Hum Genet (2001) 109 : 136–142
DOI 10.1007/s004390100550

ORIGINAL INVESTIGATION

Quasar Saleem · Anuranjan Anand · Sanjeev Jain
Samir K. Brahmachari

The polyglutamine motif is highly conserved at the *Clock* locus in various organisms and is not polymorphic in humans

Received: 2 April 2001 / Accepted: 24 May 2001 / Published online: 19 July 2001

Abstract Circadian rhythms play a central role in diverse physiological phenomena and the recent years have witnessed the identification of a number of genes responsible for the maintenance of these rhythms. One of these is the *Clock* gene, which was first identified in mouse and subsequently in a large number of organisms, including humans. The human *Clock* gene has been proposed as a possible candidate for disorders affected by alterations of circadian rhythm, including bipolar disorder and schizophrenia. This gene contains a highly conserved polyglutamine motif, that in humans is coded for by CAG repeats. In view of the involvement of CAG repeat expansion in a number of neuro-psychiatric disorders, we have sought to determine the polymorphism status of CAG repeats at the *Clock* locus in humans. Our analysis of 190 unrelated individuals, who included patients suffering from bipolar disorder and schizophrenia, indicated that the repeat, which consisted of 6 CAG triplets, was not polymorphic in humans. An analysis of the repeat in non-human primates and other organisms revealed that the glutamine stretch is shortest in humans and baboons, and longest in *Drosophila* and zebrafish. A study of various *Drosophila* species revealed that the repeat number is highly polymorphic, ranging from 25 to 33 pure glutamine repeats. Unlike most other microsatellites, the CAG repeat stretch at the *Clock* locus in humans is smaller than its homologues in non-human primates. We propose that glutamine repeat size is functionally important in this gene and thus tightly regulated. The variation in repeat number is probably deleterious to the individual, resulting in the maintenance of a short and invariable repeat structure in the human population.

Q. Saleem · S.K. Brahmachari (✉)
Functional Genomics Unit,
Centre for Biochemical Technology (CSIR),
Delhi University Campus, Mall Rd., Delhi–110 007, India
e-mail: skb@cbt.res.in,
Tel.: +91-11-7416489 or 7667298,
Fax: +91-11-7416849 or 7667471

Q. Saleem · S.K. Brahmachari
Molecular Biophysics Unit, Indian Institute of Science,
Bangalore, India

A. Anand
Molecular Biology and Genetics Unit,
Jawaharlal Nehru Centre for Advanced Scientific Research,
Bangalore, India

S. Jain (✉)
Department of Psychiatry,
National Institute of Mental Health and Neurosciences,
Bangalore, India

Introduction

Circadian rhythms underlie diverse biological functions in living systems and help organisms synchronize their activities with the 24-h periodicity of night and day. The recent years have witnessed significant progress in the elucidation of the molecular mechanisms responsible for circadian rhythms, brought about primarily by the identification of genes that play a central role in their generation and maintenance (Wilsbacher and Takahashi 1998; Wager-Smith and Kay 2000). The first of such genes to be discovered in a mammalian system was the mouse *Clock* gene (Antoch et al. 1997; King et al. 1997). Subsequent to its discovery in mice, homologues of the gene were cloned in a variety of organisms, including *Drosophila* and humans, and showed a remarkable degree of conservation across species (Allada et al. 1998; Whitmore et al. 1998; Abe et al. 1999; Steeves et al. 1999). The human *Clock* gene codes for an 846-amino-acid protein and is 96% identical to its mouse orthologue (Steeves et al. 1999). The gene encodes a novel member of the basic helix-loop-helix (bHLH)/PAS family of transcription factors, and the *Drosophila* homologue of this gene has been shown to induce transcription of two other circadian rhythm genes, *per* and *tim* (Darlington et al. 1998; Steeves et al. 1999). The C terminal portion of this protein contains a glutamine-rich region, part of which, in humans, is encoded by a continuous stretch of 6 CAG repeats (Steeves et al. 1999).

Trinucleotide repeats, in particular CAG repeats, have been the focus of a great deal of interest in the recent years, because instability of these sequences have been implicated in an ever-growing list of neurological disorders (Cummings and Zoghbi 2000). Unlike conventional mutations, where changes in the DNA sequence are stably inherited, these so-called "dynamic mutations" result in an alteration of the number of repeat units when transmitted from one generation to the next. While the normal population shows polymorphism for repeat number, instability and subsequent disease is observed only when the repeat number exceeds a threshold value of approximately 40 repeat units (Brahmachari et al. 1995; Reddy and Housman 1997). In cases where the CAG repeats are part of the coding regions of the disease genes, they code for glutamine, and repeat expansion results in long polyglutamine tracts, which are thought to play a central role in the pathogenesis of these disorders (Ross et al. 1999). Normal variations in glutamine repeats have also been implicated in psychiatric disorders, such as schizophrenia and bipolar disorder (Chandy et al. 1998; Saleem et al. 1998; Saleem et al. 2000; Vincent et al. 2000). While there have been few reports of expansions in the range seen in disorders caused by trinucleotide repeat expansions, most reports have indicated a trend towards larger repeats within the normal range in patients with psychiatric disorders (Vincent et al. 2000).

The human *Clock* gene is an attractive candidate for a large number of diseases because of the importance of circadian rhythms in human physiology. Alterations of circadian rhythms have been reported in a number of psychiatric disorders, such as schizophrenia, major depression and bipolar disorder, and Jain (1995) proposed that studying genes involved in circadian rhythms may be crucial to our understanding of bipolar disorder (Wehr and Rosenthal 1989; Wirz-Justice et al. 1997; Jain 1995). A seasonal increase in admissions for mania is often reported from temperate zones, where fluctuations of light and temperature are larger, while this is less commonly reported from tropical areas (Wehr and Rosenthal 1989; Jain et al. 1992). Phototherapy and artificial alteration of the zeitgeber can treat depression and cause a switch into mania in susceptible individuals (Wehr et al. 1979). The range of disturbances seen in bipolar disorder suggests that disruption of an endogenous rhythm may be a basic pathology in its origin (Mitterauer 2000). A single nucleotide polymorphism (SNP) in the 3' untranslated region (UTR) of the human *Clock* gene was reported to be associated with diurnal preferences (Katzenburg et al. 1998). This SNP was also studied in patients with major depression; however, no association was found (Desan et al. 2000). The presence of a CAG-repeat stretch may also make this gene susceptible to repeat expansion, as seen in the other disorders caused by trinucleotide repeat instability. An essential feature of all these disease genes is the polymorphic nature of CAG repeats in the normal population (Cummings and Zoghbi 2000). The first step to assess a locus as a candidate for repeat expansion is to determine whether the repeats at this locus are polymorphic in normal individuals. Furthermore, it has also been suggested that these repeat sequences are a source of quantitative variation, and alterations in glutamine number may also modulate the functioning of the *Clock* protein, as occurs in other proteins containing glutamine repeats (Tut et al. 1997; Irvine et al. 2000). We have, therefore, analysed the range of CAG repeats at the human *Clock* locus in the normal population. We have also looked for CAG repeat variation in patients suffering from bipolar disorder and schizophrenia, as these diseases are associated both with alterations in circadian rhythms and perhaps also with CAG repeat instability. To understand the mechanisms involved in the evolution of the CAG repeats at this highly conserved locus, we have analysed the CAG repeat structure in non-human primates and other organisms in which this gene has been identified.

Materials and methods

CAG-repeat polymorphisms were tested in 190 unrelated Indian individuals, comprising 111 normal individuals, 50 bipolar patients and 29 patients suffering from schizophrenia. Bipolar and schizophrenia samples were obtained from the National Institute of Mental Health and Neurosciences, Bangalore. These patient samples have been used in earlier published studies (Saleem et al. 1998; Saleem et al. 2000) and informed consent was obtained from all individuals before extraction of blood.

The following species of non-human primates were used in the analysis of CAG repeats at the *Clock* locus: chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), baboon (*Papio hamadryas*), bonnet monkey (*Macaca radiata*), rhesus monkey (*Macaca mulatta*) and langur (*Presbytis entellus*). DNA samples from one chimpanzee and one gorilla were obtained from the Centre for Cellular and Molecular Biology, Hyderabad, India. The remaining monkey samples were obtained from the Primate Research Centre at the National Institute of Immunology, Delhi and Indian Institute of Science Bangalore. Samples from 22 individual bonnet monkeys and 2 individuals of other species were analysed.

DNA was isolated from peripheral blood leukocytes of humans and monkeys, using a modification of the salting-out process (Miller et al. 1988). Initially, PCR was carried out using primers FP1 (5'-CATATCCTACTTTTGCTACACA-3') and RP1 (5'-ATGAGTTGAGTTGAGGGATTC-3'), derived from the cDNA sequence of the *hClock* gene, and was expected to generate a 193 bp fragment. However, when these primers were used on a human genomic DNA template, a product of ~3 kb was generated, indicating that the primers spanned an intron-exon boundary. [At the time when this work was carried out, the report by Steeves et al. (1999), describing the genomic structure of this gene, had not yet appeared.] This product was completely sequenced and the intron-exon boundary was identified. The sequence was deposited as accession number AF260318 in the GenBank database. Another reverse primer, RP2 (5'-CTGTTCACTTAATGCTTAATTTCA-3'), was synthesized from the intronic sequence, which generated a product of 227 bp. All subsequent PCR reactions on human and monkey samples to study repeat length polymorphisms were carried out using the FP1 and RP2 primers. This PCR reaction was carried out using the following cycling conditions: initial denaturation of 94 °C for 3 min, 35 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s, followed by a final extension of 72 °C for 10 min. The PCR products were fluorescently labelled by the addition of Fluoro-Green dUTP (Amersham-Pharmacia, Buckinghamshire, UK) to the reaction mixture. Fluorescent products were then sized, using the Genescan software on the ABI Prism 377 Automated DNA sequencer (Perkin Elmer, Foster City, Calif., USA). All sequencing was carried out using the dideoxy chain terminator chemistry on an ABI Prism 377 automated DNA sequencer.

Table 1 Glutamine repeat size variation at the *dClock* locus in various species and strains of *Drosophila*

Drosophila species	Strain	Geographical location	Repeat size of 'pure glutamine' repeats
D. melanogaster	Oregon-R	Lab stock	33
	1004	Berlin	33
	1006	Canton-S	29
	Canton S	Lab-stock	25
D. simulans	2001	Varanasi (India)	28
	2002	Australia	28
D. annanassae	11001	Mysore (India)	29
	11002	Sringeri (India)	29
	11003	Namakkal (India)	29

The glutamine repeats at the *Clock* locus in various species of *Drosophila* were estimated by sizing PCR products that encompassed the repeat region. Different *Drosophila* species and strains were obtained from the *Drosophila* Stock Centre, Dept. of Studies in Zoology, Manasagangotri, Mysore, India and were derived from wild-type strains from various geographical locations. The list of species and strains used, together with the geographical location of origin and sizes of the 'pure glutamine repeat' are given in Table 1. In all cases, 5 flies from each strain were used to determine the size variation. To extract DNA for PCR, a single fly was taken in a 1.5 ml microfuge tube and mashed with a pipette tip containing 50 µl of the extraction buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml of proteinase K). The mix was incubated at 37 °C for 30 min, followed by heating at 95 °C for 2 min to inactivate the proteinase K. This crude DNA extract was then stored at 4 °C, and 2 µl was used in a 25-µl PCR reaction. The PCR primers DCFP (5' ACTGAACTCGGTCGCAGATGCTT 3') and DCRP (5'-GATTACTGCCTCCATTGTAGCTT-3') were designed to flank the repeat region and generated a fragment of 414 bp for 33 pure glutamine repeats. PCR was carried out using the following conditions: initial denaturation of 94 °C for 3 min, 35 cycles of 94 °C for 45 s, 55 °C for 45 s, 72 °C for 45 s, followed by a final extension of 72 °C for 10 min. To confirm that size variations in the PCR products were due to alteration in the repeat number and not due to flanking sequences, all PCR fragments corresponding to different sizes were sequenced and glutamine-repeat variation was confirmed.

The *Clock* gene sequences from human (*Homo sapiens*, accession number. AF011568) chicken (*Gallus gallus*, accession number AF132531), rat (*Rattus norvegicus*, accession number AB019258), mouse (*Mus musculus*, accession number AF000998) and zebrafish (*Danio rerio*, accession number AF133306) were obtained using the Entrez nucleotide query at http://www.ncbi.nlm.nih.gov, using 'Clock' as the search term, and were used for the comparative analysis of glutamine repeats in these organisms.

Results

CAG repeats at the *Clock* locus are not polymorphic in humans

We did not observe any variation in the length of the CAG-repeat stretch in any of 190 individuals studied, including patients with bipolar disorder and schizophrenia. All individuals had a continuous stretch of 6 CAG repeats, as reported in the original human *Clock* sequence. Sequencing of the repeat region from 10 unrelated individuals did not reveal any sequence polymorphism in the 227 bp region amplified.

CAG repeats are longer in most non-human primates when compared to humans

The sizes of the CAG-repeats in the non-human primates examined are shown in Fig. 1. In all cases the CAG-repeat size was also confirmed by sequencing. Two unrelated samples each of baboon, langur, rhesus monkey and bonnet monkey, and one each of gorilla and chimpanzee, were used in the analysis. The CAG-repeat was longer than in humans in all of the non-human primates studied, except in the baboon samples. In the case of the baboon, an allele of 6 CAG-repeats was observed, similar to that of the repeat size observed in humans. In the rhesus mon-

Fig. 1 Comparative alignment of sequences around the CAG-repeat stretch in the human *Clock* gene and non-human primates. The number of CAG repeats is indicated next to the name of primate. The variant nucleotide is indicated in *bold*

Protein	A	T	Q	Q	Q	Q	S	Q	T	L	S	V	T	$(Q)_N$	S	S	Q	E	Q	Q	L
Human (6)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACG	$(CAG)_N$	AGC	TCC	CAG	GAG	CAG	CAG	CTC
Chimp (7)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACG	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC
Gorilla (8)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACG	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC
Langur (7)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACA	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC
Baboon (6)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACA	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC
Rhesus (7,9)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACA	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC
Bonnet (8)	GCT	ACA	CAA	CAG	CAA	CAG	TCA	CAG	ACA	TTG	TCA	GTA	ACA	$(CAG)_N$	AGC	TCC	CAG	GAA	CAG	CAG	CTC

Protein	T	S	V	Q	Q	P	S	Q	A	Q	L	T	Q	P	P	Q	Q	F	L	Q
Human	ACT	TCA	GTT	CAG	CAA	CCA	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Chimp	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Gorilla	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Langur	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Baboon	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Rhesus	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG
Bonnet	ACT	TCA	GTT	CAG	CAA	CCG	TCT	CAG	GCT	CAG	CTG	ACC	CAG	CCA	CCG	CAA	CAA	TTT	TTA	CAG

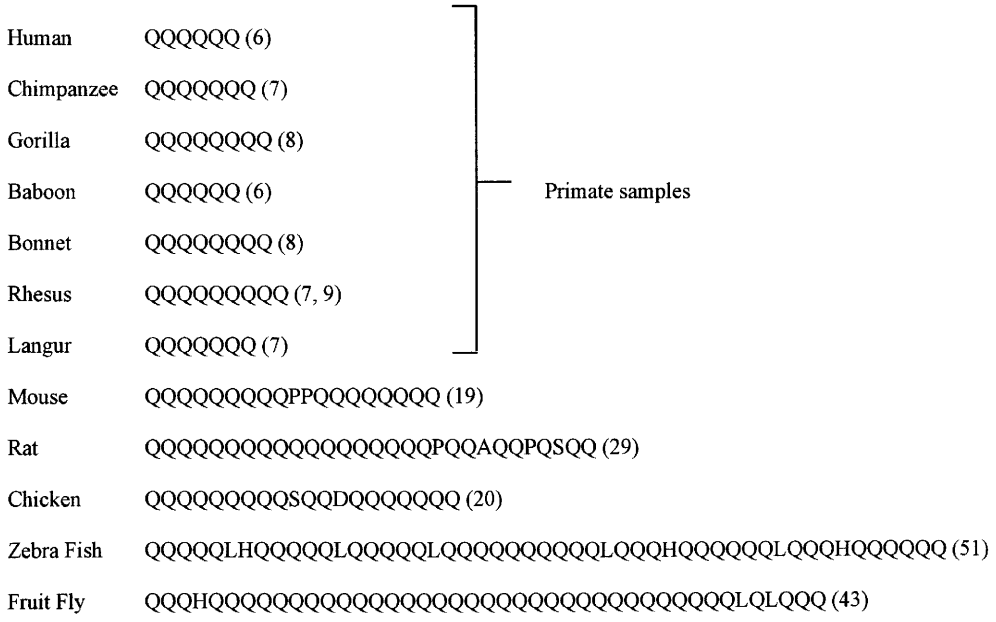

Fig. 2 Comparison of glutamine repeat number in the *Clock* gene in various organisms. The *number in brackets* refers to the length of the glutamine stretch, including non-glutamine interruptions

key sample, two alleles were observed with sizes of 7 and 9 CAG-repeats. An alignment of the sequences around the CAG-repeats for the various primates studied is shown in Fig. 1. All the non-human primates, except the gorilla and the chimpanzee, have a G nucleotide immediately preceding the CAG-repeat stretch. The gorilla, chimpanzee and human have an A nucleotide at this position. This nucleotide variation does not result in any change of amino acid.

CAG repeats at the *Clock* locus are highly conserved across species and polymorphic in *Drosophila*

An analysis of the *Clock* gene in other organisms in which the gene has been identified showed that the glutamine-rich region is highly conserved and of variable length in different species (Fig. 2). The glutamine stretch in these organisms is encoded by both CAG and CAA codons and occurs in the C-terminal segment of all the protein sequences. The longest stretches were observed in the zebrafish and *D. melanogaster* sequences, while the smallest stretches were observed in the human and baboon sequences. In the case of *D. melanogaster*, a long continuous stretch of glutamine repeats were observed, while in the case of zebrafish, the glutamine-repeat motif was found to be interrupted by non-glutamine residues. We have also determined the polymorphism status of the glutamine repeats in various species and strains of *Drosophila* derived from different geographic locations. In the different *Drosophila* species, sizes of the longest continuous glutamine repeat, or the 'pure glutamine' stretch, varied from 28 to 33 repeats. In the case of *D. melanogaster*, we were able to identify alleles of 33, 29 and 25 repeats in the various strains. In all cases, we have confirmed that size variation was due to variation in glutamine repeat number by sequencing the PCR fragments corresponding to different repeat lengths. All other species of *Drosophila* had glutamine-repeat motifs similar to that of *D. melanogaster* and variations in size were due to variations in the number of the 'pure glutamine' stretch. In the case of the *D. malerkotliana* species, we discovered that the sequence of the PCR product did not match that of the *dClock* locus and, therefore, this species was not used for further analysis.

Discussion

The human *Clock* gene is an attractive candidate for a number of disorders in which circadian rhythms play a role (Steeves et al. 1999). The presence of a CAG-repeat stretch within the gene has also raised the possibility that it might be susceptible to mutation via repeat expansion. The first step towards establishing this hypothesis was to determine whether the repeats were polymorphic in the normal population. No polymorphism was observed in a sample of 190 unrelated individuals studied and all individuals had sizes corresponding to 6 CAG repeats. This sample set also included patients suffering from schizophrenia and bipolar disorder, as these diseases are thought to be associated with both alterations in circadian rhythm and CAG-repeat polymorphism. Instability of trinucleotide repeats at all the other disease loci occur against a background of polymorphic repeats in the normal population, with the disease-causing expansions arising from the large normal alleles derived from this pool (Takano et al. 1998). The lack of repeat variation, therefore, makes it unlikely that such a stepwise expansion process is possible at the *Clock* locus.

The stability of the CAG repeats at the *Clock* locus in the normal population may be the result of one of a number of mechanisms. Either 6 CAG repeats are too small to produce slippage and subsequent length variation, or vari-

ation in repeat number at this locus is deleterious to the individual and thus not tolerated in the normal population. However, other trinucleotide repeat loci exist with repeat lengths similar to, or even smaller than, that in the *Clock* locus and exhibit high degrees of polymorphism. These include the *SCA1* locus with a range of 6–44 CAG repeats, *SCA6* with 4–18 repeats, *SCA7* with 4–35 repeats and myotonic dystrophy with 5–37 repeats (Cummings and Zoghbi 2000). Thus, the small length of the CAG repeats at the *Clock* locus may not be the sole reason for the lack of variation in repeat number. This would imply that variation of CAG repeats does indeed affect the function of the protein and that such variation may be deleterious. In such a scenario, one cannot exclude the involvement of CAG repeats in the disease process. In the case of oculopharyngeal muscular dystrophy, a small GCG expansion from 6 repeat units to 8 repeat units is associated with the disease phenotype (Brais et al. 1998). In *HOXD13*, a gene essential for normal limb development, an alanine repeat not polymorphic in the normal population, shows a modest expansion in individuals suffering from synpolydactyly (Goodman et al. 1997; Muragaki et al. 1996). Thus, small expansions in other non-polymorphic loci have been shown to have deleterious effects. Recombination has been proposed as a mechanism in the pathological expansion seen in synpolydactyly (Warren 1997), and a similar event could also occur involving the CAG-repeat/polyglutamine stretch at the *Clock* locus. Although we have not observed any CAG repeat variations in schizophrenia and bipolar patients, it does not rule out the involvement of other types of polymorphisms in this gene that may play a role in these disorders.

The conservation of the polyglutamine stretch in a large number of organisms indicates that this motif is probably essential for functioning of the clock protein. There is a striking variation in the length of the glutamine repeats in different species and it is noteworthy that the primate lineage has the smallest lengths amongst all the organisms. It is possible that in humans, and primates in general, long glutamine repeats are not conducive to the proper functioning of the *Clock* protein and hence these repeats are smaller when compared to other organisms. We have observed a great degree of polymorphism for glutamine repeat number in various species of *Drosophila*, where repeat number varied from 25 to 33 repeats. Polymorphism of CAG repeats at the *Clock* locus is thus tolerated in other species. The report that originally identified the *Clock* gene in *D. melanogaster* demonstrated alleles of 25 or 33 repeats for the 'pure glutamine' repeat stretch. While these variations did not affect locomotor activity rhythms, subtle effects on such activity, or an influence on other circadian outputs, could not be excluded (Allada et al. 1998). The glutamine motif is important for the functioning of the *Clock* protein, as in vitro studies have shown that deletion of part of the glutamine stretch resulted in a reduction of the transcription activation efficiency of the protein (Darlington et al. 1998). The longer glutamine repeats in the *Drosophila Clock* protein are in contrast to another important transcription factor gene that contains glutamine repeats, TATA-binding protein (TBP). The *Drosophila* TBP contains two stretches of continuous glutamine repeats of 6 and 9 residues in its N-terminal region (accession number U35147). This is much shorter than its orthologue in humans, which contains a stretch of 26–42 continuous glutamine repeats, which are also highly polymorphic (Rubinsztein et al. 1996).

There have been a number of reports that indicate there is directionality in the evolution of microsatellites like the CAG repeats. It has been shown that in the human lineage the number of repeats is usually higher than their homologues in non-human primates (Rubinsztein et al. 1995; Djian et al. 1996). However, in the case of the *Clock* locus, we find that the repeat number is lower in humans compared with most of the non-human primates, the reverse of the trend of longer repeats in humans. There is strong circumstantial evidence to suggest that microsatellites, such as CAG repeats, show a mutational bias in favour of longer, rather than shorter, allele lengths (Rubinsztein et al. 1994; Amos et al. 1996; Primmer et al. 1996). The human lineage would, thus, have begun with shorter microsatellite alleles, comparable to those seen in non-human primates, and in the course of evolution a gradual increase in microsatellite allele length would have occurred. As longer microsatellites are more likely to be polymorphic, the gradual increase in repeat number would also result in the locus exhibiting a greater degree of polymorphism. In the case of the *Clock* gene, the ancestral CAG repeat appears to be longer than that seen in humans, as evidenced by the repeat number in chimpanzees and other non human primates. Thus, although there is a tendency for an increase in the repeat number at several other loci, the CAG-repeat size at the *Clock* locus has undergone a contraction during the course of evolution in the human lineage. Such an event could readily be explained if one assumes that an increase in the glutamine number at the *Clock* locus were to have a deleterious effect on the individual, reducing fitness, and thus be excluded under pressure of natural selection. As shorter repeats are less likely to be polymorphic, the end result of this process would be a short, monomorphic allele at this locus. It is interesting that one of the rhesus monkey samples showed alleles of 7 and 9 repeat units, indicating that the CAG repeat is polymorphic in this species. The fact that we have not observed a polymorphism in the large number of human samples, in spite of the fact that this repeat motif is polymorphic in the rhesus monkey samples, further strengthens our assertion that any CAG-repeat variation at this locus may be tightly regulated and thus invariant in humans.

In conclusion, we have shown that the glutamine repeat stretch in the *Clock* locus is not polymorphic in humans and is shorter than their homologues in non-human primates. The repeats at this locus are unlikely to produce disease via the step-wise expansion process proposed in the case of other trinucleotide repeat disorders. The lack of polymorphism may indicate that maintenance of the polyglutamine repeat length is essential for the normal functioning of the protein and that repeat variation at this

locus is deleterious to the individual and thus not tolerated.

Acknowledgements We would like to thank the Primate Research Facilities of the National Institute of Immunology, New Delhi, the Indian Institute of Science, Bangalore and Dr Lalji Singh of the Centre for Cellular and Molecular Biology, Hyderabad for providing the monkey samples, the *Drosophila* Stock Centre, Dept. of Studies in Zoology, Manasagangotri, Mysore for providing the different fly strains. Ms Arjumand Ghazi and Mr Yograj Gowda are thanked for help in procuring the fly strains. We would also like to thank Ms M. Uma and Ms R. Jaya for help with the Genescan analysis and Tessi Sherrin and Charu Gandhi for help in DNA extraction. This work was supported by the Department of Biotechnology, Government of India, and the Council of Scientific and Industrial Research, India.

References

Abe H, Honma S, Namihira M, Tanahashi Y, Ikeda M, Yu W, Honma K (1999) Phase-dependent induction by light of rat *Clock* gene expression in the suprachiasmatic nucleus. Brain Res Mol Brain Res 66: 104–110

Allada R, White NE, So WV, Hall JC, Rosbash M (1998) A mutant *Drosophila* homolog of mammalian *Clock* disrupts circadian rhythms and transcription of period and timeless. Cell 93: 791–804

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nat Genet 13: 390–391

Antoch MP, Song EJ, Chang AM, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS (1997) Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. Cell 89: 655–667

Beilin J, Ball EM, Favaloro JM, Zajac JD (2000) Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. J Mol Endocrinol 25: 85–96

Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, Lafreniere RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korczyn AD, Heutink P, Mathieu J, Duranceau A, Codere F, Fardeau M, Rouleau GA, Korcyn AD. (1998) Short GCG expansions in the *PABP2* gene cause oculopharyngeal muscular dystrophy. Nat Genet 18:164–167

Brahmachari SK, Meera G, Sarkar PS, Balagurumoorthy P, Tripathi J, Raghavan S, Shaligram U, Pataskar S (1995) Simple repetitive sequences in the genome: structure and functional significance. Electrophoresis 16: 1705–1714

Chong NW, Bernard M, Klein DC (2000) Characterization of the chicken serotonin N-acetyltransferase gene. Activation via *Clock* gene heterodimer/E Box interaction. J Biol Chem 275: 32991–32998

Chandy KG, Fantino E, Wittekindt O, Kalman K, Tong LL, Ho TH, Gutman GA, Crocq MA, Ganguli R, Nimgaonkar V, Morris-Rosendahl DJ, Gargus JJ (1998) Isolation of a novel potassium channel gene *hSKCa3* containing a polymorphic CAG repeat: a candidate for schizophrenia and bipolar disorder? Mol Psychiatry 3: 32–37

Cummings CJ, Zoghbi HY (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. Hum Mol Genet 9: 909–916

Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TDL, Weitz CJ, Takahashi JS, Kay SA (1998) Closing the circadian loop: *CLOCK*-induced transcription of its own inhibitors *per* and *tim*. Science 280: 1599–1603

Desan PH, Oren DA, Malison R, Price LH, Rosenbaum J, Smoller J, Charney DS, Gelernter J (2000) Genetic polymorphism at the *CLOCK* gene locus and major depression. Am J Med Genet 96: 418–421

Djian P, Hancock JM, Chana HS (1996) Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. Proc Natl Acad Sci USA 93: 417–421

Goodman FR, Mundlos S, Muragaki Y, Donnai D, Giovannucci-Uzielli ML, Lapi E, Majewski F, McGaughran J, McKeown C, Reardon W, Upton J, Winter RM, Olsen BR, Scambler PJ (1997) Synpolydactyly phenotypes correlate with size of expansions in *HOXD13* polyalanine tract. Proc Natl Acad Sci USA 94: 7458–7463

Irvine RA, Ma H, Yu MC, Ross RK, Stallcup MR, Coetzee GA (2000) Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. Hum Mol Genet 9: 267–274

Jain S, Kaliaperumal V, Chatterji S, Rao S, Murthy RS (1992) Climate and admissions for mania in the tropics. J Affect Disord 26:247–250

Jain S (1995) The biology of mania. In: Koslow S, Murthy RS (eds) The Decade of the Brain. ADAMHA-NIMHANS.

Katzenberg D, Young T, Finn L, Lin L, King DP, Takahashi JS, Mignot E (1998) A *CLOCK* polymorphism associated with human diurnal preference. Sleep 21: 569–576

King DP, Takahashi JS (2000) Molecular genetics of circadian rhythms in mammals. Annu Rev Neurosci 23: 713–742

King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TD, Vitaterna MH, Kornhauser JM, Lowrey PL, Turek FW, Takahashi JS (1997) Positional cloning of the mouse circadian clock gene. Cell 89: 641–653

Koshy BT, Zoghbi HY (1997) The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis. Brain Pathol 7: 927–942

Larkin P, Baehr W, Semple-Rowland SL (1999) Circadian regulation of iodopsin and clock is altered in the retinal degeneration chicken retina. Brain Res Mol Brain Res 70: 253–263

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16: 1215

Mitterauer B (2000) Clock genes, feedback loops and their possible role in the etiology of bipolar disorders: an integrative model. Med Hypotheses 55: 155–159

Muragaki Y, Mundlos S, Upton J, Olsen BR (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in *HOXD13*. Science 272: 548–551

Primmer CR, Saino N, Moller AP, Ellegren H (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13: 391–393

Reddy PS, Housman DE (1997) The complex pathology of trinucleotide repeats. Curr Opin Cell Biol 9: 364–372

Ross CA, Wood JD, Schilling G, Peters MF, Nucifora FC, Jr., Cooper JK, Sharp AH, Margolis RL, Borchelt DR (1999) Polyglutamine pathogenesis. Philos Trans R Soc Lond B Biol Sci 354: 1005–1011

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Ramesar RS, Old J, Bontrop R, McMahon R, Barton DE, Ferguson-Smith MA (1994) Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. Nat Genet 7: 525–530

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution – evidence for directionality and variation in rate between species. Nat Genet 10: 337–343

Rubinsztein DC, Leggo J, Crow TJ, DeLisi LE, Walsh C, Jain S, Paykel ES (1996) Analysis of polyglutamine-coding repeats in the TATA-binding protein in different human populations and in patients with schizophrenia and bipolar affective disorder. Am J Med Genet 67: 495–498.

Saleem Q, Vijayakumar M, Mutsuddi M, Chowdhary N, Jain S, Brahmachari SK (1998) Variation at the *MJD* locus in the major psychoses. Am J Med Genet 81: 440–442

Saleem Q, Sreevidya VS, Sudhir J, Vijaya Savithri J, Gowda Y, Rao CB, Benegal V, Majumber PP, Anand A, Brahmachari SK, Jain S (2000) Association analysis of CAG repeats at the *KCNN3* locus in Indian patients with bipolar disorder and schizophrenia. Am J Med Genet 96: 744 –748

Steeves TD, King DP, Zhao Y, Sangoram AM, Du F, Bowcock AM, Moore RY, Takahashi JS (1999) Molecular cloning and characterization of the human *CLOCK* gene: expression in the suprachiasmatic nuclei. Genomics 57: 189–200

Takano H, Cancel G, Ikeuchi T, Lorenzetti D, Mawad R, Stevanin G, Didierjean O, Durr A, Oyake M, Shimohata T, Sasaki R, Koide R, Igarashi S, Hayashi S, Takiyama Y, Nishizawa M, Tanaka H, Zoghbi H, Brice A, Tsuji S (1998) Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. Am J Hum Genet 63: 1060–1066

Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L, Yong EL (1997) Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. J Clin Endocrinol Metab 82: 3777–3782

Vincent JB, Paterson AD, Strong E, Petronis A, Kennedy JL (2000) The unstable trinucleotide repeat story of major psychosis. Am J Med Genet 97: 77–97

Wager-Smith K, Kay SA (2000) Circadian rhythm genetics: from flies to mice to humans. Nat Genet 26: 23–27

Warren ST (1997) Polyalanine expansion in synpolydactyly might result from unequal crossing-over of *HOXD13*. Science 275: 408–409

Wehr TA, Rosenthal NE (1989) Seasonality and affective illness. Am J Psychiatry 146: 829–839

Wehr TA, Wirz-Justice A, Goodwin FK, Duncan W, Gillin JC (1979) Phase advance of the circadian sleep–wake cycle as an antidepressant. Science 206: 710–713

Whitmore D, Foulkes NS, Strahle U, Sassone-Corsi P (1998) Zebrafish *Clock* rhythmic expression reveals independent peripheral circadian oscillators. Nat Neurosci 1: 701–707

Wilsbacher LD, Takahashi JS (1998) Circadian rhythms: molecular basis of the clock. Curr Opin Genet Dev 8: 595–602

Wirz-Justice A, Cajochen C, Nussbaum P (1997) A schizophrenic patient with an arrhythmic circadian rest-activity cycle. Psychiatry Res 73: 83–90