LINDA J. KO,<sup>1</sup> MASAYUKI YAMAMOTO,<sup>2</sup> MARK W. LEONARD,<sup>1</sup> KATHLEEN M. GEORGE,<sup>1</sup> PAUL TING,<sup>1</sup> and JAMES DOUGLAS ENGEL<sup>1\*</sup>

Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208-3500,<sup>1</sup> and Department of Biochemistry, Tohoku University School of Medicine, 2-1 Seiryomachi, Aoba-ku, Sendai 981, Japan<sup>2</sup>

Received 14 December 1990/Accepted 26 February 1991

A family of transcriptional activators has recently been identified in chickens; these transcriptional activators recognize a common consensus motif (WGATAR) through a conserved  $C_4$  zinc finger DNA-binding domain. One of the members of this multigene family, cGATA-3, is most abundantly expressed in the T-lymphocyte cell lineage. Analysis of human and murine GATA-3 factors shows a striking degree of amino acid sequence identity and similar patterns of tissue specificity of expression in these three organisms. The murine and human factors are abundantly expressed in a variety of human and murine T-cell lines and can activate transcription through a tissue-specific GATA-binding site identified within the human T-cell receptor  $\delta$  gene enhancer. We infer that the murine and human GATA-3 proteins play a central and highly conserved role in vertebrate T-cell-specific transcriptional regulation.

Cellular differentiation during hematopoiesis is the result of a series of hierarchical responses to a complex extracellular, cytoplasmic, and nuclear signaling pathway finally leading to the induction of genes unique to the erythroid, myeloid, or lymphoid cell lineages. Terminal lineage specification appears to be mediated by the binding of nuclear *trans*-acting regulatory protein factors, some of which are uniquely expressed in distinct hematopoietic cell types, to *cis*-regulatory elements of a given lineage-specific gene. One of the best understood of these lineage-specific loci is that encoding  $\beta$ -globin, where both distal and proximal *cis*regulatory sequences have been demonstrated to bind tissuespecific and ubiquitous *trans*-acting factors to elicit the characteristic erythroid-specific expression of this gene.

Recently a great deal of interest has arisen from the discovery that transcription factor GATA-1 (29) (previously designated GF-1, NF-E1, Eryf1, or EF-1 [8, 22, 31, 44]) appears to be a major regulator of erythroid-specific transcription. GATA-1 was originally identified as an erythroidspecific DNA-binding protein (8, 22) and implicated in the transcriptional activation of both globin and nonglobin (24, 30, 31) erythrocyte-specific genes. The cloning of GATA-1 cDNAs has provided further insight into the properties of this factor (7, 43). The protein binds to a WGATAR consensus sequence site through two similar C-X<sub>2</sub>-C-X<sub>17</sub>-C-X<sub>2</sub>-C zinc fingers which differentially contribute to the specificity for and affinity of DNA binding (21). Domains both amino and carboxy terminal to these zinc fingers contribute to transcriptional activation by mGATA-1 (21). Although the factor was originally identified as erythroid specific, it has subsequently been found to also be expressed in megakaryocytes and mast cells of myeloid origin (23, 35).

GATA-1 has recently been shown to be but one member of a multigene family in chickens (originally designated cNF-E1a, cNF-E1b, and cNF-E1c; now termed cGATA-1, cGATA-2, and cGATA-3, respectively [29]), with each family member exhibiting a unique pattern of tissue-restricted expression (47). Members of the cGATA (chicken GATA) family of proteins are related by virtue of their zinc fingers, which show a high degree of identity to other members of the family both within species and between species (42, 47, 48). All of the factors recognize the common consensus binding site (WGATAR) and function as transcriptional activators in vivo (47).

Of particular interest was the observation that cGATA-3 mRNA is predominantly expressed in chicken T lymphocytes and embryonic brain tissue (47). Here we demonstrate that the human and murine homologs of this factor (hGATA-3 and mGATA-3, respectively) are abundantly expressed in a variety of T-lymphocyte cell lines and that the conceptually translated proteins share a striking degree of amino acid sequence identity with cGATA-3 and with one another. We show that the mRNA expression patterns of these factors are also highly conserved, suggesting that the proteins fulfill a critical function in the activation of T-lymphocyte-specific genes. The identification of GATA-binding sites in the human T-cell receptor (TCR)  $\delta$  gene enhancer, as well as the further demonstration that these sites are capable of conferring GATA-3-dependent transcriptional induction to reporter genes linked in cis, underscores the likelihood that GATA-3 plays a highly conserved, functionally important role in T-lymphocyte-specific transcriptional regulation in vertebrates.

## **MATERIALS AND METHODS**

Cell lines. Cell lines were purchased from the American Type Culture Collection (T-cell lines A0DH, BW5147.3, MOLT-3, CCRF-CEM, CCRF-HSB-2, and HuT78) or were gifts from V. Patel (murine lines MEL, BaF3, 38B9, PD31, and MPC11), S. K. Pierce (murine lines TPc and CH27), or S. T. Rosen (human B-lymphoma lines IM9 and U266). The murine TCR  $\alpha\beta^+$  T-cell lines A0DH and TPc and the TCR

<sup>\*</sup> Corresponding author.

 $\alpha\beta^-$  T-cell line BW5147.3 were grown in Dulbecco's modified Eagle's medium (DMEM). The human TCR<sup>-</sup> T-cell lines MOLT-3 and HuT78 were grown in RPMI 1640 plus 10% fetal bovine serum (FBS), and the human TCR<sup>-</sup> T-cell line CCRF-HSB-2 was grown in Iscove's DMEM plus 10% FBS; the TCR  $\alpha\beta^+$  human T-cell line CCRF-CEM was grown in RPMI 1640 plus 20% FBS. QT6 (transformed quail fibroblast) cells were grown in DMEM plus 5% FBS plus 2% chicken serum.

Isolation and characterization of mGATA-3 and hGATA-3 cDNAs. cDNA libraries prepared from RNA derived from the murine EL4 and human Jurkat TCR  $\alpha\beta^+$  T-lymphocyte cell lines (Stratagene and Clontech, respectively) were screened (standard hybridization, but reduced washing stringency: 0.5× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] plus blot wash buffer at 37°C [47]) with random primer-radiolabeled cGATA-3 (p30c [47]) as the hybridization probe (2). Approximately  $5 \times 10^6$  PFU of each library was screened; 68 and 93 human and murine positives, respectively, were identified. Twenty-four of each were plaque purified, and those with the largest cDNA inserts were subcloned and sequenced. The DNA sequence of the largest isolated murine (cDNA clone mc5b) and human (cDNA clone p8h) homologs of the cGATA-3 factor was determined by the dideoxy-chain termination method on serially deleted subclones in pGEM7Zf(+) (Promega, Inc.) (17, 47).

**RNA blot analysis.** Either 8  $\mu$ g of total RNA isolated from various mouse organs, tissues, or cell lines or 10  $\mu$ g of total RNA from either human B-lymphoma or T-lymphocyte cell lines was denatured, electrophoresed, and blotted to nylon filters (34). The filters were then hybridized to random primer-labeled mGATA-3 cDNA clone mc5b or hGATA-3 clone p8h2, respectively, washed at moderate stringency (50 mM Tris-HCl [pH 7.5]–1 mM EDTA–1 mM sodium phosphate–0.1% sodium dodecyl sulfate at 55°C three times for 1 h each time), and exposed for autoradiography. Exposure times are given in the figure legends.

Gel mobility shift assays. Whole-cell extracts were prepared as previously described (43, 47) and incubated with radiolabeled oligonucleotides (amounts specified in figure legends) corresponding to the GATA site sequences of either the mouse  $\alpha 1$  globin promoter (M $\alpha$ P [31]) or the human TCR  $\delta$  gene enhancer ( $\delta$ E4 [33]; see Results). The  $\delta$ E4 sequence is as follows:

# TCGACACTTGATAACAGAAAGTGATAACTCT GTGAACTATTGTCTTTCACTATTGAGAAGCT

DNA binding reactions and electrophoresis were performed as described previously (47).

Cotransfection trans-activation assays. The rabbit  $\beta$ -globin TATA box and transcription initiation site was used to direct transcription of the human growth hormone gene (see Fig. 6A) in plasmid p0GH (Nichols Institute Diagnostics, San Juan Capistrano, Calif. [43]). To this minimal promoter construct, either one or three copies of an oligonucleotide corresponding in sequence to the TCR  $\delta$  gene enhancer footprint 4 ( $\delta$ E4 [33]) or three copies of an oligonucleotide corresponding to the chicken  $\beta$ -globin enhancer footprint 4  $(\beta E-F4 [11])$  were inserted directly 5' to the TATA box in the orientation specified in Fig. 6 to produce the H18GH, H38GH and C3BGH reporter plasmids, respectively (see Fig. 6A). Activator plasmids R/mGATA-3 and R/hGATA-3 were prepared by sense insertion of the cDNA segments of mGATA-3 clone mc5b or hGATA-3 clone p8h2, respectively, into the unique HindIII site of RSV · CAT (14) after the chloramphenicol acetyltransferase-coding sequences were removed. The transcription of the recombinant GATA-3 cDNAs is therefore directed by the Rous sarcoma virus promoter and enhancer.

Immortalized quail fibroblasts (QT6 cells [25]) were used as recipient cells in activation assays as previously described (13, 47). The cells were transfected with 8  $\mu$ g of reporter gene plasmid and 2 µg of (potential) activator plasmid by the CaPO<sub>4</sub> method, and the supernatants were assayed for secreted human growth hormone 4 days posttransfection by radioimmunoassay (RIA) (using the Allegro hGH system from Nichols Institute Diagnostics [47]). After the supernatants were collected for RIA, the cells were trypsinized, counted, and lysed to prepare extracts for gel mobility shift assays (43). Counts derived from the hGH RIA were normalized for cell numbers, and trans-activation was calculated (47) according to the equation trans-activation = [(A +(R) - M, where A + R represents the counts recovered per minute in the RIA after transfection with activator and reporter plasmids, M represents the counts per minute in the medium alone, and R represents the counts recovered per minute after transfection with the reporter plasmid alone.

Nucleotide sequence accession numbers. The nucleotide sequence in Fig. 2 is listed in the GenBank data base under accession number X55123; the sequence in Fig. 3 is listed in GenBank under accession number X55122.

## RESULTS

Murine and human T lymphocytes contain an abundant GATA-binding protein. With the identification of T lymphocytes as the major sites of cGATA-3 expression (47), we initially sought to determine whether human and murine T lymphocytes express a GATA-binding protein. Extracts were prepared from a variety of murine T-lymphocyte cell lines and incubated with a 30-bp oligonucleotide probe containing a consensus GATA-binding site from the mouse  $\alpha$ 1 globin promoter (M $\alpha$ P [43]) in a gel mobility shift assay (9, 12). The complex produced with a protein present in the extracts has a slightly lower mobility than that formed with the cGATA-3 protein and is specifically eliminated by unlabeled M $\alpha$ P oligonucleotide in the binding reaction (Fig. 1A). Extracts were also prepared from several human T-cell lines and tested in mobility shift assays with the M $\alpha$ P probe. The strong signal produced by complex formation indicates the presence of an abundant GATA-binding protein in human T lymphocytes; this complex is also specifically inhibited by an excess of unlabeled  $M\alpha P$  oligonucleotide in the binding reaction (Fig. 1B). Thus, murine and human T-lymphocyte cell lines expressing either an  $\alpha\beta$  TCR or no TCR synthesize an abundant protein which binds with high affinity to the GATA consensus sequence.

**Cloning and characterization of murine and human GATA-3 cDNAs.** With the knowledge that a prominent GATA-binding protein is expressed in mammalian T cells, we then isolated the murine and human homologs of cGATA-3. cDNA libraries prepared from murine EL4 and human Jurkat T-lymphocyte cell lines were screened with the full-length cGATA-3 cDNA clone p31a (2, 47). Independent clones were isolated from each library, and mGATA-3 and hGATA-3 cDNA clones of 2,054 and 2,030 nucleotides, respectively, were sequenced (Fig. 2 and 3, respectively) (17). Both cDNA clones contain complete coding regions; the open reading frame of the mGATA-3 cDNA clone mc5b is predicted to encode a protein of 443 amino acids with a



FIG. 1. A protein binding to the WGATAR motif is prevalent in murine and human T-lymphocyte cell lines. (A) Whole-cell extracts were prepared from murine T-cell lines and examined by gel mobility shift assay for binding to 1 ng of radiolabeled  $M\alpha P$ oligonucleotide (43, 47). The T-lymphocyte cell lines used for extract preparation were A0DH (lanes 2 and 6), BW5147.3 (lanes 3 and 7), and TPc (lanes 4 and 8). For comparison, cGATA-3 as expressed in chicken embryo fibroblast cells from a recombinant retrovirus (6a) is shown (lanes 1 and 5). Lanes 1 to 4 contained 1.5 µg of nonspecific poly(dI-dC)/poly(dA-dT) competitor, while lanes 5 to 8 contained the same plus 150 ng of unlabeled  $M\alpha P$  oligonucleotide. B and F refer to the positions of bound and free  $M\alpha P$ oligonucleotide, respectively. (B) Extracts were prepared from various human T-cell lines (43) and examined for specific DNA binding to the radiolabeled MaP oligonucleotide by gel mobility shift assay. The cell lines used were MOLT-3 (lanes 1 and 5), CCRF-CEM (lanes 2 and 6), CCRF-HSB-2 (lanes 3 and 7), and HuT78 (lanes 4 and 8). Lanes 1 to 4 contained 1.5 µg of nonspecific competitor DNA, while lanes 5 to 8 contained the same plus 150 ng of unlabeled  $M\alpha P$  oligonucleotide. Once again, B and F indicate the positions of bound and free radiolabeled oligonucleotide, respectively.

molecular mass of 47,964 Da (pI, 9.89), and the hGATA-3 clone p8h contains a single long open reading frame encoding a 444-amino-acid protein with a predicted molecular mass of 47,973 Da (pI, 9.84). Both of these are similar in size and isoelectric point to the 48,194-Da molecular mass and pI of 9.89 predicted for the cGATA-3 protein (47).

Vertebrate GATA-3 proteins are highly conserved. A comparison of the deduced amino acid sequences of the chicken, human, and murine proteins (Fig. 4) illustrates the striking identity between the GATA-3 factors of these species. cGATA-3 and hGATA-3 have >92% overall identity; the amino-terminal residues (1 to 215) share 85% identity, while the carboxy-terminal amino acids (216 to 444) share 97% identity. Similarly, cGATA-3 and mGATA-3 show 92% overall sequence identity. In contrast to this high degree of conservation, the erythroid/megakaryocytic/mast (23, 35) cell-specific family member GATA-1 is far less well conserved between these same species, having only 41% overall identity between chickens and humans (42). The GATA-3 proteins also share far greater identity across species boundaries than they do with the GATA-1 or GATA-2 protein within the same species (data not shown).

Tissue distribution of the murine and human GATA-3 transcripts. Expression of murine and human GATA-3 was examined by RNA blot analysis. As shown in Fig. 5A, mGATA-3 is expressed as an approximately 4-kb mRNA

70 TTAACTGCAAAACAAACAATCACTAAACCACCCCTCTCTCGCGCCCCCCACGCCGCCGCAGCAGCCCCCCGCGCCCCAGCCCCGCCCCCGCCCCCGCCCCCC	1	GCTA	AACT	TATCO	CCGC/	AAGA	TTT	TCT	тсст	rccci		ссто	CTT	TTG	тсто	CTT	TCT	TACCC
139 AGGCCACAGCCCTCCCTCTACCCCCCAGGTTCCGCGCCCCCCAGCAGCGCAGCCACAGCCAGC	70	TTAA	CTG	CAAAC	CAAAC	CAT	TAAAC	GAC	ссто	TCCT	rgggg	CTCC	GACO	GCAC	GAGT	CCGC	GGA	CTCCC
208 ATG GAG GTG ACT GCG GAC CAG GCG CGC TGG AGC CAG CAC CAC CCC GGC H E V T A D Q P E V V S H H P C A L G H S 310 TAC ATG GAA GGT CAG TAT CGC GTG ACC GAA CAG GTG GAC GTA CTT TT AAC Y H E A Q Y P L T E E V D V L F N 361 ATC GAT GGT CAG GGC AAC CAC GTC CGC TAC TAC GGA AAC TCC GTC AGG I D G Q G N H V P S Y Y G N S V R 412 GCT ACG GTG CAG AGC TAT CCT CGG ACC CAC CGG AGC CAG GTA TCC AGC A T V Q R Y P P T H H G S Q V C R 413 GCT AGC GTG CAG AGC TAT CCT CGG ACC CAC CGG AGC CAG GTA TCC CCC A T V Q R Y P P T H H G S Q V C C R 453 GCG CCT CTG CTG CAC GCC TGT GG CCT GG AGC GTG GAT GGC GCA AAA CC CTG P P L L H G S L P V L D G G K A L 514 AGC AGC GAC CAC ACC GCC TGG GAC CTC AGC CTC TGC AGA GG S S H H T A S P V N L S P F S K T 565 TGC ATC CAC CAC GGC TGT CG CGC CAC TCC AGT CTC TGC AGC GC TAC S I H H G S P G P L S V Y F P A K C 5616 TGC TGT TGT CGG GGC GGC GGC CAC TGC AGT GGT GTC TGC GGC GGC AGA GGC 78 S L A A G S A A G S S P L S Y Y F P A 18 TC CGC GGC GGC GGC GAC AGT GGA CGT GG CG TG TGC GTT TAC CCT CGG GGA 718 TCC CGC GGC GGC GGC GAC AGT GGA CAT GGA CGT GG CTC TGC GCG GGA 718 TCC GGC GGC GGC GGA GAA GAA GAG GGA GGA G	139	AGGO	CCGA	CAGCO	CTCC	CTCI	LUCCO	CCCC	GGT	CCCC	GCCC	GCCC	GAGAC	GCCC	CGAC	CAC	GCCC	AGGAC
259       GTC CAAC GGT CAG CAC CAC CAC CAC CGG GGC CT CGC CAT TGG         V       L       N       G       Q       H       P       D       T       H       H       F       G       L       G       G       L       S       S       L       G       G       L       G       G       L       G       G       L       G       G       L       G       L       T       T       T       T       A       G       L       T       T       T       T       T       A       G       L       T<	208	ATG	GAG	GTG	ACT	GCG	GAC	CAG	CCC	cec	TGG	GIG	AGC	CAC	CAT	CAC	ccc	GCG
<pre>vic of of an of a control and and and and and and and and and and</pre>	250	CTC.	E		T	A	500	Q.		R.	N.	N.	S	H	H	H	P	A
310       TAC ATG GAA GCT GAG TAT CCC CTC AC GAA GAG CTG GAC GTA CTT TTT AAC         361       ATC GAT GGT GAA GGC AAC CAC GTC CGC TAC CGA AC CTC GTC TAC TTT TT AAC         371       ATC GAT GGT GAA GGC AAC CAC GTC CGC TAC CGA AC GCA GTC TGC GTC TAC TCC GAC GGA TCC GTC ACG         11       D       G       Q       G       N       H       Y       P       E       E       V       D       V       L       P       N       S       Y       Y       G       N       S       Y       Y       G       N       S       Y       Y       G       N       S       Y       P       Y       H       H       G       S       Q       Y       P       P       T       H       H       G       S       Q       Y       P       P       T       H       H       G       S       Q       Y       P       P       T       H       H       G       S       C       P       Y       P       P       AC       AC       C	239	510	10	AAC N	661	CAG 0	UAC 1	ULA B	GAU	ACG	UAU	UAU		GGU	CIC.	GGC	CAI	TCG
<pre>130 NO NO NO OT AGE AND ON OT NO OT NO OT AGE ON OTA OTA CTA CTA AGE 361 ATC GAT GGT CAA GGC AAC CAC GTC CGC TC TAC TAC GGA AAC TCC GTC AGC 1 D G Q G N H V P S Y G N S V R 412 GCT ACG GTG CAC AGC TAT CTC CGC ACC CAC GGC AAG CGC CAC A T V Q R Y P P T H H G S Q V C R 453 CGC CCT GTC CTC CTC CAC GGC ATC GTC GTG GAT GGC GGC AAA GCC CTC P P L L H G S L P V L D G G K A L 514 AGC AGC CAC CAC ACC GCC TCT GCC GTG GAT GGC GCC TTC TCC AAG AGC S S H H T A S P V H L S P F S K T 555 TCC ATC CAC CAC GGC TCT CCG GGC CAT GTC GTC TAC CTC CGG GGT TAC S I H H G S P G P L S V Y P F A S 616 TCC TCT TGT GGC GGC GGC CAC TCC AGT CCT CAT CTC TCC CAG AGC S S S L A A G H S S P H L F T F P C C C CGC CGC AAA GAC GTC TC CC CCA GAC CGT CTC CCC CG GGT TC S S S L A A G G H S S P H L F T F P C T P P K D V S P D P S L S T P G 718 TCC GCC GGC GGC GCA CAT GAC AAA GAC GTC TC ACC CCC GGC GGC AG S S C S C C G AAA GAC GTC TC CC CCAC CC TC ACT CTC CAC CCC GG GGC S A G S A R Q D E K E C L K Y Q V 769 CAG CTG CCA AGT AGC AGT GAC AAC GTG GAC CCC TAT GC ACC TT G C CCC TAT GTG GGG GCC CAG CTT GGA CCT TAG CC CCC AGC ACC TT T T L G G G A S S S A H H P I T T Y 871 CCG CCC TAT GTG GGC CCC CAT CC TG GAC CCC CAT ACC ACC TT T T Y 871 CCG CCC TAT GT GG GCC CAC CT GGA CCT TT CC ACC CCC AGC ACC TT Y T T L G G G A S S S A H H P I T T Y 871 CCG CCC TAT GTG GCC CCA AGC AGC GAC TT CT CC AC CC CAT CT TC A G C S P T G F S C L F P P S S L 22 CTG GGA GGA TCC CCT AC CTG CGA CTT TGC GAA TCT TAC CAC CTT L W R R B G O R S S A H H P I T T T Y 871 CCG GCC TAG GGC CCC TAC CTG CGA CTT TGC GAA TCT TAC CAC CTT L T R R B G O R S S A H H P I T T T Y 871 CCG GCC TAG GGC GCC GCG GGC TT CTG GAA CTC TAC CAC CTG L G G S S P T G G F C C C A S C C A T TS T F 177 AG GAA GGA TGA GGT AGC GGG GGC TT GTG GAA CCC CAA TT T C A G A GA GGA TGA GGT ACC GGG GGC TT GTG GAA CCC CAA TT T C A C K K V H B G O B R C L T C C A T T S T F T 177 AG GAA GGA TGA GGT AGC GGG GGC GGC GTT AGC GAA GCC AGC CAA CC L G G C T TG TAC TAC AGG GGA GGA GGG GGC GTT TGC GGA GGA AGC GAA CC ACC C L S A A A B R A A C T S C A S S</pre>	310	ТАС	ATC	GAA	CCT	CAC.	TAT	~~~~	cric.	100	~~~		CTC.	~~~	CT.	~~~~	TTT	
361 ATC GAT GT GAA GC AAC CAC GTC CG TC TAC TAC GA AAC TCC GTC AGC I D G Q G N H V P S Y Y G N S V R 412 GT AGG GTC CAG AGC TAT CTC CCG ACC CAC CAC GGA AGC AGC GTA TGC CGC A T V Q R Y P P T H H G S Q V C R 463 CCC CTC CTC GTG GTC GC GC GC GC GC GG GGC AAA GCC CTG P P L L L H G S L P V L D G G G K A L 514 AGC AGC CAC CG CC CTC CC CC GC TG GTC TGC CTC TCC CAG AC GC S S H H T A S P V N H L S P F S K T 565 TCC ATC CAC GG CC CC CCC CC GC GTG GTC TGC CCC CCC	510	Y	M	E		0	₩.	P	T.	T	R	R	V	n	V	T.		N
I       D       G       Q       R       H       Y       R       Y       R       Y       R       Y       P       F       H       H       G       G       Q       V       C       R         463       CCG       CCT       CTG       CGG       GGG       AGG       GGC       GGC       GGC       GGC       GGC       GGC       CG       R       A       L         144       GGC       CGC       CGC       GGC       GGC       GGC       GGC       CGC       R       A       L       D       G       G       K       A       L       L       L       G       S       L       P       V       L	361	ATC	GAT	GGT	CAA	GGC	AAC	CAC	GTC	ccg	тсс	TAC	TAC	GGA	AAC	тсс	GTC	AGG
412 GCT ACG GTG CAG AGG TAT CCT CCG ACC CAG CAG CGA GCG AGG CAG TAT GC CCC A T V Q R Y P P T H H G S Q V C R 463 CCG CCT CTG CTG CAG GAG TCT CTG CCC TCG CTG GAT GGC GGC AAA GCC CTG P P L L H G S L P V L D G G K A L 514 AGG AGC CAG CAC C GCC TCG CCG CTG GAG CTG CTC TCT CAG AGG S S H H T A S P V N L S P F S K T 555 TCC ATC CAC GG CT CTG CCG GGG CAG CTG TTG CTC TCC AGG AGG S S H H T A S P V N L S P F S K T 565 TCC ATC CAC GG CTC TCG CGG GGC CAG CTG TTG CAGC CCT TCG AGG S S S L A A G G H S S P H L P T P A S 616 TCC TCT TCT TG GCG GCG CAG TCT CC CAG CCC CAT CTC TC ACC ATC CCG S S S S L A A G G H S S P H L P T F F 667 CCC ACC CCG CGG CAG CAG CGA GAT GAG AAA GAG TGC CTC AAC TT CAG CTG C CA CC CCG CGG TGG GCC AGG CAA GAT GAG AAA GAG TGC CTC AAC TT CAG CTG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GCT GG CGC CA TC TCC TCA GCC CCC AGC CAC AGC ATT T T L G G A S S S A H H P I T T Y 871 CCG CCC AGT GGC GGC CAG TAC ACC TT GGA CTC TTC CAA GCAC CTG C T T C GG G TGG GCC GAG TAC ACC TT GGA CTC TGC CAC CCC AGC AGC AGC 9 2 C TG GGA GGA TGC CTA ACC TT GGA CTC TGC CAC CCC AGC AGC ATG 9 2 C TG GGA GGA TGC CTA ACC TT GGA CTC TGC CAC CCC AGC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC GGG TTC GGA TGT AGA CTG AGG CCC AGC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC GGG TTC GGA TGT AGC CGA GCC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC GGG TTC GGA TGT AGC CGA GCC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC GGG TTC GGA TGT AGC CGA GCC AGC CCC AGC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC GGG TTC GGA TGT AGC CGA GCC AGC CCC AGC AGC CTG 9 2 C TG GGA GGA TGC CCT ACC CGA TTA GCC AGC CCC AGC ACC TT ACC CAC 9 3 TCC ACC ACC AGA GC GAG GCT TG GGA TGT TGC AAC CCC AGC ACC TT ACC 1 3 3 3 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3		I	D	G	Q	G	N	H	V	P	S	Y	Y	G	N	S	v	R
A T V Q R Y P P T H H H G S Q V C R 463 CCC CCT CTC CTC CC CC CC CTC TCT CTC CCC CTC GCT CC CC CTC TCT CCC AGC ACC P P L L L H G S L P V L D C G GC AA GC CTC AG ACC S S H H T A S P V N L S P P S K T 565 TC ATC CAC CAC GC TCT CCG GGC CT CTG TC GT CC GT TCC CCC C	412	GCT	ACG	GTG	CAG	AGG	TAT	CCT	CCG	ACC	CAC	CAC	GGG	AGC	CAG	GTA	TGC	CGC
463 CCG CCT CTG CTG CAC GCA TCT CTG CCC TGG CTG CAT GCG CGA CACA AC P L L H G S L P V L D G G G K A L 514 AGC ACC CAC CAC CCC CCC TCG CCC TGG AAC CTC AGC CCC TTC TCC AAG ACC S S H H T A S S P V N L S P F S K T 565 TCC ATC CAC CAC GCC TCT CCG GGC CT CTG TCC GTT TAC CCT CCG GCT TCA S I H H G S P G P L S V Y P P A S 616 TCC TTC TCT GG GCG GCC GCC TCC CAC CC CAT CTC TC ACC TTC CCG S S S L A A G G H S S P T L F T F P 6718 TCC ACC CGC GCA AC AC GTC TCC CAC ACC CCC CAT CTC TTC ACC CTC CGG C CC CGC CGC ACA GCA GCT TCC CAC ACA CCT CAT CTC TC ACC CTC CGG S S S L A A G G H S S P D P S L S T P G 718 TCC GCC GGC TGC CCA CGC CAC CTC GA CTC CTC AG CTC CAC CCC GGA G C CC CCC CG GC CC ACC ACA GA AG AA GA CA CTC TC CCA GC CAC CCC GCA 769 CAC CTC CCA CAT ACC ATG AAG CTC CAG CAC CTC TC CCA GC CAC CTC Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GT GGC GCC TA TCC TCA GCC CAC CAC CAC TT ACC AGC TT T T L G G A S S S A H H P I T T T Y 811 CCG CCC TAT GTG CCC CAG CTC TCC GAG CTC TTC CCA CCC CAC CAC CAC CAC C TT GTG CCC CAG TA ACC TT ACC GGC TTC GGA CTT TCC AAG CCAC CAC TT 77 T L G G A S S S G L F P P S S L 922 CTG GCA GAA GCC CC CAC CAC CAC CAC CCC AAT ACC AGC AGC AC C ACC CTC GT AGC GCA GC GG GC TTC GGA CTT TCA AAC CGA GC AGC CCC A G S P T G C R E C V M C G A T S S T F 1024 CTC CAGC ACA GAG GCA GC GGA GTT GTG AAC TGO GCC CAAG CCC CAAC CCC ACA 7 T T L G GC GA CAC GCC GAG ATC CGC GC CAC AGC CCC AAG CCA CGA 1 G G G G G A TC CCT ACC GGC TTC GGA TTA AC CG AGC CCC AAG CCA CGA 1 G G G G G A T CC CT ACC GGC GC TTC GGA TTA AC TGO GCC CCA AAC CCC CCA 7 T T C C GCC CAC AA AGC CAG CAG CTTC GGA TTA AC TGO GCC CCA AAC CC CCA CAC CCC CAA 1 G R R R R R G Q R R R R A G C T R C R R R R A R R R R R R R R R R R R R		A	T	۷	Q	R	Y	P	P	T	H	H	G	S	Q	V	С	R
P P L L H G S L P V L D G G K A L 514 AGC AGC CAC CAC GC CTC GC CC CG GA C CTC AGC CCC TC TCC AGC AGC S S H H T A S P V N L S P F S K T 555 TCC ATC CAC CG CC CC CC GC GC CC TCT TCC CTC CC CG GC TTA S I H H G G S P G P L S V Y P P S S 616 TCC TCT TCT CTG GCG GCC GC CAC TCC AGT CTC TTC ACC TTC CCC GC CC CCC CG GAA GAC GTC TCC CCA GAC CG TCG TTC ACC TTC CCC P T P P K D V S P D F S L S T P G 718 TCC GCC GGG TG GCC AGG CAA GAC GG AT GAC AAA GAC TCC TCT CAA GAC TAC GCG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GCG GCC GCC TCA TCC TC GA CCC ATT ACC ACC TT 747 T L G G A S S S A H P I T T Y 871 CCG CCG CTG GCC GCG GCC TCA TCC TCA GCC ACT CAC CCC ATT ACC ACC TT 748 T L G G A S S S A H P I T T Y 759 CAG CTG CCA GAT GAC ATG AAC CTG GAC ACT GAC CAC CCC CATT ACC ACC TT 7 T J C G G G GC GCG GCG CTCA TCC TCA GCC CAC CAC CAC CAC CAC CAC TT 7 T J L G G A S S S A H P I T T Y 871 CCG CCC TAT CTG CCC GAG TTC AGC TT GGA CTC TCC CAC CCC ATT ACC ACC CTG P P Y V P E Y S S G L F P P S S L 222 CTG GGA GGA TCC CT ACC GCA TT GA AT CGA GCC AGC ATC 2 G G G S P T G F C G C K S R P C K A R 7 T S T E G R R E G V H C G A T S S T F 1024 CTG CGC GAG AAT GCC GAG GAT GAA CTT GC AAC CCC AAC CTT ACC CAC ACC CTG F P Y V P E Y S S G L F P P S S L 7 T T T T T T T T 7 T T T G G C GC GAG AT GCC GAG TTC GA CTT AG ACC TT CC AAC CCA ACC CTA 1 G G G S P T G F G C C K S R P C K A R 7 T CTG CGC CGA GAT GGC GAG ATC CTA ACC CTT ACC AAC CTT ACC CAC CC 7 T T T T C G CGC CAG AT AGC CAT GAG CTT ACC CAT ACC ACC ACC TT ACC CAC CC 7 T T T T T R A R R R C G T G G R C G T T G A T GC A A T S T F 1107 TAC CAC ACC ACA CC TAC CT TG CA ACC TAC TT CI AC ACC ACC TT ACC CAC ACC TT ACC CAC ACC CT TC CC C C C	463	CCG	CCT	CTG	CTG	CAC	GGA	TCT	CTG	ccc	TGG	CTG	GAT	GGC	GGC	AAA	GCC	CTG
SIA AGE AGE CAC AGE GGE GGE TEG GGE GGE GGE GGE GGE GGE G		P	P	L	L	H	G	S	L	_P	W	L	D	G	G	K	A	L
So S H H H T A S P V N L S F F S K T So S C ATC CAC GGC TG T A S P V N L S V Y P P A S So S L H H G S P G P L S V Y P P A S So C TG TG TG TG CG CGC GC GC CA TG CG TT GA CG TG CG CG S S S L A A G G H S S P F H L P T F P 667 CGC AGC CGC CG AA GAC GT TG CG CG CA GC CG TG CT CC CAC CC CG S A C S A R Q D V S P D P S L S T P G 718 TCC GGC GGG TG G CG AGG CAA GAT GAG AAA GAG TGC TC TA AG TT CAG CTG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GGT GG CG CA GG CTG GGC GGC TG CTC TG CAG CAC GG AGC AGC CTG GCA GA GAC AG AA GAG TG GAG AGA GAG TGC CAA GAT TAC GAG TG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GGT GGG GGC TCA TCC TG AGC CTC TG CGA GCC AGC AGC T T T L G G A S S S A H H P I T T Y 871 CCG CCC TAT GTG CCC GAG TAC ACC TGT GGA CTG TAG CGC AGC AGC AGC L G G S P T G P C K A S S S A H H P I T T Y 871 CCG CCC TAT GTG CCC GAG TAC ACC TGT GGA CTG TAG GCC CAAC ACC CTT AGC ACC ACC CTG GT GGG GGG TG GGA TGT AAG TGG GGC CCA AGC AGC CGA L G G S P T G P C C K S C L P P S S L 922 CTG GGA GGA TGC CCT ACC GGG TTC GGA TGT AAG TGG GGC CCA AGC AGC AGA L G G S P T G P C C K S R P K A R 973 TCC AGC AAA GAG GA GG GG TGT GA CT TGG AAC CTG TAGC TAC CGA S S T F E G R E G C V H C G A T S T F 1075 TAG CGT AGC GAA GAC GAA GGA GGC GGT TAG GT TAG AT TGG AT CGC TAC CGA AT GG 1177 ACC ACC AAC AAA GG AAG CA GGA GG TGT TAG AT GGG CGC GGA AGC AGC TG C T T T T L W R B B G T S H T T C G A T G C G A T S T F 1177 ACC ACC ACC AAG AGC AGT GGA GAC GGT ATT GGG AT GCC AGC GGA GT TG C T T T T L W R B B G T S C H T T L T T A C C G L 128 A A R R G Q B B B C C T S C C C C C G C GA AT CC C C C C C C C C C C C C C C C C	514	AGC	AGC	CAC	CAC	ACC	GCC	TCG	ccc	TGG	AAC	CTC	AGC	ccc	TTC	TCC	AAG	ACG
Solve are the first order for the first order	646	S	S ATC	E Chec		T	A	S	r	w.	N	L	S	P	F	S	K	T
616 TCC TCT TCT GC GC GC GC GC GC GC GC GC GT CC GT CT CTC TC ACC TTC CCG S S S L A A G G H S S F H L F T F F 667 CCC ACC CCC GC GC AA GAC GT TCC CA GAC CA GA CCC GC GC CCC GCC GCG S S S L A A G G T TCC CCA GAC CAC GC GC GTC TCC CA GC CCC GCA P T P P K D V S P D P S L S T P G 718 TCC GCC GC GC GC GC GC GA GAT GAA GA AA GAC GTC TC AG TAT CAG GTC S A G S A R Q D E K E C L K Y Q V 769 CAG CTC CCA GAT ACC ATC AAG GTC GAA ACG GTC CTC AAG TAT CAG GTC T T L G G A S S S A H H P I T T Y 810 CCG CCC TG GTG GG GCC TCA TCC TCA GCC CAC CAC CCC ATT ACC ACC TAT T T L G G A S S S A H H P I T T Y 811 CCG CCC TAT GTG CCC GAG TAC AACC TTC GAA CC CC ATT ACC ACC GCT GC GCA GGA TCC CT ACC GGG TCC GGA TC AGC TTC GAA CCC CAC GCC CAG GCA GCA L G G S P T G F G C X S R P K A R 973 TCC AGC ACA GAA GCC AGC GGA TCT GGA ACC TCT GCA AGC CC TT ACC GCA S S T E C R E C R E C V H C G A T S T F 1024 CTC GGC GCA GAA GGC AGG GAG TGT GTG AAC TOC GCC GAA ACC TTT ACC CCA S S T E C R E C R E C V H C G A T S T F 1024 CTC GGC GCA GAA GGC AGG GAG TGT GTG AAC TOC GCC GCA ACC TCT ACC CCA S S T E C R E C R E C C Y H C G A T S T F 1024 CTC GGC GCA GAA GGC AGG GAG GAG GAG GCA TGC ATT GC ATC CGC AAC CTC L W R R D C T C W Y L T C R A C C C G T T CC AAC CACC L W R R D C T C W Y L C C A T A C C C C G C A CT 1177 ACC ACC ACC ACA GAA GCC GCA GCT ACT TCC AAC CACC L H A A C A C A C A C A C A C C C A C C A TO TCC CC C AAC ACC C T T C C C CA CA CA AGC ACC CGA ACC CAC TAT CCC CC AAC ACC C T T C C C C C CA CA CA CA CC CA AC CC C A C C C C C C C C TTC CAA AAC G C C C T C TCC AAC CACC CT TCC AAC ACC CT ATT ACC AAA ACC ACC	505	e	T	1	11	600	101	2000	000	2001	10	100	511	TAC	201	1000	GUI	ICA C
13 S S S S S S S S S S S S S S S S S S S	616	тсс	TCT	TCT	CTG	202	222	222	CAC	тсс	ACT	сст	CAT	crc	TTC	ACC	TTC	°°°
667 CCC ACC CCC CCC AA GAC GTC TTC CA GAC CGC TTG CTC AC ACC CGG GA P T P P R D V S P P F S L S T P G 718 TCC CGC GG TGG GCC AGG CAA GAT GAC AAA GAC TGC CTC AAG TAT CAC GTG S A G S A R Q D E K E C L K Y Q V 769 CAG CTG CCA GAT AGC ATG GAG CAG GT CT CAC TTC TGG AGC AGC ATG Q L P D S H K L E T S H S R G S A 820 ACC CTG GGT GG GCC CTC TGC CAC TCC CAC TCC CGA GCC AGC ATG T T L G G A S S S A H H P I T T Y 871 CCG CCC TAT GTG CCC GAG TAC AGC CTC TGG AGC CAC CCC CAC CCC TAT T T L G G A S S S A H H P I T T Y 871 CCG CCC TA TGT GC CCG GG TTC GGA CTC TTC CAC CCC CAC GCC TA L P P S S L L P P S S L L 92 CTG GGA GGA TCC CT ACC GGG TTC GGA CTC TTC CAC CCC CAGC CGC TA C G G S P T G F G F G C K S S P F K A R 973 TCC ACC ACA GAA GGC AGC GGA TTG GTT GTG AAG TCC AGC CCC AGC CTC CCC 1075 TAG CAT AAA ATC AAT GOC GAG TGT GTG AAG TCC GGA GCA AGC TCT ACC CAA S S T E G R E C V M C G G A T S T F 1024 CTG TGG CGC GAA GAT GOT ACC GGG GCA CTA GC TTA GGC AGC GGA AGC 1075 TAG CAT AAA ATC AAT GOC CAG GCA CTA GCTT TGG AAT GOC TGC GGA ACC 1075 TAG CAT AAA ATC AAT GOC CAG GCA GTA GTT TAG AAT GOC AGC GAA CAC 128 M R R R A G Q N R R L L R T C AG CCA ACC ACC CTC TGC AAC 128 TT GGC CTC TCC AGA GGC AGC GAC ACC ATT TAC AAG CCA ACC ACC 1 T T C GCC CAC AGA GGC AGC GAC GAC TA TT TAC AGC AAA GGC AAC 128 TCT GGC CTC TCC TAG AGC AGC AGC ACC TTT TAC AAC CCC TGC AAC ACC 1 T T T C C CCC CAC AGC AGC AGC ACC TA TT TAC AGC AAA CCC ACC 1 T T T C C CCC CAC AGC GAC ACC TT TAC AAA CCC ACC ACC 1 T T C C CCC CAC AGC GAC AGC ACC CTA ATT ACC AAA ACC ACC 1 T T T C C CCC CAC AGC GCC GAC ACA TTO TTC ACC AAA ACC ACC 1 T T T C C CCC CAC AGC GCC GAC ACA TTO TTC ACC AAA ACC ACC 1 T T T C C CCC CC CAC ACC CGA CCC CT ATT ACC AAA ACC ACC ACC 1 T T C CCC CCC CCC TT CTC CCC AAC ACC CT TTC CCC AGC ACC ACC TTC TCC ACC ACC 1 T T C CCC CCC CT TAC AAG CTT CAT AAT ATT AAC AGA CCC CCC TT TTC ACC AAA ACC 1 S C L Y Y K L H N I N R P L T M K 130 TCC AAA AAG GT CT ACC CCC CT TAC CAC ACC CCC TT TCC CCC ACC A		S	s	s	L		Å	G	H	s	s	P	H	L	F	T	F	P
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	667	ccc	ACC	CCG	CCG	AAA	GAC	GTC	тсс	CCA	GAC	ccg	TCG	CTG	тсс	ACC	222	GGA
718 TCC GCC GCG TCG GCC AGG CAA GAT GAG AAA GAG TGC CTC AAG TTC GAG GTG S A G S A R Q D E K E C L K Y Q Y 769 CAG CTG CCA GAT AGC ATG AAG CTG GAG AGG TCT CAC TCT CGA GGC AGC ATG Q L P D S M K L E T S H S R G S H 820 ACC ACC CTG GGT GGG GCC TCA TCC TCA GCC CAC CCC CATT ACC ACC TAT T T L G G A S S S A H H F I T T Y 811 CCG CCC TAT GTG CC GAG TCA AGC TCT GGA CTC TC CCA CCC CAC TAT C C C TA GTG CC CTA AC TCA CC TCA GC C CAC CCC ATT ACC ACC TAT T T L G G S P T G F G C GGG TCT CAG TCT AGC CCA ACC CCC AGG CCA CC 2 C G GG GA CC CTA ACC CCT ACC GGA TCT AAG CCG AGC CCA AGC CAC L G G S P T G F G C C GA TC AGC TT ACC CCA ACC TT ACC ACC TAT 7 T L G G C R F G C C GA CT T ACC CCC AGC CCC AGG CCC CAG 2 C GTG GGC CCA AG GCA CCC CT ACC GGA TCT AAG CCG CGC GCA ACC TT ACC ACC CA 1 G G S P T G F G C C K S R P C K A R 973 TCC AGC CAA AG GCA ACC GCC GCC CCC GTT ATC AGC GCC GCA ACC TTC ACC CCA 1 G G C CA CAT GOT ACC COG CGC CCC CTT ATC AGC GCC AGC CTC CC 1 Y R R R R G Q R R C Y R C C A R C C C A T S R R R 1206 CTG TCC CCC AGC AGC GCC GCC GCC GTT ATC AGC GCC AGC ACC ACC TC 1 Y R R R R G Q R R R C Y R C C A R C C Y T R C ACC ACC ACC 1 T T T T W R R R A G T R C C A R C C C C A R C C C AT R CC 1 Y R R R R A G Q R R R C Y R C A R C C C C A AC CC ACC ACC ACC 1 Y R R R R R G Q R R R C R C C X R C TT CAC AAC CC AC ACC ACC 1 Y R R R R A G R T R C A R T R C A A T CC AAC ACC ACC ACC 1 T T T T W R R R A G AT S CC C A AC CCC AAC ACC ACC ACC 1 T T T T W R R R A G AC CCC AAC ACC CCC AAA TCC AAA AAG 1 C AG AAC GCC ACC GTC TCC GA ACC CCC AAC ACC CCC CTG ACT ATG AAG 1 C C AAA AAG GGC ATC CAG ACC CCG AAC CCG AAG ATG TCT AGC AAA AAG 1 C A A AAG GGC ATC CAG ACC CGC AAC CCG AAG ATG TCT AGC AAA AAG ACC CC CTG CCC AGC ATC CTT CCC 1 R A AAA GGC ATC CAG ACC CCG CCG CAG ACG ACC TCC TTC CCC 1 R A AAA GGC ATC CAG ACC CCG CCC CAG CAG CCC CTG CAC ATC TCT CCC 1 R A AAAG GG CAT CCAG ACC CCG CCC CAG ACG CCC CTG CAC ACT CTT CCC 1 R A AAA GGC ATC CAG ACC CCT CCCCAAC ACC CTT CCC AAC ACC TCC T		P	Т	P	P	K	D	V	S	P	D	P	S	L	S	T	P	G
S A G S A R Q D E K E C L K Y Q V 769 CAG CTG CAA CAT ACG ATG CAAG CTG CAG CAG CTG CAC TGC CAG CGC ATG CAG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GGT GGG GGC TCA TCC TCA GCC CAC CAC CCC ATT ACC ATT T L G G A S S A H H P I T T Y 871 CCG CCC TAT GTG CCC CAG TAC AGC TCT GGA CTC TTC CAC CCC AGC CGG P Y V P E Y S S G L F P P S S L 922 CTG GGA GGA TCC CCT ACC GGG TTC GGA CTC TTC CAG CCC CAG CGC GGG L G G S P T G F G C C S S R P K A R 973 TCC ACC ACA GAA GGC AGC GGC TTC GGA TGC TAC GGG AGC CTC ACC CGA S S T E G R E G V F C S S C C A T C G A GT ACC CTC ACC CGC ACC CGA C TC TGG CGC GAA GGA GGC AGC GGA TGC GGC GGA TGC GGA AGC TGT AGC AGA GGC AGC GGA AGC TGT AGC GGA AGC TGT AGC AGA GGA AGC TGT AGC GGA AGC TGT AGC AGA GGA AGC AGC AGC AGC AGC AGC AGC	718	TCC	GCC	GGG	TCG	GCC	AGG	CAA	GAT	GAG	AAA	GAG	TGC	CTC	AAG	TAT	CAG	GTG
769 CAG CTG CCA CAT AGC ATG AAG CTG GAG ACG TCT CAC TCT CCA GCA AGC ATG Q L P D S M K L E T S H S R G S M 820 ACC ACC CTG GGT GGG GGC TCA TCC TCA GCC CAC CA CTT ACC ACC TAT T T L G G A S S S A H H P I T T Y 871 CCG CCC TAT GTG CCC GAG TAC AGC TCT GGA CTC TTC CCA CCC AGT ACC CC CA P P Y V P E Y S S G L F P P S S L 922 CTG GGA GA TCC CT ACG CAT T GGA TCT AGG CCA CAC CCA AGC CTG P P Y V P E Y S S G L F P P S S L 973 TCC AGC ACA GAA GC AGC GTT GGA TGT AAG TCC AGC CCA AGC CCA C C A S S T E G G S P T G F C C K S R P K A R 7 T CG CGC CA AC CCA ACC GGA TTC TAG AT CGC TCT AGC CCA S S T E G R E C V M C G A T S T F 1024 CTC TGC GGC CGA GAT GTAC CCC GGA CT TAG AT CGC TCC GGA CTC L Y R R R B G T S H Y L C G A T S T F 1126 CTG CGC CGA GAT AGC AGC GGA GTA TTC TAG AT CGC AGC AGC AGC T T T T T T T T T T T T T T T T T T T		S	A	G	S	A	R	Q	D	E	ĸ	E	С	L	K	Y	Q	۷
Q L P D S N K L L E T S H S R G S M 20 ACC ACC CTG GG CG GG CA TCA TCC TCA TC CCA CCC CAC CCC ATT ACC ACC	769	CAG	CTG	CCA	GAT	AGC	ATG	AAG	CTG	GAG	ACG	TCT	CAC	TCT	CGA	GGC	AGC	ATG
820 ACC ACC CTC GOT GGC GCC TCA TCC TCA GCC CAC CAC CCC ATT ACC ACC TAT T T L C G G A S S S A H H P I T T T Y 871 CCG CCC TAT GTG CCC CAG TAC ACC TT GGA CTC TTC CCA CCC CAC CAGC GGC P Y V P E Y S S G L F P P S S L 922 CTG GGA GGA TCC CCT ACC GGC TTC GGA TTC TAAC TCG AGG CCC AAG GCA CGA L G G G G G C C CT ACC GGC TTC GGA TTC TAAC TGG AGG CCC AAG GCA CGA L G G G G G C C C TACC GGC TTC GGA TTC TAAC TGG GCC CCA AG GCA CGA C T C CACC ACA GGC AGG GG GG TCT GGA TGT AAC TGG GCC GCA AG CC TAC CTC L W R I D G T G C C C C A C C C C C C C C C C C C C C		Q	L	P	D	S	M	K	L	E	T	S	H	S	R	G	S	М
871 CCG CT AT GTG CCC GA SL S S S A H H F I T T Y 871 CCG CT AT GTG CCC GA TAC AGC TT GGA CT TTC CCA CAC AGC CTG P P Y V P E Y S S G L F P P S S S L 973 TCC AGC CAC AGC CCT TT GGA TCG AG CCC AAG GCA CCA L G G S P T G F G C K S R P K A R 973 TCC AGC CAC AG GCA GCT GTG AG TGG GCG GCA ACC TT ACC CCA S S T E G R E C V H C G GA CC TT AC AG CCC TT ACC CA S S T E G R E C V H C G GA CT TT CC AC TGG GCA CC TT ACC CCA 1 Y R R H C G R E C V H C G GA CC TT ACC AG CC CC CT C 1 Y R R H R G Q B R P L T C C R A C C G L I 1075 TAC CAT AGA ATC GAT GOT ACC GOG GCA TAG CTT TGG AAC TGG GCA ACC TTC CC CC 1 Y R R H R G Q B R P L T C C R A C C G I L 1075 TAC CAT AGA ATC GAT GOT AGC GOG GCA TAG CTT CG AAC TGC CG CCA CT C 1 Y R R H R G Q B R P L T C C R A C C G I R 1126 CTC TGC GCC ACC AGC AGC GCG GCG GCA CT TT C C AC ACC ACC 1 T T T T W R R B A G T T S C C A H C C C A T C C C AAT GCC T T T T W R R B A G T T S C C A B C C C TT ATC AGA GCC ACC ACC TT 127 AG ACC ACC GTC TGC AGC AGC GCG AAC CGC AAC TGT CAC ACC ACC 1 T T T W R R B A B G D P Y C R A 1288 TGT GGC CTG TAC TAC AAG CTT CAT AAT ATT AAC AGA CCC CTG ACT TATC AAG 1288 TGT GGC CTG TAC TAC CAG ACC CGG AAC CGG AAG ATG TCT AGC AAA AGC K C G I U Y K L H N I N R P L T M K 1300 TGC AAA AAG GGC ATC CGA ACC CGC AG GA GA CTG TCT AGC AAA TGG AAA AAG K C G G L Y Y K L H N S N K M S S K S K K 13100 TGC AAA AAG GGC ATC CGA CAC CGC AGC GA CAC TCC TTC CC C K K V H D A L L E D F P K S S S S F 1483 TCC GGC CTC TC TC CAG CAC CTG TC TC TC CC TAG CAC ATC TCT CCC N P A A L S R H M S S L S H I S P 1433 TCGC GCC CTC TC TC CAG CAC TG CT TC CAC AGC ACC TCC TTC CCC N P A A L S R H M S S L S H I S P 1433 TCGC CAC TCC CAG CAC CTT CCAC ACC CTT CCC CAG CAC TC TC TC CCC N P A A L S R H M L T T P T P M H P P 1433 TCCC CAC TCC TC TC CCC AGC CTT CCCCAAGCGCTGTGCATGCGAGAGCCCCCAAGCGCTGTCTCCGATTCGAGCCCCCAAGCGCTGCGAGGGACAGAGCCCCCATCCGTCTCGAATCGCCTCCAACCCTTGCATTCGCAATTCGCCAATTGCCCAATTGCCGAAATCGCGAAATCTTGCTTTTTGCAAGCGAGCCCCAAGCGCCCAAGCGCTGCCCCCACGTTCCCCCCCC	820	ACC	ACC	CTG	GGT	GGG	GCC	TCA	TCC	TCA	GCC	CAC	CAC	ccc	ATT	ACC	ACC	TAT
For every set of the order order of the order order of the order	071	T CCC	T	L	G	G		S	S	S		H	H	P	1	T	T	Y
922 CTG GGA GGA TCC CT ACC GGG TTC GGA TGT AG TCG AGG CC AAG GGA CGA L G G S P T G P C C K S R P K A R 73 TCC ACC AAA GGC AGG CGA TGT GAA TGT AGC TCC AGG CCA AG CCA CA S S T E G R C R C G V H C G A T S T P 74 CTG TGG CGC GGA AG CGA CGA TGT AGC TGC GGA GAC TT AGC CCA S S T E G G R C G R C G V H C G G A T S T P 74 CTG TGG CGC GGA AG CGA CGA CGA CT AGC TT GGA AT CGC TCC CGA S S T E G G C C G C AG AGC CGC CGA CT AGC TT GGA AT CGC CGA CA 1075 TAG CAT AAA ATC AAT GGC GGA GAG CGC CGC CT AGC CAC CAC CAC T T C C CGC CGA ACC AGGA CGA GAC CGC CGC CT AGC CAC CAC CAC T T T T T T T T T T T T T T T T T T T	8/1	CCG B		IAI	GIG	100	GAG	TAC	AGC	TCT	GGA	CTC	TTC	CCA	CCC	AGC	AGC	CTG
122 CC C	922	crc	CCA	CCA	TCC	CCT.	ACC	CCC	TTC	CCA	TOT	, L, C	TCC	ACC	ccc.		5	CCA
973 TCC AGC AC AGA GGA AGC AGG GG TGT GTG AAC TGC GCC GCA ACC TTT ACC CCA S S T E G R E G V R C G C A T S T F 1024 CTC TGG GGC GGA GAT GGT AGC GGG CAT TGG TT TGG AAT GGG TC GGG CA L W R R D G T G G V R C G G A T GG TG C L W R R D G T G G V R C G G A T G C TG C T T C AGC AGG AGC GGG CAG GAC GGC CC CTT TG AAT GGC AAG GGA AGG Y B K N R G Q B B R T L T C A G C G C A GG Y B K N R G Q B B R T L T C T A C G G A C G T T C G CG CA GC AGA AGC GGG AGG CGC CG C TG C T	722	010	<b>UGA</b>	GON	100		AUU	666	110	GON	101	AAG	106	AGG		AAG	GUA	CGA
S S T E G CR CAL AGT GOT AGT GOT AGT GOT AGT GOT TOT AGT GAL TO T F 1024 GTG TOG COG CAL CAT GOT AGT GOT GOT GOT GOT GOT GOT GOT GOT GOT G		Τ.	G	G	s	P	т	C		C	c	r	c	ъ	P	v		Ð
1024 CTC TCC CCC CCA GAT GOT ACC GOC GAC CTA CTT TCT AAT GOC TCC GGA CTC L Y R B G G T S H Y L C T TCT AAT GOC AAG GGA CTC 1075 TAG CAT AAA ATC AAT GOC CAG AAC CCC CCC CTT ATC AAG GCC AAG GGA TCC Y R C X R G G C CCA ACG AAC GCC CCC CTT ATC AAG GCC AAG GCA ACG L T A A R T A G O R L T C C A T C TC CC AAC ACC ACC L T A A R T A G O R L C C A TC TC CC CT ATC CCC ACC ACC T T T L W R R T A G C T C TC CC AAC CCC CTC TCT CAC AAC ACC C T T T L W R R W A W G D F V C R A 1228 TCT GCG CTC TCC ACG ACC CCG ACC CCC CTC TCC AAT GCC G C L Y Y K L H N I N R F L T A C C T AT GC AAA ACG C G L Y Y K L H N I N R V L T H K 1300 TCC AAA GCG ATC CTC CAC ACC CCG AAC ACC CCC ACC ACC ACC C C K K V H D A L E D F P K S S K S K K 1300 TGC AAA AG GTC CTC TCC CGA ACA CCG CTC TCC CCA AAA ACG N P A A L S R H M S S L S H I S P 1432 TTC AGC CAC TCC CAC CCC ACC CTC ACC CCC ACC AC	973	L TCC	G AGC	G ACA	S GAA	P GGC	T AGG	G GAG	F	G CTC	C AAC	K TOC	S	R	P ACC	K TCT	ACC	R
L W K K G AT CAR AT GO CAG AG CHO CHO CHO THE AL C C AG AG CHO	973	L TCC S	G AGC S	G ACA T	S GAA E	P GGC G	T AGG R	G GAG E	F TCI C	G GTC V	C AAC	K TGC C	S CCC G	R GCA	P ACC T	K TCT S	ACC T	R CCA P
1075 TAG CAT AA AT GEC CAG AAC CEG COC CIT ATC ALC SOC CAG CGA AC Y B K H B G G C B A C D B F C L K L K C K B F 1126 CTG TCG CCA CLA ACC ACA CLA CEG CAC ATC TCT CAG ACC ACC L S A A R R A C C T T T C C A B C C C C C C C C C C C C C C T T T 1177 ACC ACC ACC CTC TCG ACC CAC CAC C C A T C C C C AC T C C C C	973 1024	L TCC S CTG	G AGC S TCC	G ACA T CCC	S GAA E CGA	P GGC G GAT	T AGG R COT	G GAG E ACC	F IGI C	G GTC V CAC	C AAC N TAC	K TQC C CTT	S CCC C C TCC	R GCA A AAT	P ACC T	K ICI S ICC	A ACC T GGA	R CCA P CTC
1126 CTC TCC CCC CCC ACC ACC ACC ACC ACC ACC	973 1024	L TCC S CIG L	G AGC S TOG	G ACA T COC	S GAA E CGA R	P GGC G GAT D	T AGG R GOT	GAG E ACC T	F IGI C CCO Q	G GTG V CAC	C AAC N TAC	K TOC CTT L	S CCC CCC TCC TCC	R GCA A AAT	P ACC T COO	K ICI S ICC C	ACC T GGA G	R CCA P CTC L
1126 CHO THE CLA HEA HAR GEA HAR GEA HAR THE THE CAS ALC ACC L C A C ACC STC THE AGE AND ALC SET ANY GOE CAC COE CHT TO CAA ACC ACC T T T T L W B B A B C T F C C A F C C C T T T 1177 AGE ACC ACC GTC THE AGE AND ALC SET ANY GOE CAC COE GTC THE AT F T T L W B B A B C D P Y C A A 1288 TGT GGE CTG TAC TAC AAG CTT CAT AAT ATT AAC AGA CCC CTG ACT ATE CAG G L Y Y K L H N I N R P L T H K 1279 AAA GAA GGE ATC CAG ACC CGA AAC CGE AAG ATG TCT AGC AAA TGG AAA AAG K E G I Q T R N R K M S S K S K K 1330 TGC AAA AAG CTG CAT CAGC CGC GAG GAC CTC TCC CC AAG ACC ACC TCC TTC C K K V H D A L E D F P K S S S F 1381 AAC CGG GCC GTC TC CAG CAC CTG CAT CTA CAGC ACC TCC TTC C K K V H D A L S R H M S S L S H I S P 1433 TGC CAC TCC CAG CAC CTG CAT CAC CAC CTT C CC CTG ACC TGC CTC CCC N P A A L S R H M S S L S H I S P 1433 CGC CCT CT CT CC CAC CAT CGC ACC CTT CCA CAC ACT GT CT CCC CC F S H S S S H M L T T P T P M H P P 1534 GGT LA SAGGGAGGGCCCTGCTCCACATGCGGAGGAGTGTGCGAAGAGTCTGTGTTTTGA 6 G TAG AGAGGGAGGCCCTGCTCCACATGCGGAGGGGCGAGGGGACAGATCTGTGTTTTGG 6 G L S F G P H H P S S H V T A H 1534 GGT TAG AGAGGGAGGAGCCCTCCACATGCGGAGGAGCGGGACAGATCTGTGTTTTGA 1601 CCCCTTCTGCGAGAGCCCTGCTCCACATGCGGAGGAGCGGGACAGATCTGTGTTTTGAGGAAGACT 1739 AATCCGGAAATCTTTGCTAGAGGAGCACTTCAAGCAGCCCGAAGGGGACAGATCTGTGTTTTGAGGAAGACT 1877 TTATTGTACCTGGATATCGTTATTATTGTAATGGAATAACGGGACAGATCTGTGCTTTTGAGGAAGACT 1877 TTATTGTACCTGCAAATGCTATTAGCCATTAAGGCACTCAATGGGCGACAGATCTGTGCTTTTGAGGAAGACCCCTTATTGCAATGGCGAAATGGCGACAGATCGCGTTATAGCGATTATGCGAATTTGCGCAATTATAGCGCATTTGCGAAGCCCTAAGCGCCCAAGGGACAGATCGCGTTTTGAGGAAGACCCTTATTGCGCAATTGCGCAATTGCGCAATATGCCCAATGGCGACAGATCGTTGTGCTTTTTGAGGAAGACCCTTATTGCGCAATTGGCGCAAGGCCCAAGGGCGACAGGGCCCAAGGG	973 1024 1075	L TCC S CTC L TAC	G AGC S TOO V CAT	G ACA T COC R	S GAA E COA B ATC	P GGC G GAT D AAT	T AGG R GGT GGG	G GAG E ACC T CAC	F IGI C C C C C C C C C C C C C C C C C C	G GTG V GAC H GAC	C AAC H TAC Y CCC	K 10C CTI L	S CCC CCC TCC ATC	R GCA A AAT B AAD	P ACC T GOO A GOC	K IGI S IGC C AAG	A ACC T CCA CCA CCA	R CCA P CTC L AGG
1177 ACC ACC ACC CCC TC TCC ACC ACC ACC CCT AT CCC CCC	973 1024 1075	L TCC S CIG L TAG	G AGC S TOO V CAT B	G ACA T COC R AAA K	S GAA E COA R ATO N	P GGC G CAT D AAT	T AGG R GOT GOC GOC	GAG E ACC T CAG	F ICI C C C C C C C C C C C C C C C C C		C AAC N TAC TAC TAC		S CC C C C C C C C C C C C C C C C C C	R GCA A AAT B AAG	P ACC T COO A COC Z	K IGI S ICC C AAG	A ACC T CCA CCA CCA R	R CCA P CTC L AGG R
1217 TATCEGAAAAGGAGGCAGAAAGGAGGAGGAGGAGGAGGGGGAGAGGCGCAAAGGGAGAGGCCCAGAGGCCCCAGAGTTTTAGGAAAAGGAGCAGAGAGCAGCAAAGGCAGCAGAAGGCAGCA	973 1024 1075 1126	L TCC S CIG L TAG Y CIG	G AGC S TOG CAT H TCG	G ACA TCC AAA CCA AAA CCA	S GAA E CGA E CGA A TO B CCA	P GGC G GAT D AAT B AGG	T AGG R GGT GGG GGG AGA	G GAG E ACC T CAG GCA	F IGI C C C C C C C C C C C C C C C C C C		C AAC H TAG Y COC TCC	K TGC CTI L CL L CC	sgegegegeg.	R GCA AAT B AAT B AAC	P ACC T GOD A GOD A GOD 2 TOT	K ICI S ICC C AAG K GAG	A ACC T CGA C CGA C CGA ACC	R GCA P GTC L AGG R AGG
1228 TET GEG CTC TAC TAC AAG CTT CAT AAT ATT AAC AGA CCC CTG ACT ATG AAG G G L Y Y K L H N I N R P L T N K 179 AAA GAA GEC ATC CAG AAC CGA AA CGC AAA TCG AAA AAG K E G I Q T R N R K H S S K S K S K 130 TGC AAA AAG GTC CAT GAC GAC GCG CTG GAC ATC TCT AGC AAA TCG AAA AAG K E G I Q T R N R K H S S K S K S K S 1381 AAC CGG GCC GCT CTC TCC AGA CAC GTG CAC CCC AAC ACC CAC CTC CTC N P A A L S R H M S L S H I S P 1432 TTC AGC CAC TCC CAG CAC ATG TGA CCC CAC ACC CTC CCC CC N P A A L S R H M S L S L S H I S P 1433 TGC GAC CAC CTC CCC CAC CCC CTC CCC AGC CCC ATG CAT CCC CCC CCC N P A A L S R H M S L T T P T P M H P P 1433 TCC GGC CTC TCC CGA CCC CCC CCC CCC ACG CCC ACG CCC CCC	973 1024 1075 1126	L TCC S CTC L TAC Y CTC L	G AGC S TGC V GAT TCC S	G ACA T COC AAA CCA AAA CCA	S GAA E COA E ATC ATC A COA	P GGC G GAT D AAT B AGG R	T AGG R GGT GGC GGC GGC G AGA L	GAG E ACC T CAG GCA	F IGI G G G G G G G G G G G G G G G G G	c Brangagaga	C AAC B TAG TAG TCC B	K TGC HILLEC C	S GGC G TOC ATC C C C C C C C C	R GCA AAT B AAT B AAT B AAT B AAT B C C C A	P ACC T GOO A GOO A GOO A GOO A GOO A GOO A GOO A GOO A GOO A GOO A GOO A CC T GOO A CC T GOO A CC T GOO A CC T GOO A CC T GOO A CC T GOO A CC T GOO A CC T GOO A CC T C C C C T C C C C C C C C C C C	K IGT S TOC AAG C AC C AC	A ACC T CGA CGA CGA CGA R ACC T	R CCA P CTC L ACC R ACC T
<b>C G C C C A G Y Y K C L H H H H H H H K C A G C C A A A G C C C C A G C C C C C C C C C C</b>	973 1024 1075 1126 1177	L TCC S CTC L TAC TAC L ACC	G AGC S TGG CAT B TCG ACC *	G ACA T COC R AAA K CCA ACC	S GAA E OCIA A TO H CCA CTC	P GGC G GAT D AAT H ACC R TUC	T AGG R GGC GGC GGC GGC AGA R AGG	GAG E ACC T CAG Q GCA ACC	F ICI CCC CCC AAC B CCC AAC		C AAC TAC TAC TAC TAC TAC TAC	K Stoff I Hango Sto	S CCC G TCC ATC L CCC ATC L CCC ATC L CCC C CCC C CCC C CCC C CCC C CCC C CCC C	R GCA AAT AAT AAT AAT CCC	P ACC T GOG A BOC P C GTC GTC	K ICI S ICC C AAC K C C C C C C C C C C C C C C C	A ACC T GGA GGA CCA ACC T AAT	R GCA P CTC L AGG R ACC T GCC
1279 ARA GAA GEC ATC CEG ACC CEA ALC CEG ALG ATC TET ACC AAA TEG AAA ARG $\mathbf{x} \in \mathbf{z}$ $\mathbf{G}$ $\mathbf{i}$ $\mathbf{Q}$ $\mathbf{T}$ $\mathbf{R}$ $\mathbf{N}$ $\mathbf{R}$ $\mathbf{R}$ $\mathbf{K}$ $\mathbf{K}$ $\mathbf{S}$ $\mathbf{x}$ $\mathbf{x}$ $\mathbf{S}$ $\mathbf{x}$ $\mathbf{k}$ 1330 TEC AAA AAG GTE CAT CAC GEC CTG GAG GAC TTC CCC AAG AGC ACC TTC TTC $\mathbf{C}$ $\mathbf{k}$ $\mathbf{k}$ $\mathbf{v}$ $\mathbf{h}$ $\mathbf{D}$ $\mathbf{A}$ $\mathbf{L}$ $\mathbf{E}$ $\mathbf{D}$ $\mathbf{F}$ $\mathbf{P}$ $\mathbf{k}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{S}$ 1381 AAC CGG ECC GCT CT TC CC AGA CAC ATT CT CA TCC CTG AGC CAC ATC TTC CCC $\mathbf{N}$ $\mathbf{P}$ $\mathbf{A}$ $\mathbf{A}$ $\mathbf{L}$ $\mathbf{S}$ $\mathbf{R}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{L}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{S}$ 1422 TTC AGC CAC TCC AGC CAC ATC GTG CAC CAC CAC CTG CAC CAC ATC TCT CCC $\mathbf{F}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{L}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{P}$ $\mathbf{P}$ 1433 TCC GCC TCT CTC CC CC CTC CGC CAC CAC CAC CAC CTG CAC CAC CCC CCC $\mathbf{F}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{L}$ $\mathbf{T}$ $\mathbf{T}$ $\mathbf{P}$ $\mathbf{T}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{P}$ $\mathbf{P}$ 1433 TCC GCC TTC CTC TCC GCAC CAC CCC CAC CAC CCT GC AGC CATE CCG CCC $\mathbf{F}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{P}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{V}$ $\mathbf{T}$ $\mathbf{A}$ $\mathbf{H}$ 1534 GET TAG AGAGEGAGAGCCCTCCTCCCCACATECGFAGCAGCTCCAAAGTCGCGAAGAGTCTGTGTTCTTCCTCTCTCCTCCAC $\mathbf{S}$ $\mathbf{G}$ $\mathbf{L}$ $\mathbf{S}$ $\mathbf{F}$ $\mathbf{G}$ $\mathbf{P}$ $\mathbf{H}$ $\mathbf{H}$ $\mathbf{P}$ $\mathbf{S}$ $\mathbf{S}$ $\mathbf{H}$ $\mathbf{V}$ $\mathbf{T}$ $\mathbf{A}$ $\mathbf{H}$ 1573 GET CTACTTGCCTTTCGCATTTTCCGAGAGCAGCTCCAATCAGCGAGAGCGCCAAAGCACATCTCTGCATTTCCTCCTCCACTTCCCACCTCCCACCCCCAAGCGCCCCAAGCGCCCCAAGCAGATCTCTGCCTTTTTCCACCACCTCC 1601 CCCCTTCTACTTGCCTTTTTCGCAGCAGCCCTTCAATCTCCAAGCCCCAAGCGCCCAAGCAGATCTCGCCTTCTGCCATTCCTCCTCTCTCT	973 1024 1075 1126 1177 1228	L TCC S CTC L TAC T CTC L ACC T	GAGC S TOG S GAT S ACC S ACC S CGG	G ACA T COC R AAA K CCA ACC T CTC	S GAA E CCA E ATC M SCA A CTC L TAC	P GGC G AAT B AGG B TCC B TCC B TCC B TCC	T AGG R GGT GGG GGG GGG G AGA R AGC R AAG	GAG E ACC CAG CAG CAG CAG CAG ACC A CAG CAG CAG CAG CAG CAG E CAG E CAG E CAG E CAG E CAG E CAG E CAG E CAG E CAG E CAG CAG E CAG CAG E CAG CAG E CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG	F IGI GGG AAC B GGG AAC B GGG AAC B CAT	G GTG GAC H CHO H CAC H CHO H	C AME TAG	K TOC CII L CII L CCC CCC CCC CCC CCC CCC CC	S CCC C C C C C C C C C C C C C C C C C	R GCA AAT SAG SAG SAG SCG FCC	P ACC TO A CO P IC CT V	K IGT S IGC AAG C C C C C C	A CGA CGA CGA CGA A CC T A A T C	R CCA P CTC L ACC R ACC T GCC A
<b>K E G I Q T R N R K H S S K S K S K S K S K S K S K S K S S K S S S S S S S S S S</b>	973 1024 1075 1126 1177 1228	L TCC S CTG L TAC TAC TAC TCT C CTC C C C C C C C C C	G AGC S TGG W CAT B TCG S ACC S GGG G G	GACA T COC AAA CCA ACC T CTG L	S GAA E CGA A TC H CCA CTC L TAC Y	P GGC G GAT D AAT B AGG TGG Y TAC Y	T AGG R GOT G GOC G AGA AGC R AGC R AAG	GAG E ACC CAG GCA ACC CAG GCA ACC CTT L	F GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	G GTG V CAC H CAC H CAC A CA T GCL A A A T N	C AAC TAC TAC TAC TAC TAC TAC TAC TAC TA	K TOC CIT L C C C C C C C C C C C C C C C C C C	S GGG G G G G G G G G G G G G G G G G G	R GCA AAT B AAT B AAT B AAT B CCC P	P ACC T GOG A COC P TC GTC C TC C TC L	K IGI S IGC AAC C C C C ACT T	A GGA GGA GGA A CCA B A CCA T A A T G A A T H A T G H	R CCA P CTC L AGG R AGG T GCC A AAG K
1330 TGC AAA AAG GTG CAT GAC GCG CTG GAG GAC TTG CCC AAG ACC AGC TCC TTC C K K V H D A L E D P P K S S S F 1381 AAC CGG GCC GGT CTC TGC AGA CAC ATG TCA TGC CTG AGC CAC ATC TCT CCC N P A A L S R H M S L S L S H I S P 1432 TTC AGC CAC TCC AGC CATG CTG CAC ACG CG CAC GCC GC ATG CAT CCG CCC F S H S S H H L T T P T P M H P P 1433 TGC GGC CTC TC TC GC AGC CCT CAC ACC CG CAC CC GC CCC ATG S G L S F G P H H P S S N V T A H 1534 GGT TGC AGGGGGAGGGCCTGCTGCAGAGGGGTGGTGCTGCAGAGGGTCTCCCCGA G * 1601 CCCCTTCTCTGCGTTTTGCCAGGAGCAGTATCATCAGAGATCGTGCGAGGAGTCTCCGATGCCTTTTGAG 1670 AGGCAGAAACAAATGTTGCTTGTTTTTCAAAGGAGCTCGAGATGGTGCGGCAGAGAGCCCTCTAGTCCTC 1887 TTATTGTACCTGGATAGCGTGTAATAGCATTCAATTGCGAATGGCTGCAGTTTTGAGGATATACGGGGTNT 1896 TAAAAAAATTGCGGATGCGTGTAGGATGCGTCGAGCGCGCAAAGGGCGCGAGAGCCCCTTTTGGGCATT 1897 TTATTGTACCGGATGCGTGTAGCATTGAATGGCGCAAATGGCTTAGCTTCT 1946 AAACTAGTGCGAATAGCATTAAAAAAAAAAAAAAAAAAA	973 1024 1075 1126 1177 1228 1279	L TCC S CTC L TAC TCC L CTC L CTC L C TCC S CTC L C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S S C TCC S S S TCC S S S S	G AGC S TCG V CAT H TCG S ACC G G G G AA	GACA T COC AAA K GCA ACC T CTG L GGC	S GAA E CCA R ATC M A CCC L TAC Y ATC	P GGC G GAT D AAT H AGC B TGC TAC Y CAG	T AGG R GGT GGC G G G G G G G G G G G G G G G G	GAG EACC CAG GCA ACC CAG CTT L CGA	F IGI GOG AAC B CC C C C C C C C C C C C C C C C	G CTC V CAC H CTC A CA T CCC A AAT N CCCC	C AAC TAG TAG TOC TOC S AAT AAT I AAG	K TGC C L L L L C C C C C C C C C C C C C C	S CCC G TCC A TCC A CA CA CA CA CA CA CA CA CA CA CA CCCC	R GCA AAT B AAT B AAC B CCC P AGC	P ACC T GOG A GOC T GTC C T GTC C T G T C T G T C T G T C C C C	K IGI S IGC AAC K GAG C C C C C C C C T T C G	A GGA GGA A GGA A A C C A A T A A T G A A C 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	R CCA P CTC L ACG R ACC T CCC A AAG K AAG
C K K V H D A L E D F P K S S S F 1381 AAC CGC GCC GCC CT CTC TCC AGA CAC ATG TCA TCC GCG AGC CAC ATG TCT TCC CC N P A A L S R H M S S L S H I S F 1432 TTC AGC CAC TCC AGC CAC ATG GTG AGC ACA CCG ACG CAC CTG CAT CCG CCC F S H S S H M L T T P T P M H P P 1433 TCC GCC GCT CTC TCC GCA CCC CAC CAC CCC TCC AGC CAT CGC ACC CCC ATG S G L S F G P H H P S S H V T A M 1534 GGT TAG AGAGGCAGAGCCCTGCTCCACATGCGTGAGGAGCTGCTCCAAGCGTGTGCGAAGAGTTCCTCCGA G $\star$ 1601 CCCCTTCTACTTGCGTTTTTGCGTGTTTTTTCCAAGCAGCCCCGAAGCGGACAGATCTGTGTTTTTGGA 1670 AGCGAACAAATGTTTGCGTTGCTTTTTTTCAAAGCAGCTCGAAGTAGCGTCTCAAGCGCTCCAACCAGTG 1739 AATCGGAATCACATTGCATAGCAATCAGCATCTAAGCTAATTGTAAGAATACTGTACATTTGGGAAGAGTT 1877 TTATTGTACCTGGAAGACGTCTAAAAATAATGGCAATCAAGCGCCCGAAATGGCGACAGATTTGGCGATAGCACTCAAGCGCTCGAATGGCGACAGATTAGCGGATGATT 1876 AAAAAAATATTGCTGAAAGCATTAGCAATTAAGCAATTAAGCGCATCAAATGGCGACGCCCAAGCGCTGTGCGCACGCTGTGCTGCGTTTTGCGCTTTTAGCGAATTATGCGGATGCATTAACTTAAATGGCGAATAAGTGGCACTGAAATGGCGCCCAAGCGCCCAAGCGCTGTTTTGCGCATTTAACGAATATGCGCGTTTTGCATTTGCGATTGCAATGCGCATTAAAATAAAGAAAAGAAAG	973 1024 1075 1126 1177 1228 1279	L TCC S CTC L TAC TC L CTCC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTCC L CTCC L CTCC L CTCC L CTCC L CTCC CTCC L CTCC CTCC L CTCC L C C C C	G AGC S TOW CAT H CAT H CO S CAT H CO S CAT H CO S CAT C S CAT C S CAT C S CAT S C C C C C C C C C C C C C C C C C C	GACA TCORAAK GCAACC ACC CTG LCTG LCTG GGC G	S GAA E CCA A TC M A CTC L TAC Y ATC I	P GGC GAT D AAT B ACC R TGC R TGC Y CAG Q	T AGG R GGT GGC G G G G G G G G G G G G G G G G	G GAG E ACC T CAG GCA A CTT L CGA R	F IGI GGG AAC B GGC C AAC C AAC C AAC N	G CTC CAC CAC H CTA ACA CC A A A T CCC A A T CCC R	C AAC TAG	K TOC LI LI LICO CO C AAC N ATG N	S CCC TCC TCC TCC TCC CC CCC TCC CCC TCC CCC TCC CCC TCC CCC TCC CCC TCC CCC CCC TCC CCCC	R GCA AAT B AAT B AAT CCC P AGC S	P ACC T C C C C C C C C C C C C C C C C C	K IGT S IGC AG C C AG C C C ACT T C G S	A CGA CGA CGA CGA CGA CGA CGA CGA CGA CG	R CCA P CTC L AGG R AGG C C C C C C C C C C C C C C C C C C
1381 AAC CGC GCC GT CTC TCC AGA CAC ATG TCA TCC CTG AGC CAC ATC TCT CCC N P A A L S R H M S L S H I S P 1432 TTC AGC CAC TCC AGC CAC ATG CTG ACC ACG CCC ATG CAT CCG CCC P S H S S H M L T T P T P M H P P 1483 TCC GGC CTC TCC TTC GGA CCT CAC CAC CCT TCC AGC ATG CAT CCG CCC S G L S F G P H H P S S H V T A M 1534 GGT TAG AGACGCAGAGCCCTCCCCCACTGCGAGCGCTCTCCCAATGCATCCCCAACGCTCTCCCCCCAC G $\star$ 1600 CCCCTTCTACTTGCGTTTTTCCCAGGAGCAGTATCATGAAGCCCCGAAAGCGACAGATCTGTGTTTTTCC 1739 AATCCGGATCCCATTGGTAATAACCCATTAGTCAATATCGCAATTTCGCTGAGTCCTGAGCCCTTAGTCCTC 1808 TGAAAAAATTTCGTGAATAGCCATTGAATTAGTCAATTGATGAGCACCGCAAATGCGACAGTTTGGGGATCCCATTGCGCATTCCATTGCAGCATTGCGCATTGCGCATTCCATTGCAATGCGCCATTGCGCGCCCCCAATGCGCCCCAAGGCTTTTGAGGAATACGCGCGTTTGGGGATCCGCATTGCGCATTGCGCATTGCGCATTGCGCATTGCGCGCCCCCAAGGCTTTTAACGGCGTTTTGCGGCATTGCGCATTGCGCATTGCGCATTGCGCATTGCGCCATGTGCGCCCCAATGGCGCCCGCC	973 1024 1075 1126 1177 1228 1279 1330	L TCC S CTG L TAC TCC L CTC L CTC L CTC C L CTC C L CTC S CTG L TCC S CTG L TCC S CTG L TCC S CTG L TCC S CTG L TCC S CTG S C TCC S C TCC S C TCC S C TCC S C TCC S C TCC S C TCC S C TCC S C TCC S C TCC S S C TCC S S C TCC S S C TCC S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S S C TCC S C TCC S C TCC S S C TCC S C S	G AGC S TOW CAT B CAT B CAT B CAT B CAT B CAT B CAT B CAT B CAT B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C B CAT C C C C	G ACA T COR AA K CA T COR AA CA T COR AA CA T COR AA A CA T COR AA A CA T COR AA A CA T COR AA A CA T COR AA A CA T COR T C COR T COR T C COR T C COR T C COR T C COR T C COR T COR T C COR T C COR T C COR T C COR T C COR T C COR T C C C C C C C C C C C C C C C C C C	S GAA E COA ATC ATC TAC Y ATC I GTG	P GGC GAT D AAT B ACC TCC TCC TAC Y CAG Q CAT	T AGG R COC G ACA AGC ACA AGC R ACC T GAC	GAG EAC ACC GCA ACC CTT L CGA CTT L CGA R GCG	F IGI GGG AAC B GGC AAC C C AAC B C AAC C C T G C C C C C C C C C C C C C C C	G GTC V GAC H GAC A A T C C C H GAC A A T N C C G C C H C H C H C H C H C H C H C H	C AAC TAG TOC AAT AAT AAG K GAC	K RC CLL CLC CC CC C C C C C C C C C C C	S CCC B CCC B CCCC B CCC B CCC B CCC B CCCC B CCC B CCC B CCC B CCC B CCC B CCC B CC	R GCA AAT AAT AAT AAT CCC P AGC S AAG	P ACC TC CO CC CC CC CC CC CC CC CC CC CC CC CC	K IGI S IGC AAC C C C C C C C C C T C C C C T C C S ACC S C C C C C C C C C C C C C C	A GGA GGA CCA CCA CCA CCA CCA CCA CCA CC	R CCA F CTC L ACC R ACC T CCC A AAC T C CCC A AAG K TTC
<b>N F A A L S R H H S S L S H H F S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H S S L S H H H F F F H S S H H H F F S H H H F F S H H H F F S H H H F F S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H F S S H H H H S S H H H H H H H H H H</b>	973 1024 1075 1126 1177 1228 1279 1330	L TCC S L TAC TAC L ACC T TCC L ACC TCC L ACC TCC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC C L CTC L C CTC L C CTC L C CTC L C CTC L C CTC L C CTC L C C CTC L C C C C	G AGC S TUG CAT TCG ACC TCG GGG GAA E AAA K	G ACA T COC R AAA K CCA ACC T CTG L GGC G AAG K	S GAA E CGA A TC A TC L TAC Y ATC I GTG V	P GGC G GAT D AAT H ACC TCC TCC TCC CAG Q CAT H	T AGG R GGC G GGC G AGA R AGC R AGC T GAC D	G GAG E ACC T GCA ACC GCA ACC CAC CCA CTT L CGA R GCG A	F GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	G GTG V GAC H GAG H GAG E GAG E	C AAC TAG TAG TOC TOC TOC AAT AAT AAG K GAC D	K CC CHL CL CC CC C AAC N ATG H TTC F	S CCC TCC ATC ATC ATC ACA ACA R TCT S CCC P	R GCA AAT AAT AAT AAT AAT AAT AAT AAT AAT A	P ACC CC CC CC CC CC CC CC CC CC CC CC CC	K ICI S ICC AAG K CO C C AAG ACT ICC S AGC S	A ACC T CCA CCA ACC T AAT AAA AAA K TCC S	R CCA F CTC L ACC R ACC T CCC A AAC K AAG K TTC F
<b>F S H S S H H L T T P T P H H P P</b> <b>F S H S S H H L T T P T P H H P P</b> <b>1</b> 483 TCC GEC CTC TCC TTC GEA CCT CAC CAC CCT TCC AGC ATC GTC ACC GCC ATC <b>S G L S F G P H H P S S H V T A H</b> <b>1</b> 534 GGT TAG AGAGGCAGAGCCCTGCTCCACATGCGTGAGGAGTCTCCAAGTGTGTGCGAAGAGTTCCTCCGA <b>G *</b> <b>1</b> 601 CCCCTTCTACTTGCGTTTTTGCCTTGTTTTTTCAAAGCAGCCCCGAAGAGCGACAGATCTGTGTTTTGGAT <b>1</b> 670 AGGCAGAAAGCAAAATGTTTGCTTGCTTTTTTTCAAAGCAGCTCGAGTGTGTGCTGCGAACGAGTCCTCCAACCACTC <b>1</b> 739 AATCGGAATCCAATTGCGTATGCAATTAACCTATATTGTAAGAAATACTGTACATTTGGCGAATGAGCT <b>1</b> 870 TTATTGTACCTGGAAGCATTGCATTGCAATAACCTTAAATGGCAACAAGCGCCCAAGAGTTTTTAAGGAATATAGCGGATTCAATTGCAATAACGCATTAAATGAAATGGCCCATCAAGCGCCCCAAGGCTGTTTTAAGCAATTAGCGCATTAACTTAATTGTAAGAAATGGCCACTGAAGTGCCCCCAAGGCCCCAAGCAGTTTTTAAGCGAATTAACCGCATTAACTGAATGCGCCACAAGTGCGCCCAAGAGTTTTTGCCGCATTTAACGAATATGCGCATTAATTGCTAATTGCTGCAATCAGCGATTCAATGCGCATTAACTGAATGCGCCCCAAGCGCCCAAGCGCCCAAGCGCTTTTTGCAATTAGCGCATTATTGCCATTTGCAATGCGCATTAACTGCAATCGCCCACACTGCCCCACGCCCCAACGCCCCACGCCCCACGCTTTTTGCCCCTTTGCATTGCCAATGCGCATCAATGCGCCACGCCCAAGCGCCCCACGCCCCACGCCTTTTTGCCCCCCCC	973 1024 1075 1126 1177 1228 1279 1330 1381	L TCC S CIC L TAC TCC L CICIC L CIC CIC	G AGC S TOG TO CAT H TO S CAT H TO S CAT H CO G G G A A C C S H C C TO S H C C TO S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C C S H C S H C C S H S S H S S H S S H S S H S S H S S H S S H S S H S S H S S H S S H S H S S H S S H S S H S H S S H S S H S S S H S S H S S H S S H S S H S S H S S S H S S S S S H S S H S	G ACA T COC R AAA K CCA ACC CTG CTG GGC G GGC G AAG K GCC	S GAA E CCA A TC A TC A TC TAC Y A TC I GTG GCT	P GGC G GAT D AAT AGC TCC TCC CAT H CTC	T AGG R GGC GGC GGC GGC AGA AGC T GAC T CC	G GAG E ACC GCA ACC GCA ACC CAC GCA CTT L CGA R GCG ACC CAC GCA CAC GCA CAC CAC GCA CAC CAC	F GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	G CTC CAC H CTC ACA CC CC CC CC CC CC CC CC CC CC CC C	C AAC TAG CCC TCC AAT AAT AAG K GAC D TCA	K CC HILD LC CC C AAC N ATG H TTC F TCC	S CCC TCC TCC ATC ATC ACA ACA ACA R TCT S CCC P CTG	R GCA AAT AAT AAC BCB CCC P AGC S AAG AGC	P ACC CC CC CC CC CC CC CC CC CC CC CC CC	K ICI SCC ARC CC ACT TCG ACT TCG ACC S ACC S ACC	A ACC T CCA CCA CCA CCA CCA CCA CCA CCA CCA CCA	R CCA P CTC L ACC R CCC T CCC A ACC T CCC F CCC
1483 TCC GCC CTC TCC TTC CAC CTC CAC CCT TCC AGC ATG GTC ACC GCC ATG $S$ C L S F C P H H P S S H V T A H 1534 GCT TAC AGAGGCAGAGCCCTGCTCCACATGCGTGAGGAGCTGTCGCAGAGGGTGTCCTCCACA C * C C C C C C C C C C C C C C C C C C	973 1024 1075 1126 1177 1228 1279 1330 1381	L TCC S CTC L TAT TCC L ACT TCC L ACT TCC AAA K TCC C AAA	G AGC S TOW GAT TOW ACC GGG GAA E AAA K CCG P	G ACA T COM AAX C ACC ACC C C C C C C C C C C C C C	S GAA E CCA A TC A TC C L TAC Y A TC I GTG V GCT A	P GGC G GAT A M C C C C C C C C C C C C C C C C C C	T AGG R GGC GGC GGC GGC AGA AGC T GAC T CC S S S	G GAG E ACC T CAG Q CA A ACC T L CGA R CTT L CGA R GCG A AGA R CTT L	F GGG GGG GGG GGG GGG GGG GGG GGG GGG G	G GTG CAL H COM ACA T COM ACA T COM ACA T COM ACA T COM CAL H COM ACA T COM CAL H COM COM COM COM COM COM COM COM COM COM	C AAC TAG TC TC AAT AAT AAG K GAC D TCA S	K TOC CILL CL CC C	S CCC G TCC AT C C A C C A C A C A C A C A C A	R GCA AAT AAT AAT AAT AAT AAT AAT CCC P AGC S AAG AGC S	PACC COASCATC CTC CTC AAA AGC CAC H	K TGC G G G G G G G C C C C C C C C C C C C	A ACC T CGA C C CGA C C C C A C C C C C C C C C	R CCA P CTC L ACC R ACC T CCC R AAG K AAG K TTC F CCC P
<ul> <li>S G L S F C P H H P S H V T A H</li> <li>S G T AG GAGGCAGAGGCCTGCTCCACATGCGTGAGGAGTCTCCAAGTGTGCGAAGAGTCCTCCGA</li> <li>G *</li> <li>1601 CCCCTTCTACTTGCGTTTTTCGCAGGAGGAGGAGTATGATGAAGGCGCCGAAGCGGTCTCCAACGCTC</li> <li>1601 CCCCTTCTACTTGCGTTTTTGCTTCTTTTTTTTCAAAGCGCTCGAGGTGCTGCTGCGTCCCAACGCCT</li> <li>1601 CCCCCTTCTACTTGCGTTTTGCATTGCTTTTTTTCAAAGCGCTCGAGTGCTGTCCTCCATTCGCAACGGCTCTAAGCTGTTTTGCAATAGCCATTGAAGCCATCAAGCCCCAAGGCTCCTAGTGTCTCAACGTGCTGTCTGAACTGCCTTTGAAGAAAGCCATCAAGCCTCTAATTGTAACGGGTCTCTACGTTTGAAGGAGCGTCAAGCGCTCAAGGGCCCAAGAGCTCTAATACTGTACATTGCAACATAGCGATCAATGCGCACGAAGGGCCCAAGAGCGCCAAGAGCGTCTAAGCGATCAATGCGCAATAAGCGATCAATGCGCACTGAAGCGCCCAAGCGCCCAAGGGTTTTAAGGAATATGCGGCATTTAAGCAATAGCGATCAATGCGCACTGAATGCGCACCAATGCGCCACACGCTGTTGGTTTCCCCTTTGAATGCGAATAAGGAAGCGCCCCAAGGGCTCTGAGTGTTTGCATTGCATTGCAATAAAAAAAA</li></ul>	973 1024 1075 1126 1177 1228 1279 1330 1381 1432	L TCC S CTC L TAG T CTC L ACC T CTC C C AAA R TCC C C C C C C C C C C C C C C C C C	G AGC S TOP G AGC G G G G G G G G A C C G G G G G A C C G G G G	G ACA T CCC R AAA X CCA ACC T CTG CTG C CTG C CTG C CTG C C C C C	S GAA E CCA I ATC A CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I C C C C	P GGC G GAT U AAT NCC X TCG V CAT H CCC L AGC	T AGG R GGC GGC GGC AGA AGC AGC AAG K ACC T GAC D TCC S CAU	G GAG E ACC T CAG Q CA A ACC T L CGA R CTT L CGA R GCG A AGA R AGA	F GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	G GTE V CAC H B A GAC A A C C C C C C C C C C C C C C C	C AAC TAG TCC TCC TCC TCC AAT AAT I AAG K GAC D TCA S ACA	K TOC C C TIL C C C C C C C C C C C C C C C C C C C	S CCC G THE C C C C C C C C C C C C C C C C C C C	R GCA A AAT H AAC R CCC P AGC S AAG K AGC S CCC	P ACC T CCA A ACC T C C C C C C C C C C C	K IGT S CCC C C C C C C C C C C C C C C C C	A ACC T CGA C C CA C C C C C C C C C C C C C C	R CCA P CTC L ACC R CCC A ACC T CCC A AAG K TTC F CCC P CCC
1534 GET TAG AGAGGCAGAGGCCCTGCTCCAATGCTGAGGAGTCTCCAACTGTGCGAAGAGTTCCTCCGA G T TAG AGAGGCAGAGCCCTGCTCCAATGCAGGGCTCTCCAACTGTGCGTAAGAGTTCCTTGCGT 1610 CCCCTGCATGCCATTTGCTTTTTCCAGGAGGCGCGAGGTGGTGCTGCGTCGCAACCAGTG 1670 AGGCAGAAGCAAATGTTTGGCTTAGGAAGAGCGTCAGGGCGCGAGGTGGTGGTGGCGACGAGTG 1739 ATACGGATCGCATTTGGATAGCCATTGAGAGCGTCATAGGAATAGCTGTAGTTAGGAGGTCTGTAGTGGGGT 1808 TGAAAAAATATTGCTGGAACATTGCATTAGAAAGCCATCAAGGCGCCAAGAGTTTTAAGAGGGTCTGTAGTGGGG 1877 TTATTGTACCTGGAAGCATTGCATATAACTTATATTGTAGGAATGTGCAGTTTGGGGAGGCC 1877 TTATTGTACCTGGAAGATGCATTGGATAAACTAATGTCAGCGCGCCAAGGGTTTTGAGGAAGAT 1946 AAAGTAGGGGATAGCGATGAGGAGCAGTAATGGGGCACAATGGGGCACGAATGGGGCACGATGTGGCGTTTGTTT	973 1024 1075 1126 1177 1228 1279 1330 1381 1432	L TCC S CTC L TAG T CTC L ACC T CTC C L ACC T CTC C C C C C C C C C C C C C C C	G AGC S TOU CAT H TCC T CAT H TCC T CAT H TCC T CAT H CAT H CAT H CAT S CAT S S CAT CAT S S CAT S S CAT S S CAT S S CAT S S CAT S S CAT S S S CAT S S S CAT S S S S S S S S S S S S S S S S S S S	G ACA T CCC R AAA K CCA ACC CTG GCCA AACC GCCA AACC GCCA ACC CTCG GCCA CACC H	S GAA E CIA A TG A CIC L TAC Y A TCC S TCC	P GGC G G AAT AGC R AGC V CAT H CAC Q CAT H CTC L AGC S TTC	T AGG R GOG GOG GOG GOG GOG GOG GOG AGA R AGG R AGG T GAC D TCC S CAC H	G GAG E CAG G GCA A ACC C GCA C CAG C GCA A ACC C CAG C GCA A ACC C A C CAG C A C CAG C A C A C C A C A C C C A C C 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 A C C C A C C C A C C C A C C C C C A C 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 A C	F IGI C C C C C C C C C C C C C C C C C C	G CACC H CACC H COME L ACA T COME L ACA C CCC R CCC R CCC R CCC R CCC R CCC C CCC C CCC C C C C C C C C C C C C	C AACC H C C C C C C C C C C C C C C C C	K TGC C TL L C C C C C C C C C C C C C C C C	S CCC G C C C C C C C C C C C C C C C C C	R GCA A AAT H AAT AAT CCC P CCC P CCC P CCC P CCC S AAG S CCC P AGC S CCC P AAC	P ACC T COLE A COC C C C C C C C C C C C C C C C C C	K IGT S CC C AAG C C C C C C C C C C C C C C C	A ACC T CCA C C CCA C C CCA C C CCA C C C C C C C C CCA C	R CCA P CTC L ACC T CCC P CCC P CCC P CCC P CCC
G * 1601 CCCCTTCTACTTGCGTTTTTGCGAGGAGCAGTATCATGAAGCCCCGAAAGCGACAGATCTGTGTTTTGA 1670 AGCCAGAAAGCAAAATGTTTGCTTGTTTTTGCAAGGGAGCGCGAGGTGGTGCTGCTGCTGCTGCTGCT 1739 AATCCGGATCCATTTGTGAATAAGCCATTCAGACTCATATTGCGAATTACGGGGTCTGTGCTGATGCT 1888 TGAAAAAATTCTGCGAATAAGCCATTCAGCACCCAAGCAGCACGACTTTAAGCAATATCGGGAATATAGCGGATGTT 1986 AAAGTATGCGAGTAGCGATGAAGCCACTGAAGGACCCCAAGGGTTTTAAGGAATATAGCGGTTTTGTTG 1946 AAAGTATGGGGATAAGCAGTAAAAATAAAAAAAAGGAAGC 2015 CTTCAGTTGTTGATGCATTAAAAAAAAAAAAAAAGGAAG	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483	L TCC S CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC C C C	G AGC S TCC TCC T CAT H TCC T CAT H TCC T CAT H TCC T CAT H TCC T CAT H TCC T CAT H TCC TCC TCC TCC TCC TCC TCC TCC TCC	G ACA T CCC R AAA K CCC R ACC C C C C C C C C C C C C C C C	S GAA E CAA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I CCA I I I CCA I I I CCA I I I CCA I I I CCA I I I CCA I I I I	P GGC G G AAT AGC R AGC R TOC W TAC Y CAG Q CAT H CTC L AGC S TTC C S	T AGG R GOC G G G G G G C G C G C C C C C C C C	G GAG E ACC CAG Q GCA A ACT L CGA R CTT L CGA R CTT L CGA R CTT L CGA R CTT CAG Q Q CA C CAG Q Q CA C CAG Q Q CAG T C CAG Q Q CAG T CAG Q Q CAG Q Q CAG T C CAG Q Q CAG Q Q CAG T C CAG Q Q CAG Q Q CAG T C CAG Q Q CAG A C C CAG Q Q CAG A C C CAG Q Q CA A C C CAG Q Q CA A C C C C Q Q CA A C C C C A C C C C	F IGI C C C C C C C C C C C C C C C C C C	G CAC H CAC H CAC H CAC T CAC R AAT N CCG R CAC C CG R AAT AAT N CCG R AAT AAT N CCG C R CAC C C C C C C C C C C C C C C C C	C AACC TAC TAC TC TC TC TC TC TC TC TC TC TC TC TC TC	K TGC C TL C C C C C C C C C C C C C C C C C	S CCC G C C C C C C C C C C C C C C C C C	R GCA A AAT H AAT AAT AAT AAT AAT CCCC P AAC S CCCC P AAGC S CCCC P AAGC S CCCC AAG AAC AAC AAT AAT AAT AAT AAT AAT AAT AAT	P ACC T COLE A COC C C C C C C C C C C C C C C C C C	K IGT S CC C C C C C C C C C C C C C C C C C	A ACC T CGA C CGA CGA CGA CGA CGA CGA CGA CGA C	R CCA P CTC L ACC T CCC P CCC P CCC P ATC P CCC P ATC
1601 CCCCTTCTACTTGCGTTTTTCGCAGGAGCAGTATGATGAAGCCCGAAAGCGACAGATCTGTGTTTTGA 1670 AGCCACAAAGCAAAATGTTTGCTTCTTTTTTTGAAAGCGCGCGAGGTGGTGTCTCCCATTCGAACCACTG 1739 AATCGCGATCCCATTTGTGAAATAGCCATTGAGACTCATATTGCTACGGGCTCCTAGGGCGTCCTAGGGCGTCCTAGGCGTCCTAGGCGCTCAAGGCGCTCAAGCCATGAAGCACTAAGCTATATGTGAAAAGCAATAGCGGATCA 1877 TTATTGTACCTGGATAGCTGTAAGAAAGCCATGAAGGGACGCCAAGGCTTTAAGGAATATAGCGGATCT 1946 AAAGTATGGAGATACAGAAGCAATCAGCATGAAGGGACGCAAATGGGCACACTGTCGGTTTCGC 2015 CTTCAGTTGTTGTAGCAGCATTAAAAAAAAAAAAAGGAAG	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534	L TCC S CTC L TAC TCC L CTC L CTC L CTC L CTC L CTC C CCC T TCC C CAAA K TCC C C CCC C CTC C CTC L TCC C CTC C CTC L TCC C CTC L TCC C CTC L TCC C CTC C C CTC C C CTC C C C C	G AGC S TOG G AGC T G G G G G AAA E AAA C C G G G G C G G C G C G C G C G C	G ACA T COC A AA C C C A C C C C C C C C C C C	S GAA E CGAA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I CCA I C CCA I C C C C	P GGC G D D AAT B AGC B CAT B CAG Q CAT H CAG Q CAT H CTC L AGC C TTC C C G T TCC C T G G G T D D C AT D D C AT D D C AT D D C C T D D C AT D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C T D D C C C T D D C C C T D D C C C C	T AGG R GCC GCC GCC GCC GCC GCC CCC CCC GCC G	G GAG E CAG GCA A CTT L CGA R CTT L CGA R GCG A A CTT L CGA R CTT L CGA R CTT L CGA C CTT C CGA C C C C C C C C C C C C C C C C C	F IGI GCC GCC GCC GCC GCC GCC GCC GCC GCC	G CAC CAC B CAC B CAC B CAC CAC CAC CAC C	C AAC TAG TAG TAG TAG TAG TAG TAG TAG TAG TAG	E TGC C TTL L TGC C C C C C C C C C C C C C C C C C C	S CCC C C C C C C C C C C C C C C C C C	R GGA AAT H AAT R AAT R AC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S AAT	P ACC T C C C C C C C C C C C C C C C C C	K IGI S C C C C C C C C C C C C C C C C C C	A ACC T CCA CCA CCA CCA CCA CCA T CCA CCA	R CCA P CTC L ACC R ACC T C C C C C C C C C C C C C C C C C
1670 АСССАСАЛАССАЛАЛТСТІТІССТІСТІТІТСАЛАССАССТССАССТСССАТСССА	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534	L TCC S CTC L TAC T CTC L CTC L ACC T C TC C C C C AAAA K TGC C C C C C C C C C C C C C C C C C C	G AGC S TCG G G G G G G G G G G G G G G G G G G	G ACA T CCC R AAA CC R CC R CC R CCC CC CC CC CC CC CC CC	S GAA E COA A ATC N A COA A A COA A A COA A COA A COA A COA A A COA A A COA A A COA A A COA A A COA A A A	P GGC G T D AAT B ACC B T CA CA CA CA CA CA CA CA CA CA CA CA CA	T AGG R COT C COT C COT C C COT C C C C C C C C	G GAG E CAG GCA A CTT L CGA R GCG A A GCG A A GCG A CTT L CGA R CTT L CCA CTT L CCA CTT C CCA C CTT C CCA C C C C C C C	F IGI G G G G G G G G G G G G G G G G G	G CAC CAC B COM B ACA T COM B ACA T CCG R AAT N CCG R GAG E ATG M ACC T CCC C CCC C CCC C C C C C C C C C	C AAC TAG TAG TAG TAG TAG TAG TAG TAG TAG TAG	K TGC C TTL C TGC C C C C C C C C C C C C C C C C C	S CCC C C C C C C C C C C C C C C C C C	R GGA A AAT B AAG C C C C P AGC S C CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S AAG K AGC S CCC P AGC S A C C C C C P A C C C C C C P A C C C C	P ACC T C C C Z T C T C C Z C T C T C C C C	K IGI S C C C C C C C C C C C C C C C C C C	A ACC T CGA C CGA C CGA C CGA C CGA C CGA C C CGA C C CGA C C CGA C C CGA C C CGA C C CGA C C CGA C C CGA C C C C	R CCA P CTC L ACC R ACC T CCC A AAG K TTC P CCC P CCC P CCC P CCC C P CCC C C C C C C C C C C C C C C C C C C
1739 AATCCGGATCCCATTTGTGAATAAGCCATTCAGAGTCATATTCCCTATTTAACAGGGTCTGTAGTGCTG 1808 TGAAAAAATATTCCTGAACATTCCATATAACTTATATTGTAAGAATAACTGTACATTGGGGAGAGC 1877 TTATTGTACCTGGATAGCTGTAACAAGCGATCAAGGGACAAGGATTTAAGCAATATAGCGACAT 1946 AAAGTATGGAGATACAGAAAGCAATCGACAGAGTGTCAAGGACAACTGTCAGTTTTGTTCCC 2015 CTTCAGTTGTTTGATCGATTTAAAAAAAAAAAAAAAAAGAAAG	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534	L TCC S CTC L TAC TCC L TAC TCC L AACC T TCC C AACC N TCC C AACC N TCC S CCC S CCC S CCC C C C C C C C C C C	G AGC S TOE G AGC S CAT B TOE S CAT B CAT B CAT C S C C C C C C C C C C C C C C C C C	G ACA T COC B AAA K CCA ACC C C C C C C C C C C C C C	S GAA E CGA A CCA A A CCA A A CCA A A CCA A CCA CA	P GGC GAT B AAT B AGC B TOC B TOC B TOC C B TOC C C TOC C C TOC C C TOC C C TOC C C TOC C C T C C T C C T C C T C T	T AGG C GOT G GOT G GOT G GOT G G G G G G G G G	G GAG E CAG G G G A ACC T CAG G G G C CAG C C C C C C C C C C C C	F IGI C CCC C CCC C C C C C C C C C C C C	G CIC V CAC H CAC H CAC H CAC H CAC H CAC H CAC H CAC H CAC H CAC T CAC T CAC T CAC T T CAC H C C CAC H C C C C	C AAC H TAG T TAG T TAG T TAG T TAG T TAG T TAG T C T TAG T T C T C C C C C C C C C C C C C C C	K RGC C TIL C C C C C C C C C C C C C	S CCC C C TOPE C C C C C C C C C C C C C C C C C C C	R GCA A AAT B AAC P CCC P CCC P CCC P CCC P CCC P CCC S AAG K AGC S CCC P AGC S CCC P AGC S AAT AC AC AAT B AC AC AC AC AC AC AC AC AC AC AC AC AC	P ACC T C C C C C C C C C C C C C C C C C	K IGI S CAC C C C C C C C C C C C C C C C C C	A ACC CCA CCA CCA CCA CCA CCA CCA CCA ACC TA ACC CCA CCA	R CCA P CTC L ACG R ACG R CCC A AAG K AAG K CTC C CCC P CCC P ATG H CTCCCA CCC CCC CCC CCC C CCC C CCC C C C
1808 TGAAAAAATATTCGTCAACATTGCATATAACTTATATTGTAAGAATACTGTACATTGGGGGGGG	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534 1601	L TCC S CTC L TAG TG CTG C C AACA TG C C AAAA R TGC C C AACA N TTC F G G T G C C C C C C C C C C C C C C C C	G AGC S CAT H CAT H CAT S C CAT S CAT S CAT S CAT S CAT S CAT S CAT S CAT S CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C CAT S C C C C C C C C C C C C C C C C C C	G ACA T COC B AAA K GCC A ACG C C C C C C C C C C C C C C C	S GAA E CIA E CIA E CIA E CIA E CIC E CIC E CIC E CIC E CIC E CIA E CIC E CIA E CIC E CIA E CIC E CIA E CIC E CIA E CIC E CIA E CIC E CIA E CIC E CIC E CIA E CIC E CIC E CIA E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC E CIC CIC	P GGC G GAT B AAT B ACC B T CAC Q CAT T CAG Q CAT T CAG Q CAT T CAG Q CAT T T CAG C C T C C C C C C C C C C C C C C C C	T AGG C GOC G G G G G G G G G G G G G G G G	G GAG E CAG G G G CAG G G G C C C C C C C C C C	F IGII G G G G G G G G G G G G G G G G G	G GAC V CAC V CAC B CAC A T CAC A T CAC A A T CAC A A T CAC C T CAC C T CAC C CAC C CAC C CAC C CAC C C CAC C C CAC C C CAC C C CAC C C CAC C C C C C C C C C C C C C C C C C C C	C AACC N TAG TAG N TA TAG N TA TA T TAG N TAG N T TA TAG N TAG N T	E TGCC C C TC C C C C C C C C C C C C C C	S CCC C C TOC C C C C C C C C C C C C C C C C	R GCA A A A A A C C C C P A A C C C P A C C C P A C C C P A C C C P A C C C P A C C C P A C C P A C C P A C C P A C C P A C C P A C C P A C P A C C P A C P C P	P ACC T C C C C C C C C C C C C C C C C C	K IGI S CAG C C C C C C C C C C C C C C C C C C	A ACC CCA CCA CCA CCA CCA CCA CCA CCA ACC CCA ACC CCA ACC CCA ACC CCA ACC CCA	R CCA P CTC L ACG R C C C C C C C C C C C C C C C C C C
18// TTATTGTACCTCGATAGCTGTAAGAAGCGATGAAGGACGCCAAGAGTTTTAAGGAATATAGCGGNTTT 1946 AMAGTATGGAGATAAGGAAGAMACCGACTAAGTGTCGATATGCGAAATGGGCAACACTGTCAGTTTTGTTTCC 2015 CTTCAGTTGTTTGATGCATTTAAAAAAAAAAAAAAAAAA	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1433 1534 1601 1670 1739	L TCC S CTC L TAG T CTC L ACC C C C C C C C C C C C C C C C C	G AGC S THE G G G G G G G G G G G G G G G G G G G	G ACA T CCCC A AAA K CCA C CCC C C C C C C C C	S GAA E CIA CIA E CIA CIA CIA CIA CIA CIA CIA CIA CIA CIA	P GGC G GAT B ACC B TCC C C CAT B CAG Q CAT H CTC C CAG C CAT TCC C C CAT TCC C C C C C C C C C	T ACG R GOT G G G G G G G G G R ACA R ACG T G ACA R ACG T G G C T C C C G C C C C C C C C C C	G GAG E ACC T CAG G GCA A ACC G GCA A ACC C C C C C C A C C T C C C C A C C T C C C C C C C C C C C C C C C C C	F IGII G G G G G G G G G G G G G G G G G	G CTE V CAC V CAC B CAC B CAC B CAC B CAC B CAC B CAC CAC	C AACC N TAG TAG TAG TAG TAG TAG TAG TAG TAG TAG	E TGC C CTI L CTC C CCC GCC GCC GCC GCC GCC GCC GCC GCC	S CCC C C C C C C C C C C C C C C C C C	R GCA A A A A A C C C C C P A C C C C P A C C C P A C C C P A C C C P A C C C P A C C P A C C P A C C P A C P A C P A C P A C P A C P A C P A C P A C P A C P A C P A C P A C P A C C P A C C P A C P A C C P C P	P ACC T C C C C C C C C C C C C C C C C C	K IGI S CAC C C C C C C C C C C C C C C C C C	A ACC CCA CCA CCA CCA CCA ACC T AAT AAC T CCA CCCCA CCCC	R CCA P CTCC L ACC R ACC R CCC R ACC T CCCC A AAG K TTC F CCC CCC P CCC CCC P CCCC CCC R TTTCCA CCCC T CCCC CCC T CCCC T CCCC T CCCC T CCCC T CCCC T C CCCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCC T C CCCC T C CCC T C CCC T C CCCC T C CCC T C CCC T C CCC T C CCCC T C CCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC T C CCCC C CCCC T C CCCC C C CCCC C CCCC T C CCCC C C CCCC T C CCCC C C CCCC C CCCC C C C CCCC T C CCCC C C C C C C C C C C C C C C C C
1946 AAAGTATUGAGATACAGAAGAAGCAGTAAGTGTGATGTGCAAATGGGGCACAGTGTCAGTTTTGTTTCC 2015 CTTCAGTTGTTTGATGCATTTAAAAAAAAAAAAAAAAGGAAG	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534 1670 1739 1808	L TCC S CTC L ACC L ACC T C C C C C C C C C C C C C C C C C	G AGC S TOE S TOE S C TOE S C C TOE S C C C C C C C C C C C C C C C C C C	G ACA T CCC R AAA C C C C C C C C C C C C C C C	S GAA B CCA B CCA B CCA C C C C C C C C C C	P GGC G G T U G G C T U G G G C C T C C G G G G G G G G G G G G	T AGG R GOT G G G G G G G C R AGA R AGC T G G C C C C C C C C C C C C C C C C	G GAG E ACC T CAG Q GCA ACC Q CA ACC Q CA ACC Q CA CCT L CGA R CCT L CGA R GCG A A ACC CCA R CCG CCA CCA CCA CCA CCA CCA CCA CCA CCA	F IGI G G G G G G G G G G G G G C I C C I C C C G C I C C C G C G	G CTC V CAC H ACA A CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R CCC R C CCC C C C CCC C C C C C C C C C C C C	C AACC TAG TAG TAG TAG TAG TAG TAG ATA AATT TATT CCT TAG ACA T CATT TATT	E TOCC C C C C C C C C C C C C C C C C C	S CCC G C C C C C C C C C C C C C C C C	R GCAA A AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R AAT R R AAT R R AAT R R AAT R R AAT R R AAT R R R R	P ACC T C C C C C C C C C C C C C C C C C	K IGI S C C C C C C C C C C C C C C C C C C	A ACC T CGA CGA CGA CGA C CGA T T CC CC CC CC CC CC CC CC CC CC CC CC	R CCA P CTCC L ACC R CCC R ACC T CCC A AAG K AAG K AAG K CCC P CCC P AATG R AAG CCC T TTCA CCC A CCC T CCC AACC AAC
2015 UTTCAGTIGITIGATGCATTTAAAAAAAAAAAAAAAAAAAA	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534 1601 1670 1739 1808 1877	L TCC S CTC L CTC L CTC L CTC L C T CTC C C AAA K TCC S G G T T C S G G C C C C C C C C C C C C C C C C	G AGC S CAT T C G G G G G G G G G G G G G G G G G	G ACA T CCC Z AAAA X CC Z G CCA CCC G AAG CCC G AAG CCC L GGC C CCC C G AAG X CCC CCC CCC CCC CCC CCC CCC CCC CCC	S GAA E CUA B CUA B CUA CUA CUA CUA CUA CUA CUA CUA CUA CUA	P GGC G G M M M M M G M M M M M M M M M M	T AGG R GC GC GC GC GC GC GC CC GC GC GC GC GC	G GAG E ACC CAG G CA CA CA CA CA CA CA CA CA CA CA CA CA C	F IGI C C C C C C C C C C C C C C C C C C	G GAC B ACA. T GCC B ACA. T CCC B AAT AAT N CCC B CAC B ACC T CAC B CCC T CAC CAC T CAC CCC T CAC CCC CCC C	C AACC TAC TAC TAC TAC TAC TAC TAC TAC T	K TCC C C C C C C C C C C C C C C C C C	S CCCC G ATTCC G C C C C C C C C C C C C C C C C C	R GCA AAT R AAT R CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S CCC P AGC S AT T CCC S CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC S CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC C CCC P AGC S CCC CCC P AGC S CCC CCC P AGC AGC S CCC CCC P AGC S CCC CCC P AGC S CCC CCC P AT C AGC S CCC CCC P AT C CCC CCC P AT C ACC CCC CCC P AT C ACC CCC CCC P AT C ACC ACC CCC CCC P A C CCC CCC CCC CCC	P ACC T C C C C C C C C C C C C C C C C C	K IGI S C C C C C C C C C C C C C C C C C C	A ACC CCA GCA CCA CCA CCA CCA CCA CCA CC	R CCCA CTC CTC L ACC R CCC T CCC CCC R AAAG K AAAG K AAAG K CCCC P CCCC P CCCC P CCCC P CCCC P CCCC P CCCC TTCCCA C CCCC TTCCCA C CCCC T T CCCC T T T CCCC T T T CCCC T T T CCCC T T T T CCCCC T T T T T T T T CCCC T
	973 1024 1075 1126 1177 1228 1279 1330 1381 1432 1483 1534 1670 1739 1808	L TCCS CTC L CTC L CTC L CTC L CTC L CTC L CTC L CTC T CTC S CTC C CTC L CTC S CTC C L CTC S CTC C L C CTC L C CTC S C CTC C L C C C C C C C C C C C C C C C	G AGC S TCC S TCC S ACC S C G G G G G G G G G G G G G G G G G	G ACA T CCC R AAA CCA A CCA A CCA CCG G CCA CCC CCC C	S GAA E CCA T CCA CCA CCA CCA CCA CCA CCA T CCA CCA	P GGC G G D AAT B AGC B CAT B CAT B CAT C CAT C CAT C CAT C CAT C CAT C CAT C CAT C CAT C CAT C C C C	T ACC R C C C C C C C C C C C C C C C C C	G GAG E ACC CAC CAC CAC CAC CCA CCA CCT CCA CCT CCA CCT CCA CCA	F IGI G G G G G G G G G G G C A C C H C C G C G C G C G C G C G C G C G	G CTC V CAC H ACA ACA T CAC A CAC A CAC A CAC C CAC C CAC C CAC T CAC C C C C C C C C C C C C C C C C C C C	C AACC J TAG J TAG J TAG J TAG J TAG J AATT J AATT J AAG K GACG D TCA S ACA T CCT P GAGG G ATGA CCT T CCT TAG J COC COC	K RCC CIT L CIT L CIT CCC CCC CCC CCC CCC CCC CCC	S CCC G C C C C C C C C C C C C C C C C	R GCAA A AAT B AAT C C C C C C C C C C C C C C C C C C	P ACCC T CCC A A CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C CCC T C C CCC T C	K IGI S CCC C C C C C C C C C C C C C C C C	A ACC T CCA G CCA G CCA C CCA T T T T T T T T C C C C C C C	R CCA P CTC L ACC R CCC R CCC A AAG K C CCC A AAG K C CCC P CCC P CCC P CCC C C C C C C C

FIG. 2. Sequence of the mGATA-3 cDNA. The confirmed nucleotide sequence of cDNA clone mc5b is shown. Numbers on the left indicate nucleotides beginning at the 5' end of the insert. The shaded area corresponds to the zinc finger domain and the sequences encoding it.

and accumulates to the greatest abundance in T lymphocytes, as is the case with cGATA-3 (47). A large number of other tissues and cell types, including the murine B-cell lines 38B9 and PD31 and the myeloma MPC11 cell line, do not express detectable mGATA-3. mGATA-3 mRNA is found in murine placenta and fetal brain at an approximately threefold lower abundance than in T cells, whereas it is undetectable in RNA isolated from adult brain. Thus, the GATA-3 expression profiles in murine and chicken cells are similar in that the sites of the most abundant expression are the embryonic brain and T lymphocytes (47).

A more restricted examination was undertaken to investigate the activity and tissue specificity of the human GATA-3 mRNA. A 3.6-kb mRNA is detected in human T, but not B, lymphocytes by RNA blot analysis (Fig. 5B). hGATA-3 is also expressed in at least one human neuroblastoma cell line, SKNSH (data not shown). Thus, on the basis of examination of a limited number of human cell types, we conclude that the expression characteristics of chicken, human, and murine GATA-3 mRNAs appear to be highly conserved.

hGATA-3 and mGATA-3 trans-activate expression from a

1 CCCCCCCCCCAAGATAATTITTAAAAACCTTCCCCTTCCCTACCCTTTGCTCCCAGCCTTCCCAACCC 70 CCCACCGAAAGCAAATCATTCAACGACCCCCGACCCCCCCACGCGAGGGCCCCCCGACCTCCCAGGCG 139 GACCGCCCTCCCCTCCCCGCCGCGCGCCCCGCGCGAGAGGGCCCCGAGCACAGCCGAGGCC 207 GAG GTG ACG GCG GAC CAG CCG CGC TGG GTG AGC CAC CAC CCC GCC GTG E V T A D Q P R V V S H H H P A V 258 CTC AAC GGG CAC CCG GAC ACG CAC CCG GGC CTC AGC CAC CCT AGC L N G Q H P D T H P G L S H S Y L N G Q H F D T H F G L S H S Y 309 ATG GAC GCG GCG CAG TAC CCG CTG CCG GAG GAG GTG GAT GTG CTT TTT AAC N D A A Q Y F L P E E V D V L F N 360 ATC GAC GGT CAA GGC AAC CAC GTC CCG CCC TAC TAC GGA AAC TCG GTC AGG I D G Q G N H V P P Y G N S V K 411 GCC ACG GTG CAG AGG TAC CCT CCG ACC CAC GGG AGC CAG GTG TCC CCC A T V Q R T P P T H H G S Q V C R 462 CCG CCT CTG CTT CAT GGA TCC CTA CCC TGG CTG GAG GCC CCC AAA CCC CTG P P L L H G S L P W L D G G K A L 513 GGC AGC GAC GAC GCC TCC CCC TGC AAT CTC AGC CCC TTC TCC AAG ACG G S H H T A S P W N L S P P S K T G S H H T A S P V H L S P F S K T 564 TOC ATC CAC GGC TCC CCG GGG CCC CTC TCC CTC TAC CCC CCG CCC TCC S I H H G S P G P L S V Y P P A S 615 TCC TCC TCC TTG TCG GGG GGC CAC GCC AGC CCG CAC CTC TTC ACC TTC CCG S S S L S G G H A S P H L F T F P 666 CCC ACC CCG CCG AAG GAC GTC TCC CCG GAC CCA TCG CTG TCC ACC CCA GGC K D D s P L P T P P K D V S P D P S L S T P G 717 TCG GCC GGC TCG GCC GGC GAG GAA GAA GGC TGC CTC AAG TAC CAG GTG S A G S A R Q D E K E C L K Y Q V 768 CCC CTG CCC GAC AGC ATG AAG CTG GAG TCG TCC CAC TCC CGT GCC AGC ATG P L P D S N K L E S S H S R G S M 819 ACC GCC CTG GGT GGA GCC TCC TCG TCG ACC CAC CAC CCC ATC ACC ACC TAC **SID** ACC GCC CIC GGT GGA GCC TCC TCG TCC ACC CAC CAC CAC TAC CAC CAC TAC **T** A L G G A S S S **T** H H P I **T T T 7 70** CCG CCC TAC GTC CCC AGC TAC AGC TCC GGA CTC TTC CCC CCC AGC AGC CTC **P T V P E T S S C** L **F P P S S** L **921** CTG GCC GCC TCC CCC ACC GCC TTC GGA TCC AAG TCC AGG CCC AAG GCC CCG **L** G G S **P T** G **F** G **C X S R P K** A **R 972** TCC AGC AGA GAA GCC AGG GCC GCG TTC TGTA TCT GTC AGC TCC AGC CCA E V H G E 1023 R R D G ANN ATC AND COA R R R G 8.... 1074 CX. E K P K R R AG TGT CAG ACC ACC 2 1125 1176 GGG CTC TAC TAC AAG CTT CAC AAT ATT AAC AGA CCC CTG ACT ATG AAG 1227 N N R 
 Image: Sign of the state of C K K V H D S L E D F F K N S S F 1380 AAC CCG GCC GCC TCC TCC GAG AGC ATG TCC TCC GGG GAC ATC TCC CCC N P A A L S R H H S S L S H I S P 

FIG. 3. Sequence of the hGATA-3 cDNA. The confirmed nucleotide sequence of cDNA clone p8h is shown. Numbers on the left indicate nucleotides beginning at the 5' end of the cDNA insert. The shaded area corresponds to the zinc finger domain and the sequences encoding it.

tissue-specific human TCR  $\delta$  gene enhancer GATA-binding site. Given the degree of amino acid sequence identity between the chicken, human, and murine GATA-3 proteins, we anticipated that they should each be capable of stimulating transcription from consensus GATA-binding sites. Redondo et al. have described detailed genetic and biochemical experiments examining the human TCR  $\delta$  gene enhancer (33). Within that *cis*-regulatory domain, one footprint ( $\delta$ E4) lies over two consensus GATA-binding sites. We therefore sought to assess the ability of the GATA-3 proteins to stimulate transcription of a reporter gene containing  $\delta$ E4 to ask whether this site is capable of being regulated by the hGATA-3 protein in vivo.

For cotransfection *trans*-activation assays, the human and murine GATA-3 cDNAs were placed under the transcriptional control of the Rous sarcoma virus long terminal repeat. Reporter plasmids with GATA-binding sites upstream of the rabbit  $\beta$ -globin TATA box were cotransfected with presumptive activator plasmids into quail fibroblast (QT6) cells. We then determined whether the encoded



FIG. 4. Amino acid sequence comparison of the chicken, human, and murine GATA-3 factors. The predicted amino acid sequences of cGATA-3, hGATA-3, and mGATA-3 were aligned, and identities with the conceptually translated cGATA-3 protein (47) are shaded. Numbers on the right correspond to the position of the last amino acid on each line.

proteins (mGATA-3 or hGATA-3) could stimulate expression of the human growth hormone gene in *trans*.

The H3 $\delta$ GH and H1 $\delta$ GH reporter gene constructs contain either three or one of the  $\delta$ E4 sites of the human TCR  $\delta$  gene enhancer, respectively; each  $\delta$ E4 oligonucleotide contains



FIG. 5. Expression of mGATA-3 and hGATA-3 mRNAs. (A) RNA isolated from mouse organs, tissues, and cell lines was probed with random primer-labeled mGATA-3 cDNA clone mc5b. The RNA samples shown were isolated from the following tissues or cells: uterus (lane 1); ovary (lane 2); testes (lane 3); placenta (lane 4); liver (lane 5); heart (lane 6); fetal brain (lane 7); maternal brain (lane 8); ES cells (lane 9); 10-day-old embryoid bodies (lane 10); uninduced MEL cells (lane 11); MEL cells induced with dimethyl sulfoxide for 4 days (lane 12); BaF3 (pre-B/pre-T lymphocytes; lane 13); the B-cell lines CH27 (lane 14), 38B9 (lane 15), PD31 (lane 16), and MPC11 (lane 17); normal murine B lymphocytes (lane 18); and the T-cell lines A0DH (lane 19), BW5147.3 (lane 20), and TPc (lane 21). The position of hybridization of the upper band to the murine β-tubulin internal control is indicated by the arrow. The exposure time was 48 h. (B) Blot analysis was performed with RNA isolated from either human B-lymphoma (IM9 [lane 1] and U266 [lane 2]) or T-lymphocyte (CCRF-CEM [lane 3], CCRF-HSB-2 [lane 4], HuT78 [lane 5], and MOLT-3 [lane 6]) cell lines. The position of hybridization of radiolabeled cDNA clone p8h is indicated at the right. Hybridization to the upper band of the human β-tubulin RNA (probed with the murine  $\beta$ -tubulin cDNA clone) internal marker is indicated by the arrow in the lower panel; the size of hGATA-3 was also determined by comparison with an RNA ladder (Bethesda Research Laboratories). The exposure time was 46 h.

two GATA-binding sites in the same orientation (as encountered in the human TCR  $\delta$  enhancer [33]). The C3 $\beta$ GH reporter construct has three copies of the chicken  $\beta$ -globin enhancer footprint 4 region (which also contains two GATAbinding sites in inverted orientation [6, 11, 47]). The levels of growth hormone secreted by the transfected cells into the tissue culture medium were determined by RIA.

Both mGATA-3 and hGATA-3 are capable of stimulating transcription of the growth hormone reporter gene linked in *cis* to GATA-binding sites from each of the different reporter constructs (Fig. 6A). Although six GATA-binding sites (as found in H3 $\delta$ GH and C3 $\beta$ GH) activate transcription to a greater extent than two sites (in H1 $\delta$ GH), the number of sites alone cannot be the sole determinant of activity, since the degree of stimulation was significantly greater with the C3 $\beta$ GH reporter. The ability of the mammalian GATA-3 factors to *trans*-activate the various reporter genes was comparable to that of the chicken GATA-3 protein tested in parallel experiments (data not shown and reference 47).

In order to demonstrate that both hGATA-3 and mGATA-3 proteins were expressed at comparable levels in transfected cells, extracts were made and incubated with a radiolabeled  $\delta E4$  oligonucleotide; the resulting DNA-protein complexes were then assayed in gel mobility shift experiments. Specific complexes were formed with proteins present in extracts of cells transfected with the hGATA-3 and mGATA-3 cDNAs, but not with extracts from control transfectants (Fig. 6B). These complexes were inhibited by an excess of unlabeled  $\delta E4$  oligonucleotide, thereby demonstrating that hGATA-3 and mGATA-3 can specifically bind to the GATA sites present in the human TCR  $\delta$  gene enhancer.

#### DISCUSSION

GATA-3 is a vertebrate T-cell transcription factor. GATA-3 DNA-binding proteins are abundantly expressed in chicken, mouse, and human T-lymphocyte cell lines (Fig. 1 and data not shown). The GATA-3 protein sequences deduced from conceptual translation of the cDNA clones (Fig. 2 and 3) (47) from these three vertebrate species show a remarkable degree of amino acid sequence identity, indicating that the structure of this protein, even outside the zinc finger DNA-binding domain, has been strongly conserved through evolution (Fig. 4). With the most abundant levels of expression found in T lymphocytes and the developing brain (Fig. 5), the tissue distribution of GATA-3 mRNA expression has also been conserved between species, implying that it fulfills a critical function in those cells in which it is expressed.

The trans-activation data presented here (Fig. 6) further demonstrate that hGATA-3 and mGATA-3 can recognize and stimulate transcription from a pair of consensus WGAT AR binding sites found in the human T-cell receptor  $\delta$  gene enhancer. These binding sites lie within a previously identified human T-lymphocyte-specific footprint (TCR footprint  $\delta E4$  [33]), consistent with a central role for this factor in transcriptional regulation of the TCR  $\delta$  gene. The WGATAR consensus has also been identified within murine and human TCR  $\alpha$  and  $\beta$  gene enhancer footprints (15, 18, 41, 46), although it is not possible to determine the tissue specificity of the interactions of the presumptive GATA proteins with these regulatory sequences from the data presented in these reports. The observation that the factor is abundantly expressed in human and murine TCR  $\alpha\beta^+$  T lymphocytes (for example, in cell lines CCRF-CEM, A0DH, and Jurkat) and



FIG. 6. trans-activation by the mGATA-3 and hGATA-3 factors. (A) mGATA-3 and hGATA-3 factors activate transcription from the chicken  $\beta$ -globin enhancer or the human TCR  $\delta$  gene enhancer GATA-binding sites. mGATA-3 and hGATA-3 cDNAs placed under the transcriptional control of the Rous sarcoma virus promoter and enhancer (see Materials and Methods) were cotransfected with reporter plasmid C3BGH, H38GH, or H18GH (see text) into QT6 cells. trans-activation was calculated on the basis of human growth hormone secretion (as assayed by RIA) into the cell culture medium upon activation of the reporter gene plasmid (see Materials and Methods). Numbers given for trans activation are the results obtained after averaging at least two independent transfections. (B) mGATA-3 and hGATA-3 are expressed at similar levels in transfected cells. Extracts were prepared from the same transfected cells used for the RIA described for panel A and then examined for GATA-3 expression by gel mobility shift assays (see Materials and Methods) using 0.5 ng of the  $\delta E4$  oligonucleotide as a probe (lane 1). The extracts used were prepared either from cells transfected with mGATA-3 (lanes 2 to 4 and 9 to 11) or hGATA-3 (lanes 5 to 7 and 12 to 14) expression plasmids plus reporter genes or from cells transfected with reporter plasmids and RSV · CAT (lanes 8 and 15). The reporter genes were H36GH (lanes 2, 5, 8, and 15), C3βGH (lanes 3 and 6), and H18GH (lanes 4 and 7). Lanes 2 to 8 contained 0.5 µg of nonspecific competitor DNA in the binding reaction, while lanes 9 to 15 contained the same plus 50 ng of unlabeled  $\delta E4$  oligonucleotide. B and F indicate the positions of bound and free radiolabeled oligonucleotide, respectively.

TCR<sup>-</sup> T lymphocytes (e.g., CCRF-HSB-2, HuT78, and MOLT-3) strongly implies that the TCR  $\delta$  gene enhancer site identified here is but one element within T cells which serves as a target for regulation by this transcription factor.

Vertebrate GATA-3 factor expression in the nervous system. In addition to their expression in T cells, mGATA-3 and cGATA-3 mRNAs are also expressed in embryonic (fetal) brain, while hGATA-3 mRNA is strongly expressed in the human neuroblastoma cell line SKNSH (data not shown). While we have no indication at present as to how this factor may be utilized in the nervous system, there is precedence for tissue-specific transcription factor expression in the brain. Transcription factor Oct-2, originally thought to be exclusively expressed in B lymphocytes, has since been shown to be expressed in the brain as well (16). There are also other examples of temporally distinct developmental patterns of expression similar to that which we find for GATA-3. The *Drosophila* homeodomain gene fushi tarazu (5) plays a role in early pattern formation and in later development is expressed in the nervous system; furthermore, Pit-1, originally identified as a pituitary gland-specific regulatory factor, is expressed in the embryonic brain but not in the adult brain, much like mGATA-3 (16).

GATA gene family regulation during development. The identification of multiple members of a transcription factor family is not a unique observation. A number of other transcription factor gene families (including Oct [3, 28, 36, 39, 40], CCAAT [4, 19, 20, 26, 27, 38], and AP-1 [32, 37]) in which individual members of the family display distinct tissue and temporal expression profiles have also been identified. The cloning of hGATA-3 and mGATA-3 adds two to the growing list of transcription factors containing a  $C-X_2-C-X_{17}-C-X_2-C$  zinc finger motif. This DNA-binding domain has been evolutionarily conserved from fungi (1, 10) to humans (42, 48) and is utilized repeatedly within individual organisms as the conserved element defining the GATA transcription factor family. GATA protein families have now been identified in Xenopus laevis (47a), chickens (47), mice, and humans (references 42, 45, and 48 and this report). These vertebrate factors all bind to the WGATAR sequence by using this highly conserved, duplicated C<sub>4</sub> zinc finger structure (21, 47). As in chickens, the various human proteins that bind to the GATA consensus sequence are expressed in distinct cell types and are therefore inferred to direct tissue-specific transcription in those lineages. mGATA-1 and hGATA-1 have been shown to be expressed specifically in the erythroid, megakaryocytic, and mast cell lineages (23, 35). A distinct hGATA protein found in endothelial cells may, on the basis of its pattern of expression, be the human homolog of cGATA-2 (45, 47). hGATA-3 and mGATA-3 (identified here), like cGATA-3 (47), are most abundantly expressed in T lymphocytes.

We conclude from this study that the vertebrate GATA-3 transcription factors, having virtually the same DNA sequence recognition properties, primary amino acid sequence, and mRNA developmental expression profiles, act to stimulate specific sets of genes in T lymphocytes and the developing brain in chickens, mice, and humans. The differential expression of specific members of the vertebrate GATA gene families may, by cooperative activity with other cell-specific and/or ubiquitous factors, act to direct the tissue-specific transcription of different genes in a number of distinct and developmentally unrelated cell lineages. Functional analysis of the activities of each member of the GATA gene family of activator proteins within an appropriate in vitro or in vivo transcriptional regulatory environment should determine whether or not this expectation is correct.

#### ACKNOWLEDGMENTS

We thank F. Grosveld, D. I. H. Linzer, and V. Patel for the generous donation of murine RNA samples and S. K. Pierce and S. T. Rosen for the gifts of a variety of human and murine cell lines. We also thank M. Davis for T-cell tutelage and S. H. Orkin for continued enlightening discussions.

This work was supported by NIH grants GM 28896 and HL 24415 (J.D.E.), by an NIH NRSA predoctoral training award (GM 08061) to Northwestern University (L.J.K.), and by a postdoctoral fellow-ship from the Leukemia Society of America (M.W.L.).

## REFERENCES

- 1. Arst, H. N., Jr., B. Kudla, and N. Martinez-Rossi. 1989. Aspergillus and mouse share a new class of 'zinc finger' protein. Trends Genet. 5:291.
- Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science 196:180–183.
- Clerc, R. G., L. M. Corcoran, J. H. LeBowitz, D. Baltimore, and P. A. Sharp. 1988. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. Genes Dev. 2:1570– 1581.
- 4. Descombes, P., M. Chojkier, S. Lichtsteiner, E. Falvey, and U. Schibler. 1990. LAP, a novel member of the C/EBP gene family, encodes a liver-enriched transcriptional activator protein. Genes Dev. 4:1541-1551.
- 5. Doe, C. Q., Y. Hiromi, W. J. Gehring, and C. Goodman. 1988. Expression and function of the segmentation gene *fushi tarazu* during *Drosophila* neurogenesis. Science 239:170–175.
- 6. Emerson, B. M., J. M. Nickol, P. D. Jackson, and G. Felsenfeld. 1987. Analysis of the tissue-specific enhancer at the 3' end of the chicken adult  $\beta$ -globin gene. Proc. Natl. Acad. Sci. USA 84:4786-4790.
- 6a. Engel, J. D. Unpublished data.
- 7. Evans, T., and G. Felsenfeld. 1989. The erythroid-specific transcription factor Eryf1: a new finger protein. Cell 58:877-885.
- Evans, T., M. Reitman, and G. Felsenfeld. 1988. An erythrocytespecific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. USA 85:5976-5980.
- 9. Fried, M. G., and D. M. Crothers. 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. 9:6505-6525.
- Fu, Y.-H., and G. A. Marzluf. 1990. nit-2, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. Mol. Cell. Biol. 10:1056-1065.
- Gallarda, J. L., K. P. Foley, Z. Yang, and J. D. Engel. 1989. The β-globin stage selector element factor is erythroid-specific promoter/enhancer binding protein NF-E4. Genes Dev. 3:1845– 1859.
- Garner, M. M., and A. Revzin. 1981. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res. 9:3047– 3060.
- Giguere, V., S. M. Hollenberg, M. G. Rosenfeld, and R. M. Evans. 1986. Functional domains of the human glucocorticoid receptor. Cell 46:645-652.
- 14. Gorman, C. M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. H. Howard. 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA 79:6777-6781.
- 15. Gottschalk, L. R., and J. M. Leiden. 1990. Identification and functional characterization of the human T-cell receptor  $\beta$  gene transcriptional enhancer: common nuclear proteins interact with the transcriptional regulatory elements of the T-cell receptor  $\alpha$  and  $\beta$  genes. Mol. Cell. Biol. 10:5486–5495.
- He, X., M. N. Treacy, D. M. Simmons, H. A. Ingraham, L. W. Swanson, and M. G. Rosenfeld. 1989. Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature (London) 340:35-42.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351– 359.
- Ho, I.-C., L.-H. Yang, G. Morle, and J. M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3' of Cα in the human T-cell receptor α locus. Proc. Natl. Acad. Sci. USA

**86:**6714–6718.

- Landschulz, W. H., E. Y. Adashi, B. J. Graves, and S. L. McKnight. 1988. Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev. 2:786-800.
- Lum, L. S. Y., L. A. Sultzman, R. J. Kaufman, D. I. H. Linzer, and B. J. Wu. 1990. A cloned human CCAAT-box-binding factor stimulates transcription from the human hsp70 promoter. Mol. Cell. Biol. 10:6709–6717.
- Martin, D. I. K., and S. H. Orkin. 1990. Transcriptional activation and DNA-binding by the erythroid factor GF-1/NF-E1/Eryf1. Genes Dev. 4:1886–1898.
- 22. Martin, D. I. K., S.-F. Tsai, and S. H. Orkin. 1989. Increased  $\gamma$ -globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature (London) 338: 435-438.
- Martin, D. I. K., L. I. Zon, G. Mutter, and S. H. Orkin. 1990. Expression of an erythroid transcription factor in megakaryocytic and mast cell lineages. Nature (London) 344:444-447.
- Mignotte, V., J. F. Eleouet, N. Raich, and P.-H. Romeo. 1989. Cis- and trans-acting elements involved in the regulation of the erythroid promoter of the human porphobilinogen deaminase gene. Proc. Natl. Acad. Sci. USA 86:6548-6552.
- Moscovici, C., M. G. Moscovici, H. Jimenez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell 11:95-103.
- Mueller, C. R., P. Maire, and U. Schibler. 1990. DBP, a liver-enriched transcription activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell 61:279-291.
- Muller, M. M., S. Ruppert, W. Schaffner, and P. Matthias. 1988. A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells. Nature (London) 336:544–551.
- Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu, and H. Hamada. 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60:461-472.
- 29. Orkin, S. H. 1990. Globin gene regulation and switching: circa 1990. Cell 63:665–672.
- Perkins, N. D., R. H. Nicolas, M. A. Plumb, and G. H. Goodwin. 1989. The purification of an erythroid protein which binds to enhancer and promoter elements of haemoglobin genes. Nucleic Acids Res. 17:1299–1314.
- Plumb, M., J. Frampton, H. Wainwright, M. Walker, K. Macleod, G. Goodwin, and P. Harrison. 1989. GATAAG; a ciscontrol region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. Nucleic Acids Res. 17:73-92.
- 32. Rauscher, F. J., III, D. R. Cohen, T. Curran, T. J. Bos, P. K. Vogt, D. Bohmann, R. Tjian, and J. B. R. Franza. 1988. Fos-associated protein p39 is the product of the *jun* protooncogene. Science 240:1010–1016.
- Redondo, J. M., S. Hata, C. Brocklehurst, and M. S. Krangel. 1990. A T cell-specific transcriptional enhancer within the human T cell receptor δ locus. Science 247:1225–1229.
- 34. Riddle, R. D., M. Yamamoto, and J. D. Engel. 1989. Expression

of  $\delta$ -aminolevulinate synthase in avian cells: separate genes encode erythroid-specific and non-specific isozymes. Proc. Natl. Acad. Sci. USA **86**:792-796.

- Romeo, P.-H., M.-H. Prandini, V. Joulin, V. Mignotte, M. Prenant, W. Vainchenker, G. Marguerie, and G. Uzan. 1990. Megakaryocytic and erythrocytic lineages share specific transcription factors. Nature (London) 344:447–449.
- Rosner, M. H., M. A. Vigano, K. Ozato, P. M. Timmons, F. Poirier, P. Rigby, and L. M. Staudt. 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature (London) 345:686–692.
- Ryder, K., L. F. Lau, and D. Nathans. 1988. A gene activated by growth factors is related to the oncogene v-jun. Proc. Natl. Acad. Sci. USA 85:1487-1491.
- Santoro, C., N. Mermod, P. C. Andrews, and R. Tjian. 1988. A family of human CCAAT-box-binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature (London) 334:218–224.
- Schöler, H. R., S. Ruppert, N. Suzuki, K. Chowdhury, and P. Gruss. 1990. New type of POU domain in germ line-specific protein Oct-4. Nature (London) 344:435–439.
- 40. Sturm, R. A., G. Das, and W. Herr. 1988. The ubiquitous octamer protein Oct-1 contains a POU domain with a homeo subdomain. Genes Dev. 2:1582-1599.
- Takeda, J., A. Cheng, F. Mauxion, C. A. Nelson, R. D. Newberry, W. C. Sha, R. Sen, and D. Y. Loh. 1990. Functional analysis of the murine T-cell receptor β enhancer and characteristics of its DNA-binding proteins. Mol. Cell. Biol. 10:5027– 5035.
- Trainor, C. D., T. Evans, G. Felsenfeld, and M. S. Boguski. 1990. Structure and evolution of a human erythroid transcription factor. Nature (London) 343:92–96.
- 43. Tsai, S.-F., D. I. K. Martin, L. I. Zon, A. D. D'Andrea, G. G. Wong, and S. H. Orkin. 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature (London) 339:446–451.
- 44. Wall, L., E. deBoer, and F. Grosveld. 1988. The human β-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. Genes Dev. 2:1089–1100.
- 45. Wilson, D. B., D. M. Dorfman, and S. H. Orkin. 1990. A nonerythroid GATA-binding protein is required for function of the human preproendothelin-1 promoter in endothelial cells. Mol. Cell. Biol. 10:4854-4862.
- 46. Winoto, A., and D. Baltimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3' end of the T cell receptor α locus. EMBO J. 8:729-733.
- Yamamoto, M., L. J. Ko, M. W. Leonard, H. Beug, S. H. Orkin, and J. D. Engel. 1990. Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. Genes Dev. 4:1650-1662.
- 47a.Zon, L., and S. Orkin. Personal communication.
- Zon, L. I., S.-F. Tsai, S. Burgess, P. Matsudaira, G. A. P. Bruns, and S. H. Orkin. 1990. The major human erythroid DNAbinding protein (GF-1): primary sequence and localization of the gene to the X chromosome. Proc. Natl. Acad. Sci. USA 87:668-672.