present in peripheral lymphocytes and therefore could be assumed to represent new germline mutations. There were nine cases where the origin of the mutant chromosome could be inferred using chromosome heteromorphisms, and in eight of those cases the mutant chromosome was derived from the father. Combining these data with ours, new germline mutations appeared in the paternally derived chromosome in 16 out of 17 informative cases. This result would be very improbable ($P < 0.0005$) should germinal mutations of the retinoblastoma gene be equally likely in maternally and paternally derived chromosomes.

The paternal origin of most new germline mutations is easily explained if they arise predominantly during spermatogenesis. Such a hypothesis would be supported by a paternal age effect, in which new cases of bilateral retinoblastoma might occur more frequently in the offspring of older fathers. Although we detected no paternal age effect in our small set of patients (see Table 1), a review of published data allowed Vogel and Rathenberg to conclude that a weak paternal age effect does exist for bilateral retinoblastoma[13]. Furthermore, if germline mutations arise predominantly during DNA replication, then the observed excess of germline mutations in paternal versus maternal copies of chromosome 13 (16:1) should correlate with the number of cell divisions from embryonic development to meiosis in males versus females. Vogel and Rathenberg provide the following estimates: 380 divisions for sperm from a 28-year-old male; 23 for ova from an adult female. The ratio (380:23 or 16.5:1) agrees with the data.

In summary, our investigations of the parental origin of mutations of the retinoblastoma gene cast serious doubt on the notion that initial somatic mutations occur predominantly on the chromosome 13 derived from the father, and they eliminate a theoretical requirement for genomic imprinting as a factor in somatic mutagenesis at the retinoblastoma locus. A comparison of our results with studies of chromosome 11p in Wilms tumour[14,15], where the paternally derived chromosome was preferentially retained in tumour cells, is complicated by the fact that the locus governing hereditary susceptibility to Wilms tumour is not on that chromosome arm[16,17]. Furthermore, those studies included patients with bilateral disease, indicating a constitutional, if not heritable, susceptibility to the tumours. On the other hand, our results and those from a previous study[12] convincingly demonstrate that the majority of new germline mutations of the retinoblastoma gene arise in the paternal chromosome 13. Future investigations of the high mutation rate of the retinoblastoma locus should explore mutagenic factors peculiar to fathers. □

Received 29 March; accepted 11 May 1989.

1. Knudson, A. G. Jr *Proc. natn. Acad. Sci. U.S.A.* **68,** 820–823 (1971).
2. Comings, D. E. *Proc. natn. Acad. Sci. U.S.A.* **70,** 3324–3328 (1973).
3. Friend, S. H., Dryja, T. P. & Weinberg, R. A. *New Engl. J. Med.* **318,** 618–622 (1988).
4. Dryja, T. P., Rapaport, J. M., Joyce, J. M. & Petersen, R. A. *Proc. natn. Acad. Sci. U.S.A.* **83,** 7391–7394 (1986).
5. Friend, S. H. *et al. Nature* **323,** 643–646 (1986).
6. Lee, W.-H. *et al. Science* **235,** 1394–1399 (1987).
7. Fung, Y-KT. *et al. Science* **236,** 1657–1661 (1987).
8. Toguchida, J. *et al. Nature* **338,** 156–158 (1989).
9. Dryja, T. P. *et al. New Engl. J. Med.* **310,** 550–553 (1984).
10. Vogel, F. *Hum. Genet.* **52,** 1–54 (1979).
11. Reik, W. & Surani, M. A. *Nature* **338,** 112–113 (1989).
12. Ejima, Y., Sasaki, M. S., Kaneko, A. & Tancoka, H. *Hum. Genet.* **79,** 118–123 (1988).
13. Vogel, F. & Rathenberg, R. *Adv. hum. Genet.* **5,** 223–318 (1975).
14. Schroeder, W. T. *et al. Am. J. hum. Genet.* **40,** 413–420 (1987).
15. Mannens, M. *et al. Hum. Genet.* **81,** 41–48 (1988).
16. Grundy, P. *et al. Nature* **336,** 374–376 (1988).
17. Huff, V. *et al. Nature* **336,** 377–378 (1988).
18. Wiggs, J. *et al. New Engl. J. Med.* **318,** 151–157 (1988).
19. Bookstein, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **85,** 2210–2214 (1988).
20. Dryja, T. P., Rapaport, J. M., Weichselbaum, R. & Bruns, G. A. P. *Hum. Genet.* **65,** 320–324 (1984).
21. Cavenee, W., Leach, R., Monandus, T., Pearson, P. & White, R. *Am. J. hum. Genet.* **36,** 10–24 (1984).
22. Yandell, D. W. & Dryja, T. P. *Am. J. hum. Genet.*, in the press.

ACKNOWLEDGEMENTS. We thank Tracey Campbell, Siri Dayton and Terri McGee for Fig. 1. This research was supported by the NIH and the American Cancer Society, by the Alcon Research Institute Inc., and by gifts to the Taylor R. Smith Laboratory. T.P.D. is a Research to Prevent Blindness-Dolly Green Scholar.

The *Neurospora* clock gene *frequency* shares a sequence element with the *Drosophila* clock gene *period*

C. Robertson McClung*, Barbara A. Fox & Jay C. Dunlap†

Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03756, USA

THE isolation and characterization of single gene mutations affecting the circadian biological clocks of several organisms (reviewed in ref. 1) has left little doubt that circadian rhythms can be subjected to classical genetical analysis. Many of these mutations occur at the same few genetic loci (*frequency* (*frq*) in the fungus *Neurospora*[2], and *period* (*per*) in fruit fly *Drosophila*[3]); these loci represent the best studied clock-affecting genes known. Mutant strains are usually affected in more than one basic clock property[1,4], suggesting an inter-relatedness at the molecular level among these basic properties that would not have been predicted *a priori*. The extensive background information available concerning the *frq* locus[4] provides a basis for the molecular dissection of the *Neurospora* circadian clock—the most minimal circadian system thus far described. We report here the cloning and analysis of the *frq* locus and show it to be larger and more complex than would have been predicted from the available genetic data. Complete rescue of all of the pleiotropic mutant phenotypes[5,6] of the recessive frq^9 allele requires transformation with a 7.7-kilobase (kb) region of DNA encoding at least two transcripts. Sequence analysis of this region has allowed the identification of a common element between *frq* and *per* which, given the background similarities in their classical genetic characteristics, suggests the possibility of a common element in the clock mechanisms of these two organisms.

The *frq* locus of *Neurospora crassa* lies on the right arm of linkage group VII (VII R), ~2 map units distal to *oligomycin resistance* (*oli*) and 2.5 map units proximal to *formate* (*for*) (Fig. 1*a*). A chromosomal walk was undertaken starting from the *oli* gene[7] and covering ~190 kb corresponding to ~8 map units[8] along VII R (Fig. 1*b*). Individual phage and cosmids arising from the walk were assayed by transformation[9] for their ability to rescue the circadian banding pattern of the recessive frq^9 allele (Fig. 1*b*). Four partially overlapping cosmids arising from the right-hand side of the walk were capable of rescuing the frq^9 phenotype, providing strong presumptive evidence that these four cosmids each contained *frq*. This was confirmed in two ways. First, in agreement with expectations based on the genetic map, the next rightward cosmid (31:5E) was found to be capable of rescuing *for* mutations (Fig. 1). Second, DNA fragments arising from cosmids 2:10A, 8:3B and 23:9D were mapped by restriction fragment length polymorphism analysis[10], confirming for all three a genetic map location on VII R (1/38 recombinants with *ars-1* near the centromere, 6/38 with *nic-3* on VII L; data not shown). We have thus established a physical map covering ~8 map units along the right arm of linkage group VII, shown that the physical map corresponds well to the known genetic map, and identified both the *frq* and *for* genes.

At this resolution the *frq* locus was defined by the 13 kb that span the region of overlap of the four frq^9 complementing cosmids. This region was subjected to restriction analysis, subcloning, and retransformation of subclones into frq^9 (Fig. 1*c*). Based on previous experience with the *frq*-containing cosmid clones, *for*, and other genes, two outcomes of these transforma-

* Present address: Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA.
† To whom correspondence should be addressed.

© 1989 Nature Publishing Group

tion/rescue experiments were anticipated: either a complete complementation of the pleiotropic *frq*^{9} mutation or a complete lack thereof. Surprisingly there appeared a third intermediate phenotype (denoted ±) in which some of the mutant characteristics of *frq*^{9} were complemented (see Fig 1*c*, legend) but in which there was still an absence of circadian conidial banding; the reverse phenomenon (complementation of circadian banding in the absence of robust conidiation) was not observed. These studies have shown that the smallest region of DNA capable of complete complementation of the circadian regulation phenotype is the ~7.7 kb insert of pKAJ101. In the transformation assay, *frq* generally displays a variable phenotype in that the circadian banding pattern of *frq*^{9} strains transformed to benomyl resistance and containing *frq*$^+$ DNA is phenotypically transformed only ~20% of the time. But in nearly all cases where we have examined this subset of transformed strains, the period length of the circadian oscillation has been indistinguishable from wild-type (average period length among transformants 20.9 ± 1.8 h (s.d.), $n = 60$ versus 21.6 ± 0.5 h, $n = 5$ for wild-type), and the pattern of temperature compensation appears to be completely normal (Fig. 2). There is a suggestion, however, that the variability of the period length was consistently greater among transformants than in wild-type strains. Transformants could be entrained normally to daily 5-minute light pulses but in some cases the stable phase of entrainment (that is, the time of day at which the centre of the conidial band occurred) was different from wild-type.

The *frequency* locus gives rise to at least two processed transcripts, one ~1.5 kb in size arising from the left-hand side of the region and one or more large overlapping transcripts of ~5 kb arising from the right-hand side (Fig. 3). The origin of these transcripts within the *frq*-complementing region, as

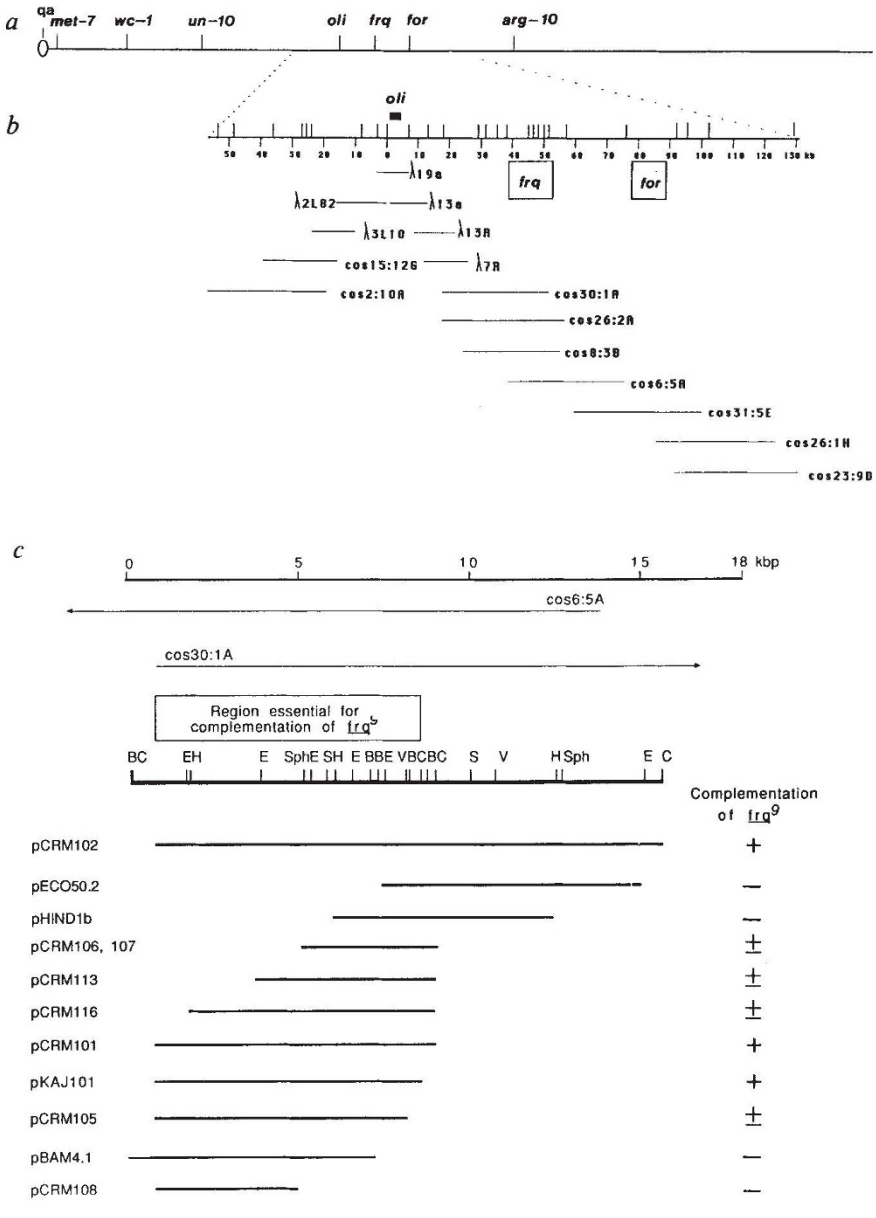

FIG. 1 Localization of *frq* gene sequences on *Neurospora crassa* linkage group (LG) VIIR: correlation of the genetic map and the physical map via chromosomal walking from *oli*. *a*, Genetic map of LG VIIR, with the centromere indicated by the open circle at left. Genetic distances are *oli-frq* ~2 map units and *frq-for* ~2 map units. *b*, Physical map of the region of LG VIIR around *oli* (solid box). Dotted lines extending from the genetic map (*a*) to the physical map indicate the approximate correlation of the two maps. Vertical lines above the map indicate *Eco*RI restriction sites; restriction sites for other enzymes are not shown for the sake of clarity. Lines below the map indicate the positions of inserts of phage and cosmid clones identified in the walk. Open boxes beneath the physical map indicate the initial positioning of *frq* and *for* genes. *c*, Expansion of the region defined as common to the four cosmids complementing *frq*^{9} (open box labelled *frq* in *b*). Restriction enzymes are *Bam*HI (B), *Cla*I (C), *Eco*RI (E), *Eco*RV (V), *Hin*dIII (H), *Sal*I (S) and *Sph*I (Sph). Inserts of plasmid subclones used to transform *frq*^{9} recipients are indicated by the narrow lines below the restriction map. Insert end points correspond to restriction sites indicated in the restriction map with the exception of the right ends of pCRM101, pCRM102, pCRM105, pCRM108 and pKAJ101 which are defined by the insert–vector junction of cos 30:1A. The ability of the subclones to rescue the *frq*^{9} phenotype is indicated at the right; − denotes no phenotypic rescue, + denotes full restoration of circadian banding with wild-type period length, and restoration of wild-type levels of pigmentation and conidiation, and ± denotes restoration of wild-type pigmentation and conidiation but no restoration of circadian banding (see text).

METHODS. Steps were taken in a λ genomic library (M. Orbach and C. Yanofsky, Stanford University) and later in an ordered cosmid library[9]; chromosomal sequences between cos 15:12G and cos 30:1A are not represented in the Vollmer-Yanofsky cosmid library. Cosmids and subclones[24] were used to transform *frq*^{9} *Neurospora* to benomyl resistance (Bm^r); phage clones were co-transformed with pSV50 (ref. 9). Primary transformants (~500 per μg DNA) were identified after three days growth at 30 °C, transferred to Horowitz Complete[25] slants supplemented with 500 ng ml^{-1} benomyl (Bm), initially screened for phenotypic rescue by visual inspection of conidial pigmentation and subsequently screened on race tubes[6]. Media for race tubes of Bm^r transformants did not contain Bm because the presence of as little as 100 ng ml^{-1} of the drug interfered with conidial banding pattern even in resistant strains (data not shown). For the identification of *for* containing cosmids and subclones, primary transformants of *for*$^-$ were identified on Horowitz Complete medium plates supplemented with 500 ng ml^{-1} Bm and 500 μg ml^{-1} sodium formate, and individual transformants then screened in Vogel's minimal liquid cultures with or without sodium formate. The *for* gene was placed between 74 and 77 kb distal from *oli* by the finding that the *for*-complementing ability unique to 31:5E was preserved following digestion with *Hin*dIII and *Bgl*II (data not shown).

© 1989 Nature Publishing Group

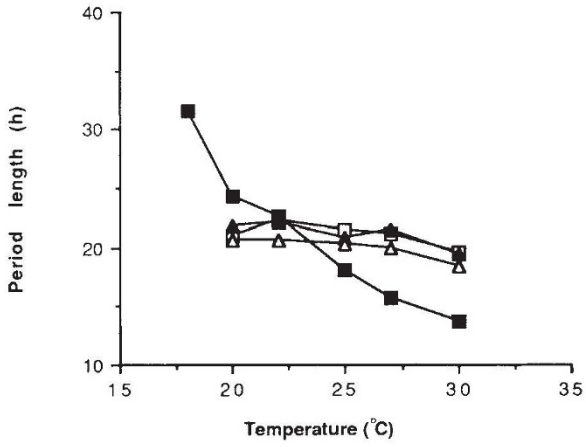

FIG. 2 Period length and temperature compensation characteristics of the clock are completely restored in phenotypically rescued *frq*^{9} transformants. Period length of the rhythm is plotted versus temperature for *bd,frq*$^+$ (open squares), *bd,frq*^{9} (closed squares) and for *bd,frq*^{9} strains transformed with pCRM101 (closed triangles) or pCRM102 (open triangles). For comparison, *bd,frq*^{9} data are replotted from ref. 5.
METHODS. A single race tube having six to eight circadian cycles was considered to yield a single estimate of period length[26]. Points here were the average of between three and 25 race tubes, depending on the experiment, and standard deviations, omitted for the sake of clarity, ranged from 0.3 h to 1.5 h; error bars at ±1 s.d. overlap for all points except *frq*^{9}.

determined by hybridization analysis using different subcloned pieces (Fig. 3), agrees well with the limits of the *frq* region as determined by phenotypic analysis of transformants receiving different regions of DNA (Fig. 1*c*). It thus appears that DNA sufficient to allow expression of all of the transcripts must be present to rescue fully *frq*^{9} recipients, whereas to a first approximation, clones containing only the region encoding the long transcript(s) are sufficient for partial rescue.

Figure 4*a* contains the DNA sequence for an 8,657 base-pair (bp) region of genomic DNA containing the inserts of pKAJ101 and pCRM101, the interval containing the transcription unit(s) encoding the *frq* gene products. Within the first (5′) 200 bp of pKAJ101 (that is, within the region shown to be required for *frq* complementation; compare pCRM116 with pKAJ101), there exists an extended inverted repeat (18/22 matches separated by 100 bp of potential loop) which if transcribed could yield a stable stem loop ($\Delta G(25\,^\circ C) = -23.7$ kcal; ref. 11) of the type implicated in translational control under some conditions[12]. Computer analysis of the entire sequence 3′ of this reveals the presence of 11 open reading frames (ORFs) greater than 100 amino acids in length, one of which deserves special comment. Within the region corresponding to the long transcript(s), there is a single very long ORF of 2,364 bp that extends from nucleotide 5174 to 7538 and appears in only one of the three possible reading frames (Fig. 4). Predicted codon usage within this region, based on a conceptual translation, is strongly biased in a manner typical for expressed genes in *Neurospora*[13]. This would not be expected in randomly chosen nucleotide triplets and suggests that at least part of this ORF may be expressed and can be used to predict a part of the amino-acid sequence of the *frq* gene product (see below).

The entire *frq* sequence shown in Fig. 4*a* (as four overlapping 3,000-nucleotide windows) was used to search GENBANK (July, 1988 release) using FASTA[14]. These searches failed to identify any gene or protein with extended homology to *frq* and, with one exception, none of the genes or gene products identified suggested any biologically relevant relationship. The region of DNA extending from nucleotides 4001 to 7000, however, did identify a sequence element in the *per* gene of *Drosophila* similar (Z value of 3.5, ref. 15) to a region within *frq* (extending from nucleotides 5776 to 5830). The long ORF referred to above lies within this segment of DNA. Although this similarity would be classified only as possibly statistically significant[16], we find it noteworthy for the following reasons: (1) Although several other sequences bearing weak similarities to *frq* were detected, these were generally based on small regions of highly biased nucleotide composition and were of marginal significance; (2) The similarity with *per* apparently extends beyond the nucleic acid to the protein sequence. When the conceptual translation of the region of similarity within the long ORF was used to search the NBRF (National Biomedical Research Foundation) database using FASTA[14], the homology to *per* reappeared, again among the strongest of a number of weak similarities that included several proteoglycans. Part, but not all, of the region of *per* sharing similarity with the *frq* long ORF contains a repeating unit of threonine and glycine (TG repeat[17,18]). This region is contained within the longest exon in the *per* transcription unit; (3) part of the region of similarity in the *frq* long ORF is a modified TG repeat (Fig. 4*b*) where there is considerable diversity in the third position among the codons used. Thus it does not appear simply as an oligonucleotide repeat. Furthermore, the similarity between *frq* and *per* is not confined to the direct TG repeat; it extends both 3′ and 5′ beyond the direct repeat region of both genes (Fig. 4*b*).

The region containing the TG repeat is one of the salient features of the *per* gene product in *Drosophila*, where it may act as the site for some of the extensive glycosylation thought to be present in the *per* protein[17,19]. As in *per*, the region in the *frq* long ORF within and around the similarity region contains several possible sites for both N-linked and O-linked

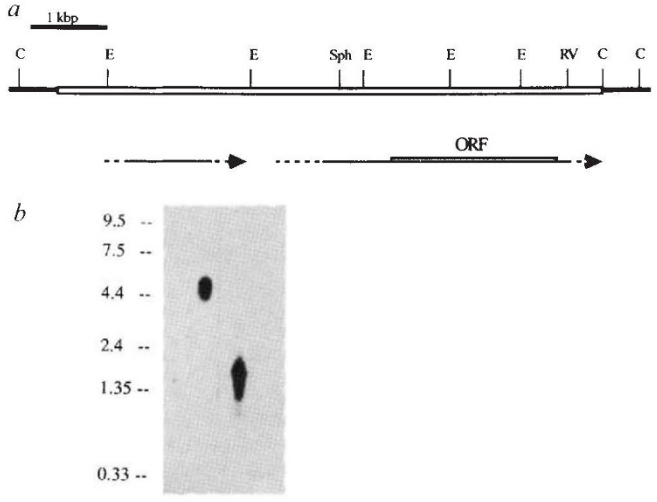

FIG. 3 Transcription of the *frq* locus region. *a*, Map of the region required for rescue of the *frq*^{9} phenotype. Sites are as in Fig. 1*c*. Arrows beneath the map indicate transcripts detected in poly(A)$^+$ RNA from *bd,frq*$^+$ mycelia; open box above the long transcript indicates the position of the long ORF (see text). Dotted lines indicate imprecision concerning transcriptional initiation and termination sites. *b*, RNA blot analysis of transcripts prepared from *bd,frq*$^+$ mycelia. Left hand panel shows hybridization to the *Cla*I/*Sph*I fragment extending from nucleotide 4,604 to 8,180 in Fig. 4*a*; right hand panel shows hybridization to the *Eco*RI fragment extending from 1,201 to 3,207.
METHODS. Blots using RNA isolated from *bd,frq*$^+$ mycelia were prepared and processed for autoradiography as described previously[27]. Direction of transcription and localization of the transcripts were determined using single-stranded DNA probes[28] generated from each of the *Eco*RI fragments shown in the restriction map in *a*.

© 1989 Nature Publishing Group

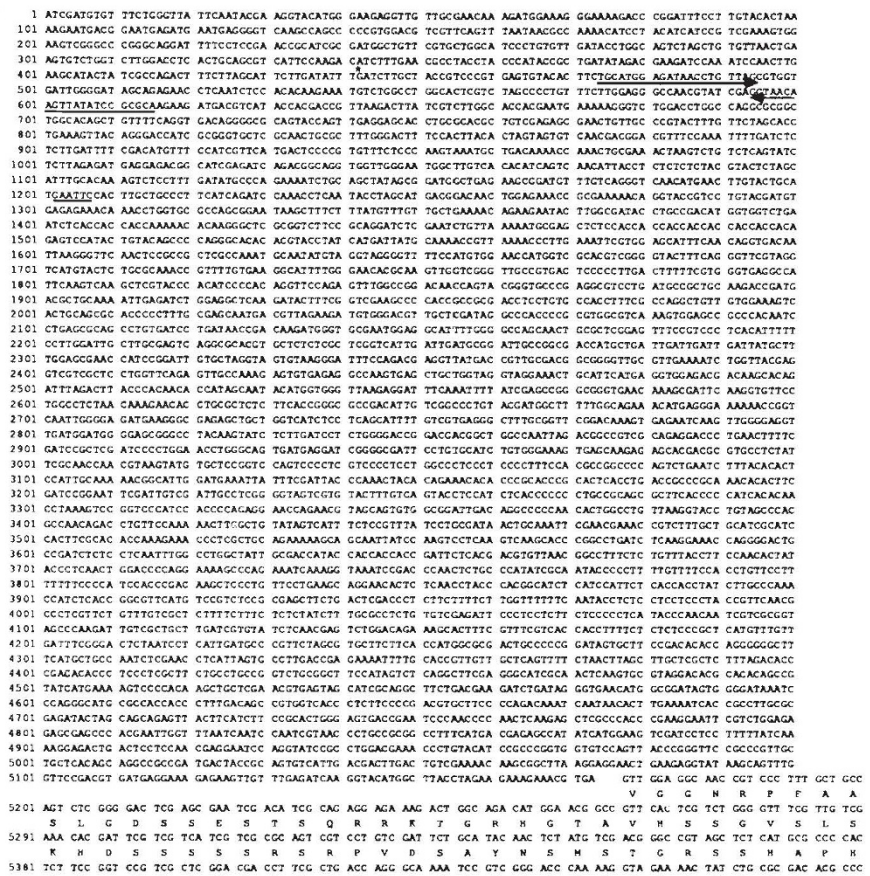

FIG. 4 *a*, DNA sequence of the *frq* gene region. Position 1 of the sequence corresponds to the *Cla*I site at the left hand of the restriction map in Fig. 1*c*. The location of an inverted repeat near the beginning of the sequence (see text) is noted by arrows below the sequence and the end of the insert of cosmid 30:1A is noted by an asterisk above the sequence at position 442. Deletion 3′ of this to the *Eco*RI site (underlined) at position 1,202 (pCRM116) results in fragments capable of only partial rescue. The *Cla*I site (underlined) at position 8,179 marks the end of pKAJ101. Deletion 5′ of this to the *Eco*RV site (underlined) as position 7,765 (pCRM105) results in fragments capable of only partial rescue. Deletion further 5′ to the *Bam*Hi site (underlined) at position 6,619 (pBAM4.1) completely eliminates all phenotypic rescue (see text and Fig. 1*c*). Conceptual translation product of a 788-amino-acid polypeptide encoded by the open reading frame starting at position 5,174 is indicated below the DNA sequence, and the region bearing strongest similarity to the *per* gene is underlined. The numbers in the right margin indicate nucleotide number from the beginning. *b*, Alignment of the core of the region of similarity between the conceptual translations of *frq* and *per*. Amino-acid identities are boxed and residues generally described as conservative replacements[29,30] are denoted by a dot.

METHODS. *a*, Ordered sets of overlapping deletion subclones[31] were generated from *Neurospora* genomic DNA cloned into the Bluescript series of vectors (Stratagene), and gaps in the sequence were resolved by cloning specific subregions. The nucleotide sequence was determined for both strands by the dideoxy method[32]. *b*, Alignment was generated using the Align program[33] as implemented in the FASTA program package[14].

© 1989 Nature Publishing Group

glycosylation, thus suggesting that the *frq* gene product could, like *per*, be glycosylated. Additionally, analysis of the conceptual protein sequence both 5′ and 3′ of the similarity region reveals the presence of several potential kinase A and kinase C phosphorylation sites[20], although in the absence of a complete complementary DNA sequence, speculation as to the potential importance of these sites seems premature.

The apparent size and diversity of the *frq* products are noteworthy, particularly in light of previous genetic studies. Fine structure mapping of the existing *frq* alleles[6] has suggested that they lie within ~0.003 map units of each other, a region comprising a few as 100 bp. Preliminary studies on the *frq* mutant alleles have identified no differences on Southern blots (data not shown), thus eliminating the presence of gross structural distortions in the DNA (which might reduce recombination frequencies) as the molecular basis of these mutations. We suggest therefore that there may be a region of *frq* which is either particularly mutable or particularly important to its function as a part of the circadian clock.

In addition to the surprising size and molecular complexity of the *frq* locus, the identification of a region of similarity with *per* was unexpected. In selecting *per*, the search identified the only other clock gene that has been cloned to date, a gene whose mutations confer phenotypes similar in many ways to the documented phenotypes of the various *frq* alleles[1,2]. The sequence similarity is not extensive so that its existence may be merely coincidental, and given our present state of knowledge of the products of *frq* it is clearly impossible to conclude that the two genes share a common ancestor. Alternatively it has been something over 1×10^9 years since plants and fungi diverged from animals during the Precambrian era[21], plenty of time for genetic drift to eliminate similarities. Because a part of the region of similarity in each gene contains similar structural features including several potential glycosylation sites, this similarity could reflect convergent evolution to protein structural elements necessary for the function of these and perhaps other clock proteins. Indeed, if the similarity is biologically relevant, it is noteworthy and suggests that 'clock' genes in other organisms might be identified through hybridization analysis. In light of this, the identification of genes in other organisms based on hybridization to *per*[22,23] may take on a renewed significance. In hybridization experiments at moderate stringency, DNA from the *frq* locus has been used to screen genomic DNAs from a number of vertebrate, invertebrate, fungal and plant systems. The identification in this way of potential homologues in yeast, *Arabidopsis* (J.C.D., S. Kay, A. Miller, N.-H. Chua; unpublished data), soybean and several mammalian species, combined with the novel ability to transform and genetically manipulate these organisms, suggests the possibility of extending the molecular genetic analysis of circadian clocks to these species as well. □

Received 22 February; accepted 8 May 1989.

1. Hall, J. C. & Rosbash, M. *A. Rev. Neurosci.* **11,** 373–393 (1988).
2. Gardner, G. F. & Feldman, J. F. *Genetics* **96,** 877–886 (1980).
3. Konopka, R. J. *Fed. Proc.* **38,** 2602–2605 (1979).
4. Feldman, J. F. & Dunlap, J. C. *Photochem. Photobiol. Rev.* **7,** 319–368 (1983).
5. Loros, J. J. & Feldman, J. F. *J. biol. Rhythms* **1,** 187–198 (1986).
6. Loros, J. J., Richman, A. R. & Feldman, J. F. *Genetics* **14,** 1095–1110 (1986).
7. Viebrock, A., Perz, A. & Siebald, W. A. *EMBO J.* **1,** 565–571 (1982).
8. Perkins, D. D., Radford, A., Newmeyer, D. & Bjorkman, M. *Microbiol. Rev.* **46,** 426–570 (1982).
9. Vollmer, S. J. & Yanofsky, C. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1–5 (1986).
10. Metzenberg, R. L., Stevens, J. N., Selker, E. U. & Morzycka-Wroblewska, E. *Proc. natn. Acad. Sci. U.S.A.* **82,** 2067–2071 (1985).
11. Tinoco, I. *et al. Nature New Biol.* **246,** 40–41 (1973).
12. Kozak, M. *Molec. cell. Biol.* **8,** 2737–2744 (1988).
13. Orbach, M. J., Porro, E. A. & Yanofsky, C. *Molec. cell. Biol.* **6,** 2452–2461 (1986).
14. Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85,** 2444–2448 (1988).
15. Lipman, W. J. & Pearson, W. R. *Science* **227,** 1435–1441 (1985).
16. Doolittle, R. F. *Of URFs and ORFS* (University Science Books, Mill Valley, 1986).
17. Jackson, F. R., Bargiello, T. A., Yun, S. H. & Young, M. W. *Nature* **320,** 185–188 (1986).
18. Citri, Y. *et al. Nature* **326,** 42–47 (1987).
19. Reddy, P., Jacquier, A. C., Abovich, N., Petersen, G. & Rosbach, M. *Cell* **46,** 53–61 (1986).
20. Gonzalez, G. A. *et al. Nature* **337,** 749–752 (1989).
21. Dayhoff, M. O. in *Atlas of Protein Sequence and Structure* Vol. 5, suppl. 3. (ed. Dayhoff, M. O.) 1–8 (Natn. biomed. Res. Found., Silver Spring, 1978).
22. Shin, H. S., Bargiello, T. A., Clark, B. T., Jackson, F. R. & Young, M. W. *Nature* **317,** 445–448 (1985).
23. Weber, M. L., deGroot, E. J. & Schweiger, H. G. *Molec. gen. Genet.* **209,** 1–7 (1987).
24. McClung, R.. M., Phillips, J. D. Orbach, M. J. & Dunlap, J. C. *Expl. Mycol.* (in the press).
25. Davis, R. L. & deSerres *Meth. Enzym.* **71A,** 79–143 (1970).
26. Dunlap, J. C. & Feldman, J. F. *Proc. Natn. Acad. Sci. U.S.A.* **85,** 1096–1100 (1988).
27. Loros, J. J., Denome, S. & Dunlap, J. C. *Science* **243,** 385–388 (1989).
28. Hu, N. & Messing, J. *Gene* **17,** 271–277 (1982).
29. Feng, D. F., Johnson, M. S. & Doolittle, R. F. *J. molec. Evol.* **21,** 1112–125 (1985).
30. Schwartz, R. M. & Dayhoff, M. O. in *Atlas of Protein Sequence and Structure* Vol. 5, Suppl. 3 (ed. Dayhoff, M. O.) 353–358 (Natn. biomed. Res. Found, Silver Spring, 1978).
31. Henikoff, S. *Gene* **28,** 351–359 (1984).
32. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5463–5467 (1977).
33. Myers, E. & Miller, W. *CABIOS* **4,** 11–17 (1988).

ACKNOWLEDGEMENTS. We thank J. Phillips and S. Denome for technical assistance, Jennifer Loros and B. Masucco for help in analysis of race tubes, S. Brody for pointing out potential phosphorylation sites, and J. Loros, J. Hall and J. Feldman for comments on the manuscript. This work was supported by the NIH and the Norris Cotton Cancer Center (J.C.D.).

RNA helicase activity associated with the human p68 protein

H. Hirling, M. Scheffner, T. Restle & H. Stahl

Fakultät für Biologie, Universität Konstanz, D-7750 Konstanz, FRG

IT has been proposed that p68, a nuclear protein of relative molecular mass 68,000, functions in the regulation of cell growth and division[1]. A complementary DNA analysis of the protein[2] has revealed extensive amino-acid sequence homology to the products of a set of genes recently identified in organisms as diverse as *Escherichia coli* and man, which include the eukaryotic translation initiation factor eIF-4A (refs 3–9). The protein products of the new gene family have several motifs in common which are thought to be involved in nucleic acid unwinding[10–12]. As yet, however, only eIF-4A, through its effect on RNA, has been shown to possess unwinding activity[13,14]. Here we report that purified p68 also exhibits RNA-dependent ATPase activity and functions as an RNA helicase *in vitro*. The protein was first identified by its specific immunological cross reaction with the simian virus 40 large T antigen[1], the transforming protein of a small DNA tumour virus[15]. Surprisingly, T antigen also has an RNA-unwinding activity[16]: the homology between the two polypeptides, although confined to only a small region resembling the epitope of the cross-reacting antibody (PAb204) (ref. 2), should therefore be of functional significance. Furthermore, the RNA-unwinding activity may be involved in the growth-regulating functions of both proteins.

Cell fractionation analysis has revealed a close association of the p68 protein with the nuclear matrix (unpublished data). Therefore, conditions were developed to quantitatively extract native p68 from cell nuclei, followed by a conventional chromatographic purification program (see Fig. 1*a* legend for details). Essentially homogenous p68 protein was obtained which sedimented at 4–5 S in a sucrose gradient (Fig. 2*b*).

During protein purification we made use of a single-stranded (ss) DNA cellulose column to which p68 bound with high affinity. The binding of purified p68 to nucleic acid was further tested using radiolabelled RNA in a nitrocellulose filter assay. Efficient RNA binding (resistant to salt up to 250 mM; data not shown) was observed irrespective of the presence or absence of Mg^{2+} and ATP, and could be best competed with RNA (total cytoplasmic RNA from HeLa cells) or ss M13 DNA (Fig. 1*b*). Double-stranded (ds) DNA competed only partially for RNA binding, and no binding of p68 to ds DNA was found at 100 mM NaCl (data not shown). RNA binding has also been reported for the *E. coli* SrmB protein, another member of the eIF-4A-like family, and this activity may be a function of the relatively basic C-terminus of the two proteins[6]. Factor eIF-4A lacks this region and the smaller eIF-4A protein binds RNA only in the presence of a second protein[17].

The SrmB protein, like eIF-4A, is an RNA-dependent

© 1989 Nature Publishing Group