J. Biochem. 126, 1033-1042 (1999)

Molecular Cloning and Splicing Isoforms of Mouse p144, a Homologue of CA150[1]

Midori Shimada,* Masashi Saito,* Tomoya Katakai,† Akira Shimizu,† Tohru Ichimura,‡ Saburo Omata,*,‡ and Tsuneyoshi Horigome*,‡,[2]

*Course of Advanced Material Science, Graduate School of Science and Technology, Niigata University, 2-Igarashi, Niigata 950-2181; †Center for Molecular Biology and Genetics, Kyoto University, Syogoin Kawaharacho-53, Sakyo, Kyoto 606-8507; and ‡Department of Biochemistry, Faculty of Science, Niigata University, 2-Igarashi, Niigata 950-2181

Received July 27, 1999; accepted September 20, 1999

We previously characterized p144 bearing *N*-acetylglucosamine residues in a rat liver nuclear matrix fraction. Based on partial amino acid sequences of rat p144, mouse p144 cDNA was cloned and sequenced, and its amino acid sequence was predicted. The sequence revealed that p144 is a rat homologue of CA150, which is a transcription factor involved in Tat-activated human immunodeficiency virus type 1 transcription. The reported human CA150 consists of 1098 amino acids and has a leucine zipper-like motif in its carboxyl-region. However, a clone of mouse p144 cDNA encoded a CA150 consisting of 1,034 amino acids. The mouse CA150 was shorter by 64 amino acids than hitherto known human CA150 and lacked the leucine zipper-like motif. We designated the longer and shorter CA150 species as CA150a and CA150b, respectively. The partial nucleotide sequences of other mouse p144 cDNA clones were examined and it was found that some clones encode CA150a having a leucine zipper-like motif. It was suggested that CA150a and CA150b are splicing isoforms. All rat and mouse tissues examined contained transcripts for both CA150a and CA150b. Both transcripts were detected in human blood and Jurkat cells as well as mouse CD4$^+$ T-cells, which are the HIV-1-sensitive counterpart in humans.

Key words: CA150, HIV-1, p144, splicing isoform, Tat-activated transcription.

Human immunodeficiency virus type 1 (HIV-1), the etiologic agent of AIDS, is a complex retrovirus (*1, 2*). The HIV-1 promoter is located in the 5′ long terminal repeat (LTR) and contains a number of regulatory elements important for RNA polymerase II transcription. The sites for several cellular transcription factors are located upstream of the start site, including those for NF-κB, Sp1 and TBP (*1*). These cellular factors help control the rate of transcription initiation from the integrated provirus, and their abundance in different cell types or at different times likely determines whether a provirus is quiescent or actively replicating. Despite the importance of these factors, transcription complexes initiated at the HIV-1 promoter are rather inefficient at elongation and require the viral protein, Tat, to enhance the processivity of transcribing polymerases. Under some conditions, Tat may also enhance the rate of transcription initiation. Tat increases the production of viral mRNAs about 100-fold and consequently is essential for viral replication.

Unlike typical transcriptional activators, Tat binds not to a DNA site but rather to an RNA hairpin known as TAR (*trans*-activating response element), located at the 5′ ends of nascent viral transcripts (*1, 2*). The Tat-TAR complex is the master switch that controls transcription from the HIV-1 LTR. Extensive studies have suggested the involvement of cellular coactivators in Tat-mediated activation (*3-9*). The occurrence of interactions between Tat and general transcription factors (GTFs) (*10, 11*), and as well as between Tat and RNA polymerase II (*12, 13*) has been suggested. Zhou and Sharp (*14*) described the characterization and cloning of Tat-SF1, a human protein that may be involved in Tat *trans* activation.

Suñé *et al.* developed a functional assay based on transcriptional activation *in vitro* to identify cellular factors required for Tat-mediated transcription activation and showed that a factor was depleted by passage through a Tat affinity column (*8*). Then, they recently purified and molecular-cloned the CA150 factor, a nuclear protein that is associated with the human RNA polymerase II holoenzyme and is involved in Tat-dependent HIV-1 transcriptional activation (*15*). Immunodepletion of CA150 abolished Tat *trans* activation *in vitro* (*15*). Moreover, overexpression of a mutant CA150 protein specifically and dra-

[1] This work was supported by Grants-in-Aid for Scienticic Research on Priority Areas from the Ministry of Education, Science, Sport and Culture of Japan, and a grant from the Biodesign Research Project. Phone: +81-25-262-6160, Fax: +81-25-262-6165, E-mail: thori@sc.niigata-u.ac.jp

[2] To whom correspondence should be addressed at: Department of Biochemistry, Faculty of Science, Niigata University, 2-Igarashi, Niigata 950-2181.

Abbreviations: EST, expressed sequence tag; GTF, general transcription factor; HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; Lys-C, lysylendopeptidase; PKA, protein kinase A; PMSF, phenylmethanesulfonyl fluoride; RT-PCR, reversed transcription-polymerase chain reaction; TAR, *trans*-activating response element; WGA, wheat germ agglutinin.

matically decreased Tat-mediated activation of the HIV-1 promoter *in vivo*, strongly suggesting a role for CA150 in HIV-1 gene regulation (*15*). Immunoprecipitation experiments demonstrated that both CA150 and Tat associate with the RNA polymerase II holoenzyme. They suggested that the Tat action is transduced *via* an RNA polymerase II holoenzyme that contains CA150 (*15*). Other cellular factors that interact with Tat have been identified and characterized in some detail (*5, 16-20*), but none of these proteins has been conclusively shown to play a role in the Tat function.

On the other hand the presence of a family of proteins with single *N*-acetylglucosamine residues *O*-linked to serine and threonine residues is known. These proteins are abundant in the nucleus (*21*). Important nuclear proteins such as nucleoporins p62, p60, and p54 (*22*), RNA polymerase II (*23*), and transcription factors Sp1 (*24*) and HNF1 (*25*) have been shown to be *O*-linked *N*-acetylglucosamine (*O*-GlcNAc)-bearing proteins. Thus, we have interest in the nuclear glycoprotein family. We have isolated *O*-GlcNAc proteins from a rat liver nuclear matrix fraction with wheat germ agglutinin-Sepharose (WGA-Sepharose) and characterized it (*26*). This fraction contained many *O*-GlcNAc-proteins, such as p144 (*27*), p39 (*28*), and a nucleoporin complex comprising p62, p60, and p54 (*26*). Non-glycosylated proteins, such as karyopherin β (*29*), and RuvB DNA helicase-like proteins p50 and p47 (*30, 31*) were also bound indirectly to WGA-Sepharose, purified, and characterized.

In this study, the partial amino acid sequences of p144 were analyzed to extend the structural analysis and it was suggested that p144 is a rat homologue of CA150. Then, we cloned cDNA for mRNA of mouse p144, analyzed the nucleotide sequence, and deduced the amino acid sequence. The obtained sequences revealed that p144 is a rat homologue of CA150 and that mouse liver contains two kinds of transcripts for CA150: one encoding a longer CA150 similar to human CA150 previously reported (*15*), and a shorter one in the carboxyl-terminal region. It was suggested that these two kinds of mRNAs for CA150 are generated through alternative splicing. Moreover, very interestingly, it was shown that the shorter CA150 is also expressed in human cells and mouse $CD4^+$ T-cell clones.

EXPERIMENTAL PROCEDURES

Buffers and Solvents—Buffer A: 10 mM Tris-HCl buffer, pH 7.4, at 4°C, containing 0.2 mM $MgCl_2$; buffer B: 50 mM Tris-HCl buffer, pH 7.2, at 4°C, containing 0.1 mM $CaCl_2$, 500 mM NaCl, and 0.2 mM PMSF; buffer C: 50 mM Tris-HCl buffer, pH 7.2, containing 0.1 mM $CaCl_2$, 1% Triton X-100 and 0.2 mM PMSF; solvent A: 60% (v/v) formic acid; and solvent B: 56% (v/v) isopropanol, 10% (v/v) acetonitrile, and 20.4% (v/v) formic acid.

Preparation of a Rat Liver Nuclear Matrix Fraction—Rat liver nuclei were isolated from fasting rats by the established method (*32*), with the exception that all buffers were supplemented with proteinase inhibitors: 1 mM PMSF, 2 mM benzamidine, 10 μg/ml of leupeptin, and 5 μg/ml each of antipain, chymostatin and pepstatin A. The "nuclear matrix fraction" was prepared from the nuclei as described (*26*). Briefly, isolated nuclei were suspended in 50 mM Tris-HCl buffer, pH 7.4, containing 250 mM sucrose, 5 mM $MgSO_4$, and proteinase inhibitors. The suspension was treated with DNase I and RNase A (final 125 μg/ml of each enzyme) at 4°C for 2 h, and then centrifuged at $800\times g$ for 15 min. The nuclear matrix fraction was obtained as a pellet. This fraction contained nuclear membrane proteins and insoluble nuclear matrix proteins.

Preparation of a WGA-Bound Fraction—The nuclear matrix fraction suspended in buffer A containing 500 mM NaCl, 4% 2-mercaptoethanol, and proteinase inhibitors was incubated at 4°C for 30 min, and then centrifuged at $10,000\times g$ for 15 min. The supernatant was designated as the "salt extract fraction" of the nuclear matrix. This fraction was dialyzed against buffer B and then applied to a WGA-Sepharose column (Seikagaku, Tokyo) equilibrated with buffer B. The column was washed with buffer B and then buffer C containing 1 M NaCl. Proteins bound to the column were eluted with buffer B containing 500 mM GlcNAc. The eluate thus obtained was designated as the "WGA-bound fraction" of the nuclear matrix salt extract.

Purification and Partial Amino Acid Sequencing of Rat p144—Reversed-phase HPLC in 60% formic acid was carried out as described previously (*30*). The WGA-bound fraction was added to a 1/9 volume of 10% SDS and a 1/20 volume of 2-mercaptoethanol, followed by incubation at 4°C for 30 min. After that, the solution was mixed with 1.5 volumes of formic acid, diluted with an equal volume of solvent A, and then centrifuged at $6,500\times g$ for 10 min. The supernatant was applied to a Poros 10 R1 column (4×80 mm) (PerSeptive Biosystems, USA), and eluted with a 60-min linear gradient of 0 to 65% (v/v) solvent B at 0.5 ml/min. Then the concentration of solvent B was brought to 100% in 1 min and kept at 100% for 15 min. The fractions containing rat p144 were combined and then subjected to 8% SDS-PAGE. For internal amino acid sequences, the gel was stained with Coomassie Brilliant Blue R-250, and then gel slices containing rat p144 (86 μg) were subjected to lysylendopeptidase digestion as described previously (*33*). The slices containing peptides were centrifuged after homogenization. Then the supernatants were collected, concentrated by lyophilization, and applied to a reversed-phase HPLC equipped with a Silica-base C8 column ($4.6\times$ 250 mm, Capcel Pak C8 column; Shiseido, Tokyo). Peptides were eluted with a linear gradient of 5-75% acetonitrile containing 0.1% trifluoroacetic acid at 0.5 ml/min. The sequences of the isolated peptides were determined with a Protein Sequencer 470A (Applied Biosystems, USA).

cDNA Cloning and Sequencing of Mouse p144—A search of EST databases revealed that several partial amino acid sequences of rat p144 matching parts of those deduced from mouse, rat, and human EST clones seemed to be derived from transcription factor CA150. Clones containing the 3′-terminal region of CA150 were found in rat and mouse EST databases. However, clones containing the 5′-terminal region were only found in mouse databases, *i.e.* not in the rat EST ones. Then, based on two mouse EST sequences (GenBank accession numbers AA530507 and AA1556203) in these EST clones, two primer oligonucleotides were synthesized for cloning of the entire mouse p144 cDNA: N, 5′-TGTAATGGCGGAGCATGGCG-3′, as a 5′ primer; and B, 5′-GCTTCTCCTTCTGTTNGTCAG-3′, as a 3′ primer (B is complementary). Then the reverse transcription-polymerase chain reaction (RT-PCR) was carried

out with mouse liver total RNA using these oligonucleotide primers and LA Taq (Takara Shuzo, Kyoto) DNA polymerase. The 3.5 kb PCR product was cloned into the Bluescript II SK(−) plasmid vector, and its sequence was determined by the dideoxy chain termination method (*34*).

PCR Conditions—Total cDNA was prepared using SuperScript II reverse transcriptase (Gibco BRL, USA) according to the manufacturer's instructions. cDNA for mouse p144 was amplified from mouse liver total cDNA using the N and B primers. PCR was performed using LA Taq (Takara Shuzo, Japan). The PCR conditions were 95°C (1 min) [98°C (10 s), 62°C (1 min), 72°C (3.5 min)] × 30 and 72°C (7 min), 4°C, with a GeneAmp PCR System 2400 thermal cycler (Perkin Elmer, USA). For analysis of the presence of CA150a and CA150b cDNAs, the PCR was performed with LA Taq, and the conditions were 95°C (1 min) [95°C (30 s), 62°C (1 min), 72°C (30 s)] × 30, and 72°C (7 min), 4°C. The primers used for this PCR were: S, 5′-GCCAAAGCTGACTTCAGGAC-3′, as a 5′ primer, and B, as a 3′ primer. The thus obtained PCR products were analyzed by 1.5% agarose gel electrophoresis.

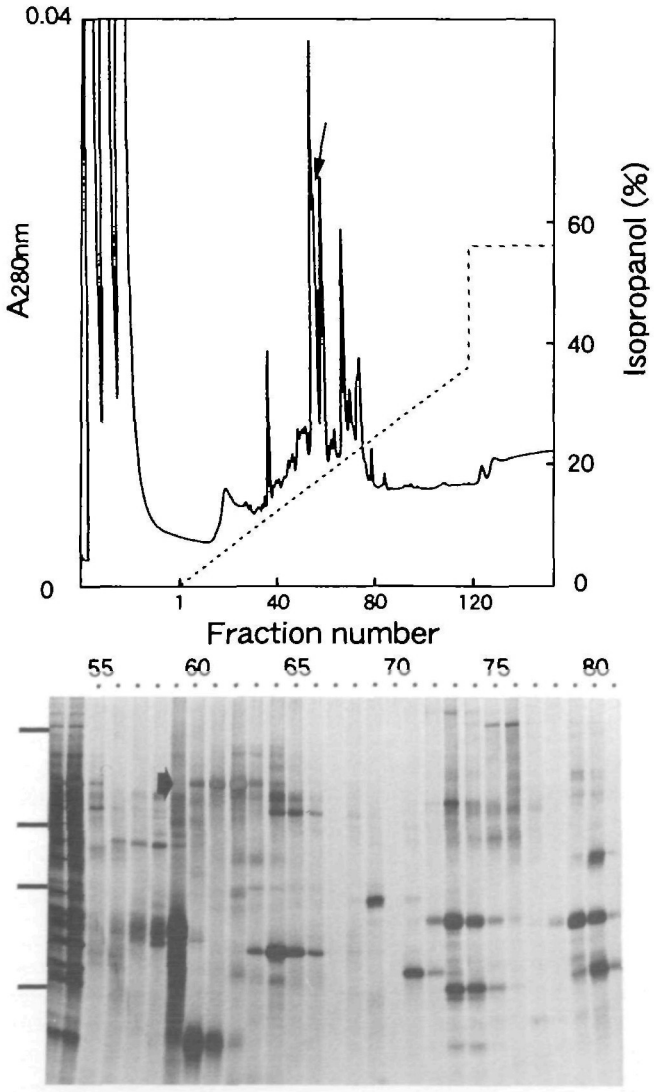


Fig. 1. **Purification of p144 from the WGA-bound fraction of a rat liver nuclear matrix extract.** The WGA-bound fraction of a nuclear matrix salt extract, 266 μg, was treated with formic acid, and then applied to a Poros 10R1 column and eluted with a 60-min linear gradient of 0 to 56% (v/v) isopropanol. The eluate was fractionated as indicated in the elution profile. Aliquots of the fractions were analyzed by SDS-PAGE and silver staining. The bars at the left of the gel indicate the positions of marker proteins of 200 k, 97.4 k, 66.3 k, and 43 k, from top to bottom. Arrows indicate the positions of rat p144.

Phosphorylation of Rat p144—Calmodulin-dependent kinase II was purified from a bovine brain cytosol fraction (*35*). Protein kinase A (PKA) was purchased from Sigma Chemical. Phosphorylation of proteins in a WGA-bound fraction (4 μg) by calmodulin-dependent kinase II was carried out in a reaction mixture comprising 50 mM Hepes, pH 7.6, 5 mM $Mg(CH_3COO)_2$, 0.1 mM $CaCl_2$, 0.05 mM ATP, 10 μCi of [γ-^{32}P]ATP, 0.5 μg of calmodulin, and 0.7 μg of calmodulin-dependent kinase II, in a final volume of 80 μl. The reaction was performed at 30°C for 20 min and terminated by the addition of 1 mM ATP. After the addition of 8 μg of carrier ovalbumin, protein in the reaction mixture was precipitated with a methanol-chloroform mixture and then analyzed by SDS-PAGE. The gel was stained with Coomassie Brilliant Blue R-250 and then subjected to autoradiography. Phosphorylation of proteins in the WGA-bound fraction (4 μg) with PKA was carried out as in the case of calmodulin-dependent kinase II except that the incubation mixture comprised 50 mM Hepes, pH 7.6, 5 mM $Mg(CH_3COO)_2$, 0.1 mM $CaCl_2$, 0.05 mM ATP, 10 μCi of [γ-^{32}P]ATP, 2 mM dithiothreitol (DTT), and 0.5 μg of PKA.

RESULTS

Purification and Partial Amino Acid Sequencing of Rat p144—A nuclear membrane and matrix fraction prepared from purified rat liver nuclei by treatment with DNase I and RNase A was extracted with 500 mM NaCl. The salt extract, which contained peripheral membrane proteins of nuclear envelopes, and proteins of the nuclear pore complex and nuclear matrix, was applied to a WGA-Sepharose column and eluted with a buffer containing 500 mM GlcNAc. The thus obtained WGA-bound fraction was separated by reversed-phase HPLC in 60% formic acid. An aliquot of each fraction was analyzed by SDS-PAGE (Fig. 1). Then, the remaining fractions 61 and 62, which contained the 144 k protein (p144), were combined and electrophoresed. The protein band of p144 on the gel was excised and digested with lysylendopeptidase. The peptides of p144 generated were purified by C8 reversed-phase HPLC (data not shown), and the amino acid sequences of the peptides were determined. The thus obtained seven partial amino acid sequences are shown in Table I.

Molecular Cloning of Mouse p144 cDNA—Searches of amino acid databases showed that a recently reported human transcription factor, CA150, contains amino acid sequences similar to those of peptides derived from rat p144. We also searched nucleic acid databases and found that seven partial amino acid sequences of rat p144 match parts of the deduced amino acid sequences of several clones of "mouse, rat, and human expressed sequence tags (EST)." In these clones, mouse but not rat ESTs containing nucleotide sequences corresponding to the 5′- and 3′-end noncoding regions of CA150 cDNA were found. Then we performed cloning of cDNA of mouse p144 by the PCR

method to compare the amino acid sequences of human CA150, mouse p144, and rat p144. Based on two mouse-EST sequences (accession nunbers: AA530507 and AA-1556203), two primer oligonucleotides were synthesized for PCR. Using these primers, several PCR conditions were optimized. Then, a nucleotide fragment of the expected length, about 3.5 kbp, was obtained by means of an RT-PCR involving mouse liver total RNA. The fragment was cloned into the Bluescript II SK(−) plasmid vector and the restriction enzyme fragments were subcloned into the same vector for sequencing (Fig. 2). The nucleotide sequences of the subcloned fragments were determined by the dideoxy chain termination method, and these sequences were combined, as shown in Fig. 3. This clone contained an ORF encoding a M_r 115,934 polypeptide comprising 1034 amino acid residues.

Homology searches of databanks revealed that the cloned mouse p144 is 97% identical in amino acids and 90% in nucleotides to the human transcription factor, CA150 (Fig. 4). From these results we concluded that the cDNA encodes a mouse CA150. We designated the CA150 as CA150b for the reasons given later. The amino acid sequences of rat p144 peptides 1 to 5 shown in Table I correspond to those of mouse CA150b residues 401-417, 690-702, 799-807, 907-926, and 959-974 with the 4 exceptions of 401, 417, 690, and 920 (Fig. 3). The amino acid sequences of peptides 6 and 7 could be explained as mixtures of peptides 6a and 6b, and 7a and 7b, respectively, by comparison with that of mouse CA150 with the exceptions of positions 152 or 657, 161 and 662 (Fig. 3). From these results we concluded that rat p144 is a homologue of mouse CA150.

TABLE I. **Partial amino acid sequences of peptides obtained on lysylendopeptidase digestion of rat p144.** A lysylendopeptidase-digest of rat p144 was subjected to reversed-phase HPLC and the isolated peptides were sequenced with an amino acid sequencer. Peptides 6 and 7 were each a mixture of two kinds of peptides. The two kinds of amino acids detected are shown in parentheses for each step.

Peptide 1:	AGVLPGMAPPIVPMIHN
Peptide 2:	LVFDPRYLLLNPK
Peptide 3:	SDFFELLSN
Peptide 4:	ALLSDMVRSSDVSCSDTRRT
Peptide 5:	KREHFRQLLDETSAI
Peptides 6a/6b:	(A,V)(A,Y)(Y,R)(E,Y)(R,N)(A,P)(R,I)(V,T)(R,P)(E,L)(S,T)(A,X)(R,X)(M,T)(K,D)
Peptides 7a/7b:	(L,D)(R,H)(H,R)(P,X)(A,E)(P,S)(G,T)(M,S)(L,X)(L,X)(E,I)(R,Q)(E,K)

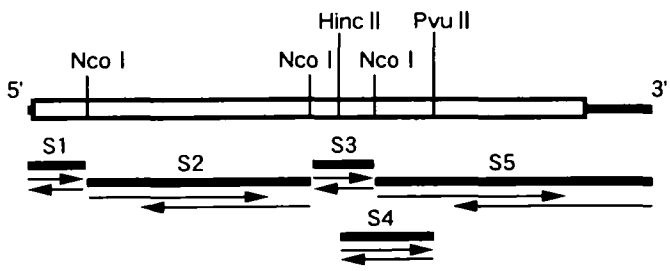


Fig. 2. **Sequencing strategy for mouse p144.** The length of the mouse p144 PCR product is 3.5 kbp and the ORF is represented by an open box. Three *Nco*I digest fragments and a *Hinc*II/*Pvu*II fragment of mouse p144 were subcloned into the Bluescript II SK(−) plasmid vector for sequencing. Arrows indicate the sequencing directions and obtained lengths.

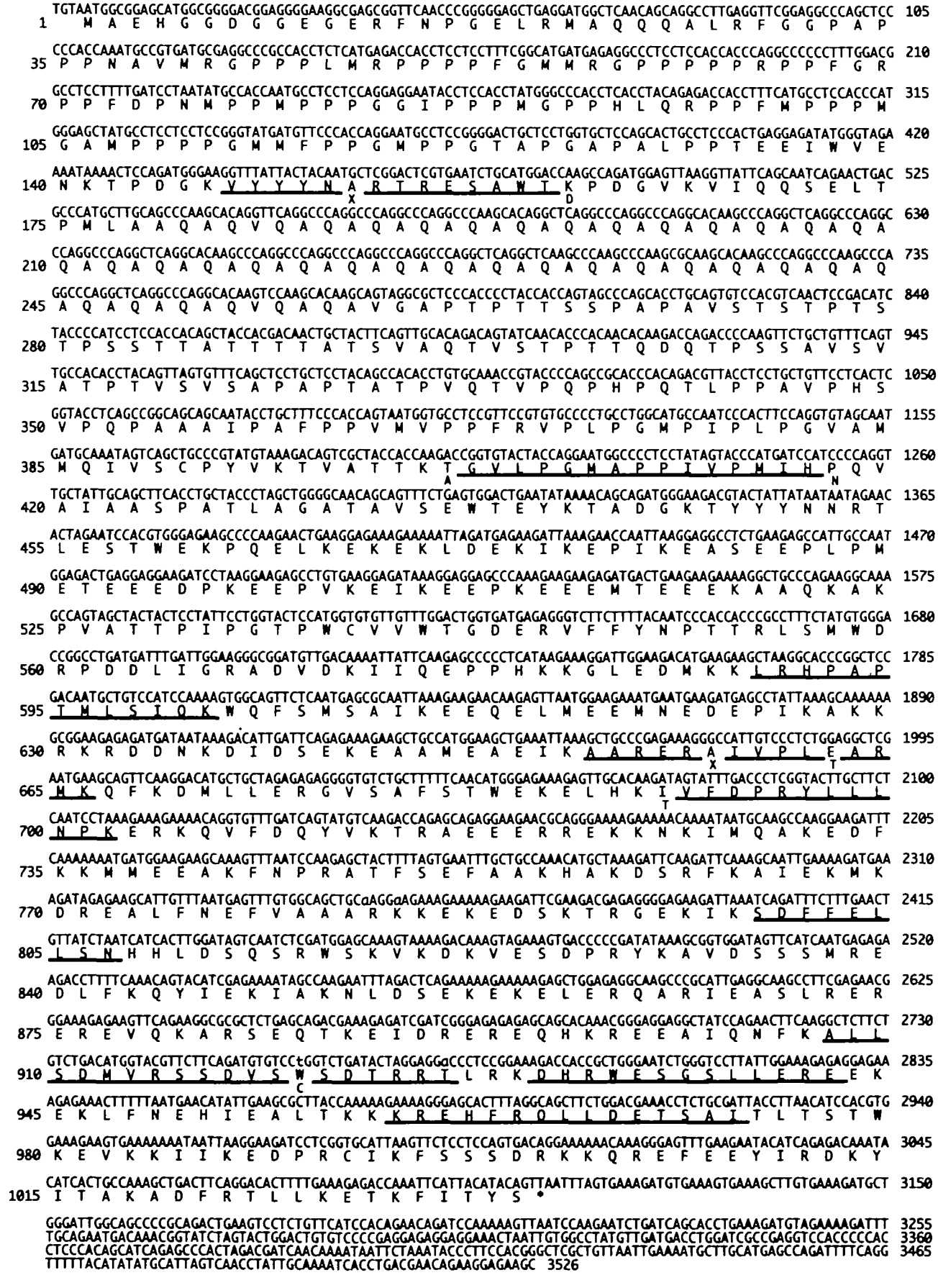
The deduced amino acid sequences of mouse CA150 revealed that the glutamine-alanine (QA) repeat, WWP domains, polyproline region, and STP region found in human CA150 are conserved (Fig. 4). However, the leucine zipper-like structure found in the carboxyl-terminal region of human CA150 was missing from this mouse CA150b due to shortening by 66 amino acids of the carboxyl-terminal (Fig. 4).

Two Splicing Isoforms of CA150—The comparison of the cDNA sequences of mouse CA150b shown in Fig. 3 and human CA150 reported (*15*) revealed that human CA150 is shorter by 95 bp than mouse CA150b in the 3′-terminal part (data not shown). The difference could be explained as a result of alternative splicing. Because, (i) the cDNA sequence of mouse CA150b contained alternative splicing consensus sequences, (C/A)AGGT(A/G)AGT....TTNCAG, and (ii) the sequences of mouse CA150b cDNA flanking the putative alternative splicing region are almost the same as those of the human CA150 cDNA. To examine the possible alternative splicing, PCR was performed with four clones encoding mouse CA150 cDNA obtained by RT-PCR (Fig. 5). Two clones gave PCR products of 470 bp in length expected for mouse CA150b (Fig. 5, lanes 1 and 2). The other two clones gave ones of 370 bp in length expected for the human CA150-type one (Fig. 5, lanes 3 and 4). Moreover, PCR products of both lengths were obtained from cDNA of human blood total RNA (Fig. 5, lane 5). These results strongly suggested that both types of CA150 mRNAs exist in human blood and mouse liver, respectively. Therefore, we designated these two types of mRNAs as CA150a and CA150b mRNAs: CA150a is the spliced type and CA150b is the non-spliced type of an alternative exon.

The cDNA sequences of the splicing site were examined by sequencing of the 470 bp and 370 bp fragments in Fig. 5 after insertion into a vector. The nucleotide sequences of these fragments derived from mouse CA150a and CA150b cDNAs completely matched each other except in the intron region (Fig. 6A). The same results were obtained for human CA150a and CA150b cDNAs (Fig. 6A). Furthermore, the nucleotide sequences of the alternative exons of the mouse and human CA150s were also completely the same as each other. These results confirm the above idea: CA150a and CA150b are splicing isoforms. The deduced amino acid sequences of the carboxyl-terminal regions of CA150a and CA150b of mouse and man are shown in Fig. 6B. The most prominent difference between CA150a and CA150b is that CA150a has a leucine zipper-like motif.

PCR Analysis of CA150a and CA150b mRNAs in Various Tissues and Species—Whether both types of CA150 mRNA are generated in various tissues and species or not was examined by PCR analysis (Fig. 7A). The results show that alternative splicing of pre-mRNA to generate CA150a and CA150b mRNAs occurs in all animals and tissues examined. It was also shown that CA150 mRNA is

Fig. 3. **Nucleotide and amino acid sequences of a mouse p144, CA150b.** Nucleotides and amino acids are numbered on the right and left of each line, respectively. Amino acid sequences identical with those determined for the Lys-C-digest of rat p144 are underlined, and amino acids different from each other are indicated in the figure. "X" means A (alanine) or P (proline). This sequence has been submitted to Genbank, under accession number AB023485.



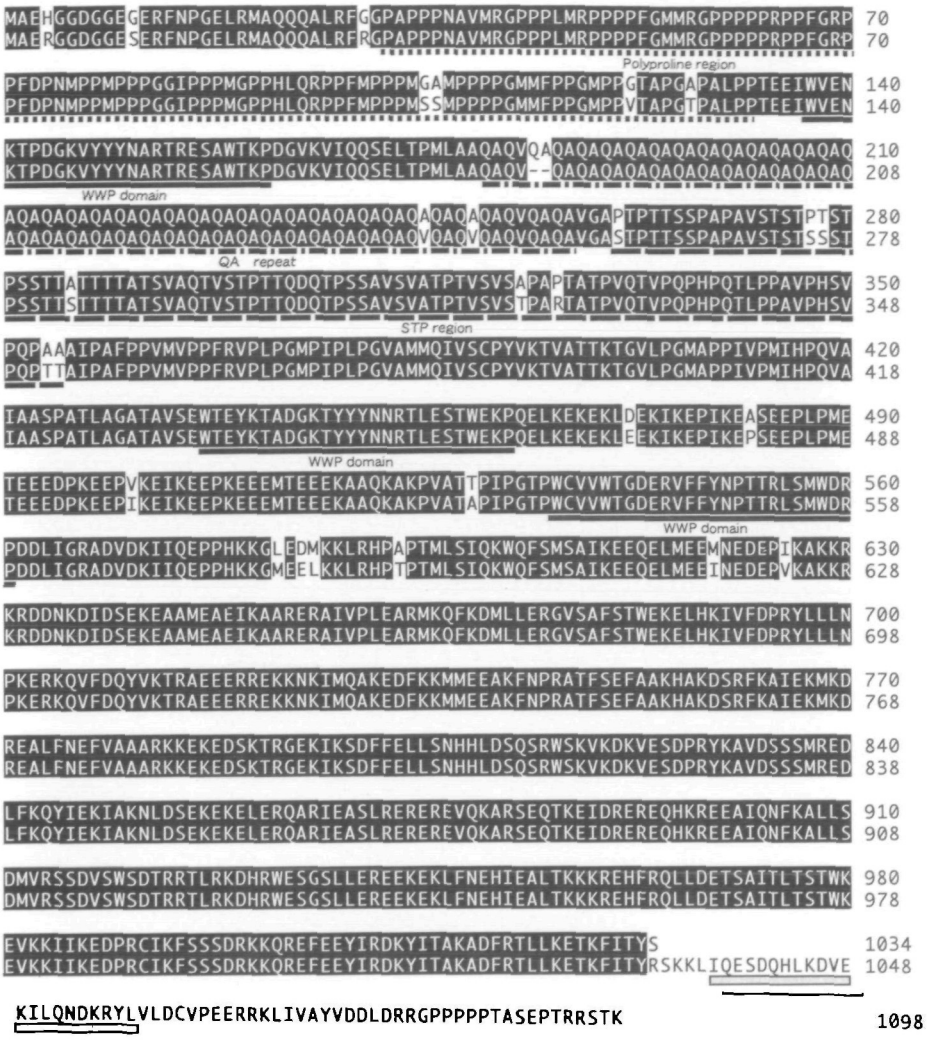


Fig. 4. **Comparison of the amino acid sequences of a mouse p144, CA-150b, and a human transcription factor, CA150a.** Upper: mouse CA150b; lower: human CA150a. Black boxes in the figure indicate identical amino acids in the two sequences. Interesting domains and motifs are underlined. The proline-rich region in the amino-terminal portion of the protein (----); three WWP domains (——); the glutamine-alanine (QA) repeat (——); the region rich in serine, threonine, and proline residues (STP) (——); the putative leucine zipper in the carboxyl terminal sequence (▩).

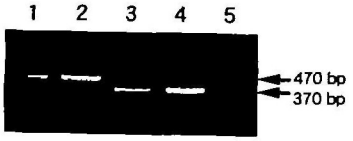


Fig. 5. **Two kinds of CA150 clone.** Cloned plasmid DNA encoding mouse CA150 (1, 2, 3, and 4) and cDNA reverse transcribed from human blood total RNA were amplified by PCR using the S and B primers. Then the thus obtained PCR products were analyzed by agarose gel electrophoresis. The 470 bp and 370 bp bands were generated from the b and a type CA150 clones, respectively.

present in a wide range of vertebrates and in various tissues. Then, we further examined the alternative splicing of CA150 pre-mRNA in lymphocytes and related cell lines (Fig. 7B). It was found that all cDNAs prepared from lymphocyte-related cell preparations shown in Fig. 7B contain both types of CA150 mRNAs, although the ratios of the CA150a and CA150b types are different. This suggests that the alternative splicing also occurs in all these lymphocyte populations. It should be noted that human cells related to blood cells: human whole blood cells (Fig. 5, lane 5) and a human T-cell clone, *i.e.* Jurkat cells (Fig. 7B, lane 6), also contain mRNA for not only CA150a but also CA150b.

Phosphorylation of Rat p144—The mouse CA150b amino acid sequence contains 8 and 2 consensus sequences for phosphorylation by PKA and calmodulin-dependent kinase II, respectively. However, no sequence for phosphorylation by cdc2 kinase was found. We examined whether or not rat p144 is phosphorylated by PKA and calmodulin-dependent kinase II *in vitro*. The WGA-bound fraction of a rat liver nuclear extract was incubated with these protein kinases in the presence of [γ-^{32}P]ATP. As shown in Fig. 8, lanes 2 and 3, the rat p144 protein was phosphorylated by calmodulin-dependent kinase II but not by PKA. It was also confirmed that rat p144 in the fraction was not phosphorylated by cdc2 kinase (data not shown). These results suggested that CA150 was phosphorylated by at least calmodulin-dependent kinase II. The function of CA150 may be regulated by calmodulin-dependent kinase II *in vivo*.

DISCUSSION

Cloning of Mouse p144, CA150—We cloned the cDNA for mouse p144, CA150b. For the cloning, various DNA polymerases for PCR were examined under various conditions, *i.e.*, LA Taq, Pyrobest, Pfu Turbo, Pwo, and KOD

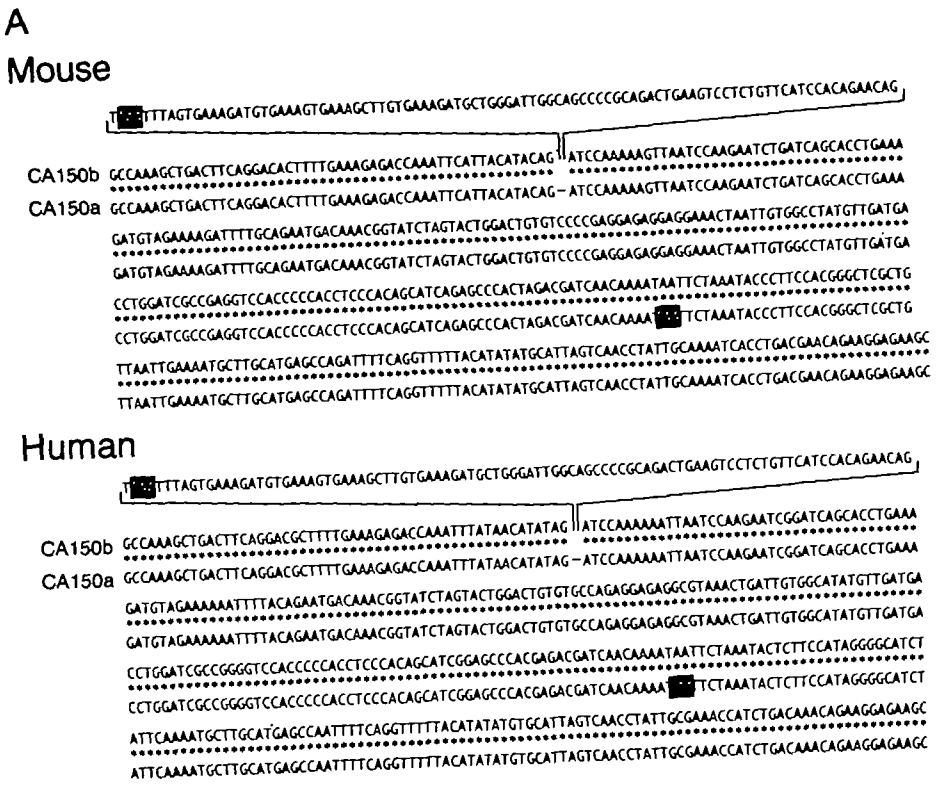


Fig. 6. **Comparison of the nucleotide and amino acid sequences of the splicing isoforms, CA150b and CA150a.** (A) Comparison of the nucleotide sequences around the splicing sites of mouse CA150a and CA150b, and human CA150a and CA150b. cDNAs for sequencing were prepared by PCR from cloned mouse CA150 and a reversed-transcribed product of human blood total RNA. Black boxes indicate the stop codons. (B) Amino acid sequences of the carboxyl terminal regions of mouse and human CA150a and CA150b. The leucine zipper-like motifs found in mouse and human CA150a are underlined.

Dash DNA polymerases. The expected PCR product, about 3.5 kb, was obtained only when LA Taq DNA polymerase was used. We cloned mouse p144 cDNA using this DNA polymerase and sequenced it. To sequence the mouse p144, CA150b, cDNA, it was subcloned, as shown in Fig. 2, and all clones were sequenced in both strands. The sequence of sixteen nucleotides from the 5′-terminal of CA150b cDNA, and coupling of S1 to S2 and S2 to S3 were verified by utilizing mouse EST sequences (GenBank accession numbers AA530507, AA553135, and AA153474). The nucleotide sequences of regions for which only one strand was determined were verified by using mouse EST sequences (GenBank accession numbers AA553135, AA530507, AA087465, AA153474, AA895438, AA561204, AA231515, AA162719, AA271545, W88230, and W10098). Furthermore, whole sequences were confirmed by re-sequencing of another CA150b cDNA clone by the same procedures as above. Sequence analysis revealed an open reading frame encoding a 1034-amino-acid protein (Fig. 3). We could not obtain a peptide containing the amino terminal residue of rat p144 from its lysylendopeptidase digest. The amino acid sequence of the amino terminal region was also not obtained on direct sequencing of intact p144 isolated from rat liver. Therefore, we deduced the AUG start codon of mouse CA150b, as shown in Fig. 3, from the following indirect suggestive data. First, the position of the first methionine Met^1 is the same as that of human putative Met^1. Second, the DNA sequence around Met^1, GTAATGG, is similar to the Kozak consensus sequence for translational initiation, (A/G)CCATGG (*36*). However, the possibility of the translation starting from Met^{21} or another methionine could not be excluded.

Splicing Isoforms of CA150—Cloning and nucleotide sequencing of cDNA for mouse transcription factor CA150 revealed the presence of two types of mRNAs: for mouse CA150a and CA150b. They are splicing isoforms. The "b" type uses an alternative exon containing a stop codon. It was shown that these two types of mRNA are present in all

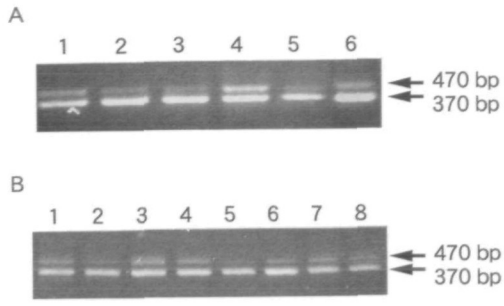


Fig. 7. **PCR analysis of CA150a and CA150b cDNAs in various tissues, species, and cell lines.** (A) Total RNAs from 1, mouse liver; 2, rat liver; 3, *Xenopus* liver; 4, human blood; 5, mouse lung; and 6, mouse brain were reversed transcribed. The resulting cDNAs were amplified by PCR using the S and B primers, and then analyzed by agarose gel electrophoresis. The 470 bp and 370 bp bands were generated from the CA150b and CA150a types, respectively. (B) Total RNAs from 1, BALB/c mouse naive $CD4^+$ T cells ($CD4^+$, CD45RBhi); 2, BALB/c mouse memory $CD4^+$ T cells ($CD4^+$, CD45-RBlo); 3, mouse $CD4^+$ Th1 cells; 4, mouse $CD4^+$ Th2 cells; 5, mouse EL4 lympoma cells; 6, human Jurkat lymphoblastoma cells; 7, mouse cytotoxic T cells derived from L8313 leukemic mice (STIL-3 C5, $CD8^+$); and 8, mouse helper T cells derived from leukemic mice (STIL-3 DF, $CD4^+$) were reversed transcribed. The resulting cDNAs were treated the same as in (A).

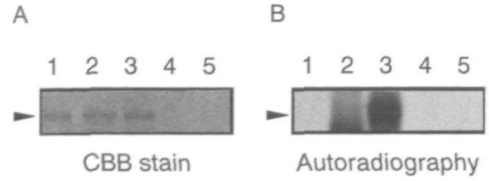


Fig. 8. **Phosphorylation of rat p144 with calmodulin-dependent kinase II and PKA.** The WGA-bound fraction of a rat liver nuclear matrix extract (4 μg proteins in lanes 1, 2, and 3) was incubated with protein kinases in the presence of 0.05 mM [γ-^{32}P]-ATP at 30°C for 20 min. The thus treated sample protein was precipitated with methanol-chloroform and carrier protein ovalbumin, and then subjected to SDS-PAGE. The gel was stained with CBB-R250 (A) to detect the protein bands and then exposed to an X-ray film for 10 h (B). 1, mock experiment in the absence of kinase; 2, PKA; 3, calmodulin-dependent kinase II; 4 and 5, the same as 2 and 3, respectively, except that the protein was omitted. Arrowheads indicate the positions of rat p144.

animal tissues examined, from amphibians to mammals (Fig. 7A). Suñé *et al.* reported only one type of cDNA for human CA150, which was isolated from an oligo(dT)-primed HeLa cell cDNA library and two random-primed human T-cell cDNA libraries. The sequence corresponds to that of our CA150a. However, we detected both transcripts in human normal blood cells (Fig. 5, lane 5, and Fig. 6) and in a human T-cell derived clone, *i.e.* Jurkat cells (Fig. 7B, lane 6). Moreover, mouse $CD4^+$ T-cell clones also contained both types of transcript (Fig. 7B, lanes 1–4, and 8). These results strongly suggest that human normal $CD4^+$ T-cells, which are HIV-1 sensitive cells, contain both types of CA150 transcript. It was reported that human CA150 is associated with the RNA polymerase II holoenzyme and is involved in Tat-dependent HIV-1 transcriptional activation (*15*). Therefore, it would be very interesting in the future to determine whether the two types of CA150s act differently in Tat-activated transcription in human $CD4^+$ T cells or not. The ratio of cDNAs for CA150a and CA150b differed from tissue to tissue and from cell clone to cell clone (Fig. 7). In most preparations the amount of cDNA for CA150a is highest. However, the amounts of cDNA for CA150a and CA150b were almost the same in the preparation of human whole blood cells (Fig. 5). This may mean that some type blood cells other than T cells contain more mRNA for CA150b than that for CA150a. Quantitative PCR is necessary for further analysis.

The most prominent difference between CA150a and CA150b is that the former has a leucine zipper-like motif (*37*) in its carboxyl-terminal region. In the leucine zipper motif, leucine residues appear in an α-helix conformation more than four times at every seventh position (*37*). The α-helix conformation is often stabilized by charged residues located at positions suitable for ion pairing ($i\pm3$ or $i\pm4$) (*37*). Then, we analyzed the structural characteristics of the carboxyl-terminal region of mouse CA150a. The region from K^{1036} to D^{1063} was predicted to be in an α-helix conformation by the Robson method (*38*). Two series of charged residues, K^{1037}-E^{1041}-H^{1045}-D^{1048}-K^{1051}, and D^{1043}-K^{1047}-E^{1050}, are located at positions suitable for ion pairing to stabilize the α-helix conformation. I^{1039}, L^{1046}, L^{1053}, and L^{1060} appear at every seventh position. Destabilization of the interdigitated leucine zipper caused by occupation of position 1039 by isoleucine instead of leucine was estimated to be low, because the methyl group of the β carbon of I^{1039} seems not to inhibit the interdigitation of leucine side chains, because I^{1039} is at the outer most position of the interdigitation. These results suggested that this region of mouse CA150a acts as a leucine-zipper, by which mouse CA150a forms a dimer or binds with another protein. Landschulz *et al.* suggested that some DNA binding proteins with a leucine zipper motif also have a 30-amino-acid region containing a high proportion of basic residues immediately adjacent to their leucine zipper (*37*). They predicted that the leucine zipper causes juxtaposing of the basic regions of two polypeptides in a manner suitable for sequence-specific recognition of DNA (*37*). In mouse CA150a, a 30-amino-acid region (Y^{1009}-L^{1038}) immediately adjacent to the leucine zipper indeed contains a high proportion of basic residues: nine positively charged and three negatively charged ones. Therefore, this region of mouse CA150a may participate in sequence-specific recognition of DNA. All these results should be applicable to human CA150a because the amino acid sequence of this region is entirely the same as that of mouse CA150a. CA150b, which lacks the leucine zipper-like motif and some of the basic amino acids in the 30-amino-acid region, should function in the transcription mechanism in a very different manner from CA150a.

Other Properties of CA150—p144 purified from the WGA-Sepharose bound fraction of a rat liver nuclear matrix extract was shown to be a rat homologue of CA150. We previously purified rat p144 in the native form and characterized it (*27*). p144 binds directly with WGA and cross-reacts with an antiserum which recognizes *O*-GlcNAc bearing proteins (*27*). p144 in the WGA-bound fraction aggregated on lowering of the ionic strength of the medium (*27*). This may show that the protein undergoes self-polymerization and/or associates with other nuclear matrix proteins through ionic interactions in living cells. The amount of p144 per rat liver nucleus was estimated to be

about 8×10^4 molecules from the yield of purified p144 (*27*). Based on the large axial ratio of p144, 35 ± 4, we proposed that p144 is a rod-shaped molecule (*27*). We can not yet separate the rat p144 protein into p144a and p144b because our SDS-PAGE system is not sophisticated enough to detect them as separate bands. The amino acid sequences of mouse and human CA150 are very similar to that of rat p144 (Fig. 3). Therefore, most of these characteristic properties of rat p144 should be common to mouse and human CA150.

For *O*-GlcNAc, some common features are evident. All *O*-GlcNAc-modified proteins are also phosphorylated by specific kinases in a regulated manner (*39*). Indeed, rat *O*-GlcNAc p144 in the WGA-bound fraction was phosphorylated by calmodulin-dependent kinase II (Fig. 8). On the other hand, the location of *O*-GlcNAc at Ser(Thr) sites that are similar or identical to those used by many kinases suggested that competition between Ser(Thr)-*O*-GlcNAcylation and Ser(Thr)-*O*-phosphorylation results in the regulation of many of a cell's important proteins (*40*). It would be interesting to determine whether the *O*-GlcNAc Ser(Thr) sites of CA150 are *O*-phosphorylation Ser(Thr) sites or not. The competition of *O*-GlcNAcylation and *O*-phosphorylation may affect Tat-dependent transcriptional activation.

We wish to thank Kazuko Hasegawa for her help in preparing the figures.

REFERENCES

1. Jones, K.A. and Peterlin, B.M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**, 717-743
2. Frankel, A.D. and Young, J.A. (1998) HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67**, 1-25
3. Bohjanen, P.R., Colvin, R.A., Puttaraju, M., Been, M.D., and Garcia-Blanco, M.A. (1996) A small circular TAR RNA decoy specifically inhibits Tat-activated HIV-1 transcription. *Nucleic Acids Res.* **24**, 3733-3738
4. Carroll, R., Peterlin, B.M., and Derse, D. (1992) Inhibition of human immunodeficiency virus type 1 Tat activity by coexpression of heterologous *trans* activators. *J. Virol.* **66**, 2000-2007
5. Herrmann, C.H. and Rice, A.P. (1995) Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. *J. Virol.* **69**, 1612-1620
6. Madore, S.J. and Cullen, B.R. (1993) Genetic analysis of the cofactor requirement for human immunodeficiency virus type 1 Tat function. *J. Virol.* **67**, 3703-3711
7. Song, C.Z., Loewenstein, P.M., and Green, M. (1994) Transcriptional activation *in vitro* by the human immunodeficiency virus type 1 Tat protein: evidence for specific interaction with a coactivator(s). *Proc. Natl. Acad. Sci. USA* **91**, 9357-9361
8. Suñé, C. and Garcia-Blanco, M.A. (1995) Transcriptional *trans* activation by human immunodeficiency virus type 1 Tat requires specific coactivators that are not basal factors. *J. Virol.* **69**, 3098-3107
9. Zhou, Q. and Sharp, P.A. (1995) Novel mechanism and factor for regulation by HIV-1 Tat. *EMBO J.* **14**, 321-328
10. Kashanchi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.M., Roeder, R.G., and Brady, J.N. (1994) Direct interaction of human TFIID with the HIV-1 transactivator tat. *Nature* **367**, 295-299
11. Kato, H., Sumimoto, H., Pognonec, P., Chen, C.H., Rosen, C.A., and Roeder, R.G. (1992) HIV-1 Tat acts as a processivity factor *in vitro* in conjunction with cellular elongation factors. *Genes Dev.* **6**, 655-666
12. Keen, N.J., Gait, M.J., and Karn, J. (1996) Human immunodeficiency virus type-1 Tat is an integral component of the activated transcription-elongation complex. *Proc. Natl. Acad. Sci. USA* **93**, 2505-2510
13. Mavankal, G., Ignatius Ou, S.H., Oliver, H., Sigman, D., and Gaynor, R.B. (1996) Human immunodeficiency virus type 1 and 2 Tat proteins specifically interact with RNA polymarese II. *Proc. Natl. Acad. Sci. USA* **93**, 2089-2094
14. Zhou, Q. and Sharp, P.A. (1996) Tat-SF1: cofactor for stimulation of transcriptional elongation by HIV-1 Tat. *Science* **274**, 605-610
15. Suñé, C., Hayashi, T., Liu, Y., Lane, W.S., Young, R.A., and Garcia-Blanco, M.A. (1997) CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* **17**, 6029-6039
16. Fridell, R.A., Harding, L.S., Bogerd, H.P., and Cullen, B.R. (1995) Identification of a novel human zinc finger protein that specifically interacts with the activation domain of lentiviral Tat proteins. *Virology* **209**, 347-357
17. Jeang, K.T., Chun, R., Lin, N.H., Gatignol, A., Glabe, C.G., and Fan, H. (1993) *In vitro* and *in vivo* binding of human immunodeficiency virus type 1 Tat protein and Sp1 transcription factor. *J. Virol.* **67**, 6224-6233
18. Nelbock, P., Dillon, P.J., Perkins, A., and Rosen, C.A. (1990) A cDNA for a protein that interacts with the human immunodeficiency virus Tat transactivator. *Science* **248**, 1650-1653
19. Shibuya, H., Irie, K., Ninomiya-Tsuji, J., Goebl, M., Taniguchi, T., and Matsumoto, K. (1992) New human gene encoding a positive modulator of HIV Tat-mediated transactivation. *Nature* **357**, 700-702
20. Yu, L., Zhang, Z., Loewenstein, P.M., Desai, K., Tang, Q., Mao, D., Symington, J.S., and Green, M. (1995) Molecular cloning and characterization of a cellular protein that interacts with the human immunodeficiency virus type 1 Tat transactivator and encodes a strong transcriptional activation domain. *J. Virol.* **69**, 3007-3016
21. Holt, G.D. and Hart, G.W. (1986) The subcellular distribution of terminal *N*-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, *O*-linked GlcNAc. *J. Biol. Chem.* **261**, 8049-8057
22. Davis, L.I. and Blobel, G. (1987) Nuclear pore complex contains a family of glycoproteins that includes p62: Glycosylation through a previously unidentified cellular pathway. *Proc. Natl. Acad. Sci. USA* **84**, 7552-7556
23. Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by *O*-GlcNAc. *J. Biol. Chem.* **268**, 10416-10424
24. Jackson, S.P. and Tjian, R. (1988) *O*-Glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation. *Cell* **55**, 125-133
25. Lichtsteiner, S. and Schibler, U. (1989) A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. *Cell* **57**, 1179-1187
26. Kita, K., Omata, S., and Horigome, T. (1993) Purification and characterization of a nuclear pore glycoprotein complex containing p62. *J. Biochem.* **113**, 377-382
27. Saito, M., Kita, K., Sekiya, K., Omata, S., and Horigome, T. (1995) Purification and molecular shape of a 144 kDa protein bearing *N*-acetylglucosamine residues from rat liver nuclear envelopes. *J. Biochem.* **117**, 47-53
28. Imai, N., Saito, M., Denta, K., Ichimura, T., Omata, S., and Horigome, T. (1995) Purification of 39 kDa and 50 kDa putative nucleoporins from rat liver nuclear envelopes. *Biochem. (Life Sci. Adv.)* **14**, 59-66
29. Makifuchi, S., Takeuchi, M., Kawahire, S., Odani, S., Ichimura, T., Omata, S., and Horigome, T. (1996) Characterization of p92, karyopherin β, co-purified with *N*-acetylglucosamine-bearing nucleoporins from rat liver nuclear envelopes. *J. Biochem.* **120**, 716-724
30. Kikuchi, N., Gohshi, T., Kawahire, S., Tachibana, T., Yoneda,

Y., Isobe, T., Lim, C.R., Kohno, K., Ichimura, T., Omata, S., and Horigome, T. (1999) Molecular shape and ATP binding activity of rat p50, a putative mammalian homologue of RuvB DNA helicase. *J. Biochem.* **125**, 487-494
31. Gohshi, T., Shimada, M., Kawahire, S., Imai, N., Ichimura, T., Omata, S., and Horigome, T. (1999) Molecular cloning of mouse p47, a second group mammalian RuvB DNA helicase-like protein: homology with those from human and *S. cerevisiae*. *J. Biochem.* **125**, 939-946
32. Blobel, G. and Potter, V.R. (1966) Nuclei from rat liver: isolation method that combines purity with high yield. *Science* **154**, 1662-1665
33. Kawahire, S., Takeuchi, M., Gohshi, T., Sasagawa, S., Shimada, M., Takahashi, M., Abe, T.K., Ueda, T., Kuwano, R., Hikawa, A., Ichimura, T., Omata, S., and Horigome, T. (1997) cDNA cloning of nuclear localization signal binding protein NBP60, a rat homologue of lamin B receptor, and identification of binding sites of human lamin B receptor for nuclear localization signals and chromatin. *J. Biochem.* **121**, 881-889
34. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
35. Yamauchi, T. and Fujisawa, H. (1983) Purification and characterization of the brain calmodulin-dependent protein kinase (kinase II), which is involved in the activation of tryptophan 5-monooxygenase. *Eur. J. Biochem.* **132**, 15-21
36. Mcbratney, S. and Sarnow, P. (1996) Evidence for involvement of *trans*-acting factors in selection of the AUG start codon during eukaryotic translational initiation. *Mol. Cell. Biol.* **16**, 3523-3534
37. Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764
38. Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97-120
39. Haltiwanger, R.S., Kelly, W.G., Roquemore, E.P., Blomberg, M.A., Dong, L.-Y.D., Kreppel, L., Chou, T.-Y., and Hart, G.W. (1992) Glycosylation of nuclear and cytoplasmic proteins is ubiquitous and dynamic. *Biochem. Soc. Transact.* **20**, 264-269
40. Hart, G.W., Kreppel, L.K., Comer, F.I., Arnold, S.C., Snow, D.M., Ye, Z., Cheng, X., DellaManna, D., Caine, D.S., Earles, B.J., Akimoto, Y., Cole, R.N., and Hayes, B.K. (1996) *O*-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with *O*-phosphorylation and putative roles in protein multimerization. *Glycobiology* **6**, 711-716