[Cell Cycle 4:8, 1099-1106; August 2005]; ©2005 Landes Bioscience

Report

A Nuclear Targeting Determinant for SATB1, a Genome Organizer in the T Cell Lineage

Yuji Nakayama
I. Saira Mian
Terumi Kohwi-Shigematsu*
Takuya Ogawa

Lawrence Berkeley National Laboratory; University of California; Berkeley, California USA

*Correspondence to: Terumi Kohwi-Shigematsu; Lawrence Berkeley National Laboratory; University of California; 1 Cyclotron Rd (84-171); Berkeley, California 94720 USA; Tel.: 510.486.4983; Fax: 510.486.4545; Email: terumiks@lbl.gov

Received 05/18/05; Accepted 05/20/05

Previously published online as a *Cell Cycle* E-publication: http://www.landesbioscience.com/journals/cc/abstract.php?id=1862

KEY WORDS

nuclear targeting sequences, SATB1, SATB2, nuclear localization signal, *Drosophila dve*

ABBREVIATIONS

HMM	hidden Markov model
BUR	base-unpairing region
NLS	nuclear localization signal

ACKNOWLEDGEMENTS

We thank Mitsuo Oshimura, thesis supervisor for Y. Nakayama, for his kind support. This work was supported by US National Institutes of Health grant CA39681 to T.K.-S. and a California Breast Cancer Research Program Grant to I.S.M..

Figure 1 is available for download at http://www.landesbioscience.com/journals/cc/nakayamaCC4-8-sup.pdf.

ABSTRACT

SATB1 is a nuclear protein, which acts as a cell-type specific genome organizer and gene regulator essential for T cell differentiation and activation. Several functional domains of SATB1 have been identified. However, the region required for nuclear localization remains unknown. To delineate this region, we employed sequence analysis to identify phylogenetically diverse members of the SATB1 protein family, and used hidden Markov model (HMM)-based analysis to define conserved regions and motifs in this family. One of the regions conserved in SATB1- and SATB2-like proteins in mammals, fish, frog and bird, is located near the N-terminus of family members. We found that the N-terminus of human SATB1 was essential for the nuclear localization of the protein. Furthermore, fusing residues 20–40 to a cytoplasmic green fluorescence protein (GFP) fused to pyruvate kinase (PK) was sufficient to quantitatively translocate the pyruvate kinase into the nucleus. The nuclear targeting sequence of human SATB1 (residues 20–40) is novel and does not contain clusters of basic residues, typically found in 'classical' nuclear localization signals (NLSs). We investigated the importance of four well-conserved residues (Lys29, Arg32, Glu34, and Asn36) in this nuclear targeting sequences. Remarkably, full-length SATB1 harboring a single point mutation at either Lys29 or Arg32, but not Glu34 or Asn36, did not enter the nucleus. Our results indicate that SATB1 N-terminal residues 20–40 represent a novel determinant of nuclear targeting.

INTRODUCTION

SATB1 is expressed at a high level in thymocytes,[1] and its expression is downregulated in mature T cells.[2] However, when T cells are activated, the *Satb1* transcript is induced as an immediate early gene, during the initial G_0 to G_1 phase[3] and SATB1 is accumulated in the nuclei (data not shown). SATB1 provides a unique cage-like nuclear architecture in thymocyte nuclei, and folds chromatin by tethering to specialized DNA sequences. Such DNA sequences contain a cluster of ATC sequence stretches, where one strand consists of a mixture of Adenine (A), Thymine (T), and Cytosine (C), excluding Guanine(G).[1] SATB1 recruits chromatin remodeling/modifying factors onto the ATC sequence stretches and thus regulates the state of histone modifications and nucleosome positioning over long distances.[4,5] Analyses of SATB1 knockout mice have revealed that SATB1 is necessary for coordinating the temporal and spatial expression of a large number of genes during T cell development.[2] In its absence, many genes are ectopically expressed, including interleukin receptors genes such as *IL-2Rα* and *IL-7Rα* indicating that SATB1 acts as a repressor for these genes. Critical genes involved in cell proliferation such as *c-myc* are also directly regulated by SATB1, and in SATB1-null thymocytes, *c-myc* cannot be induced in response to mitogen stimulation. SATB1 can also act as a transcriptional activator for certain genes.[5] In SATB1-null mice, T cell development is mostly arrested at the $CD4^+CD8^+$ double-positive stage.[2] The few CD4 single positive T cells found in SATB1-null mice fail to proliferate upon mitogen stimulation.[2]

The genomic DNA sequences specifically recognized by SATB1 are highly prone to becoming continuously unpaired when subjected to superhelical strain.[6] Such sequences are designated as base-unpairing regions (BURs). BURs often contain a core unwinding element. When the core unwinding element of a BUR is mutated so as to disrupt its ATC sequence context, the DNA remains double-stranded in supercoiled plasmid DNA, even though the sequence remains AT-rich, indicating that merely being AT-rich by itself is not sufficient for the strong base-unpairing property.[6] BURs are often found in genomic regions known as matrix attachment regions (MARs) and the MARs surrounding the immunoglobulin heavy chain gene (IgH) enhancer each contain a BUR.[6,7] The biological

significance of MARs surrounding the IgH enhancer in the transcription and chromatin structure of a rearranged μ gene has been documented to have roles in tissue-specific expression[8] and in long-range chromatin interaction and modification.[9-12] SATB1 binds to double-stranded BUR sequences with high affinity and specificity both in vitro and in vivo, and binding is abrogated when the core unwinding element of a BUR is disrupted.[1,5,13] SATB1 binds in the minor groove of BUR sequences, primarily to the sugar-phosphate backbone without making direct contact with the specific bases, suggesting that SATB1 recognizes the ATC sequences indirectly through the altered phosphate backbone structure within genomic DNA.[1]

To date, multiple functional protein domains have been identified in SATB1, including a BUR-binding domain,[14] an atypical homeodomain,[15] two Cut repeats found in the Cut homeodomain protein family, and a PDZ domain.[16] These multiple domains are all necessary for SATB1 to confer its high specificity to the core unwinding element of BURs. The BUR-binding domain, a 150-amino acid polypeptide (human SATB1 residues 346–495) partially overlapping the Cut repeats, is necessary for SATB1 to distinguish the BUR from any AT-rich DNA sequences. When the 21 amino acids from either end of this domain are deleted, DNA-binding activity is lost.[14] Using a phage library displaying nonamer random peptides, a peptide exhibiting 50% identity with a segment of human SATB1 (residues 355–364) was identified and was found able to replace the native sequence in a chimeric SATB1 construct without loss of binding affinity or specificity. A single amino acid mutation of the conserved Arg or Glu residue in this SATB1 segment to Ala greatly reduced SATB1 binding to BURs.[17]

The homeodomain of SATB1 is not necessary for DNA binding per se. However, when combined with the BUR-binding domain, the homeodomain of SATB1 increases the specificity of SATB1 so that it recognizes the core unwinding element of BURs.[15] At the same time, the BUR domain and homeodomain alone are not sufficient for DNA binding. SATB1 has to be dimerized in order to bind DNA and the homodimerization of SATB1 is mediated by a PDZ-like domain (residues 90–204).[16] At an early stage of apoptosis, SATB1 is cleaved by a caspase 6-like protease, which separates the PDZ domain from the BUR domain and the homeodomain, resulting in a rapid dissociation of SATB1 from chromatin. The PDZ domain is a well-characterized protein-protein interaction domain.[16] Many PDZ domain-containing proteins identified to date are associated with the plasma membrane, and are involved in recruiting signaling proteins to protein complexes at the membrane.[18] SATB1 represents one of the few nuclear proteins with a PDZ domain. SATB1 makes protein complexes with many nuclear proteins, and the N-terminal half of SATB1 containing the PDZ domain is required for complex formation.[4] Most recently, SATB2, which is highly homologous to SATB1, was identified in the human brain as a cleft palate gene.[19] SATB2 was also found to be expressed in preB cells, and it modulates immunoglobulin μ gene expression.[20]

Here we performed computational studies to identify phylogenetically diverse members of the SATB1 protein family, and to determine previously uncharacterized regions and motifs that are conserved. Experimental studies based on these predictions led to the identification of a novel sequence motif which is necessary and sufficient for nuclear localization of SATB1. A stretch of 21 amino acids found in the N-terminus of SATB1 localizes the normally cytoplasmic pyruvate kinase to the nucleus. A single amino acid mutation of a conserved Arg or a conserved Lys located in the N-terminal region abrogates the nuclear location of the full-length SATB1 protein.

MATERIALS AND METHODS

Sequence analysis and HMM-based modeling. Members of the SATB1 family of proteins were enumerated by application of the following steps. First, the full-length human SATB1 protein sequence was used as the query for a database search using the NCBI WWW interface to the PSI-BLAST program (August/September 2004 version: http://www.ncbi.nlm.nih.gov/blast/Blast.cgi; default parameter settings were used). At convergence, sequences with statistically significant E-values and where the conservation extended across the entire length of human SATB1 were regarded as SATB1 family members (the latter criterion eliminated proteins where the nature, number and order of known protein domains was not preserved, i.e., PDZ-CUT1-CUT2-Homeobox). Second, sequences in the Ensembl ortholog predictions for SATB1 (Ensembl GeneID ENSG00000182568) and SATB2 (ENSG00000119042) were examined and those not identified in the first step were added to the family. Third, human SATB1 was used as the query for a search using the WWW interfaces to the BLAST program and model organism-specific databases (October 2004 version of the *Xenopus tropicalis* genome at http://genome.jgi-psf.org/). The sequences identified as SATB1 family members by this procedure spanned a range of vertebrates. The proteins were from mammals (*Homo sapiens, Pan troglodytes, Mus musculus, Rattus norvegicus*), a bird (*Gallus gallus*), an amphibian (*Xenopus tropicalis*) and fish (*Tetraodon nigroviridis, Fugu rubripes, Danio rerio*).

The pairwise alignments of SATB1 family members specified by the PSI-BLAST/BLAST searches were employed as a guide to estimate an HMM of the full-length sequence using the Sequence Alignment and Modeling System (SAM; http://www.soe.ucsc.edu/research/compbio/sam.html). The resultant SAM-style HMM was used to generate a multiple sequence alignment of members of the SATB1 protein family.

DNA constructs. The human SATB1 cDNA with 3'-UTR was isolated from pAT1146 vector[1] by the PCR method using AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA) and the appropriate primers containing a *Xho*I- or an *Eco*RI- linker sequences, in 5'-primer or in 3'-primer, respectively. The PCR product was double-digested with *Xho*I and *Eco*RI and was cloned into the corresponding sites of the pEGFP-C1 vector (BD Biosciences Clontech, Palo Alto, CA) to generate peGFP-SATB1. For the constructs we prepared, the term 'eGFP' was used instead of 'EGFP' for enhanced green fluorescence protein. All mutant clones of peGFP-SATB1 were generated by the two-step PCR-based overlap extension method as shown previously.[21] pEGFP-PK (a gift from WC. Greene, University of California, San Francisco),[22] containing a fragment of chicken pyruvate kinase cDNA at the carboxyl terminus of eGFP, was slightly modified with digestion with *Kpn*I, following by excision with T4 DNA polymerase, and the resulting blunt ends were allowed to be self ligated. The product thereof was designated as peGFP-PK'. A series of SATB1 N-terminal fragments were cloned as *Bgl*II- *Bam*HI fragments and ligated into the *Bam*HI site of peGFP-PK'. All constructs were verified by DNA sequencing as well as by immunoblotting with rabbit polyclonal anti-GFP (sc-8334, Santa Cruz Biotechnology, Santa Cruz, CA).

Cell culture and transfection. Both U2OS cells and HeLa cells were obtained from the American Type Culture Collection (Manassas, VA) and maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum. Semiconfluent cells grown in each well of 24-well plates were transiently transfected with the eGFP constructs using Lipofectamine 2000 Reagent (Invitrogen, Carlsbad, CA). For HeLa cells, 0.8 μg of plasmid DNA and 1.5 μl of Lipofectamine 2000 were incubated in 100 μl of Opti-MEM I (Invitrogen). After 20 to 30 minutes incubation, DNA-liposome complexes were added to 0.5 ml of serum- and antibiotics-free DMEM. After 1 h, the medium was replaced by culture medium. For U2OS cells, a reduced dose of DNA-liposome was used; 0.6 μg of plasmid DNA and 1.25 μl of Lipofectamine 2000 were mixed in 50 μl of OPTI-MEM I, added to 1 ml of serum- and antibiotics-free medium, and then the medium was removed after 4 to 6 h.

Subcellular localization assay. Subcellular localization of eGFP fusion proteins was directly visualized under the fluorescence microscope (DIAPHOT, Nikon Instruments Inc., Melville, NY). In some cases, 24 h

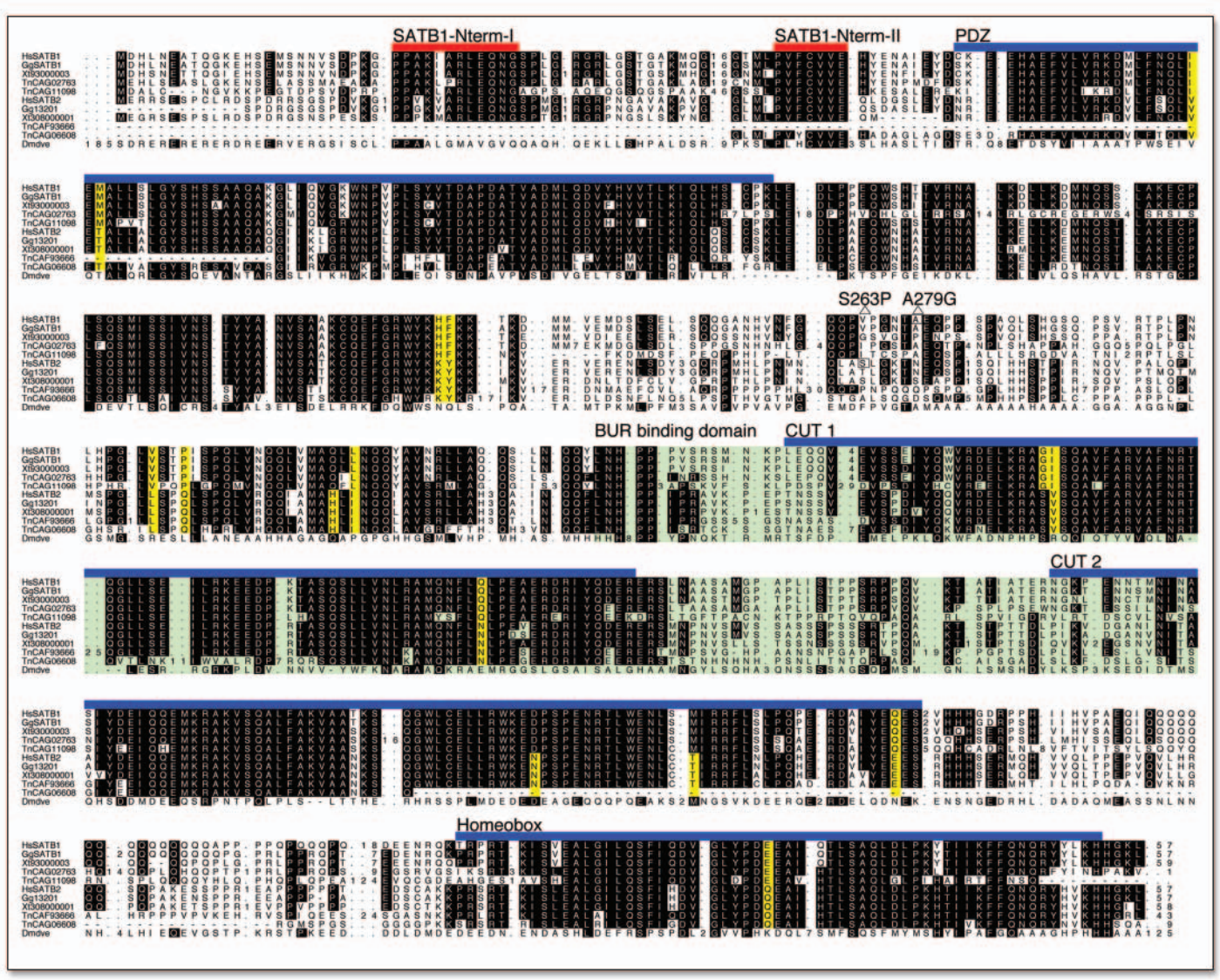

Figure 1. An HMM-generated multiple sequence alignment of selected known and predicted members of the SATB1 family of proteins. Columns in black denote positions where the amino acid is conserved in 6 of the 11 sequences shown; columns with numbers indicate the number of residues not depicted explicitly; columns in yellow signify positions where the amino acid is conserved within the SATB1-like set of proteins (HsSATB1 to TnCAG11098) and the SATB2-like set of proteins (HsSATB2 to TnCAG06608) but not between these two sets. Open red triangles and boxed residues mark the locations of non-synonymous coding SNPs observed in human SATB1 (HsSATB1 A279 to G; geneID 6304) and SATB2 (HsSATB2 S263 to P; geneID 23314) and taken from the dbSNP entry linked to the appropriate gene ID. The green box marks the location of the base-unpairing region (BUR). The blue bars denote the locations of protein domains that have been identified previously ("PDZ", "CUT1", "CUT2", "Homeobox"). The red bars indicate the locations of the amino acids deleted in the GFP-fused human SATB1 constructs characterized experimentally ("SATB1-Nterm-I", "SATB1-Nterm-II"). In a given sequence, long stretches of "-" denote a possible error made by the gene parsing program used to predict the location of open reading frames in genomic DNA. The vertebrate SATB1-like set of proteins and their database codes are as follows, HsSATB1, *Homo sapiens* SATB1 (human, accession number NP_002962); GgSATB1, *Gallus gallus* SATB1 (chicken, XP_418746); Xt93000003, *Xenpous tropicalis* homology-based gene model (frog, fgenesh.C_scaffold_93000003); TnCAG02763, *Tetraodon nigroviridis* open reading frame (ORF; freshwater puffer fish, CAG02763); and TnCAG11098, *T. nigroviridis* ORF (CAG11098). The vertebrate SATB2-like set of proteins are HsSATB2, *H. sapiens* SATB2 (NP_056080); Gg13201, *G. gallus* ORF (ENSGALT00000013201); Xt308000001, *X. tropicalis* homology-based gene model (fgenesh.C_scaffold_308000001); TnCAF93666, *T. nigroviridis* ORF (CAF93666); and TnCAG06608, *T. nigroviridis* ORF (CAG06608). Dmdve, *Drosophila melanogaster dve* (defective proventriculus; fly, NP_477272), is a putative invertebrate member of the SATB1 protein family.

after transfection, cells were trypsinized and diluted, and then plated onto Lab-Tek™ II 4-well glass chamber slides (Nalge Nunc International, Rochester, NY) precoated with poly-L-lysine. After another 24 h incubation, reseeded cells were fixed in 4% PFA for 5 min at room temperature, permeabilized by 0.2% Triton X-100 for 10 min, then double-stained with Rhodamine-Phalloidin (Molecular Probes Inc., Eugene, OR) and DAPI (Sigma-Aldrich, St. Louis, MO), at a concentration of 6.6 nM and 10 ng/ml respectively. Stained cells were visualized with a BX60 fluorescence microscope (Olympus American Inc., Melville, NY).

Flow cytometry analysis. Transfected cells in 35 mm dishes were trypsinized and resuspended in 1 ml of PBS containing 1mM EDTA, then filtered through a FALCON cell strainer cap/test tube (BD Biosciences, San Jose, CA) to remove cell aggregates. Flow cytometry for fluorescence was performed with a BD FACScan (BD Biosciences). A minimum of 20,000 events were collected for each analysis. The GFP fluorescence was detected with a 530/30 nm bandpass filter. Data acquisition and analysis were performed with CellQuest™ Pro software (BD Biosciences).

RESULTS

SATB1 protein family members are present in diverse species. SATB1 plays a critical role in the T cell lineage as a global gene regulator and potentially in other lineages including osteoblasts.[2] Although SATB1 is a nuclear protein, an NLS characterized by a cluster of Lys and Arg residues could not be found. We postulated that the hitherto unidentified NLS might be located in a region of the protein that was conserved in a wide variety of species. To address this hypothesis, we used human SATB1 as the query sequence for database searches aimed at determining SATB1-like proteins. The resultant collection of proteins was used to estimate a probabilistic model, a hidden Markov model (HMM), of the protein family. We inspected an HMM-generated multiple sequence alignment of SATB1-like proteins to identify phylogenetically conserved regions and motifs and tested these predictions experimentally. Previously, such an integrated computational-experimental strategy was used to gain new structural and functional insights about TERT, the key protein subunit of the telomerase complex.[23,24]

Figure 1 shows an HMM-generated multiple sequence alignment of representative members of the SATB1 family (SATB1-like and SATB2-like proteins) from various vertebrates: a mammal (human: HsSATB1, HsSATB2), a bird (chicken: GgSATB1, Gg13201), an amphibian (frog: Xt93000003, Xt308000001), and a fish (freshwater puffer fish: TnCAG02763, TnCAG11098, TnCAF93666, TnCAG06608). Given the high degree of conservation amongst the sequences depicted, for clarity and simplicity, other mammalian (chimpanzee, *Pan troglodytes*; mouse, *Mus musculus*; rat, Rattus norvegicus) and fish (puffer fish, *Fugu rubripes*; zebra fish, *Danio rerio*) family members are not shown. Examination of the alignment reveals that in addition to the previously characterized PDZ, CUT1, CUT2, and Homeobox domains (blue, Fig. 1), there exist segments and amino acids at the N-terminus and between known domains that are highly conserved across a diverse range of species. We investigated the functional role(s) of some of these conserved residues and regions experimentally (red, Fig. 1, discussed in more detail subsequently).

An N-terminal region is required for the nuclear localization of SATB1. To examine the biological relevance of the conserved regions of SATB1, we made enhanced green fluorescent protein (eGFP)-fused human SATB1 expression constructs containing specific deletions in these regions. In order for the eGFP-SATB1 protein to be translated efficiently, we found that the 3'UTR sequence (442bp) was essential. Without the SATB1 3'UTR, even in the presence of an SV40 early mRNA polyA signal in the expression construct, translation of SATB1 was reduced by 10-fold (Fig. 2A). Therefore, the 3'UTR was included in all constructs prepared for this study. The eGFP-SATB1 (residues 1–763) with the 3'UTR was successfully translated in full length and it was exclusively localized in the nuclei of either transfected osteosarcoma U2OS cells (Fig. 2C) or cervical carcinoma HeLa cells (data not shown). The U2OS cells express SATB1 endogenously but only at a low level or are absent in the HeLa cells. Using either cell type, we found similar results in all the experiments described below.

To identify a region responsible for the nuclear localization of SATB1, we first tested the contribution of the N-terminal region that precedes the known PDZ domain (residues 90–130; Fig. 1) and which we designate as SATB1-Nterm (residues 1–89). We made an expression construct, eGFP-SATB1Δ (1–89), which lacks the first 89 amino acids. Thus, the expressed protein is devoid of the two conserved segments (residues 26–41 and 75–87). The first segment contains a hexapeptide (R32LEQNG37) that is invariant among vertebrate SATB1- and SATB2-like proteins, while the second contains a heptapeptide

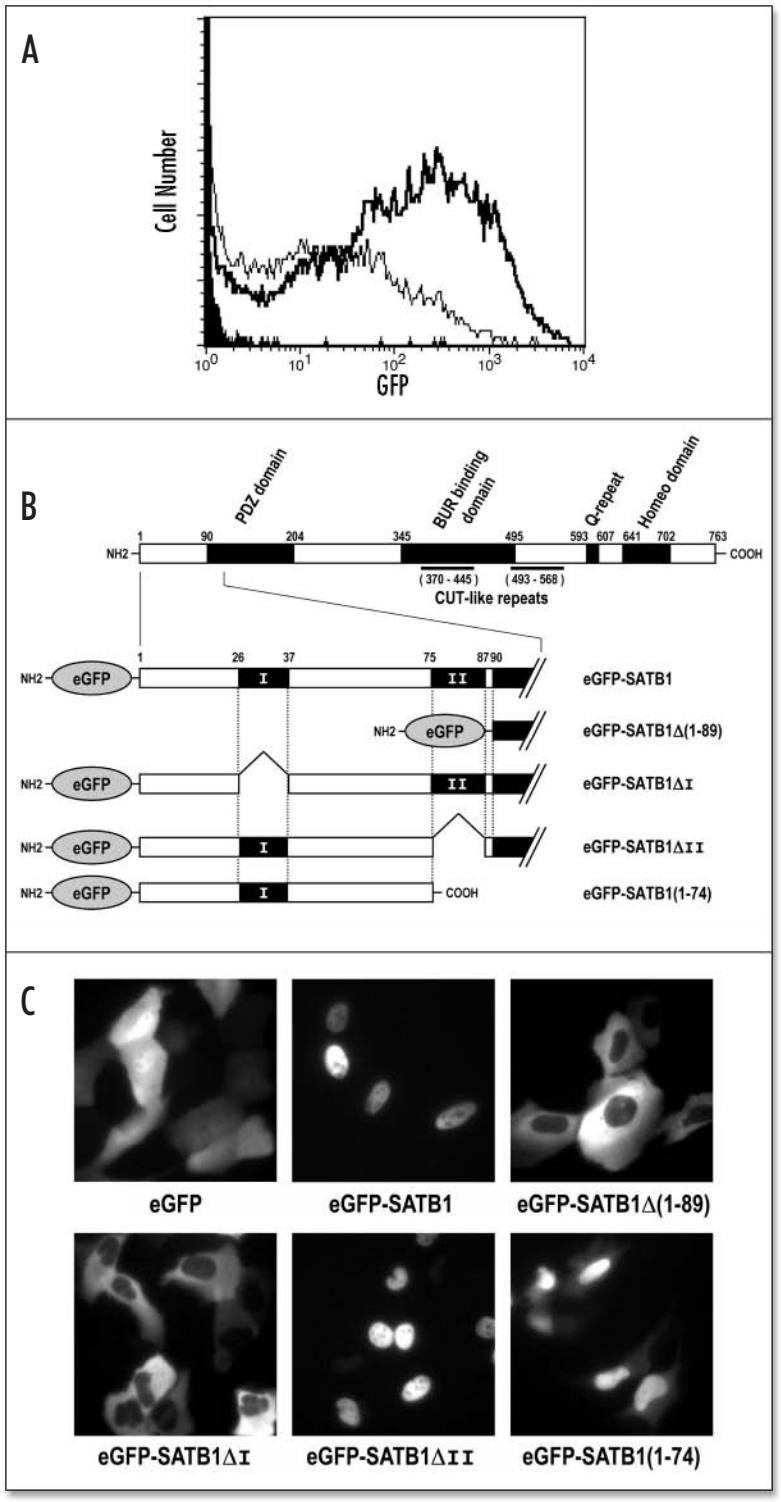

Figure 2. Mapping of nuclear import signal in SATB1. (A) The FACS analysis of the effect of 3'-UTR addition on eGFP-SATB1 expression. Insertion of 3'-UTR following eGFP-SATB1 (bold line) significantly upregulated the fluorescent intensity of eGFP-SATB1 alone (thin line). Mock transfected group (filled by black) is shown as negative control. (B) Diagram of SATB1 with positions of deleted regions. Previously characterized domains are denoted by shaded boxes along with their domain names and positions in amino acids. The two conserved regions (SATB1-Nterm-I, SATB1-Nterm-II) illustrated in Figure 1, are shown. (C) Subcellular localization of SATB1 as an eGFP fusion protein. U2OS cells were transiently transfected with indicated SATB1 variants expression vectors, and pictures were taken of living cells 24 h after transfection.

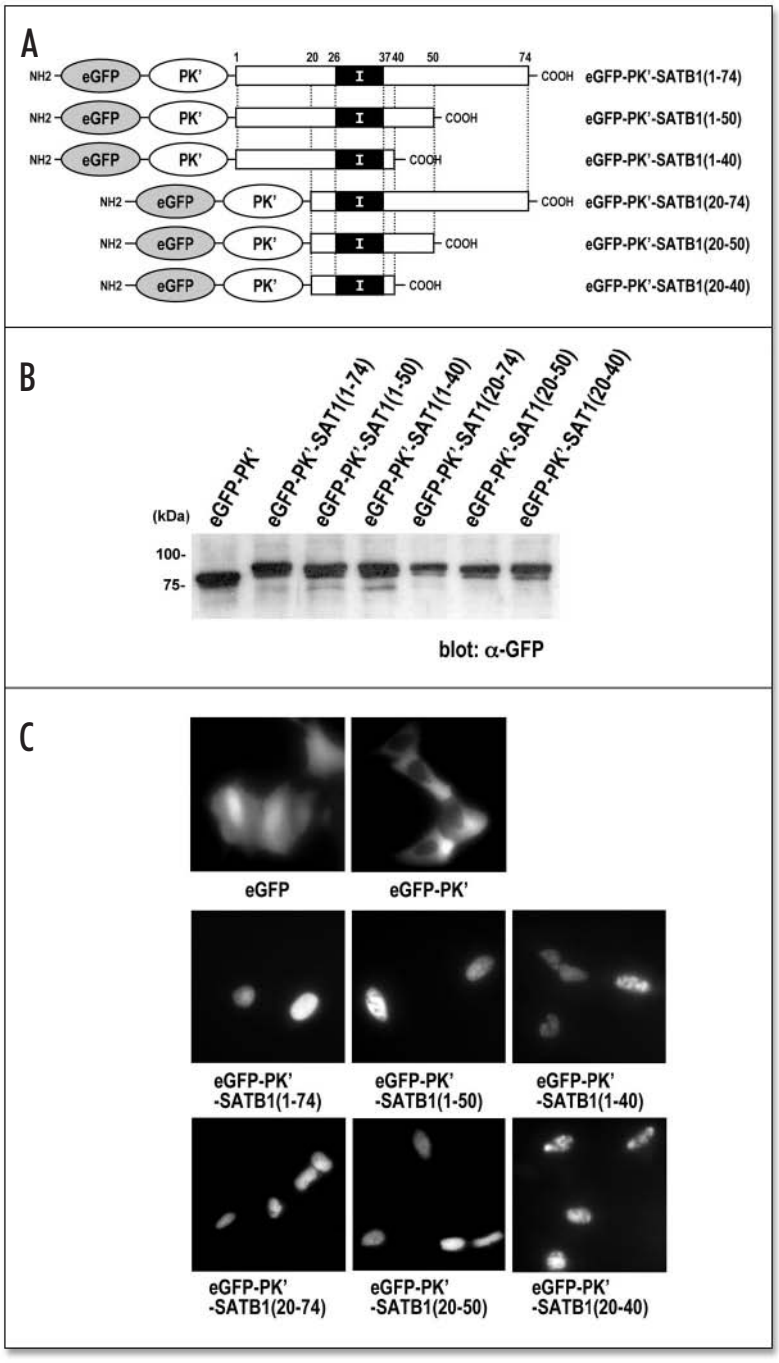

Figure 3. Delineation of SATB1 nuclear localization signal with cytoplasmic pyruvate kinase chimera. (A) Constructs used to identify SATB1 sequence necessary for targeting PK′ to the nucleus. eGFP-PK′ contains sequence corresponding to residues 18 to 442 of chicken muscle pyruvate kinase. Numbers displayed above the top construct refer to the SATB1 amino acids. (B) The expression of each PK′ fusion constructs was verified by immunoblotting with transfected HeLa cells. (C) Subcellular localization of eGFP-PK′ fusion proteins. Each construct shown in (A) was overexpressed in U2OS cells. After 24 h, localization of eGFP-signal was assessed under the fluorescent microscope.

(P75VFCVVE81) which is invariant except for F77-Y in a fish SATB2-like protein. The proteins expressed from the eGFP-SATB1Δ (1–89) construct in the U2OS cells were localized exclusively in the cytoplasm, and thus restricted the location of the SATB1 nuclear localization signal to the first 89 amino acids (Fig. 2C).

We tested whether either conserved segment of SATB1-Nterm is essential for nuclear localization. We created two constructs, eGFP-SATB1ΔI in which the hexapeptide was deleted, and eGFP-SATB1ΔII in which the heptapeptide was deleted (Fig. 2B). The proteins expressed from eGFP-SATB1ΔI, but not eGFP-SATB1ΔII, were localized exclusively in the cytoplasm of transfected U2OS cells (Fig. 2C). Furthermore, when fused with eGFP, only the first 74 amino acids of SATB1, a region that does not include the second conserved heptapeptide containing region, was located preferentially in the nuclei (Fig. 2C). This result suggests that the first hexapeptide containing segment plays a role in the nuclear localization of members of the SATB1 family.

Human SATB1 residues 20–40 are sufficient for nuclear import of a heterologous cytosolic protein. In order to determine which region of SATB1 is important for nuclear localization, we prepared fusion proteins consisting of eGFP, pyruvate kinase (PK) and a series of truncated SATB1 protein segments derived from the N-terminal 74 amino acid region. If one of these segments was a true determinant of nuclear localization, it would be expected to transport the cytoplasmic pyruvate kinase protein into the nucleus. Unlike [eGFP-SATB1(1–74)], which is a small protein of 35.6kDa in size, and which is below the threshold of ~9 nm (40–60 kDa) for passive diffusion through the nuclear pore complex,[25] eGFP-PK' chimera is a 74.7 kDa protein which is not expected to passively diffuse through the nuclear pore complex (Fig. 3B). While eGFP-PK' alone did not show any nuclear localization, the SATB1(1–74) fusion protein [eGFP-PK'-SATB1(1-74)] was found exclusively in the nucleus (Fig. 3C).

To determine the minimal region of SATB1 which is sufficient to confer nuclear localization of the chimeric protein, we deleted the first 20 amino acids from the SATB1 (1–74) region, and found that the [eGFP-PK'-SATB1(20–74)] protein was located exclusively in the nucleus. We also examined the [eGFP-PK'-SATB1(1–50)], [eGFP-PK'-SATB1(1–40)], and [eGFP-PK'-SATB1 (20–50)] proteins and found that each of these truncated proteins still exhibited nuclear localization (Fig. 3C). Finally, the [eGFP-PK'-SATB1(20–40)] protein which contains the 21 amino acids of N-terminal I was examined. This segment is sufficient to bring the eGFP-PK' protein to the nucleus. Thus, SATB1 residues 20-40, which contains the hexapeptide, act as a determinant for nuclear localization for SATB1 (Fig. 3C).

A single lysine or arginine mutation abrogates SATB1 nuclear localization. To investigate the importance of specific residues in the hexapeptide located in SATB1-Nterm-I (red, Fig. 1), a region required for nuclear localization of SATB1 (Fig. 3), we generated an alignment of all predicted and known vertebrate members of the SATB1 family (Fig. 4A). We focused on three highly conserved amino acids, Arg32, Glu34 and Asn36. To evaluate whether these residues are important for nuclear localization, we made a mutated construct in which all three amino acids were changed to Ala in the full-length SATB1 [eGFP-SATB1(3A)] (Fig. 4B). The mutated protein expressed in the transfected cells exhibited cytoplasm localization exclusively, indicating that these three amino acids contain a crucial determinant(s) for the nuclear localization of SATB1 (Fig. 4C).

We next examined single-point mutations by mutagenizing Arg, Glu, or Asn individually to Ala in the context of the full-size SATB1. Both the Glu and Asn single-point mutations [eGFP-SATB1(E34A) or (N36A)] maintained nuclear localization similar to wild-type SATB1. In contrast, the single Arg mutation in SATB1(R32A) caused SATB1 to localize exclusively in the cytoplasm (Fig. 4C).

The Nterm-I region contains another basic residue, Lys29, located three amino acids upstream of Arg32. We examined if, like Arg32, Lys29 also contributes to the nuclear import of SATB1. As shown in Figure 4C, the Lys29Ala change is sufficient to retain SATB1 in the cytoplasm. We conclude that both Lys29 and Arg32 play pivotal roles in nuclear localization of SATB1, and we hypothesize that this is also true for other members of the SATB1 protein family.

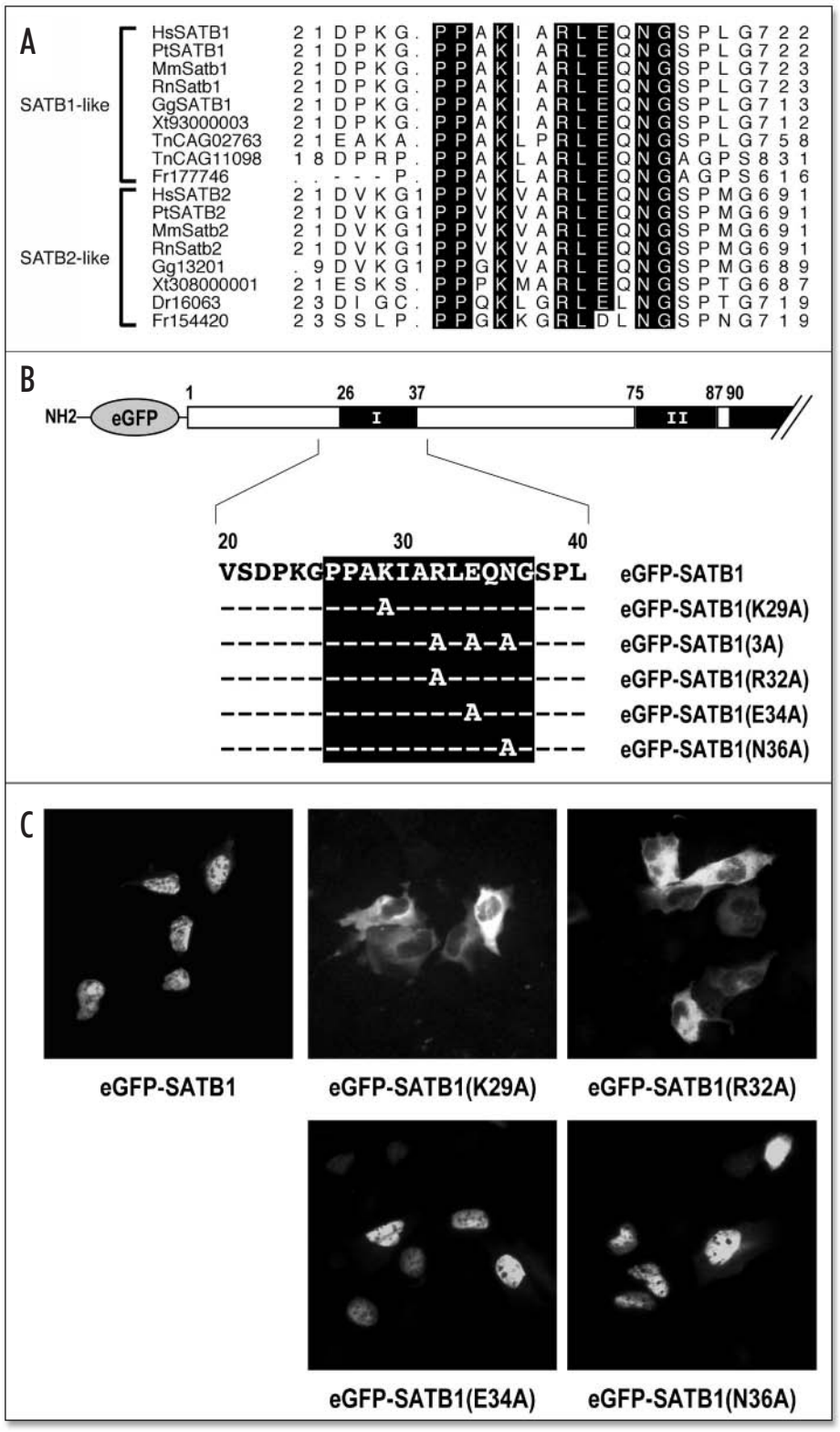

Figure 4. Identification of Lys29 and Arg32 as critical residues in the nuclear import of SATB1. (A) An HMM-generated alignment of the region labeled "SATB1-Nerm-I" in Figure 1, for vertebrate members of the SATB1 (the format is the same). The additional sequences (those not shown in Figure 1) are as follows: PtSATB1, *Pan troglodytes* SATB1 (chimpanzee, Ensembl identifier ENSPTRT00000027458); MmSatb1, *Mus musculus* Satb1 (mouse, NP_033148) RnSatb1, *Rattus norvegicus* Satb1 (rat, ENSRNOT00000017556); Fr177746, *Fugu rubripes* ORF (puffer fish, SINFRUT00000177746); PtSATB2, *P. troglodytes* SATB2 (ENSPTRT00000023677); MmSatb2, *M. musculus* Satb2 (NP_631885); RnSatb2, *R. norvegicus* Satb2 (ENSRNOT00000013781); Dr16063, *Danio rerio* ORF (zebra fish, ENSDART00000016063); and Fr154420, *F. rubripes* ORF (SINFRUT00000154420). (B) Point mutants in SATB1-Nterm-I. Positions of Ala substitutions used in this study are indicated. (C) Effects of Ala substitutions on its nuclear import activity. Plasmid expression constructs encoding eGFP-SATB1 mutants shown in (B) were transfected into U2OS cells and fixed. Images of representative cell fields were captured on a fluorescence microscope.

DISCUSSION

Using a combination of computational and experimental techniques, we identified a novel sequence motif, which is well conserved in vertebrate members of the SATB1 protein family, to be necessary and sufficient for nuclear localization. This sequence is located in the N-terminus of SATB1 and contains a previously unrecognized but well-conserved hexapeptide (R32LEQNG37). Mutation of the single Arg in this motif as well as its proximal Lys in the context of the full-length SATB1 protein is sufficient to prevent nuclear localization of SATB1.

Unlike most nuclear proteins characterized to date, SATB1 does not contain a "classical" NLS with short stretches of Arg and Lys similar to the basic type NLS previously identified in large T antigen of SV-40, nucleoplasmin,[26-28] p53,[29] hnRNP A1,[30,31] hnRNP K,[32] HuR,[33,34] red cell protein 4.1(4.1R)[35] and human immunodeficiency virus type 1 Rev.[36,37] There are several proteins in which non-classical NLSs have been identified. Nuclear localization of the trans-acting protein Tax of the human T-cell leukemia virus type I[38] and the snRNP protein U1A is mediated by large sequences.[39] However, the NLS of U1A still contains several basic amino acid clusters that are similar to the classical NLS. For Tax, the NLS which consists of the amino terminal 48 residues contains zinc finger-like cysteine-rich sequences. More recently, the two functionally independent NLSs of the αCP-RNA-binding protein have been identified, and neither of them shows any resemblance to NLSs already described.[40] In addition, nonclassical NLSs have been reported for human ADAR1,[41] the forkhead transcription factor AFX[42] and FUSE-binding protein (FBP) which binds to the far upstream element of the *c-myc* gene.[43] The SATB1 residues 20-40, which determine nuclear localization lacks similarity to any of these nonclassical NLSs as well as to other NLSs reported in the database.[44] Even though these residues do not contain a cluster of basic residues, it is significant that a single point substitution of either basic residue, Arg or Lys, within the region is sufficient to abrogate nuclear localization of the full-length SATB1

protein. Previous work has shown that the single Arg residue located in the BUR-binding domain of SATB1 is essential for its DNA-binding activity.[17] Furthermore, two adjacent N-terminal Arg residues of the homeodomain are critical for the binding specificity of SATB1. In conjunction with the present data on the nuclear targeting sequence of SATB1, it is evident that individual basic residues have critical roles in diverse functions of SATB1, and that this is most likely true for SATB2.

When T cells are induced to proliferate by mitogens, *Satb1* transcription is rapidly induced[3] and SATB1 appears in nuclei. Therefore, an active nuclear transport mechanism must be present upon T cell activation. Active protein transport from the cytoplasm to the nucleus is mediated by multiple families of adaptors and receptors.[26-28] Proteins carrying classical NLSs bind a cytoplasmic receptor, importinα, which associates with importin β, in order to be translocated to the nucleus. SATB1 residues 20-40 do not possess the previously described importin α and importin β recognition signals. We speculate that SATB1 may be recognized by novel factors of the intracellular transport machinery. Our data do not formally exclude the existence of a nuclear export signal in SATB1. In the case of p53 which contains a bipartitite NLS[29] and an NES,[45,46] both nuclear import and export activities are required for MDM2-mediated degradation of this protein.[47] It remains to be determined whether nucleocytoplasmic shuttling ever occurs for SATB1, and how SATB1 might be degraded when T cells become quiescent.

Our investigation of the phylogenetic distribution of members of the SATB1 protein family provides a number of insights and guidance for subsequent investigations. In the freshwater puffer fish, there are two SATB1-like proteins and two SATB2-like proteins, an observation consistent with recent analysis of the *Tetraodon* and human genomes which indicated that a whole-genome duplication occurred in the teleost fish lineage subsequent to its divergence from mammals.[48] The alignment reveals positions that distinguish SATB1-like proteins from SATB2-like proteins (yellow, Fig. 1) and hence residues that are good candidates for experimental studies of the functional differences between SATB1 and SATB2. The nonsynonymous, single nucleotide polymorphism in the coding region (cSNPs) observed in human SATB1 and SATB2 (open red triangles, Fig. 1) occur at similar relative locations in their respective proteins, but the significance of this remains still to be determined. Finally, a previous study identified sequence similarities between parts of human SATB2 and parts of the *Drosophila* protein *dve*.[19] The FlyBase report for *dve* (Fbgn0020307) indicates that this protein is associated with the Gene Ontology terms "AT DNA binding", "transcription factor activity", "regulation of transcription", "wing morphogenesis" and "nucleus". The results here suggest that this invertebrate protein may be the functional equivalent of vertebrate SATB1-like proteins: not only are the nature, number, and organization of known domains the same as in vertebrate members of the SATB1 protein family, but the sequence conservation extends also to the N-terminal and inter-domain regions (Dmdve, Fig. 1). Thus, the small number of amino acids that are invariant across all species (vertebrate and invertebrate) represent tractable targets for site-directed mutagenesis studies aimed at probing the biology of the SATB1 family. In addition to a motif required for nuclear localization, we expect to identify other key functional domains of SATB1 by this strategy.

References

1. Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T. A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. Cell 1992; 70:631-45.
2. Alvarez JD, et al. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev 2000; 14:521-35.
3. Beadling C, Johnson KW, Smith, KA. Isolation of interleukin 2-induced immediate-early genes. Proc Natl Acad Sci USA 1993; 90:2719-23.
4. Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T. SATB1 targets chromatin remodelling to regulate genes over long distances. Nature 2002; 419:641-5.
5. Cai S, Han HJ, Kohwi-Shigematsu T. Tissue-specific nuclear architecture and gene expression regulated by SATB1. Nat Genet 2003; 34:42-51.
6. Kohwi-Shigematsu T, Kohwi Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. Biochemistry 1990; 29:9551-60.
7. Cockerill PN, Yuen MH, Garrard WT. The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. J Biol Chem 1987; 262:5394-7.
8. Forrester WC, van Genderen C, Jenuwein T, Grosschedl R. Dependence of enhancer-mediated transcription of the immunoglobulin *mu* gene on nuclear matrix attachment regions. Science 1994; 265:1221-5.
9. Jenuwein T, et al. Extension of chromatin accessibility by nuclear matrix attachment regions. Nature 1997; 385:269-72.
10. Lichtenstein M, Keini G, Cedar H, Bergman Y. B cell-specific demethylation: A novel role for the intronic kappa chain enhancer sequence. Cell 1994; 76:913-23.
11. Kirillov A, et al. A role for nuclear NF-kappaB in B-cell-specific demethylation of the Igkappa locus. Nat Genet 1996; 13:435-41.
12. Fernandez LA, Winkler M, Grosschedl R. Matrix attachment region-dependent function of the immunoglobulin mu enhancer involves histone acetylation at a distance without changes in enhancer occupancy. Mol Cell Biol 2001; 21:196-208.
13. de Belle I, Cai S. Kohwi-Shigematsu T. The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. J Cell Biol 1998; 141:335-48.
14. Nakagomi K, Kohwi Y, Dickinson LA, Kohwi-Shigematsu T. A novel DNA-binding motif in the nuclear matrix attachment DNA-binding protein SATB1. Mol Cell Biol 1994; 14:1852-60.
15. Dickinson LA, Dickinson CD, Kohwi-Shigematsu T. An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. J Biol Chem 1997; 272:11463-70.
16. Galande S, Dickinson LA, Mian IS, Sikorska M, Kohwi-Shigematsu T. SATB1 cleavage by caspase 6 disrupts PDZ domain-mediated dimerization, causing detachment from chromatin early in T-cell apoptosis. Mol Cell Biol 2001; 21:5591-604.
17. Wang B, Dickinson LA, Koivunen E, Ruoslahti E, Kohwi-Shigematsu T. A novel matrix attachment region DNA binding motif identified using a random phage peptide library. J Biol Chem 1995; 270:23239-42.
18. Fanning AS, Anderson JM. Protein-protein interactions: PDZ domain networks. Curr Biol 1996; 6:1385-8.
19. FitzPatrick DR, et al. Identification of *SATB2* as the cleft palate gene on 2q32-q33. Hum Mol Genet 2003; 12:2491-501.
20. Dobreva G, Dambacher J, Grosschedl R. SUMO modification of a novel MAR-binding protein, SATB2, modulates immunoglobulin *mu* gene expression. Genes Dev 2003; 17:3048-61.
21. Diviacco S, et al. A novel procedure for quantitative polymerase chain reaction by coamplification of competitive templates. Gene 1992; 122:313-20.
22. Sherman MP, de Noronha CM, Heusch MI, Greene S, Greene WC. Nucleocytoplasmic shuttling by human immunodeficiency virus type 1 Vpr. J Virol 2001; 75:1522-32.
23. Xia J, Peng Y, Mian IS, Lue NF. Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. Mol Cell Biol 2000; 20:5196-207.
24. Bosoy D, Peng Y, Mian IS, Lue NF. Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. J Biol Chem 2003; 278:3882-90.
25. Peters R. Fluorescence microphotolysis to measure nucleocytoplasmic transport and intracellular mobility. Biochim Biophys Acta 1986; 864:305-59.
26. Nakielny S, Dreyfuss G. Transport of proteins and RNAs in and out of the nucleus. Cell 1999; 99:677-90.
27. Gorlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol 1999; 15:607-60.
28. Mattaj IW, Englmeier L. Nucleocytoplasmic transport: The soluble phase. Annu Rev Biochem 1998; 67:265-306.
29. Liang SH, Clarke MF. A bipartite nuclear localization signal is required for p53 nuclear import regulated by a carboxyl-terminal domain. J Biol Chem 1999; 274:32699-703.
30. Michael WM, et al. Signal sequences that target nuclear import and nuclear export of premRNA-binding proteins. Cold Spring Harb Symp Quant Biol 1995; 60:663-8.
31. Siomi H, Dreyfuss G. A nuclear localization domain in the hnRNP A1 protein. J Cell Biol 1995; 129:551-60.
32. Michael WM, Eder PS, Dreyfuss G. The K nuclear shuttling domain: A novel signal for nuclear import and nuclear export in the hnRNP K protein. EMBO J 1997; 16:3587-98.
33. Fan XC, Steitz JA. HNS, a nuclear-cytoplasmic shuttling sequence in HuR. Proc Natl Acad Sci USA 1998; 95:15293-8.

34. Peng SS, Chen CY, Xu N, Shyu AB. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. EMBO J 1998; 17:3461-70.
35. Gascard P, Nunomura W, Lee G, Walensky LD, Krauss SW, Takakuwa Y, Chasis JA, Mohandas N, Conboy JG. Deciphering the nuclear import pathway for the cytoskeletal red cell protein 4.1R. Mol Biol Cell 1999; 10:1783-98.
36. Meyer BE, Malim MH. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. Genes Dev 1994; 8:1538-47.
37. Stauber R, Gaitanaris GA, Pavlakis GN. Analysis of trafficking of Rev and transdominant Rev proteins in living cells using green fluorescent protein fusions: Transdominant Rev blocks the export of Rev from the nucleus to the cytoplasm. Virology 1995; 213:439-49.
38. Smith MR, Greene WC. Characterization of a novel nuclear localization signal in the HTLV-I tax transactivator protein. Virology 1992; 187:316-20.
39. Kambach C, Mattaj IW. Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. J Cell Biol 1992; 118:11-21.
40. Chkheidze AN, Liebhaber SA. A novel set of nuclear localization signals determine distributions of the alphaCP RNA-binding proteins. Mol Cell Biol 2003; 23:8405-15.
41. Eckmann CR, Neunteufl A, Pfaffstetter L, Jantsch MF. The human but not the *Xenopus* RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein. Mol Biol Cell 2001; 12:1911-24.
42. Brownawell AM, Kops GJ, Macara IG, Burgering BM. Inhibition of nuclear import by protein kinase B (Akt) regulates the subcellular distribution and activity of the forkhead transcription factor AFX. Mol Cell Biol 2001; 21:3534-46.
43. He L, Weber A, Levens D. Nuclear targeting determinants of the far upstream element binding protein, a c-myc transcription factor. Nucleic Acids Res 2000; 28:4558-65.
44. Nair R, Carter P, Rost B. NLSdb: Database of nuclear localization signals. Nucleic Acids Res 2003; 31:397-9.
45. Roth J, Dobbelstein M, Freedman DA, Shenk T, Levine AJ. Nucleo-cytoplasmic shuttling of the hdm2 oncoprotein regulates the levels of the p53 protein via a pathway used by the human immunodeficiency virus rev protein. EMBO J 1998; 17:554-64.
46. Stommel JM, Marchenko ND, Jimenez GS, Moll UM, Hope TJ, Wahl GM. A leucine-rich nuclear export signal in the p53 tetramerization domain: Regulation of subcellular localization and p53 activity by NES masking. EMBO J 1999; 18:1660-72.
47. O'Keefe K, Li H, Zhang Y. Nucleocytoplasmic shuttling of p53 is essential for MDM2-mediated cytoplasmic degradation but not ubiquitination. Mol Cell Biol 2003; 23:6396-405.
48. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy J, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quetier F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissenbach J, Roest Crollius H. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 2004; 431:946-57.