Positional Cloning of the Mouse Circadian *Clock* Gene

David P. King, **Yaliang Zhao**, **Ashvin M. Sangoram**, **Lisa D. Wilsbacher**, **Minoru Tanaka**[*],
Marina P. Antoch, **Thomas D. L. Steeves**, **Martha Hotz Vitaterna**, **Jon M. Kornhauser**[†],
Phillip L. Lowrey, **Fred W. Turek**, and **Joseph S. Takahashi**
National Science Foundation Center for Biological Timing

Department of Neurobiology and Physiology Northwestern University Evanston, Illinois 60208

Summary

We used positional cloning to identify the circadian *Clock* gene in mice. *Clock* is a large transcription unit with 24 exons spanning ~100,000 bp of DNA from which transcript classes of 7.5 and ~10 kb arise. *Clock* encodes a novel member of the bHLH–PAS family of transcription factors. In the *Clock* mutant allele, an A T nucleotide transversion in a splice donor site causes exon skipping and deletion of 51 amino acids in the CLOCK protein. *Clock* is a unique gene with known circadian function and with features predicting DNA binding, protein dimerization, and activation domains. CLOCK represents the second example of a PAS domain–containing clock protein (besides Drosophila PERIOD), which suggests that this motif may define an evolutionarily conserved feature of the circadian clock mechanism.

Introduction

Circadian (~24 hr) rhythmicity of biological processes is a fundamental property of all eukaryotic and some prokaryotic organisms (Takahashi, 1995). Evidence gathered over the past four decades demonstrates that these rhythms are driven by an internal time-keeping system (Pittendrigh, 1993). Changes in the external environment, particularly in the light-dark cycle, entrain this biological clock. In contrast, under constant environmental conditions devoid of time cues, rhythms driven by the clock free run with a period near, but usually not equal to, 24 hr (Pittendrigh, 1993; Turek and Van Reeth, 1996).

Considerable progress has been made toward elucidating the physiological basis of mammalian circadian rhythms. There is now substantial evidence that the bilaterally paired suprachiasmatic nuclei (SCN) of the hypothalamus contain a master "circadian clock" that regulates most, if not all, circadian rhythms in mammals (Meijer and Rietveld, 1989; Klein et al., 1991; Moore, 1995). In contrast, less progress has been made toward identifying the genetic basis of mammalian circadian rhythms (Schwartz and Zimmerman, 1990; Lynch and Lynch, 1992). An exception to this general finding is the spontaneous mutation, *tau*, found in the golden hamster (Ralph and Menaker, 1988). *tau* is a semi-dominant, autosomal mutation that shortens the circadian period by about two hours in heterozygous mutants and by about four hours in homozygous mutants. Studies of the *tau* mutation have clearly demonstrated that a single gene can affect mammalian circadian rhythms and continue to provide important insights into the physiology of circadian rhythms; however, due to the

paucity of genetic information in hamsters, this mutation has not been useful for the molecular dissection of circadian rhythmicity.

Three genes essential for circadian rhythms have been cloned: the *period* (*per*) and *timeless* (*tim*) genes in Drosophila and the *frequency* (*frq*)gene in Neurospora (reviewed in Hall, 1995; Dunlap, 1996; Rosbash et al., 1996). Considerable progress has been made in characterizing these genes at the transcriptional and translational levels (Hardin et al., 1990, 1992; Aronson et al., 1994; Edery et al., 1994; Sehgal et al., 1995; Dunlap, 1996; Rosbash et al., 1996). Unfortunately, it has not been possible to extend this work to vertebrates by cloning orthologous genes. For these reasons, we have initiated a genetic approach to the molecular analysis of circadian rhythms in mammals. Using the mouse as a tool for gene discovery, our approach involves three phases: (1) the isolation and analysis of circadian rhythm mutants using phenotype-driven N-ethyl-N-nitrosourea (ENU) mutagenesis screens (Dove, 1987); (2) the identification of the genes affected by these mutations using candidate gene and positional cloning methods (Takahashi et al., 1994); and (3) the elucidation of the function of these genes to understand the pathways in which they interact to generate circadian oscillations.

Using this forward genetic approach, we identified a single-gene mutation that dramatically alters the phenotypic expression of circadian rhythmicity in mice (Vitaterna et al., 1994). This mutation, named *Clock*, affects at least two fundamental properties of the circadian system: the length of the free-running period and the persistence of circadian rhythmicity in constant darkness. We have mapped *Clock* previously by linkage analysis to the mid-portion of chromosome 5 (King et al., 1997). Here, we report the molecular identification of a circadian clock gene in mammals and describe the nature of a mutation in this gene that alters the phenotypic expression of circadian rhythmicity in mice.

Results

Genetic and Physical Mapping of *Clock*

As an initial step toward positional cloning, *Clock* was localized by high resolution genetic mapping to the mid-portion of mouse chromosome 5, 0.7 cM distal of *Kit* (= *W, Dominant white spotting*) (King et al., 1997). The *Clock* mutation originated on a C57BL/6J (B6) strain background (Vitaterna et al., 1994). BALB/cJ (BALB), C3H/HeJ (C3H), and Mus castaneus (CAST/Ei) were used as counterstrains to generate 12 mapping cross panels segregating the *Clock* mutation. Including 1804 meioses reported in King et al. (1997), we have now genotyped over 2400 meioses obtained from these crosses. We have used simple sequence length polymorphisms (SSLPs) (Dietrich et al., 1992) to map *Clock* to a 0.2 cM interval, 0.1 cM distal from *D5Mit307* (3 recombinants/2401 meioses) and 0.1 cM proximal to *D5Mit112* (1 recombinant/845 meioses) (Figure 1A).

To identify genomic clones that map to the genetic interval containing *Clock*, we screened three yeast artificial chromosome (YAC) libraries (Larin et al., 1991; Chartier et al., 1992; Kusumi et al., 1993). We identified 42 YAC clones from one of these libraries (Kusumi et al., 1993), using SSLPs from the *Clock* region and sequence-tagged sites (STS) derived from the sequence of loci mapping close to *Clock* (*Pdgfra, Kit, Flk1, yREB16*, and *yREB14*) (Brunkow et al., 1995; King et al., 1997). Of these clones, 32 define a contig about 4 megabases in length that spans the nonrecombinant region containing *Clock* (data not shown). An additional clone that spans the entire *D5Mit307–D5Mit112* interval (YAC 55 in Figure 1B) was identified from YAC DNA pools of the other two libraries (maintained by the Cloning Core Laboratory at the Baylor College of Medicine). Long-range restriction mapping of these YAC clones with NotI, NruI, and EagI revealed two CpG islands (Antoch et al., 1997 [this issue of *Cell*]) and allowed us to estimate the *D5Mit307– D5Mit112*

interval to be about 400 kb (Figure 1B). Two YACs mapping to the *Clock* region (YACs 55 and 40 in Figure 1B) appear to be intact, nonchimeric clones. Further analysis of this YAC contig indicated that about half of these clones are chimeric (D. P. K. et al., unpublished data). To obtain clones that are more amenable to manipulation and less likely to be chimeric, we isolated bacterial artificial chromosome (BAC) clones (Shizuya et al., 1992), using the genetic markers that most closely flank *Clock* (*D5Mit307* and *D5Mit112*) as well as the nonrecombinant SSLPs and the STSs that map (by YAC content mapping) to the *D5Mit307–D5Mit112* interval. The BAC contig of the minimal genetic interval containing *Clock* consists of 11 overlapping BACs and 1 independent BAC with one gap between the distal end of BAC 50 and the proximal end of BAC 52 (Figure 1B).

To reduce the minimal genetic interval containing *Clock* further, one BAC clone (BAC 51) was used to identify new SSLPs. From two small-insert M13 sub-clone libraries of this BAC, we identified seven simple sequence repeats (SSRs) (*D5Nwu1–D5Nwu7* in Figure 1B). All of these SSRs reveal interspecific polymorphisms between the B6 strain and CAST/ Ei, and *D5Nwu4*, *D5Nwu5*, *D5Nwu6*, and *D5Nwu7* reveal intra-specific polymorphisms between the B6 and C3Hstrains as well. The closest distal recombination, detected in animal t1134, is also recombinant for five of these new markers and defines *D5Nwu2* as the closest distal marker (Figure 1B). An additional set of SSRs were later obtained from sequence analysis of the BAC 52 shotgun sequencing project (see below). These SSLPs, *D5Nwu8– D5Nwu14*, also identify interspecific polymorphisms between B6 and CAST/Ei; however, none of them detected recombinations with *Clock*.

Candidate Gene Identification

No previously identified genes had been mapped to the non-recombinant interval containing the *Clock* locus. Therefore, we took three different approaches to identify candidate genes: (1) direct screening of SCN cDNA libraries using BACs 51 and 52 as probes; (2) cDNA selection of SCN cDNA using BAC 51 as "driver;" and (3) shotgun sequencing of M13 libraries sub-cloned from BACs 52 and 54. The first two of these methods used cDNA libraries constructed from the mouse SCN. These libraries were important resources for the analysis of expressed sequences because the SCN is very small (about 16–20,000 neurons), and it is therefore difficult to obtain high quality mRNA samples.

For direct screening, two different BAC clones (BACs 51 and 52), which together cover "3/4 of the genetic region containing *Clock*, were used as complex probes to screen the SCN cDNA libraries. The cDNAs isolated using this method were characterized in two ways. The ends of the clones were sequenced, and these sequences were used to search the DNA and protein sequence data bases with the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990). In addition, these cDNA clones were used as probes on Southern blots consisting of six HindIII-digested BAC clones that map to the *Clock* region (BACs 48, 50, 51, 52, 53, and 54; see Figure 1B). We isolated 15 cDNA clones with this method, and 6 classes of clones mapped to the *Clock* region.

The second method of transcription unit identification was an adaptation of the cDNA selection protocol developed by Lovett (1994). We used biotinylated BAC 51 DNA "amplicons" as driver to enrich the SCN cDNA "amplicons" with cDNA fragments that map to the *Clock* region. Following two rounds of this selection, the DNA obtained (representing restriction-digested fragments of SCN cDNA clones) was cloned, and 576 clones were picked into six 96-well plates. To characterize these clones, we made replica filters from these plates and then screened them with the following probes: BAC 51 (positive probe), BAC 48 (negative probe), c-*fos* (a positive control for the selection method; see Experimental Procedures), and COT-1 DNA. We identified 60 clones that were positive for the BAC 51 probe and negative for the other three probes. We sequenced these 60 selected

clones and tested them for hybridization to the BAC Southern blots described above. Of the 60 clones, 38 appeared valid by sequence, 14 had repetitive sequences, and 8 were false positives (ribosomal or vector DNA). All 38 positive clones mapped to the *Clock* region on BAC Southern blots. These cDNA-selected fragments appeared to represent about 13 different classes of expressed sequences. These fragments were then used to screen the SCN cDNA libraries to obtain complete cDNA clones. From this screen, we identified 25 types of cDNA clones that mapped to the *Clock* region.

In addition to the cDNA-directed approaches described above, we used shotgun sequencing of genomic DNA as a third method of transcription unit identification. Genomic sequence of this region was also useful for several other purposes, including exon mapping of cDNAs, analysis of transcription units, interpretation of BAC rescue experiments (see Antoch et al., 1997), and *Clock* mutation identification and analysis. We chose to sequence BAC clones 52 and 54 based on genetic information that defined the interval containing *Clock* (Figure 1). We sequenced the ends of ~1300 M13 sub-clones from BAC 52 and ~1500 M13 sub-clones from BAC 54, with an average read length of ~580 bases, which yielded about 4-fold sequence coverage of each BAC. Sequences were assembled into contigs and were placed on the physical map by alignment with SSLPs and STSs that map to the *Clock* region. Each M13 sequence was searched against nucleotide and protein data bases using several BLAST algorithms (BLASTN-nr, BLASTX-nr, BLASTN-dbEST, and TBLASTX-dbEST) to find regions of sequence similarity. We identified one candidate gene from the non-redundant nucleotide data base, *pFT27* (GenBank accession no. M23568; Akagi et al., 1988), that was completely identical to exons in the genomic sequence (Figure 1C). In addition, we found candidate sequences that were homologous (i.e., similar but not identical in nucleotide sequence) to human *bendless* (D83004;Muralidharand Thomas,1993;Yamaguchi et al., 1996), mouse *adenomatous polyposis coli* (*APC*) mRNA (M88127; Su et al., 1992), the g subunit of bovine ATP synthase Fo membrane protein (S70448; Collinson et al., 1994), and human *MOP4* (U51625; Hoganesch et al., 1997). The *bendless* similarity (90% nucleotide identity) is addressed in the accompanying paper (Antoch et al., 1997). The similarity to mouse *APC* (92% nucleotide identity) occurred in the 3 untranslated region and extended off the proximal end of BAC 52 in the 5 direction. The *MOP4* similarities (69% nucleotide identity) were scattered throughout the genomic sequence, and one locus contained ATP synthase similarity (73% nucleotide identity). In addition to the candidate sequences found in the non-redundant nucleotide data base, we also identified 25 expressed sequence tags (of both mouse and human cDNAs) in the EST data base that fell into 10 classes. Seven of these EST classes aligned to a ~4 kb region of genomic sequence on the proximal end of BAC 54, and several shotgun sequences defined mouse EST identities.

To determine the expression patterns of selected candidate sequences, we probed Northern blots of total RNA collected at Zeitgeber Time 6 (ZT6, 6 hr following light onset) from hypothalamic tissue and eyes of wild-type and *Clock/Clock* mutant mice. As described in the accompanying paper (Antoch et al., 1997), data from the transgenic BAC complementation analysis indicated that the wild-type circadian phenotype was completely rescued in *Clock* mutant mice by BAC 54. Therefore, candidate gene expression analysis was limited to clones mapping to this 140 kb region. Hybridization patterns detected on the Northern blots suggested the existence of at least four different genes within this interval (data not shown). Subsequently, most of these expression patterns were determined to be the result of either chimeric cDNA clones (two cases) or non-identical clones that cross-hybridized to BAC clones from the *Clock* region (one case). One M13 shotgun sequencing clone with sequence similarity to the predicted PERIOD-ARNT-SIM (PAS) domain (Nambu et al., 1991) of a recently identified basic–helix-loop-helix (bHLH)–PAS gene (*MOP4/NPAS2*, Hoganesch et al., 1997; Zhou et al., 1997) detected two transcripts of ~8 kb and ~11 kb in size and revealed an RNA abundance difference between wild-type and

Clock/Clock mice in both hypothalamus and eye. A cDNA clone (yz50) also having sequence similarity to the *MOP4/NPAS2* gene was then used as probe. This clone also detected two transcripts, ~8 kb and ~11 kb in size, and revealed a difference in abundance between the wild-type and *Clock/Clock* RNA samples (Figure 2). No other candidate genes detected RNA differences (of abundance or size) between the wild type and *Clock/Clock* mutant genotypes with this Northern blot analysis. Based on these RNA expression differences and the sequence similarity to the PAS domain of the MOP4/NPAS2 protein, this gene became our primary candidate for the *Clock* locus.

Sequence of the *Clock* Gene

We next focused our attention on the identification of the complete cDNA sequence of this candidate gene for the *Clock* locus. During the course of candidate gene identification, we isolated several cDNA clones from the SCN cDNA libraries that mapped by hybridization to the minimal genetic interval. Alignment of 5 and 3 end sequences from these cDNA clones with the genomic sequence contigs of BACs 52 and 54 suggested that many of them were partial transcripts of our primary candidate. To characterize these clones further and to determine complete expressed sequences of the gene, all of these cDNA clones were sequenced.

We identified 11 different types of cDNA clones (Figure 3A). Based on exon mapping of this set of cDNA clones, the transcribed region of this *Clock* candidate gene spans almost 100,000 base pairs of genomic sequence and contains 24 exons. Two of the exons (exons 1a and 1b) are distal to a set of three NotI sites that identify a CpG island. There is alternative use of exons 1a and 1b in clones yz50, L8 and yz80. Interestingly, yz80, the only apparently full-length clone beginning with exon 1a, skips exons 19–22 and part of exon 23. In addition, the partial cDNA clone L7c skips exons 18 and 19. The complete cDNA sequence of this gene is ~10,000 nucleotides in length, with an open reading frame of 2565 bases extending from the middle of exon 4 to the 5 end of exon 23. The first four exons (1a, 1b, 2, and 3; Figure 3A) contain 5 un-translated sequence that originates from a CpG island. The presence of a CpG island is characteristic of the promoter regions of both constitutively expressed and tissue-specific genes (Bird, 1992; Kozak, 1996). Within the 5 untranslated region, there are multiple stop codons in all three reading frames, including seven stop codons in-frame preceding the ORF. There are two methionine codons at the beginning of the ORF (see Figure 3B). The nucleotide sequence surrounding both of these methionine codons corresponds well to the Kozak consensus translation initiation sequence, with purines at the −3 and at the +4 positions (Kozak, 1996). The scanning translation initiation hypothesis predicts that the first of these methionines is likely to be the start codon for this protein (Kozak, 1996). Following a termination codon (TAG) in exon 23, there is a very long 3 untranslated sequence ("6 kb), with a polyadenylation signal sequence at ~7450 bp (identified by the subset of cDNA clones with poly[A] tails immediately downstream: yz80, L4, yz161, and yz197). The sequence of this 3 un-translated region is AT rich, with eight ATTTA sites, which are often found in transcripts with short half-lives (Sachs, 1993). In addition, in the genomic sequence ~*2*500 bp downstream from the first polyadenylation site, there are additional polyadenylation signal sequences that could be utilized to form transcripts of ~10 kb. Several expressed sequence tags (ESTs) identified in the NCBI EST data base align to this region and appear to utilize these polyadenylation signals. The 7.5 and ~10 kb transcript sizes based on cDNA and genomic sequence correspond extremely well to the ~8 kb and ~11 kb mRNA transcript sizes estimated from Northern blot analysis.

Translation of the open reading frame from this *Clock* candidate predicts a protein of 855 amino acids, of which 112 are glutamine and 100 are serine residues (Figure 3B). This protein has an expected mass of 96.4 kDa and an isoelectric point of 6.52. Within the predicted protein sequence, there are multiple potential phosphorylation sites and two N-

linked glycosylation sites, implicated in the function of transcription factors (Mitchell and Tjian, 1989). The glutamine residues are concentrated in the C-terminal half of the protein, which is highlighted by a poly-glutamine repeat of 17 residues (interrupted by two proline residues; amino acids 751– 769).

A search of the NCBI data base reveals that the nucleotide sequence of this candidate gene is novel. There are only three nucleotide sequences with significant similarity: mouse *NPAS2* (U77969, 67% identical), human *MOP4* (U51625, 68% identical), and human *NPAS2* (U77970, 69% identical; human *MOP4* and *NPAS2* appear to be cDNAs of the same gene, as they are >99.5% identical across 1900 overlapping bases). All of these clones were recently identified in screens for new bHLH– PAS domain-containing genes (Hoganesch et al., 1997; Zhou et al., 1997). A search of the NCBI data base with the predicted protein sequence of this *Clock* candidate gene shows highly significant similarity to the predicted protein sequences of these same three clones as well as significant similarity to a large number of bHLH–PAS proteins. The candidate protein has very high amino acid sequence similarity to the mouse NPAS2 and human MOP4/NPAS2 proteins in the bHLH region as well as in the PAS-A and PAS-B domains. We also detected sequence similarity in the PAS domains to the Drosophila PERIOD protein. Based on these sequence alignments, we predict that this candidate gene encodes a novel member of the bHLH–PAS domain family of transcription factors (Figure 4A). Amino acid alignments of the conserved regions of the predicted protein to several other members of this family are shown in Figures 4B–4D.

The similarity between the predicted protein of this *Clock* candidate gene and mouse NPAS2 is quite extensive. In addition to the similarities in the bHLH and PAS domains (see Figures 4B–4D), it also shares significant similarity to NPAS2 in a 28 residue serine-rich region (amino acids 427–454; 83% identity), and a 55 residue glutamine-rich region (amino acids 515–569; 64% identity). Although the C-terminal regions of both proteins are glutamine rich, they diverge substantially in sequence. In addition, the candidate CLOCK protein, unlike NPAS2, has a poly-glutamine repeat near the C terminus of the protein (amino acids 751–769), which is characteristic of the activation domains of many transcription factors (Mitchell and Tjian, 1989).

Identification of the *Clock* Mutation

To determine whether the mutation conferring the altered circadian phenotypes in *Clock* mice affects this candidate gene, we compared the nucleotide sequence of wild-type and *Clock/Clock* genomic DNA containing exons encoding this bHLH–PAS protein. Because the *Clock* mutant allele was ENU-induced and has been maintained as a coisogenic strain (Vitaterna et al., 1994), we searched for point mutations. We used genomic sequence from shotgun sequencing to design PCR primers that flank all exons bearing coding sequence. Each exon, as well as ~100–300 bp of flanking intronic sequence, was amplified by PCR from the genomic DNA of four wild type and four *Clock/Clock* homozygous B6 coisogenic mice. Both strands of these PCR products were then sequenced using the appropriate PCR primers. From a total of 160 PCR products amplified from the coding and flanking intronic sequence of exons4–23, we detected one nucleotide change in the DNA sequence of *Clock/Clock* mice, an A T transversion at the third base position of the 5 splice donor site of intron 19 (Figure 5A). The consensus nucleotide for this position is adenosine, and furthermore, it is almost always a purine (96%) (Krawczak et al., 1992). Analysis of all of the intron–exon junctions of *Clock* revealed that an adenosine is present at this position in 20 of 23 5 splice donor sites; the remaining three have guanosine at this position (data not shown). Other purine pyrimidine mutations at this position in the 5 splice donor site have caused skipping of the exon immediately upstream (Krawczak et al., 1992). Based on the identification of this nucleotide transversion and in conjunction with the demonstration of

phenotypic rescue described in the accompanying paper (Antoch et al., 1997), we conclude that this gene encodes the *Clock* locus.

To determine if this mutation has an effect on the exon splicing of the *Clock* gene, we performed reverse transcription (RT)–PCR on RNA samples collected from hypothalamic tissue of wild-type and *Clock*/*Clock* B6 coisogenic mice. PCR primers designed from the cDNA sequence of exons 15 and 21, which flank the potentially skipped exon, were used to amplify a PCR product with a wild-type size of 761 bp. Consistent with the predicted effect of this mutation, the RT–PCR product from *Clock*/ *Clock* RNA is "150 bases smaller than the wild-type product (Figure 5B). We also detected a second, fainter band in wild-type and *Clock*/*Clock* RT–PCR products that is "100 bases shorter than the primary band. All eight of these bands were gel purified, reamplified by PCR, and sequenced. Sequence analysis confirmed that exon 19 is missing from *Clock* mRNA (Figure 5C). The additional, more rapidly migrating band resulted from the variable splicing of exon 18 (Figure 5C). The exon sizes are in close agreement with the observed band size differences: exon 19 is 153 bases long and exon 18 is 90 bases long (data not shown). Interestingly, the cDNA clone L7c (Figure 3A) is missing these exons, suggesting that this clone probably originated from *Clock* mutant mRNA.

The alternative splicing of exon 18 in both wild type and *Clock* mutant transcripts causes a 30 amino acid deletion (amino acids 484–513) of the predicted CLOCK protein without any change in the reading frame. The deletion of exon 19 in *Clock* mutant transcripts causes a 51 amino acid deletion within the glutamine-rich region of the C terminus of the predicted CLOCK protein (amino acids 514–564). As with exon 18, the deletion of exon 19 is not expected to cause any changes in the reading frame of the CLOCK protein. Interestingly, exon 19 defines the only region of amino acid sequence similarity (61% identity) between CLOCK and NPAS2 in the C-terminal half of the predicted proteins.

Clock mRNA Expression and Sequence Conservation

Our initial Northern blot analysis indicated that *Clock* is expressed in hypothalamus and eye. To characterize the spatial expression of *Clock* further, mRNA expression in mouse brain was determined by in situ hybridization. In brain tissue collected at ZT6, there is abundant expression of *Clock* mRNA within the SCN, as well as in the pyriform cortex (Figures 6A–6D). There is also strong expression of *Clock* in the hippocampus (data not shown). In addition, we detected low-level expression throughout the rest of the brain.

In order to survey the RNA expression patterns of *Clock* more broadly, we analyzed by Northern blot total RNA collected at ZT6 from several mouse tissues. This analysis indicates that expression of the *Clock* gene is widespread (Figure 6E). In addition to the hypothalamus and the eye, *Clock* is also expressed in brain (without hypothalamic tissue), testes, ovaries, liver, heart, lung, and kidney. In most of these tissues, RNA expression levels appear to be similar to those seen in hypothalamus and eye. When detectable, the hybridization pattern (i.e., two transcripts of "8 kb and "11 kb in size) is the same as that seen in hypothalamus and eye.

The amino acid sequence similarity to several bHLH– PAS genes in organisms other than mouse suggests that *Clock* may be a conserved gene. To determine if the *Clock* gene sequence is conserved in other organisms, genomic DNA samples from several vertebrate species were analyzed by Southern blot analysis. Using PCR-amplified probes from the PAS-B and C-terminal (exons 15–21) regions of *Clock*, we detected hybridization in human, golden hamster, chicken, Anolis lizard, Xenopus frog, and zebrafish DNA samples, in addition to mouse DNA (Figure 7).

Discussion

Using the circadian phenotype of an ENU-induced mutation (Vitaterna et al., 1994) as a guide for positional cloning, we have identified the molecular nature of the gene at the *Clock* locus. This finding, in conjunction with the functional rescue of circadian phenotype in transgenic mice expressing a BAC clone containing the *Clock* gene (Antoch et al., 1997), provides definitive proof that this gene is involved in the organization of the mammalian circadian clock system. The *Clock* gene is a large transcription unit that spans ~100,000 base pairs of genomic sequence and contains 24 exons (one for each hour of the day). The predicted amino acid sequence of the *Clock* gene product indicates that *Clock* encodes a novel member of the bHLH–PAS domain family of transcription factors. An A   T nucleotide transversion in a splice donor site that results in exon skipping and a deletion of 51 amino acids in the protein product appears to be the cause of a number of changes in circadian phenotype, most notably a 3–4 hr lengthening of circadian period, usually followed by a loss of circadian rhythmicity in constant darkness, in animals homozygous for the mutant *Clock* allele. That CLOCK contains both a DNA-binding domain and protein dimerization domains indicates that it could, in combination with other proteins, regulate circadian rhythmicity by regulating gene transcription.

The pattern of *Clock* expression is consistent with its role in circadian rhythms (Vitaterna et al., 1994; King et al., 1997). Of the neural tissues examined, the levels of expression were highest in hypothalamus and eye, both of which are tissues known to contain self-sustaining circadian oscillators (Moore, 1995; Tosini and Menaker, 1996). The suprachiasmatic nuclei of the hypothalamus contain a "master" circadian pacemaker, which is both necessary and sufficient for determining clock properties of overt circadian rhythmicity (Ralph et al., 1990; Klein et al., 1991). Because the *Clock* mutation affects pacemaker properties of circadian rhythms (period and persistence) (Vitaterna et al., 1994), the enriched expression of *Clock* mRNA in the SCN is consistent with its role in circadian organization. Interestingly, *Clock* is ex pressed in many tissues throughout the body. The Drosophila *period* gene has a similarly wide tissue distribution pattern. It will be interesting to determine whether *Clock* expression in non-pacemaker tissues plays a role in circadian rhythm expression in those tissues, and if, like *per*, it is involved in non-circadian oscillations as well (Kyriacou and Hall, 1980).

From an evolutionary perspective, it is significant that the *Clock* gene appears to be highly conserved among vertebrate species, including humans. It is well established that circadian organization among the vertebrates involves circadian pacemakers in the SCN, retina, and pineal gland (Takahashi, 1991). The identification of *Clock* will provide an important tool for analyzing these phylogenetic relationships at the molecular level.

A distinguishing characteristic of the *Clock* gene is the presence of several conserved domains in the amino acid sequence, including the basic–helix-loop-helix, PAS-A, PAS-B, and C-terminal Q-rich domains (Mitchell and Tjian, 1989; Murre et al., 1989; Nambu et al., 1991). Like other bHLH transcription factors, members of the bHLH–PAS domain transcription factor superfamily bind DNA and modulate transcription following dimerization (Murre et al., 1989; Swanson and Bradfield, 1993). This family of transcription factors has been shown to play important roles in both development and nuclear signaling (Nambu et al., 1991; Burbach et al., 1992; Gradin et al., 1996; Wilk et al., 1996). In Drosophila, two of these genes, *single-minded* (*sim*) and *tracheless* (*trh*), are "master regulators" of neural midline and tracheal development, respectively (Nambu et al., 1991; Wilk et al., 1996). Each of these genes, once induced, maintains expression levels during the development of its appropriate tissues by autoregulation (Nambu et al., 1991; Wilk et al., 1996). At least two other bHLH–PAS proteins, the aryl hydrocarbon receptor (AHR) and the

hypoxia inducible factor *a* (HIF*a*), respond to cellular distress by activating response genes (Burbach et al., 1992; Gradin et al., 1996). These proteins are usually bound to HSP90, but upon dioxin binding or hypoxia, respectively, they interact with the aryl hydrocarbon nuclear translocator (ARNT) (Hoffman et al., 1991), another member of the bHLH–PAS gene family, to regulate the transcription of target genes.

The knowledge that *Clock* appears to be a bHLH–PAS gene allows us to make some predictions regarding the effect of the mutant allele. The A   T transversion results in the deletion of exon 19, which is situated within the glutamine-rich C-terminal region of *Clock*. Glutamine-rich regions of transcription factors are associated with transactivation of the transcription initiation complex (Mitchell and Tjian, 1989). The mutation presumably would not have a significant effect on the N-terminal bHLH and PAS domains, leaving *Clock* dimerization and DNA binding intact. Thus, we hypothesize that the mutant CLOCK protein could still dimerize (probably as a heterodimer with another bHLH–PAS protein) and bind DNA, but it would be less able to activate transcription. This model could explain two features of the *Clock* mutation: its anti-morphic behavior and the reduced expression of *Clock* mRNA in homozygous mutant mice. We have demonstrated that *Clock* is an anti-morph, a "competitive" type of dominant-negative mutation (King et al., 1997). This conclusion comes from the analysis of the phenotypic effect of a *Clock* null allele (in this case, a ~3 cM radiation-induced deletion that contains *Clock*) in trans with the wild type or *Clock* mutant allele. The circadian period of *Clock*/null mice is significantly longer than the period of *Clock*/+ mice, indicating that the wild-type allele is interacting with the mutant allele to lessen the severity of its phenotypic effect. This is the essential feature of an anti-morphic allele (Muller, 1932). A transcription factor that is intact except for its transactivation domain has been suggested as a hypothetical dominant-negative mutation (Herskowitz, 1987). The resulting transcription factor would competitively dimerize and bind DNA, but would be less able to activate transcription, thus disrupting the activity of the wild-type allele. The mutant phenotype would presumably be worse in the absence of a wild-type copy of the gene, in keeping with Muller's description of an anti-morphic mutation. The nature of the *Clock* mutation is entirely consistent with this type of dominant-negative mutation. The reduced expression of *Clock* mRNA in *Clock*/*Clock* homozygotes could then be explained if the *Clock* protein product positively affects its own transcription.

It is of particular interest that *Clock* defines the second example of a circadian clock gene with a PAS dimerization domain. In Drosophila, the *period* and *timeless* genes constitute two basic elements of a 24hr molecular oscillation of transcription and translation that form an auto-regulatory feedback loop (Siwicki et al., 1988; Hardin et al., 1990, 1992; Zeng et al., 1994; Myers et al., 1995; Sehgal et al., 1995; Lee et al., 1996). While it is established that the PERIOD (PER) and TIMELESS (TIM) proteins heterodimerize and translocate to the nucleus (Gekakis et al., 1995; Saez and Young, 1996), neither of these proteins appears capable of DNA binding. Thus, it has been proposed that PER could act as a transcriptional regulator by interacting with a transcription factor either as a "partner" or as a dominant-negative regulator (Takahashi, 1992; Huang et al., 1993). In fact, recent evidence indicates that PER can interact with bHLH– PAS proteins to inhibit their ability to activate transcription: protein–protein interactions via the PAS domain have been demonstrated in vitro between PER and SIM, AHR, and ARNT (Huang et al., 1993; Lindebro et al., 1995). These results have led to the idea that there exist as yet unidentified genes interacting with PER (presumably via the PAS domain) that are transcription factors. The mouse *Clock* gene could be the mammalian ortholog of such a gene.

Recent analysis of the *per* promotor demonstrates that a 69 bp enhancer sequence is sufficient for high amplitude circadian cycling of *per* transcription (Hao et al., 1997). Interestingly, within this 69 bp region, there is a consensus E box sequence (CACGTG),

which is a bHLH DNA–binding motif (Murre et al., 1989). Mutations induced within this E box abolish high amplitude *per* mRNA cycling, but do not completely eliminate the rhythm. The activity of the 69 bp enhancer is also dependent upon the protein product of *per* (Hao et al., 1997), which provides strong evidence that PERIOD interacts (either directly or indirectly) with bHLH gene products to regulate the rate of *per* transcription. Among the known bHLH–PAS domain proteins, ARNT homodimers bind to the consensus E box sequence as symmetric half-sites (Swanson et al., 1995). Thus, it is conceivable that a bHLH–PAS protein such as CLOCK could be involved in regulating *per* transcription.

How might the *Clock* gene product function in the mammalian circadian system? We envisage two working hypotheses for CLOCK function. As is the case for the canonical clock genes *per*, *tim*, and *frq*, *Clock* could play a direct role in a transcription–translation feedback loop that generates circadian rhythms. On this hypothesis, one would expect to see a circadian oscillation in the activity of CLOCK (expressed as a change in abundance, localization, and/or functional activity). Alternatively, like *sim* and *trh*, *Clock* could be a master regulator of circadian pacemaker function by controlling the expression patterns of the components of the circadian oscillator, without necessarily having a circadian rhythm in its own expression or activity. In either case, *Clock* would play a central role in the circadian organization of mammals.

The cloning and molecular characterization of a novel clock gene in mammals provides an entre´ e for elucidating the genetic and molecular mechanisms underlying the entrainment, generation, and expression of circadian rhythms in higher organisms, including humans. In view of the central role played by the circadian clock in the regulation of the sleep–wake cycle, the identification of the mammalian *Clock* gene is expected to yield new insights into the control of sleep and its disorders, the effects of shift work and jet lag, as well as pathologies that occur with advanced age (Turek and Van Reeth, 1996).

Experimental Procedures

Animals

All *Clock* mice were produced in our breeding colony in the Center for Experimental Animal Resources at Northwestern University. Colony maintenance, collection of locomotor activity data, and behavioral testing were essentially as described in Vitaterna et al. (1994). The *Clock* mutation was induced in, and has been maintained as a coisogenic line by at least six generations of backcrossing to, the C57BL/6J (B6) inbred strain. Mice used for studies reported here were the progeny of one of two types of mating. For mutation detection and Northern blot experiments, we used coisogenic B6 mice. Because of the isogenic strain background, the *Clock* genotype of these animals was determined based on the free-running period of the circadian rhythm of locomotor activity when the animals were placed in constant darkness. For in situ experiments, we used the progeny of an intercross between (BALB/cJ *x* C57BL/6J)F1 or F2 *Clock*/+ heterozygous mice. Because *Clock* was associated with the B6 strain background in this cross, genotypes could be determined using strain-specific SSLP markers that flank the *Clock* locus (Dietrich et al., 1992).

Genetic and Physical Mapping

Experimental procedures for the genetic mapping of the *Clock* locus have been described elsewhere (King et al., 1997). In addition to the mapping crosses described there, four more intraspecific backcrosses were generated using C3H/HeJ as a counter strain. Details of these crosses will be provided elsewhere.

Physical mapping of the *Clock* region will be described in detail elsewhere. Most mouse genomic DNA clones were isolated from a C57BL/6J strain, large insert YAC library

(Kusumi et al., 1993), which was pooled for PCR screening by Research Genetics. We screened these pools using the SSLP and STS PCR primers for loci mapping near the *Clock* locus. We used the same protocol for PCR amplification that was used for SSLP genotyping (King et al., 1997), with the following modifications: the number of cycles was extended to 30, the template DNA was a 1:1 dilution of the YAC DNA pools, and control DNA samples were interspersed among the pooled DNA samples to provide reference-positive lanes. All clones identified that mapped close to *Clock* were obtained and tested directly by PCR and hybridization.

The mouse BAC library (also pooled by Research Genetics) was screened by PCR to obtain 384-well plates. These plates were screened in either one of two ways: dot filters of positive plates were screened using ^{32}P end-labeled PCR primer as probe, or alternatively, clones were identified by obtaining a replicate 384-well plate that was screened directly by PCR. Briefly, for SSLP identification, BAC 51 restriction-digest fragments were sub-cloned into M13 vector, and these sub-clone libraries were then screened with (CA)15 and (GT)15 oligonucleotides. Details of this screen will be provided elsewhere.

cDNA Libraries

Two SCN cDNA libraries, "DARK" and "LIGHT," were constructed using [(BALB/cJ *x* C57BL/6J) F1 *x* B6]N2 backcross mice that were either wild-type or heterozygous for the *Clock* mutation. Approximately equal numbers of wild type and *Clock*/+ heterozygous mice were used for each library. DARK library SCN tissue was collected in the dark at circadian time (CT, where CT = 0 refers to the beginning of the subjective day) 1, 7, 13, and 19 (107 animals). LIGHT library animals were transferred into ambient light at CT 1, 7, 13, or 19; SCN tissue was collected in light 30–90 min later at CT 2, 8, 14, or 20 (103 animals). cDNAs were uni-directionally cloned using the ZAP Express Kit (Stratagene). Primary libraries contained 1.7×10^6 (DARK) and 1.2×10^6 (LIGHT) pfu/*I*g vector; 1×10^6 pfu from each library were plate amplified once. Average insert sizes for the libraries were 2.3 kb and 2.2 kb for the DARK and LIGHT libraries, respectively; the insert sizes for both libraries ranged from 0.6 to 5.2 kb.

BAC DNA Isolation

BAC DNA for cDNA selection and for library screen probe production was isolated by alkaline lysis of a 50 ml overnight culture followed by phenol:chloroform extraction and ethanol precipitation (Sambrook et al., 1989). DNA was re-suspended in TE (10 mM Tris, 1 mM EDTA [pH 8.0]) containing 20 *I*g/ml RNase A (Boehringer Mannheim). BAC DNA was further purified by precipitation in 1 M NaCl and 6.5% PEG 8000 (Sigma) at 0°C for 20 min; samples were centrifuged at 4°C for 20 min at 16,000 g, washed in 70% ethanol, and re-suspended in TE. All BAC inserts were digested with NotI (New England Biolabs) and separated by either pulsed-field gel electrophoresis (PFGE) or field inversion gel electrophoresis (FIGE). Pulsed-field conditions were as described in Antoch et al. (1997). Field inversion conditions were 5 V/cm for 10 min followed by switch time ramping at 0.05 s initial reverse time, 0.01 s reverse increment, 0.15 s initial forward time, and 0.03 s forward increment for 81 cycles with 0.001 s reverse increment increment and 0.003 s forward increment increment for 24 hr.

Transcription Unit Analysis

We directly screened the SCN DARK and LIGHT cDNA libraries. After NotI digestion and separation by FIGE, BAC insert DNA was purified by *i*-agarase digestion (New England Biolabs). Radiolabeled probe was generated using the DECAprime II kit (Ambion) with gel-purified, full-length BAC insert as template; probe was purified with a ProbeQuant G-50 Micro column (Pharmacia). Labeled probe was pre-annealed to COT-1 DNA (Life

Technologies) to suppress repetitive DNA sequences (Marchuk and Collins, 1994). Filters were wiped with Kim wipes soaked in Tris-buffered saline (pH 7.5) and incubated in 20 μg/ ml proteinase K (Life Technologies) at 30°C for 30 min. Filters were pre-hybridized according to Marchuk and Collins (1994), then hybridized in fresh buffer with pre-annealed probe at 65°C for 48 hr. Filters were washed three times for 3 hr at room temperature in $2x$ SSC/0.1% SDS, 24 hr at 65°C in $2x$ SSC/ 0.1% SDS, and 30 min at 65cCin1x SSC/0.1% SDS. Positive clones were plaque purified and excised using the ExAssist helper phage (Stratagene).

The cDNA selection technique developed for this positional cloning effort will be described in detail elsewhere. Briefly, lambda DNA was obtained from plate lysates of the SCN libraries described above. cDNA inserts were excised by digestion with BamHI and XhoI and gel-purified from the lambda vector arms. BamHI adapters from the representational difference analysis (RDA) method (Lisitsyn et al., 1993) were ligated to cDNA inserts digested with DpnII. Amplicons from the cDNA fragments were then made by PCR as described (Lisitsyn et al., 1993). Genomic DNA from BAC clones was released with NotI digestion, and inserts were purified by PFGE. BAC DNA was then digested with Sau3AI, and a different set of BamHI RDA adapters was ligated. Amplicons from the BAC DNA were made by PCR using a biotin end-labeled oligonucleotide primer. cDNA and BAC amplicons were then hybridized at 65°C, with the addition of mouse COT-1 DNA, mouse ribosomal cDNA, and vector DNA to suppress nonspecific hybridization. Hybrids were captured with streptavidin-coated magnetic beads as described by Lovett (1994). Two rounds of selection were performed, and the efficiency was monitored with a positive control (BACDNA spiked with c-*fos* cDNA), a negative control (*jun-B*) and COT-1 DNA level. The selected cDNA fragments were cloned into Bluescript II KS(−) vector (Stratagene), and a plasmid sublibrary was constructed in E. coli DH5a cells (Life Technologies). BAC 51, BAC 48, c-*fos*, and COT-1 DNA were used as probes to screen this sublibrary. Clones that hybridized with BAC 51, but not BAC 48, c-*fos*,or COT-1 DNA, were sequenced and then searched against the NCBI data base using the BLAST algorithm (Altschul et al. 1990). To obtain complete cDNA clones, these BAC 51 positive cDNA fragments were pooled together to screen the SCN LIGHT cDNA library. Hybridizations were performed at 42°C in buffer containing 50% formamide, 10% dextran sulfate, 1 M NaCl, 1% SDS, and 0.1 mg/ml salmon sperm DNA. Filters were washed once in $2x$ SSC at room temperature and twice in 0.1x SSC and 0.1% SDS at 55°C.

All candidate clones obtained by direct SCN cDNA library screening and cDNA selection were confirmed to be in the *Clock* region by using them as probes on a Southern blot containing HindIII-digested BAC DNA (see below).

For shotgun sequencing of genomic BAC clones, detailed protocols for BAC DNA preparation, small-insert M13 library construction, and sequence-ready template preparation will be described elsewhere. Briefly, BAC DNA was prepared by large-scale alkaline lysis of two-liter cultures, followed by a two-step CsCl gradient purification. DNA (5 μg) was sonicated, and blunt ends were generated by mung bean nuclease (NEB). Samples were run on a 0.8% SeaPlaque GTG agarose gel (FMC Bioproducts) for size selection of insert DNA. The 1.3–1.7 kb range was gel purified using i-agarase (NEB) and blunt end ligated into M13mp19 (Pharmacia) with T4 DNA ligase (NEB). Ligation products were electroporated into E. coli XL1 Blue MRF' cells (Stratagene). M13 templates for cycle sequencing were prepared by a modified solid phase reversible isolation protocol (Hawkins et al., 1994). M13 insert sequencing reactions were performed using an Applied Biosystems (ABI) PRISM Catalyst Turbo 800 Molecular Biology LabStation with the ABI Dye Primer Cycle Sequencing Ready Reaction -21M13 FS Amplitaq Kit; sequencing reaction products were separated on an ABI 377 DNA Sequencer. Sequences were analyzed and aligned using

SEQUENCHER 3.0 (GeneCodes); each sequence was used to search NCBI databases for nucleotide, EST, and protein homology.

cDNA Sequencing

Clock gene cDNA for sequencing was prepared by alkaline lysis, followed by CsCl density gradient purification. Dideoxy sequencing reactions were performed using the ABI Dye Terminator Cycle Sequencing Ready Reaction FS Amplitaq kit. Primers for sequence walking were designed from the cDNA and genomic sequences and were synthesized by Integrated DNA Technologies (Coralville, IA). All clones were sequenced on both strands.

Mutation Detection

PCR primers were designed, using the genomic sequence obtained from shotgun sequencing, to flank each exon bearing coding sequence. Each exon sequence (and the intronic sequence immediately flanking it) was amplified by PCR from 8 B6 isogenic DNA samples obtained from tail-tips of 4 wild type and 4 *Clock/Clock* mice. These PCR products were purified using a solid phase reversible isolation protocol (Hawkins et al., 1994), and 25–100 ng of template DNA was sequenced using the ABI Dye Terminator chemistry described above.

RT–PCR was carried out on ~200 ng of hypothalamic RNA obtained both from wild type and *Clock/Clock* mice. Reactions were performed using the GIBCO-BRL SuperScript Kit. RNA was reverse transcribed by three priming methods: oligo d(T)12–18, random hexamer primers, and a gene-specific primer from the 3 end of *Clock* (5 -AGCTACCAACACTGTTAGTGCTCTTCGG-3). Following the reverse transcription, samples were amplified (forward primer:5 -GCG AGAACTTGGCATTGAAGAG-3 ; reverse primer: 5 -CTGTGTCCACT CATTACACTCTGTTG-3) using the following parameters: 94°C for 3 min; 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min; 72°C for 10 min, and a 4°C hold. PCR products were separated on 2% NuSieve agarose gels (FMC Bioproducts) containing 0.5 Ig/ml ethidium bromide and were visualized by UV transillumination. Each band of the RT–PCR reaction was gel purified, reamplified using 25 cycles of the PCR protocol described for the initial RT–PCR, and sequenced on both strands using the same protocol that was described for the genomic exon PCR samples.

RNA Analysis

For Northern blot hybridization, tissue was collected from the hypothalamus and eyes of (BALB/cJ *x* C57BL/6J)F2 or F3 (Figure 2) or C57BL/6J (Figure 6E) mice at ZT6 (ZT refers to zeitgeber time; ZT0 = lights on; ZT6 = 6 hr after lights on). The extraction of total RNA was done according to Chomczynski and Sacchi (1987). Fifteen Ig of total RNA was separated by size in a 1% denaturing formaldehyde gel, transferred to nylon membranes (Duralon-UV membrane, Stratagene) and immobilized by UV cross-linking. Probe was prepared from yz50 (a cDNA clone that contains most of the coding region of *Clock*) by the random priming method (Ready-To-Go, Pharmacia) using [a-^{32}P]dCTP (6000 Ci/mmol, 20 mCi/ml, NEN Research Products) and was purified with a Sephadex G50 spin column (ProbeQuant G-50 Micro Columns, Pharmacia). The membrane was prehybridized at 42°C in hybridization solution (5× Denhardt's, 100 Ig/ml salmon sperm DNA, 6× SSPE, 0.1% SDS, and 50% formamide) and then hybridized overnight at 42°C in hybridization solution containing 10% dextran sulfate. Following hybridization, the membrane was washed twice with 0.2× SSC/0.1% SDS at 65°C for at least 30 min each wash. Blots were normalized for the RNA loading by hybridization with a 51-mer oligonucleotide complementary to 18S ribosomal RNA: 5 -GATGTGGTAGCCGTTTCTCAGGCTCCCTC TCCGGAATCGAACCCTGATTCC-3 . The probe was labeled with [–-^{32}P]ATP using T4

polynucleotide kinase and purified by polyacrylamide gel electrophoresis. Hybridization was performed as described in Fukada et al. (1996).

For in situ hybridization, mice were killed by cervical dislocation at ZT6 after they were housed in LD12:12 for 2 weeks. Brains were rapidly removed and frozen on dry ice. Coronal sections of 20 μm thickness were thaw mounted on subbed slides. In situ hybridization was performed as described by Suhr et al. (1989). ^{35}S-labeled riboprobes were transcribed from yz50, a 2.8 kb cDNA clone in the pBK-CMV phagemid vector. This phagemid was linearized with SacI to generate an antisense template of 2.4 kb or with NotI to generate a sense template of 2.8 kb.

Southern Blot Analysis

For BAC Southern blot hybridization, DNA from BACs 48 and 50 (negative controls) and 51, 52, 53, and 54 were digested with HindIII (Pharmacia). Each BAC DNA (2.5 μg) was separated on a 1% agarose gel (Life Technologies), transferred to a nylon membrane (Gene-Screen Plus, NEN Research Products), and immobolized by UV cross-linking. Six blots were prepared. Probes were labeled by the random priming method (Ready-To-Go, Pharmacia). Hybridization was performed at 42cCinbuffer containing6× SSC, 50% formamide, 1% SDS, 0.1% Blotto, and 0.1 mg/ml salmon sperm DNA. Blots were washed once in 2× SSC at room temperature, and twice in 0.1× SSC/0.1% SDS at 45°C.

For the zoo blot, the following tissues were used for each vertebrate: whole blood (human); liver (mouse, hamster, chicken, and frog); liver and brain (lizard); whole body minus head and viscera (zebrafish). Genomic DNA from human blood was extracted according to Current Protocols in Human Genetics (Dracopoli et al., 1994). For all other samples, tissues were frozen in a dry ice/ethanol bath, and DNA was prepared using proteinase K digestion and phenol/chloroform extraction. The DNA was digested with PstI, electrophoresed in a 1% agarose gel (Life Technologies), and alkaline blotted onto a GeneScreen Plus membrane (NEN Research Products). Because of the complex intron–exon structure of the *Clock* gene, Southern blots using probes with multiple exons yielded complex hybridization patterns. The probes for hybridization were prepared by PCR amplification in the presence of [a-^{32}P]dCTP using primers from: (1) the PAS-B domain (5 - TGTGTACTGTTGAAGAACCAAATG AAGAGTT-3 and 5 - GTGACATTTTGCCAGATTTTCTAGGTCATG3) and (2) the C-terminal region of *Clock* (exons 15–21) using the PCR primers used for RT–PCR (Figure 5B). Hybridization was performed at 55°C in buffer containing 6× SSC, 5× Denhardt's solution, 0.1 mg/ml salmon sperm DNA, and 0.1% SDS. The blot was washed once in 2× SSC at room temperature, and twice in 2× SSC/0.1% SDS at 45°C (Fukuda et al., 1995) for the PAS-B probe and two additional washes in 0.5× SSC/0.1% SDS at 55° for the C-terminal probe.

Acknowledgments

References

Akagi JM, Nomiyama H, Setoyama C, Akagi M. Messenger RNA expressed in mouse teratocarcinoma stem cells and down-regulated by a tumor-promoting phorbol ester codes for a novel transmembrane protein. Biochem. Biophys. Res. Commun. 1988; 157:548–557. [PubMed: 3202867]

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J. Mol. Biol. 1990; 215:403–410. [PubMed: 2231712]

Antoch MP, Song E-J, Chang A-M, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS. Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. Cell. 1997 this issue.

Aronson BD, Johnson KA, Loros JJ, Dunlap JC. Negative feedback defining a circadian clock: autoregulation of the clock gene *frequency*. Science. 1994; 263:1578–1584. [PubMed: 8128244]

Bird A. The essentials of DNA methylation. Cell. 1992; 70:5–8. [PubMed: 1377983]

Brunkow ME, Nagle DL, Bernstein A, Bucan M. A 1.8Mb YAC contig spanning three members of the receptor tyrosine kinase gene family (Pdgfra Kit Flk1) on mouse chromosome 5. Genomics. 1995; 5:421–432. [PubMed: 7540588]

Burbach KM, Poland A, Bradfield CA. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. Proc. Natl. Acad. Sci. USA. 1992; 89:8185–8189. [PubMed: 1325649]

Chartier FL, Keer JT, Sutliffe MJ, Henriques DA, Mileham P, Brown SD. Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. Nat. Genet. 1992; 1:132–136. [PubMed: 1302006]

Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 1987; 162:156–159. [PubMed: 2440339]

Collinson IR, Runswick MJ, Buchanan SK, Fearnley IM, Skehel JM, van Raaij MJ, Griffiths DE, Walker JE. Fo membrane domain of ATP synthase from bovine heart mitochondria: purification, subunit composition, and reconstitution with F1ATPase. Biochemistry. 1994; 33:7971–7978. [PubMed: 8011660]

Dietrich W, Katz H, Lincoln SE, Shin H-S, Friedman J, Dracopoli NC, Lander ES. A genetic map of the mouse suitable for typing intraspecific crosses. Genetics. 1992; 131:423–447. [PubMed: 1353738]

Dove WF. Molecular genetics of *Mus musculus*: point mutagenesis and milliMorgans. Genetics. 1987; 116:5–8. [PubMed: 3596231]

Dracopoli, NC.; Haines, JL.; Korf, BR.; Moir, DT.; Morton, CC.; Seidman, CE.; Seidman, JG.; Smith, DR. Current Protocols in Human Genetics. New York: John Wiley and Sons, Inc.; 1994.

Dunlap JC. Genetic and molecular analysis of circadian rhythms. Annu. Rev. Genet. 1996; 30:579–601. [PubMed: 8982466]

Edery I, Zwiebel LJ, Dembinska ME, Rosbash M. Temporal phosphorylation of the *Drosophila* period protein. Proc. Natl. Acad. Sci. USA. 1994; 91:2260–2264. [PubMed: 8134384]

Fukada S, Tanaka M, Matsuyama M, Kobayashi D, Nagahama Y. Isolation, characterization and expression of cDNA encoding the medaka (*Oryzias latipes*) ovarian follicle cyctochrome P-450 aromatase. Mol. Reprod. Develop. 1996; 45:285–290.

Fukuda S, Tanaka M, Iwaya M, Nagahama Y. The sox gene family and its expression during embryogenesis in the teleost fish, medaka (Oryzias latipes). Dev. Growth Differ. 1995; 37:379–385.

Gekakis N, Saez L, Delahaye-Brown A-M, Myers MP, Sehgal A, Young MW, Weitz CJ. Isolation of *timeless* by PER protein interaction: defective interaction between *timeless* protein and long-period mutant PER^L. Science. 1995; 270:811–815. [PubMed: 7481773]

Gradin K, McGuire J, Wenger RH, Kvietikova I, Whitelaw ML, Toftgard R, Tora L, Gassman M, Poellinger L. Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. Mol. Cell. Biol. 1996; 16:5221–5231. [PubMed: 8816435]

Hall JC. Tripping along the trail to the molecular mechanisms of biological clocks. Trends Neurosci. 1995; 18:230–240. [PubMed: 7610493]

Hao H, Allen DL, Hardin PE. A circadian enhancer mediates PER-dependent mRNA cycling in Drosophila. Mol. Cell. Biol. 1997 in press.

Hardin PE, Hall JC, Rosbash M. Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. Nature. 1990; 343:536–540. [PubMed: 2105471]

Hardin PE, Hall JC, Rosbash M. Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. Proc. Natl. Acad. Sci. USA. 1992; 89:11711–11715. [PubMed: 1465387]

Hawkins TL, O'Conner-Morin T, Roy A, Santillan C. DNA purification and isolation using a solid-phase. Nucl. Acids Res. 1994; 22:4543–4544. [PubMed: 7971285]

Herskowitz I. Functional inactivation of genes by dominant negative mutations. Nature. 1987; 329:219–222. [PubMed: 2442619]

Hoffman EC, Reyes H, Chu F-F, Sander F, Conley LH, Brooks BA, Hankinson O. Cloning of a factor required for activity of the Ah (dioxin) receptor. Science. 1991; 252:954–958. [PubMed: 1852076]

Hoganesch JB, Chan WK, Jackiw VH, Brown RC, Gu Y-Z, Pray-Grant M, Perdew GH, Bradfield CA. Characterization of a subset of the basic–helix-loop-helix–PAS superfamily that interacts with components of the dioxin signaling pathway. J. Biol. Chem. 1997; 272:8581–8593. [PubMed: 9079689]

Huang ZJ, Edery I, Rosbash M. PAS is a dimerization domain common to *Drosophila* period and several transcription factors. Nature. 1993; 364:259–262. [PubMed: 8391649]

King DP, Vitaterna MH, Chang AM, Dove WF, Pinto LH, Turek FW, Takahashi JS. The mouse *Clock* mutation behaves as an antimorph and maps within the W^{19H} deletion, distal of *Kit*. Genetics. 1997 in press.

Klein, DC.; Moore, RY.; Reppert, SM. Suprachiasmatic Nucleus—The Mind's Clock. New York: Oxford University Press; 1991.

Kozak M. Interpreting cDNA sequences: some insights from studies on translation. Mammalian Genome. 1996; 7:563–574. [PubMed: 8679005]

Krawczak M, Reiss J, Cooper DN. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet. 1992; 90:41–54. [PubMed: 1427786]

Kusumi K, Smith JS, Segre JA, Koos DS, Lander ES. Construction of a large-insert yeast artificial chromosome library of the mouse genome. Mammalian Genome. 1993; 4:391–392. [PubMed: 8358173]

Kyriacou CP, Hall JC. Circadian rhythm mutations in Drosophilamelanogaster affect short-termfluctuations in the male's courtship song. Proc. Natl. Acad. Sci. USA. 1980; 77:6929–6933.

Larin Z, Monaco AP, Lehrach H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA. 1991; 88:4123–4127. [PubMed: 2034658]

Lee C, Parikh V, Itsukaichi T, Bac K, Edery I. Resetting the Drosophila clock by photic regulation of PER and a PER–TIM complex. Science. 1996; 271:1740–1744. [PubMed: 8596938]

Lindebro MC, Poellinger L, Whitelaw ML. Protein– protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor–Arnt transcription factor complex. EMBO J. 1995; 14:3528–3539. [PubMed: 7628454]

Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. Science. 1993; 259:946–951. [PubMed: 8438152]

Lovett M. Fishing for complements: finding genes by direct selection. Trends Genet. 1994; 10:352–357. [PubMed: 7985239]

Lynch, GR.; Lynch, CB. Using quantitative genetic methods to understand mammalian circadian behavior and photoperiodism. In: Goldowitz, D.; Wahlsten, D.; Wimer, RE., editors. Techniques for the Genetic Analysis of Brain and Behavior. New York: Elsevier; 1992. p. 251-268.

Marchuk, DA.; Collins, FS. The use of YACs to identify expressed sequences: cDNA screening using total YAC insert. In: Nelson, DL., editor. YAC Libraries: A User's Guide. New York: W.H. Freeman; 1994. p. 113-142.

Meijer JH, Rietveld WJ. Neurophysiology of the suprachiasmatic circadian pacemaker in rodents. Physiol. Rev. 1989; 69:671–707. [PubMed: 2664825]

Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science. 1989; 245:371–378. [PubMed: 2667136]

Moore, RY. Organization of the mammalian circadian system. In: Chadwick, DJ.; Ackrill, K., editors. Ciba Foundation Symposium on Circadian Clocks and Their Adjustment. West Sussex, England: John Wiley and Sons; 1995. p. 88-106.

Muller, HJ. Sixth International Congress of Genetics. Ithaca, New York: Brooklyn Botanic Gardens; 1932. Further studies on the nature and causes of gene mutations.

Muralidhar MG, Thomas JB. The Drosophila *bendless* gene encodes a neural protein related to a ubiquitin-conjugating enzyme. Neuron. 1993; 11:253–266. [PubMed: 8394720]

Murre C, McCaw PS, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell. 1989; 56:777–783. [PubMed: 2493990]

Myers MP, Wager-Smith K, Wesley CS, Young MW. Positional cloning and sequence analysis of the Drosophila clock gene, timeless. Science. 1995; 270:805–808. [PubMed: 7481771]

Nambu JR, Lewis JO, Wharton KAJ, Crews ST. The Drosophila *single-minded* gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. Cell. 1991; 67:1157–1167. [PubMed: 1760843]

Pittendrigh CS. Temporal organization: reflections of a Darwinian clock-watcher. Annu. Rev. Physiol. 1993; 55:17–54.

Ralph MR, Menaker M. A mutation of the circadian system in golden hamster. Science. 1988; 241:1225–1227. [PubMed: 3413487]

Ralph MR, Foster RG, Davis FC, Menaker M. Transplanted suprachiasmatic nucleus determines circadian period. Science. 1990; 247:975–978. [PubMed: 2305266]

Rosbash M, Allada R, Dembinska M, Guo WQ, Le M, Marrus S, Quian Z, Rutila J, Yaglom J, Zeng H. A Drosophila circadian clock. Cold Spring Harbor Symp. Quant. Biol. 1996 in press.

Sachs AB. Messenger RNA degradation in eukaryotes. Cell. 1993; 74:413–421. [PubMed: 7688664]

Saez L, Young M. Regulation of nuclear entry of the Drosophila clock proteins period and timeless. Neuron. 1996; 17:911–920. [PubMed: 8938123]

Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular Cloning: A Laboratory Manual. Second Edition. New York: Cold Spring Harbor Laboratory Press; 1989.

Schwartz WJ, Zimmerman P. Circadian timekeeping in BALB/c and C57BL/6 inbred mouse strains. J. Neurosci. 1990; 10:3685–3694. [PubMed: 2230953]

Sehgal A, Rothenfluh-Hilfiker M, Hunter-Ensor M, Chen Y, Myers MP, Young MW. Rhythmic expression of timeless: a basis for promoting circadian cycles in period gene autoregulation. Science. 1995; 270:808–810. [PubMed: 7481772]

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. Cloning and stable maintenance of 300 kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc. Natl. Acad. Sci. USA. 1992; 89:8794–8797. [PubMed: 1528894]

Siwicki KK, Eastman C, Petersen G, Rosbash M, Hall JC. Antibodies to the *period* gene product of Drosophila reveal diverse tissue distribution and rhythmic changes in the visual system. Neuron. 1988; 1:141–150. [PubMed: 3152288]

Su L-K, Kinzler KW, Vogelstein B, Preisinger AC, Moser AR, Luongo C, Gould KA, Dove WF. Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science. 1992; 256:668–670. [PubMed: 1350108]

Suhr ST, Rahal JO, Mayo KE. Mouse growth hormone–releasing hormone: precursor structure and expression in brain and placenta. Mol. Endocrinol. 1989; 3:1693–1700. [PubMed: 2514346]

Swanson HI, Bradfield CA. The AH-receptor: genetics, structure and function. Pharmacogenetics. 1993; 3:213–230. [PubMed: 8287061]

Swanson HI, Chan WK, Bradfield CA. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. J. Biol. Chem. 1995; 270:26292–26302. [PubMed: 7592839]

Takahashi JS. Circadian rhythms: from gene expression to behavior. Curr. Opin. Neurobiol. 1991; 1:556–561. [PubMed: 1822296]

Takahashi JS. Circadian clock genes are ticking. Science. 1992; 258:238–240. [PubMed: 1384127]

Takahashi JS, Pinto LP, Vitaterna MH. Forward and reverse genetic approaches to behavior in the mouse. Science. 1994; 264:1724–1733. [PubMed: 8209253]

Takahashi JS. Molecular neurobiology and genetics of circadian rhythms in mammals. Annu. Rev. Neurosci. 1995; 18:531–553. [PubMed: 7605073]

Tosini G, Menaker M. Circadian rhythms in cultured mammalian retina. Science. 1996; 272:419–421. [PubMed: 8602533]

Turek, FW.; Van Reeth, O. Circadian Rhythms. In: Fregly, MJ.; Blatteis, CM., editors. Handbook of Physiology. Oxford, U.K: Oxford University Press; 1996. p. 1329-1360.

Vitaterna MH, King DP, Chang A-M, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS. Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. Science. 1994; 264:719–725. [PubMed: 8171325]

Wilk R, Weizman I, Shilo B-Z. *tracheless* encodes a bHLH–PAS protein that is an inducer of tracheal cell fates in Drosophila. Genes Dev. 1996; 10:93–102. [PubMed: 8557198]

Yamaguchi T, Kim NS, Sekine S, Seino H, Osaka F, Yamao F, Kato S. Cloning and expression of cDNA encoding a human ubiquitin-conjugating enzyme similar to the Drosophila *bendless* gene product. J. Biochem. 1996; 120:494–497. [PubMed: 8902611]

Zeng H, Hardin PE, Rosbash M. Constitutive overexpression of the Drosophila *period* protein inhibits *period* mRNA cycling. EMBO J. 1994; 13:3590–3598. [PubMed: 8062834]

Zhou Y-D, Barnard M, Tian H, Li X, Ring HZ, Francke U, Shelton J, Richardson J, Russell DW, McKnight SL. Molecular characterization of two mammalian bHLH–PAS domain proteins selectively expressed in the central nervous system. Proc. Natl. Acad. Sci. USA. 1997; 94:713–718. [PubMed: 9012850]
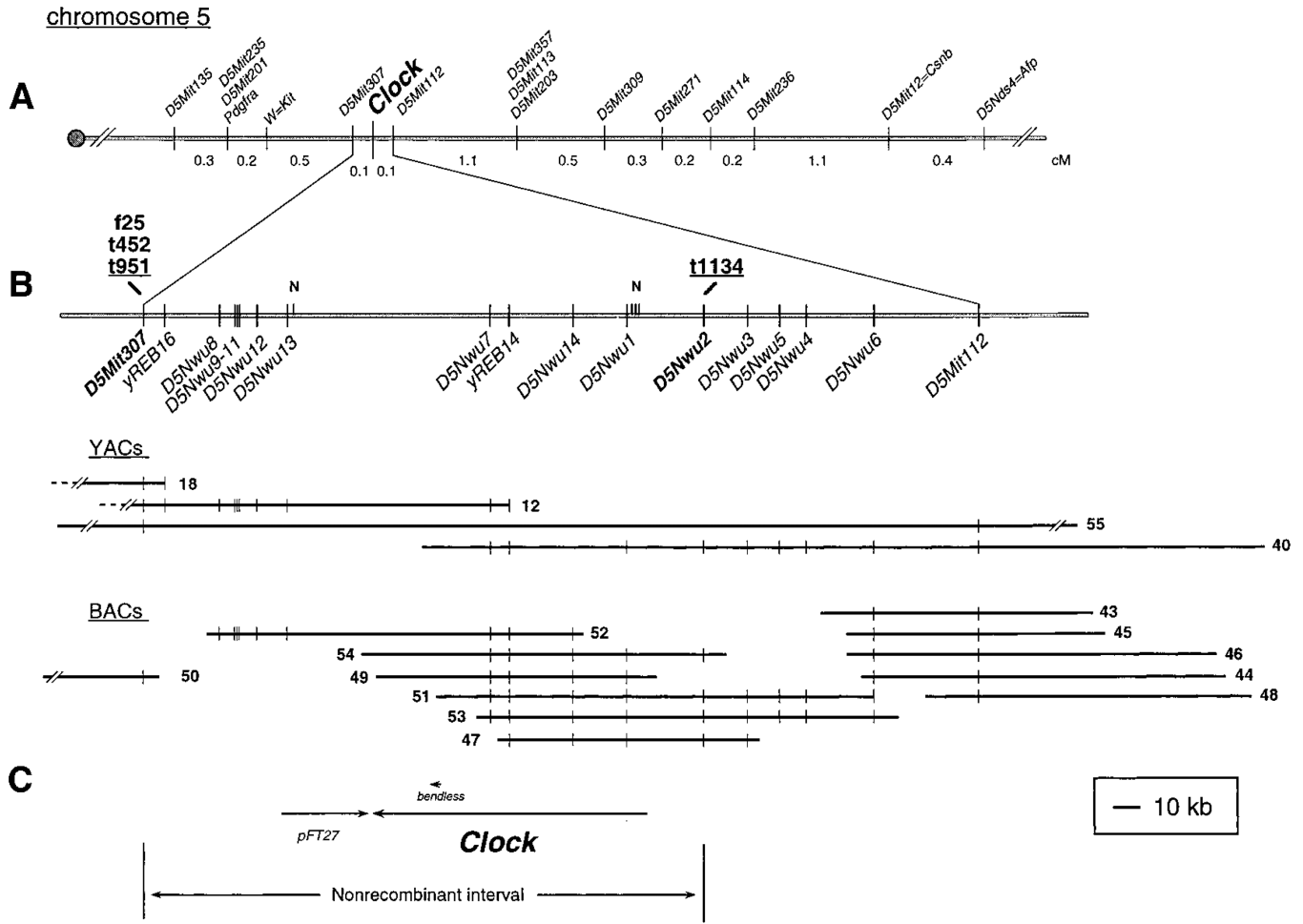
Figure 1. Genetic, Physical, and Transcription Unit Mapping of the *Clock* Locus

(A) Genetic map of mouse chromosome 5, showing the location of *Clock* on the midportion of this chromosome, 0.7 cM distal of *Kit*, flanked by *D5Mit307* and *D5Mit112*.

(B) Physical and high resolution genetic mapping of the *Clock* locus. The SSLPs and STSs in the *D5Mit307*–D5Mit112 interval are shown on the upper bar. The maximum nonrecombinant interval is defined by the SSLPs *D5Mit307* and *D5Nwu2*. The animals with recombinations between these SSLPs and *Clock* are noted. Also shown are restriction sites for NotI (depicted with an [N]), which identify two CpG islands associated with the 5 ends of *Clock* and *pFT27*. YAC and BAC contigs of the *Clock* region are shown below. Only four clones from the 32 clone YAC contig are shown. YACs 18 and 12 are chimeric clones and appear to be identical to YACs B16.S5.RE.C12 and B14.S4.RH.C8 in Brunkow et al. (1995); the part of each clone not mapping to this region of chromosome 5 is represented by a dashed line. The YACs were isolated from a library constructed by Kusumi et al. (1993), except for clone 55, which came from a library constructed by Larin et al. (1991). The BAC clones were isolated from a library constructed in Dr. Melvin Simon's laboratory.

(C) Candidate genes mapping to the nonrecombinant interval. Shown are their relative locations and transcriptional orientations.
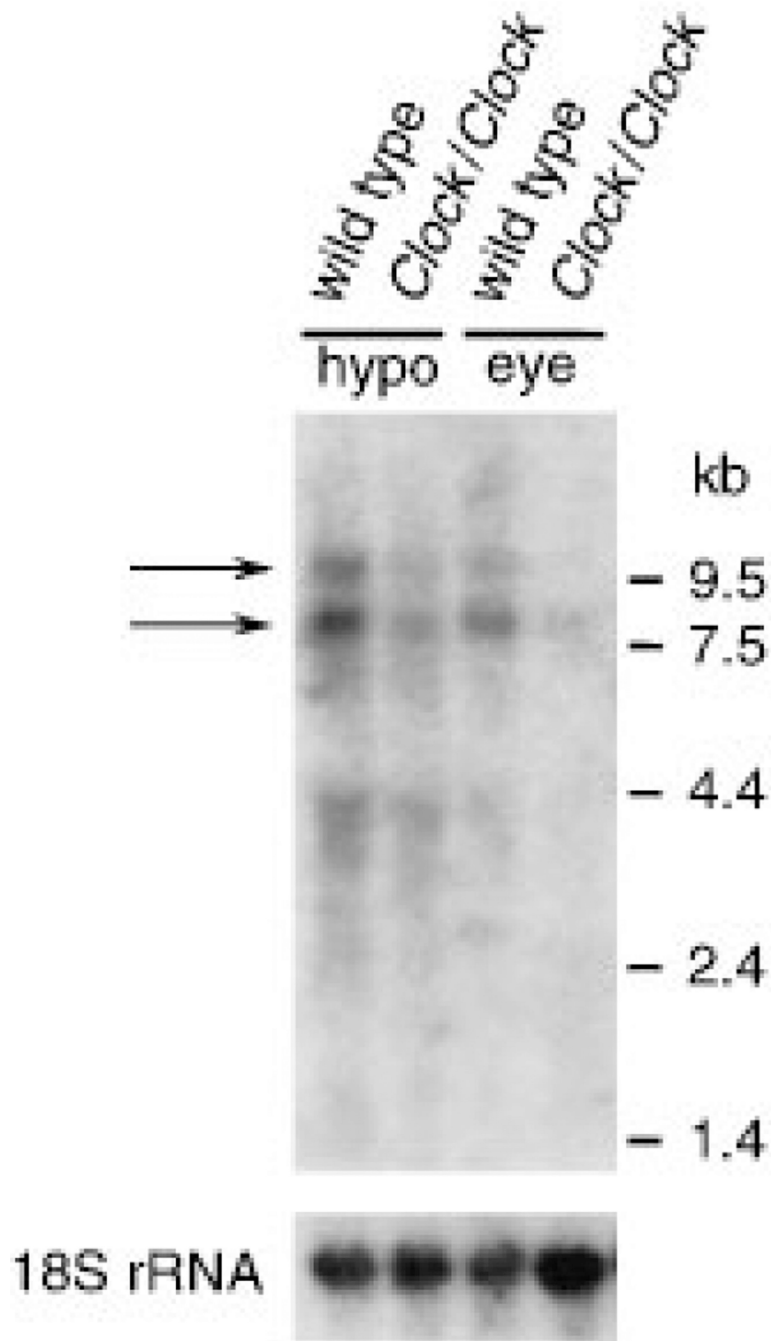
Figure 2. RNA Expression of the PAS Domain Candidate Gene in
Clock/Clock Mice Differs from the Wild Type Northern blot analysis of RNA extracted (at ZT6) from eye and hypothalamus of (BALB/cJ *x* C57BL/6J)F2 and F3 wild-type and *Clock/Clock* mice, using the yz50 cDNA clone as probe, reveals a reduction in the level of mRNA expression of both the ~8 kb and ~11 kb transcripts. The positions to which standards migrated and their sizes are shown on the right. The bottom panel shows the same filter hybridized with a ^{32}P-labeled oligonucleotide of 18S RNA for normalization.

A

B

```
  1    MVFTVSCSKMSSIVDRDDSSIFDGLVEEDDKDKAKRVSRNKSEKKRRDQF
 51    NVLIKELGSMLPGNARKMDKSTVLQKSIDFLRKHKETTAQSDASEIRQDW
101    KPTFLSNEEFTQLMLEALDGFFLAIMTDGSIIYVSESVTSLLEHLPSDLV
151    DQSIFNFIPEGEHSEVYKILSTHLLESDSLTPEYLKSKNQLEFCCHMLRG
201    TIDPKEPSTYEYVRFIGNFKSLTSVSTSTHNGFEGTIQRTHRPSYEDRVC
251    FVATVRLATPQFIKEMCTVEEPNEEFTSRHSLEWKFLFLDHRAPPIIGYL
301    PFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQWIW
351    LQTHYYITYHQWNSRPEFIVCTHTVVSYAEVRAERRRELGIEESLPETAA
401    DKSQDSGSDNRINTVSLKEALERFDHSPTPSASSRSSRKSSHTAVSDPSS
451    TPTKIPTDTSTPPRQHLPAHEKMTQRRSSFSSQSINSQSVGPSLTQPAMS
501    QAANLPIPQGMSQFQFSAQLGAMQHLKDQLEQRTRMIEANIHRQQEELRK
551    IQEQLQMVHGQGLQMFLQQSNPGLNFGSVQLSSGNSNIQQLTPVNMQGQV
601    VPANQVQSGHISTGQHMIQQQTLQSTSTQQSQQSVMSGHSQQTSLPSQTP
651    STLTAPLYNTMVISQPAAGSMVQIPSSMPQNSTQSATVTTFTQDRQIRFS
701    QGQQLVTKLVTAPVACGAVMVPSTMLMGQVVTAYPTFATQQQQAQTLSVT
751    QQQQQQQQQPPQQQQQQQQQSSQEQQLPSVQQPAQAQLGQPPQQFLQTSRL
801    LHGNPSTQLILSAAFPLQQSTFPPSHHQQHQPQQQQQLPRHRTDSLTDPS
851    KVQPQ
```

Figure 3. cDNA Clones Identifying the *Clock* Gene and the Predicted Amino Acid Sequence of the CLOCK Protein

(A) cDNA clones isolated in the cDNA selection and direct library screening efforts. Clones beginning with yz are derived from cDNA selection, and those beginning with L (LIGHT) or D (DARK) are clones isolated from the direct library screen. Their exon content was determined by alignment to the genomic sequence that was assembled from the M13 shotgun sequencing subclones of BACs 52 and 54. Exon and intron sizes are depicted schematically. Exons 1a–22 vary in size from 71 base pairs (exon 1a) to 295 base pairs (exon 22). Intron sizes vary from 107 base pairs (between exons 10 and 11) to >22 kilobase pairs (between exons 2 and 3). yz80 appears to contain a unique splice variant. L7c appears

to be a partial clone of the *Clock* mutant splice variant transcript. Poly(A) tails of clones yz80, L4, yz161, and yz197 are indicated schematically. Exons containing open reading frames are closed, and exons containing 5 and 3 un-translated regions are open. Clones L8 and yz130 were reverse transcribed from intronic poly(A) sequences. The intronic sequences of these clones are shaded. The 2565-base ORF begins at base 44 of exon 4 and ends at base 177 of exon 23. Note the long 3 un-translated region, which extends for >6 kb.

(B) When translated, the 2565-base ORF predicts a protein of 855 amino acids, dominated by glutamine (112) and serine (100) residues. This protein has an expected molecular mass of 96.4 kDa and a pI of 6.52.

A

bHLH PAS Domain Q-Rich Poly-Q

CLOCK | A | B | ... (domain diagram)
ARNT | A | B |
SIM | A | B | AAQ |
PER | A | B | TG |

B

Basic Helix-Loop-Helix Domain

Consensus: `.KEKSRN-AAR.RR.KEN..F.ELA..LPLP.A..S.LDKASV.RL..SYL` % Identity

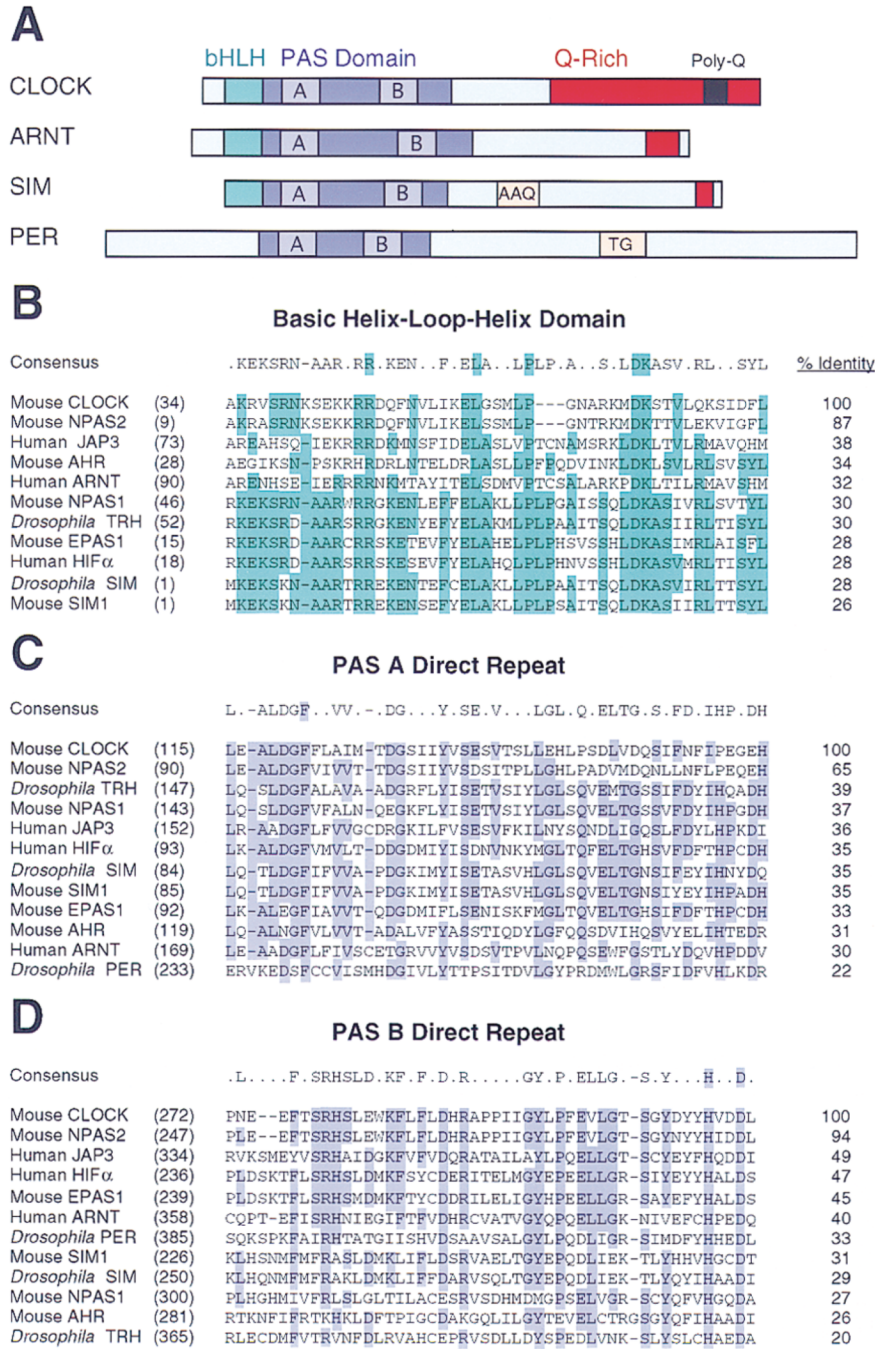| Protein	Pos	Sequence	% Identity
Mouse CLOCK	(34)	AKRVSRNKSEKKRRDQFNVLIKELGSMLP---GNARKMDKSTVLQKSIDFL	100
Mouse NPAS2	(9)	AKRASRNKSEKKRRDQFNVLIKELSSMLP---GNTRKMDKTTVLEKVIGFL	87
Human JAP3	(73)	AREAHSQ-IEKRRRDKMNSFIDELASLVPTCNAMSRKLDKLTVLRMAVQHM	38
Mouse AHR	(28)	AEGIKSN-PSKRHRDRLNTELDRLASLLPFPQDVINKLDKLSVLRLSVSYL	34
Human ARNT	(90)	ARENHSE-IERRRRNKMTAYITELSDMVPTCSALARKPDKLTILRMAVSHM	32
Mouse NPAS1	(46)	RKEKSRN-AARWRRGKENLEFFELAKLLPLPGAISSQLDKASIVRLSVTYL	30
Drosophila TRH	(52)	RKEKSRD-AARSRRGKENYEFYELAKMLPLPAAITSQLDKASIIRLTISYL	30
Mouse EPAS1	(15)	RKEKSRD-AARCRRSKETEVFYELAHELPLPHSVSSHLDKASIMRLAISFL	28
Human HIFα	(18)	RKEKSRD-AARSRRSKESEVFYELAHQLPLPHNVSSHLDKASVMRLTISYL	28
Drosophila SIM	(1)	MKEKSKN-AARTRREKENTEFCELAKLLPLPAAITSQLDKASVIRLTTSYL	28
Mouse SIM1	(1)	MKEKSKN-AARTRREKENSEFYELAKLLPLPSAITSQLDKASIIRLTTSYL	26

C

PAS A Direct Repeat

Consensus: `L.-ALDGF..VV.-.DG...Y.SE.V...LGL.Q.ELTG.S.FD.IHP.DH`

Protein	Pos	Sequence	% Identity
Mouse CLOCK	(115)	LE-ALDGFFLAIM-TDGSIIYVSESVTSLLEHLPSDLVDQSIFNFIPEGEH	100
Mouse NPAS2	(90)	LE-ALDGFVIVVT-TDGSIIYVSDSITPLLGHLPADVMDQNLLNFLPEQEH	65
Drosophila TRH	(147)	LQ-SLDGFALAVA-ADGRFLYISETVSIYLGLSQVEMTGSSIFDYIHQADH	39
Mouse NPAS1	(143)	LQ-SLDGFVFALN-QEGKFLYISETVSIYLGLSQVELTGSSVFDYIHPGDH	37
Human JAP3	(152)	LR-AADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDI	36
Human HIFα	(93)	LK-ALDGFVMVLT-DDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDH	35
Drosophila SIM	(84)	LQ-TLDGFIFVVA-PDGKIMYISETASVHLGLSQVELTGNSIFEYIHNYDQ	35
Mouse SIM1	(85)	LQ-TLDGFIFVVA-PDGKIMYISETASVHLGLSQVELTGNSIYEYIHPADH	35
Mouse EPAS1	(92)	LK-ALEGFIAVVT-QDGDMIFLSENISKFMGLTQVELTGHSIFDFTHPCDH	33
Mouse AHR	(119)	LQ-ALNGFVLVVT-ADALVFYASSTIQDYLGFQQSDVIHQSVYELIHTEDR	31
Human ARNT	(169)	LE-AADGFLFIVSCETGRVVYVSDSVTPVLNQPQSEWFGSTLYDQVHPDDV	30
Drosophila PER	(233)	ERVKEDSFCCVISMHDGIVLYTTPSITDVLGYPRDMWLGRSFIDFVHLKDR	22

D

PAS B Direct Repeat

Consensus: `.L....F.SRHSLD.KF.F.D.R.....GY.P.ELLG.-S.Y...H..D.`

Protein	Pos	Sequence	% Identity
Mouse CLOCK	(272)	PNE--EFTSRHSLEWKFLFLDHRAPPIIGYLPFEVLGT-SGYDYYHVDDL	100
Mouse NPAS2	(247)	PLE--EFTSRHSLEWKFLFLDHRAPPIIGYLPFEVLGT-SGYNYYHIDDL	94
Human JAP3	(334)	RVKSMEYVSRHAIDGKFVFVDQRATAILAYLPQELLGT-SCYEYFHQDDI	49
Human HIFα	(236)	PLDSKTFLSRHSLDMKFSYCDERITELMGYEPEELLGR-SIYEYYHALDS	47
Mouse EPAS1	(239)	PLDSKTFLSRHSMDMKFTYCDDRILELIGYHPEELLGR-SAYEFYHALDS	45
Human ARNT	(358)	CQPT-EFISRHNIEGIFTFVDHRCVATVGYQPQELLGK-NIVEFCHPEDQ	40
Drosophila PER	(385)	SQKSPKFAIRHTATGIISHVDSAAVSALGYLPQDLIGR-SIMDFYHHEDL	33
Mouse SIM1	(226)	KLHSNMFMFRASLDMKLIFLDSRVAELTGYEPQDLIEK-TLYHHVHGCDT	31
Drosophila SIM	(250)	KLHQNMFMFRAKLDMKLIFFDARVSQLTGYEPQDLIEK-TLYQYIHAADI	29
Mouse NPAS1	(300)	PLHGHMIVFRLSLGLTILACESRVSDHMDMGPSELVGR-SCYQFVHGQDA	27
Mouse AHR	(281)	RTKNFIFRTKHKLDFTPIGCDAKGQLILGYTEVELCTRGSGYQFIHAADI	26
Drosophila TRH	(365)	RLECDMFVTRVNFDLRVAHCEPRVSDLLDYSPEDLVNK-SLYSLCHAEDA	20

Figure 4. CLOCK Encodes a Novel Member of the bHLH–PAS Family of Transcription Factors
(A) The domain organization of CLOCK is similar to the original members of the PAS domain family—PERIOD, ARNT, and SIM. (B–D) The amino acid alignments of the basic–helix-loop-helix region and the PAS-A and PAS-B direct repeats of CLOCK and 10 other members of the bHLH–PAS domain family of transcription factors are shown (8 of these are mouse or human protein sequences; the remaining 2 are from Drosophila). PERIOD is included in the alignments of the PAS-A and PAS-B domains but lacks a basic–helix-loop-helix motif. Amino acid sequences were obtained through the Entrez sequence retrieval system (National Center for Biotechnology). Consensus amino acids represent residues identical among 50% of the proteins analyzed. Numbers in parentheses indicate position of

the start of amino acid fragment with respect to the initiator methionine of the corresponding protein. NPAS2 shares the highest identity with CLOCK of all sequences in the database at the time of submission. There is very high sequence identity (>90%) in the basic portion of the basic–helix-loop-helix domain and the PAS B direct repeat. Accession numbers for the PAS proteins shown: Drosophila PER (PERIOD, S17286); mouse NPAS1 (Neuronal PAS1,U77967); mouseNPAS2 (NeuronalPAS2,U77969); Drosophila TRH (trachealess, U42699); human JAP3 (bHLH–PAS protein JAP3, U60415); mouse EPAS1 (endothelial PAS domain protein 1, U81983); Drosophila SIM (single minded, 134494); mouse AHR (Aryl hydrocarbon receptor, M94623); mouse SIM1 (mSim1, D79209); ARNT (Aryl Hydrocarbon Nuclear Translocator, M69238); human HIF*a* (hypoxia-inducible factor 1 alpha, U22431).
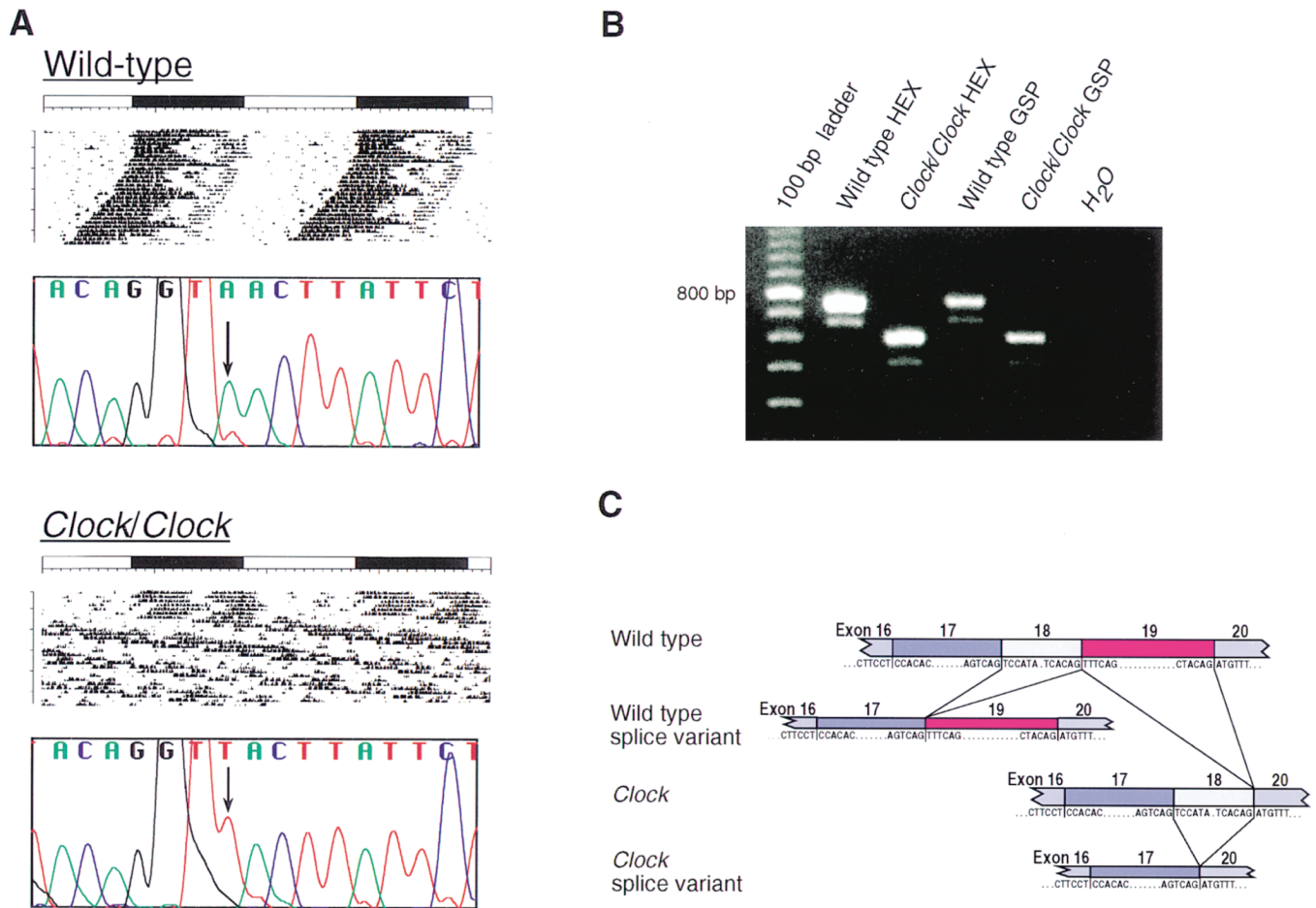
Figure 5. Identification and Analysis of the *Clock* Mutation

(A) Locomotor activity records and DNA sequence data from wild-type and *Clock*/*Clock* isogenic C57BL/6J (B6) mice. Activity records are double plotted so that 48 hr are represented horizontally, and each 24 hr interval is presented both to the right of and beneath the preceding day. Wheel-running activity is indicated by black markings. Mice were kept on a light:dark cycle (LD 12:12) for the first 8 days shown, as indicated by the bar above each record. Beginning with lights-off on day 8, the animals were kept in continuous darkness for the remainder of the record. Shown below each activity record is the DNA sequence of the exon–intron junction 3 to exon 19. The DNA used for sequencing was PCR amplified from the genomic DNA of the mice whose activity records are presented above each sequence trace. Arrows indicate the base that is altered in *Clock*/*Clock* mutant mice. This A T transversion was the only nucleotide difference detected from a comparative sequence scan of all of the coding exons (and flanking intronic sequence) of this gene.

(B) RT–PCR products. RNA samples for lanes 1 and 2 were reverse transcribed using random hexamer primers. RNA samples for lanes 3 and 4 were reverse transcribed using a gene-specific primer located in exon 23. PCR amplification was performed using gene specific primers from exons 15 and 21, which amplify a wild-type product of 761 bp. Lanes 1 and 3 are RT–PCR products amplified from hypothalamic RNA of wild-type isogenic B6 mice; lanes 2 and 4 are RT–PCR products amplified from hypothalamic RNA of *Clock*/*Clock* isogenic B6 mice.

(C) Exon content of RT–PCR products. Sequencing of the RT–PCR products revealed that exon 19 is indeed missing in the RNA transcripts of *Clock/Clock* mice. The additional band seen in the RT–PCR products is the result of the variable splicing of exon 18.
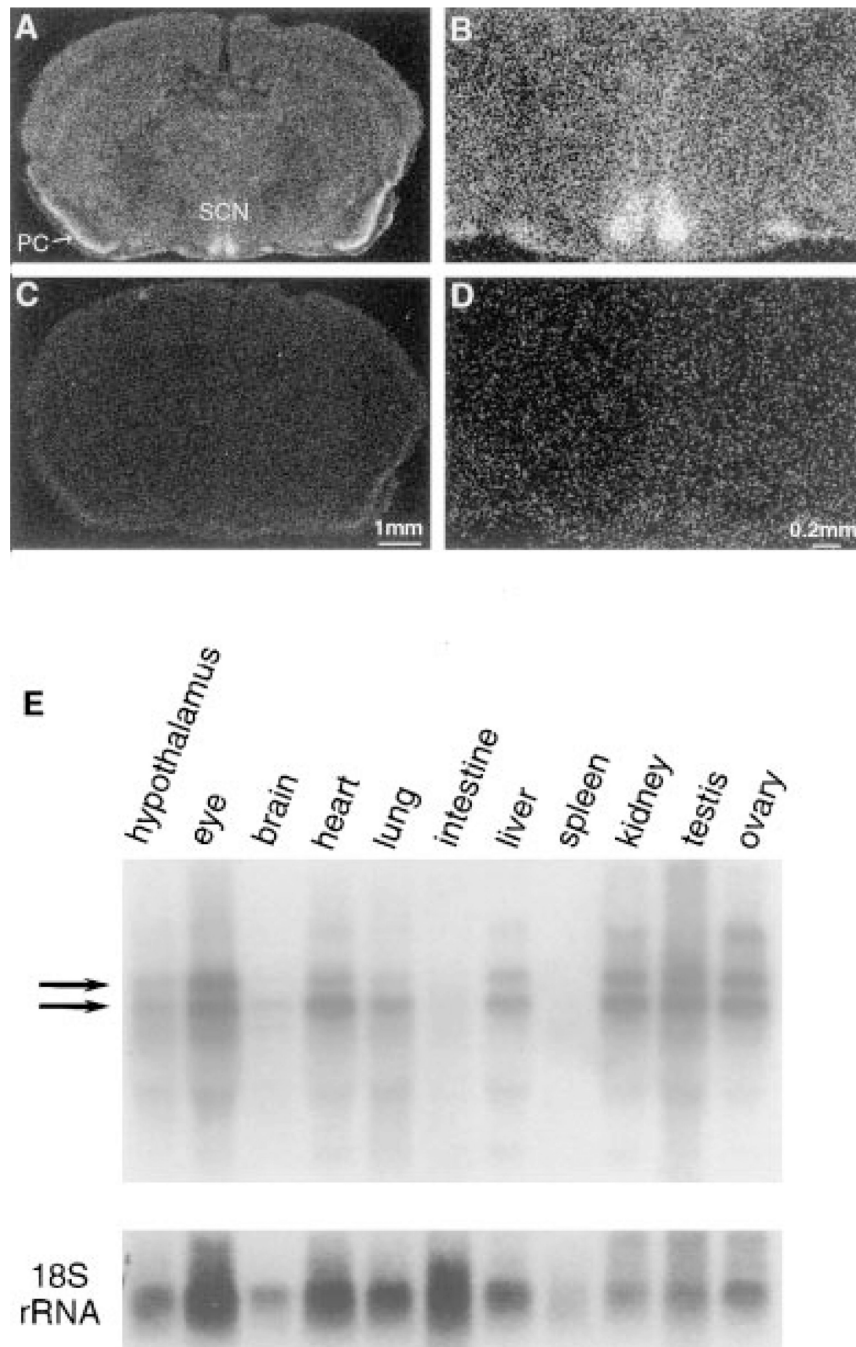
Figure 6. Analysis of *Clock* mRNA Expression
(A–D) In situ hybridization of *Clock* mRNA in the mouse brain. Coronal sections from brains of wild-type mice collected at ZT6, probed with yz50 antisense (A and B) and sense (C and D) ribo-probes.
(E) Tissue distribution blot. *Clock* mRNA expression is not limited to the SCN. Northern blot analysis of total RNA extracted from diverse tissues (all of which were collected at ZT6 from wild type C57BL/6J mice) reveals that *Clock* expression is widespread. The bottom panel shows the same filter hybridized with a ^{32}P-labeled oligonucleotide of 18S RNA for normalization.
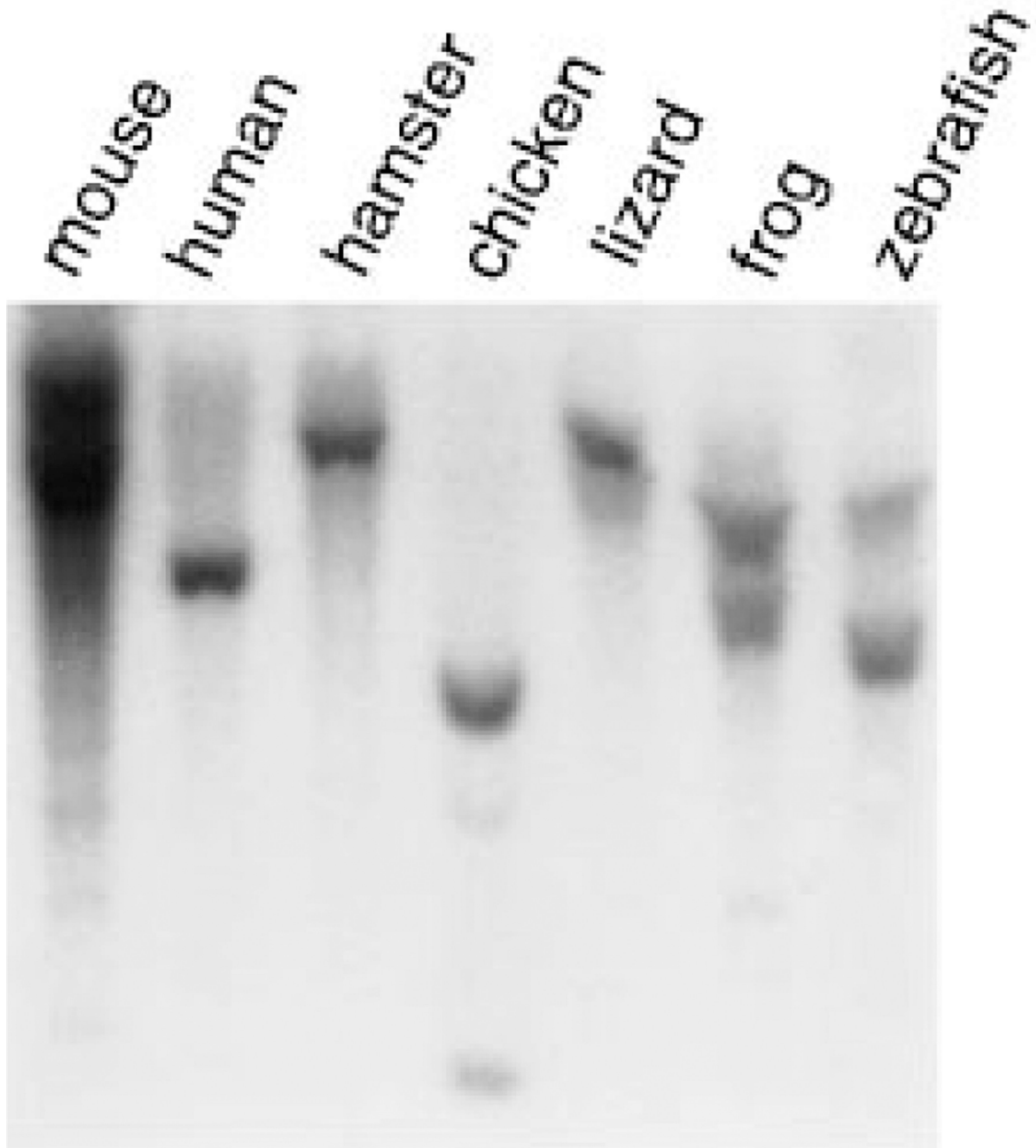
Figure 7. Evolutionary Conservation of the *Clock* Gene
Genomic DNA samples collected from mouse (C57BL/6J), human (D. P. K.), chicken, golden hamster (Mesocricetus auratus), lizard (Anolis carolinensis), frog (Xenopus laevis), and zebrafish (Danio rerio) were digested with PstI and analyzed by Southern blot using a mouse *Clock* PAS-B PCR probe (see Experimental Procedures). Similar results were obtained with a C-terminal probe (data not shown).