, F. (1987) Proc. Natl. Acad. Sci. USA 84, 1472-1476.
- [12] Bourbon, H.M., Lapeyre, B. and Amalric, F. (1988) J. Mol. Biol. 200, 627–638.
- [13] Aviv, M. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
- [14] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
- [15] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269.

- [16] Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 86, 1194–1198.
- [17] Young, R.A. and Davis, R.W. (1983) Science 222, 778-782.
- [18] Benten, W.D. and Davis, R.W. (1977) Science 196, 1880-1882.
- [19] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
- [20] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
- [21] Finer-Moore, J. and Stroud, R.M. (1984) Proc. Natl. Acad. Sci. USA 81, 155–159.
- [22] Hod, Y., Morris, S.M. and Hanson, R.W. (1984) J. Biol. Chem. 259, 25603–25608.
- [23] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 137, 266-267.
- [24] Winkler, H. and Westhead, E. (1980) Neuroscience 5, 1803-1823.
- [25] Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
- [26] Icangelo, A., Affolter, M.-U., Eiden, L.E., Herbert, E. and Grimes, M. (1986) Nature 324, 82-86.
- [27] Docherty, K. and Steiner, D.F. (1982) Annu. Rev. Physiol. 44, 625-638.

- [28] Lischwe, M.A., Richards, R.L., Busch, R.K. and Busch, M. (1981) Exp. Cell Res. 136, 101-109.
- [29] Bugler, B., Caizergues-Ferrer, M., Bouche, G. and Amalric, F. (1982) Eur. J. Biochem. 128, 475-480.
- [30] Peters, R. (1986) Biochim. Biophys. Acta 864, 305-359.
- [31] Kalderon, D., Roberta, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499-509.
- [32] Slomi, H., Shida, H., Nam, S.K., Nosaka, T., Maki, M. and Hatanaka, M. (1988) Cell 55, 197–209.
- [33] Erard, M., Belenguer, P., Caizergues-Ferrer, M., Pantaloni, A. and Amalric, F. (1988) Eur. J. Biochem. 175, 525-530.
- [34] Dhar, V.N., Miller, D.A., Kulkarni, A.B. and Miller, O.J. (1987) Mol. Cell. Biol. 7, 1289–1292.
- [35] Mamrack, M.D., Olson, M.O.J. and Busch, H. (1979) Biochemistry 18, 3381-3386.
- [36] Caizergues-Ferres, M., Mariottini, P., Curie, C., Lapeyre, B., Gas, N., Amalric, F. and Amaldi, F. (1989) Genes Dev. 3, 324-333.
- [37] Chung, S.Y. and Wooley, J. (1986) Proteins: Struct. Func. Gen. 1, 195-210.
- [38] Schleif, R. (1988) Science 241, 1182-1187.