Structure, evolution and properties of a novel repetitive DNA family in *Caenorhabditis elegans*

Adriana La Volpe*, Maria Ciaramella and Paolo Bazzicalupo

International Institute of Genetics and Biophysics-CNR, Via Marconi 10, I-80125, Naples, Italy

Received July 27, 1988; Accepted August 15, 1988

Accession nos X07681-X07684 (incl.)

ABSTRACT

We have identified a moderately repeated DNA sequence in *Caenorhabditis elegans* present at least at twenty different locations in the genome. Elements of this intermingled repetitive DNA family are made up of tandem subrepeats whose smaller unit is ten base pairs long. The occurrence of single base changes between units is reminiscent of mammalian satellite DNA. Sequence analysis has shown that the consensus of these repeats is identical to the consensus of the heat-shock element (HSE) common to all eukaryotes (C--GAA--TTC--G). This consensus in our sequences is repeated in tandem with an overlap of four bases (C--GAA--TTC--GAA--TTC...). We studied in detail one cloned element of the family and we were unable to detect transcription in the flanking regions either under normal growth or after heat induction. Nevertheless a 242 bp sequence out of this same element was sufficient, when located on a multicopy plasmid in *Saccharomyces cerevisiae*, to drive transcription from a downstream gene under heat shock conditions.

INTRODUCTION

A large fraction of eukaryotic DNA (10/30%) is present in more than one copy per haploid genome. Repeats may be interspersed in the genome or they may be arranged in tandem. Some repeated sequences have a known function: they code for abundant RNAs such as ribosomal RNAs and transfer RNAs or for abundant proteins. Nevertheless, the function and the origin of most repetitive sequences remain unknown; some repeated sequences might play a regulatory role in the control of gene expression, recombination, DNA replication, while others might be involved in the structural organization of the genome and of the chromosomes such as the sequences associated with telomeres, centromeres or sequences binding the nuclear scaffold. In addition several mobile elements, or repetitive DNA sequences in general, have been described which have no evident function beyond their own proliferation and maintenance in the genome. Many mechanisms are involved in the generation of repeated DNA families: DNA transposition, gene duplication, chromosomal rearrangements, integration of viral genomes, gene retroposition, slippage replication etc.

The size of the genome of the free living nematode Caenorhabditis elegans is small (8×10^7 bp) (1); repetitive sequences are present (14%)(2) but the degree of repetitiveness is low, ranging, for different families, from ten to not more than a few hundred. This circumstance makes the study of repetitive sequences in this organism particularly informative since the percentage of inactive members of each family should be lower than in larger genomes and it should be possible to collect and compare all the members of a given repetitive family.

In this paper we report on a moderately repeated DNA family we have isolated from C. elegans, which is both interspersed in the genome and consists of individual elements that are internally tandemly repeated. Sequence analysis reveals that the internal tandem repeats consensus is the same as the consensus for the heat-shock promoter element (HSE) (3,4), present in few copies upstream of the heat-shock genes of all eukaryotes. The entire element we studied consists of 10bp tandem repeats, the heat-shock element consensus is fourteen bp long but, in some heat-shock promoters (5,6), two HSEs may overlap by four base pairs. This occurs throughout our sequence, so that about 100 HSEs can fit in the one kb long element. Surprisingly, the agreement to the general HSE consensus (C--GAA--TTC--G) is for most of these repeats in our element better than that shown for 'genuine' C. elegans heat-shock genes (6,7). Although we failed to detect transcription, in the flanking sequences of prC9 during normal growth and under heat-shock condition in C. elegans, this repeated sequence works as a heat-shock promoter in yeast.

We have tried to understand the mechanisms that gave rise through evolution to this kind of repetitive DNA family.

MATERIALS AND METHODS.

C. elegans strains: N2-JW-H, W-PA-2 H, are wild-type, independently isolated C. elegans strains and CB1392 is a nuc001 (nuclease minus) mutant. CbrgB0 H is a C. briggsae strain, a very similar species not interfertile with C. elegans; they all were obtained from the Caenorhabditis elegans Genetics Center. Routine maintenance was as described by Brenner (8). All strains were raised at 20°C.

Parascaris equorum was isolated directly from naturally infected hosts.

Dirofilaria repens DNA was extracted from microfilariae present in the blood of infected dogs kept at the Istituto di Parassitologia, University of Rome

and were a gift from Dr. G. Cancrini. Oregon R Drosophila melanogaster strain was a gift from Dr. C. Malva.

Yeast strain: Plasmids were transformed and maintained in yeast in strain ScKY117 (a, ara3-52, tryp1-del1, his3-del200, ade2-101^{acc}, lys2-801^{am} (from Dr K. Struhl). tryp1-del1 deletes all sequences corresponding to the 1.4 kb Eco RI TRP-1 ARS-1 fragment (9), his3-del200 completely deletes chromosomal HIS3 sequences (10).

Cosmids and plasmids: The genomic C.elegans cosmid library was a gift from Dr. Silvana Gargano, of this laboratory. Colony hybridization was according to Hanahan and Meselson (11). Plasmids were transformed and maintained in E. coli in the rec A strain HB101 (F⁻, hsd S20, (r⁻ m⁻), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Sm^r), xyl5, mtl-1, supE-44, λ^-) selecting for Ampicillin resistance. The strain was checked periodically for the recA phenotype. Plasmid pPW229, a gift from Dr. P.P. di Nocera, contains the D. melanogaster wild type gene for heat-shock protein 70. Plasmid pS26-1, a gift from Dr. Susan Lindquist, contains the yeast wild type gene for heat-shock protein 26. Plasmid prC9 (a subclone of the original cosrC9 we selected for the presence of repetitive sequences) contains a 7kb HindIII fragment inserted in pEMBL8 (12). Plasmid pC9N40 contains a 242bp fragment obtained by XbaI partial digestion of the C9 element cloned in the polylinker of pGEM3 (Promega tm). Plasmids pC9XX and pCXX.1 were obtained from a partial XbaI digestion of prC9 and selected because they contain the 2.7kb XbaI-XbaI fragment at the 5' of the C9 element (as shown in fig.1b) and variable portions of the C9 element itself. Plasmid pC9XR contains the XbaI-Eco RI fragment at the 3' end of the C9 element and covers the junction with this flanking sequence. They are cloned in pGEM3. Plasmid pGM8 is the test-vector for transformation in yeast (Fig. 7). A 1.4 kb EcoRI yeast fragment containing the TRP-1 gene and an ARS element (ARS-1,9) cloned at the EcoRI site in the polylinker of PAT011 (13) simultaneously introduces a selectable marker and the ability to replicate autonomously in yeast cells. At the BamHI site of plasmid pGM8 is inserted a mutant derivative of HIS3 from which, by in-vivo recombination in a lambda vector (14), the 435 base pairs upstream of the transcription initiation site had been replaced by 260 base pairs from the bacteriophage lambda attachment site. The mutant sequence, also called his3-del4 is contained within a 1612 bp BamHI fragment (Sc2715).

Nematode DNA extraction: Worms were grown on NGM plates (8), collected, washed a few times in a bench centrifuge and cleaned on a cold 7%/35% sucrose

step gradient. Interface material was subsequently freed from sucrose on a bench centrifuge and frozen by dropping small volumes in liquid nitrogen. Clean, frozen worms can be kept at -70 for several months before DNA (or RNA) extraction. Frozen worms pellets were added to warm (65°C) lysis buffer (2%SDS, 50mM EDTA, 100mM NaCl, 100µg/ml proteinase K) and digested for 20 min. After proteinase digestion, two or three phenol extractions and one chloroform extraction were followed by precipitation at room temperature with one volume of propan-2-ol. Pellets were resuspended and treated with RNase A at 37°C. After a further phenol/chloroform/ isoamyl alcohol (24:24:1) extraction, DNA was precipitated with ethanol.

Yeast DNA extraction: 8 ml yeast cultures were grown in SD medium. DNA was prepared by the "mini-prep" procedure described in Sherman et al. (15).

Southern transfer: was performed as described by Southern 1975 (16). Hybridizations were performed in 4xSSC, 0.2%SDS, 50mM Na Phosphate buffer at 65°C for 16 hours.

C. elegans RNA extraction: Worms were collected, cleaned on a sucrose step gradient, and quickly frozen in liquid nitrogen as described for DNA extraction. Frozen worms pellets were reduced to a very fine dust in a mortar kept very cold with liquid nitrogen. Extraction buffer (100mM Na cacodilate, 2% SDS, 500mM NaCl) was added in such a way that the homogenate would immediately freeze. The frozen mixture was added to and left to melt in tubes containing phenol/chloroform kept at 4°C. After 3 phenol/chloroform extractions RNA was precipitated with ethanol and resuspended in 2M LiCl, left to precipitate o/n at 0°C, washed with more LiCl and resuspended in sterile, RNase free distilled H₂O. RNA was kept in 70% ethanol at -20°C.

Yeast RNA extraction: Cells were grown at 30° C to late log phase in SD (15) minimal medium containing all appropriate supplements but tryptophan. Cells were lysed and total RNA was prepared by a slight modification of the protocol of Struhl & Davis (17):one cold phenol/chloroform extraction and three cycles of extraction with hot phenol/chloroform followed by rapid cooling in dry-ice/ ethanol, one chloroform extraction and ethanol precipitation. RNA preparations were stored in ethanol.

Northern transfer: RNA concentration was determined at A₂₆₀. Aliquots of new preparations were monitored for degradation and the presence of DNA by running aliquots on 10ml 2% agarose minigels in acetate buffer. For hybridization analysis aliquots of 20 µg total RNA were denatured at 55° C in 6% formaldehyde, 50% formamide, loaded on 1% agarose gels containing 6% formalde-

hyde, run in MOPS (Sigma) buffer and transferred to nitrocellulose in 20xSSC. To detect the various transcripts the RNA blots were hybridized at 42° C in buffer containing 50% formamide to the following probes: C9, a 7kb HindIII fragment from prC9 (fig.1 B), hsp70, a XhoI internal fragment from pPW229 (18), HIS3, an internal HindIII fragment (nts +328 to 515) from pGM8 and hsp26, an BglII-PstI fragment from pS26-1 (19). We chose to use total RNA in our transfer experiment because polyA RNA preparation induced by heat-shock cannot be reliably normalized to non heat-shock polyA RNA: genes expressed under heat shock conditions are usually not expressed otherwise and viceversa. For total RNA we normalized by OD adsorbance and ribosomal RNA fluorescence of RNA aliquots run on test gels. Furthermore we routinely hybridized each filter with a non heat-shock protein gene probe (not shown) and with a heat-shock gene probe (fig 6 and 8).

Yeast transformation: Yeast cells were transformed by a modification of the standard LiCl protocol (20). Late log phase cells ($1-2 \times 10^7$ cell/ml) grown

Fig.1: a) Plasmid prC9, partially digested with the restriction endonuclease XbaI and hybridized with itself, shows on an agarose gel a characteristic ladder pattern. b) The HindIII fragment was end labelled, one end was removed with EcoRI and the HindIII-EcoRI fragment was partially digested with a number of restriction enzymes. The map, in this figure, shows the peculiar high density of XbaI sites in the central region of the clone.

in rich (YPD) medium were harvested, washed once in one volume of TE (10mM Tris, pH8, 1mM EDTA), and once in one volume 100mM Li Acetate in TE. Cells were resuspended in 1/100 vol of the above solution and incubated at 30°C for 20 min. DNA (1/10 µg) was added to 0.1 ml aliquots of cells together with salmon sperm DNA carrier and incubated at 30°C for 15 min.; 0.7 ml of 40% PEG 4000 in 100mM Li Acetate in TE were added and cells were incubated at 30°C for one hour. After a heat-pulse of 5 min at 42°C, cells were harvested and resuspended in 0.2ml of TE. The phenotype of yeast cells harboring plasmids was determined by standard procedures (15) on SD minimal plates lacking appropriate nutrients. In some cases the ability of recombinant plasmids to synthesize abundant imidazoleglycerolphosphate dehydratase (IGPD), the product of the HIS3 gene, was tested by growth on SD plates containing 10mM aminotriazole, a potent competitive inhibitor of IGPD.

Preparation of radioactive probes: Probes were double-stranded and were prepared from restriction fragments uniformly labelled with 32 P by nick-translation (21). End-labelled DNA was obtained using Klenow DNA polymerase and the appropriate 32 P-dNTPs. Single strand RNA probes and RNase mapping experiments were performed according to Melton et al. (22).

Polyacrilamide non denaturing gels were according to Maniatis (21).

DNA sequencing: the dideoxy chain termination method (23) was used for fragments cloned in pGEM-3 (Promega tm) plasmid vector.

RESULTS

Isolation and characterization of the C9 element

The C9 element was originally isolated from a cosmid library (as well as 13 others belonging to independent repeated DNA families: La Volpe et al. manuscript in preparation) because it hybridized when nick-translated *C. elegans* genomic DNA was used as probe. From the original cosmid a 7kb Hind III fragment was subcloned and the plasmid named prC9. This fragment contains a region about 1 kb long with a peculiar distribution of XbaI sites. The restriction map of the 7 kb long insert of prC9 is shown in fig.1b.

We sequenced a few fragments obtained by random subcloning of the XbaI rich segment of clone prC9. The complete sequence of the 242bp insert of pC9N40 is shown in fig. 2 a. As diagrammed below the sequence, the entire fragment is the reiteration in tandem of the 10bp motif C--GAA--TT. The consensus for the 10 bp repeated unit is

C⁰⁰ T⁵⁵ A⁵⁰ G⁵⁵ A¹⁰⁰ A⁵⁵ T⁵⁵ T⁵⁰ T⁰⁰ T¹⁰⁰

Fig.2: Nucleotide sequence of the 242bp XbaI fragment (cloned as partial XbaI digestion from the internal repetitive region). The diagrams show how this sequence is an arrangement of short degenerate direct repeats (10bp) and overlapping palindromes (14bp) and that there are higher order structures like 40bp almost perfect tandem and inverted repetitions. Arrows with X indicate the position of XbaI sites.

where the number adjacent to each nucleotide is the frequency with which that nucleotide is found at that position in the 24 repeats of pC9N40. This same sequence can also be read as a series of 14bp palindromes each overlapping with the preceding by four bp. Surprisingly the consensus of this palindromic repeat (C--GAA--TTC--G) is the same as that of the heat-shock element (HSE) found in the regulatory region of genes induced by stress in all eucaryotes (3,4,24). Our sequence is apparently in better agreement to the HSE general consensus than HSEs from 'genuine' heat-shock promoters isolated in *C. elegans* (6,7). There is also a strong second choice preference:

C T₆ A₆ G A A T₆ T₆ T T C T₆ A₆ G

The fourteen bp consensus of this repeated sequence (as described for the HSE consensus sequences, 24), is rotationally symmetric. Beside the 10bp tandem repeats or the 14bp overlapping palindromes, we can identify other structural features of the 242bp sequences: the XbaI sites are spaced almost regularly every 40bp (in agreement with the difference in length between bands in the

a) 10 20 30 40 50 60 70 80
AAATGTCTGCAATGTTCCAGAACTTTCTAGAAAATCGGAAATCGGCATGTTCTAGAAAATTCGGAATGGTCCAGAAATTT
TCTAGAAAATTCGGAATATCCAGAACTTTCTAAAAAGTCTAG

b) 10 20 30 40 50 60 70 80
AAAAGGCCTTTTAGCGGAAATTCGGAATTTCTAGAAAATGCAGAAATGCTCTGCAATGTTCCAGAACTTTCTAGAAA
ATTCGAGAAAATTCGGAATGGTCCAGAAATTTCTAG

c) 10 20 30 40 50 60 70 80
CTAGAAAATCTTGAAAAGTCTGCAATGTTCCAGAACTCTCTAAAAATTCGATAAAATCTGAAATGTTCAATTCGTG
GAAAATTCGAACTTCTGCAAAGTTCTACACAGGGTTCTGGCTCGACCCGCGCTTCATTGGACACTCGAAATGGCGGG

Fig.3: Nucleotide sequences of three fragments from the C9 element we studied, fragments 'a' and 'b' are internal to the element, but located near the 5' end of it, were independently cloned from a XbaI partial digestion because physi- cally linked to the 2.7kb XbaI-XbaI flanking fragment. Fragment 'c' inclu- des the junction between the C9 element and the flanking region at the 3' side.

ladder shown in fig 1a). Sequence homology among 10bp units in the sequence confirms the existence of a duplication-step every 40bp. Again, the nucleotide sequence between 40 and 80 is identical but for a single base deletion to the sequence between 160 and 199. Finally, the first 40 bp are perfectly homolo- gous with the opposite strand between nts 202 and 242. Some of these observa- tions are summarized in the diagram in fig.2 b. Obviously we cannot position the 242bp fragment within the 1kb element, but we have determined the nucleo- tide sequence of three other short segments located toward the two ends of the element: sequences 'a', 'b' and 'c' in fig.3 (for cloning of these fragments see Materials and Methods: plasmids pC9XX, pC9XX.1 and pC9XR). The 10bp consensus

is conserved in these sequences with small differences in the relative representativity of each nucleotide and with a few insertions or deletions of short segments. Long stretches of perfect homology (30 nucleotides and more) with the 242bp fragment can be identified in sequences 'a' and 'b'. The combination of all these features makes the DNA region very unusual and many alternative conformations of the DNA helices can be predicted at this locus. No homology to the HSE and to our consensus can be seen in sequence 'c' past nucleotide 108 which then represents the beginning of the flanking region.

Comparing our sequence data with the Microgenie data bank we identified strong homologies with the regulatory region of HSP83 *Drosophila* heat-shock gene (locus 63B) and with the first exon of HSP83 gene (locus 63BC), with HSP26 *Drosophila* heat-shock gene (locus 67B) and with the heat-shock promoter present in the left and right arm of the slime mold *Dictyostelium Discoideum* transposable element TDD-1.

Distribution and evolution of the C9 family

The 242bp long insert of pC9N40 (fig.2), was hybridized back to *C. elegans* total DNA digested with Hind III: at least 17 bands can be identified in a Southern transfer experiment (fig.4 lanes c and f,g,h) and many seem to be doublets. The 7kb band (corresponding to the fragment cloned in prC9) is one of these. We wanted to know if somatic unequal crossing over was abundant in a region so highly repeated and was responsible for the lack of sharpness of this band. We therefore hybridized the same filter with the 300bp Hind III-Xba I fragment present at the very 5' of prC9 insert that does not contain repetitive sequences (see fig 1b): a unique very sharp band was present in all the three *C. elegans* lanes (not shown) indicating that the broad signal with the repetitive probe is probably due to hybridization to two or three different sites and is not the result of somatic unequal crossing-over.

To confirm the gene counting we hybridized a cosmid total genomic library at low stringency with the pC9N40 insert: we obtained between 25 and 30 positive clones every 2000 cosmids corresponding approximatively to one genome size.

The copy-number of the C9 family seems relatively constant in the genomes of independently isolated *C. elegans* strains (fig.4 lanes f,g). We performed the Southern transfer experiment (not shown) also with a strain containing a high number of copies of the transposable element Tc1 (25), and the number and the sizes of restriction fragments homologous to pC9N40 are similar to those shown by N2-JW-H, W-PA-2 and CB1392 strains which all have few Tc1 elements in their genomes.

Fig.4:left side: Southern genomic transfer of different species: lane a) *Saccharomyces cerevisiae*, lane b) *Drosophila melanogaster*, lane c) *Caenorhabditis elegans*, lane e) *Xenopus laevis*. Each lane was loaded with an amount of Hind III digested DNA proportional to genome size. Right side: Southern genomic transfer of different *C. elegans* strains and nematode species: lane f,g,h) *C. elegans* strains N2-JW-H, W-PA-2 and CB1392 respectively, lane j) *C. briggsae* strain CbrgBO-H, lane k) *Dirofilaria repens*, lane l) *Parascaris equorum*. Equal amounts of DNA were loaded. Molecular weights are in kilobases.

Since the heat-shock element, to which the C9 family is related, is present in all eukaryotes that have been looked at, we wanted to see if, in other organisms and especially in other nematodes, the sequence had undergone an amplification similar to the one observed in *C. elegans*. No homology of pC9N40 was seen in *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Xenopus laevis*

Fig.5: Total DNA from *C. briggsae* and *C. elegans* was digested with XbaI, end labelled by filling-in with DNA polymerase large fragment and α - 32 PdCTP and TTP and run on a 12% non-denaturing polyacrilamide gel. Molecular weights shown on the left are in base pairs. Multiple bands around 10bp are probably due to variability in the number of nucleotides added (from one to four) and to the long protruding ends left by XbaI that affect the mobility of small fragments (already described,39).

genomic DNA (fig.4 lanes a,b and e). The same result was obtained in the Southern transfer analysis of DNA from other nematodes : *Caenorhabditis briggsae*, *Dirofilaria repens* and *Parascaris equorum* (fig.4 lanes j,k and l). In both cases filters were washed at low stringency (2xSSC, 0.2%SDS at 65°C) exposed and then washed at higher stringency (0.2xSSC, 0.2%SDS at 65°C) and exposed again. No bands could be seen in other species and no differences were observed in *C. elegans* lanes at different stringency. This latter finding indicates a good sequence conservation between members of the family in *C. elegans*. We wondered if the sequence organization and the pattern of duplications observed in the sequence of the C9 element we cloned was also common to the other members of the family. We have indirect evidence that the pattern is very similar at other loci; in fact, if total *C. elegans* DNA is partially digested with Xba I, end-labelled and run on a 12% polyacrilamide non-denaturing gel, very short discrete bands can be seen (fig. 5 right). The same experiment was performed with total DNA from *Xenopus laevis*, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Parascaris equorum* and *Caenorhabditis briggsae*. The rationale be-

hind this experiment is that since XbaI sites are expected to be a few kilobases apart in a random distribution, the presence of discrete radioactive bands at low molecular weights should indicate a high frequency of XbaI sites within a repetitive DNA family. *Xenopus* shows one discrete band about 190bp in length indicating the presence of an 'Xba family' probably of different origin (not shown). No other discrete bands of low molecular weight were present in the other lanes (not shown) with the exception of *C. briggsae* DNA in which a very clear pattern of repetition is present with many bands in common with *C. elegans* (fig.5 left).

RNA transcription analysis in *C. elegans*

Surprisingly the bases conserved among the HSE elements present 5' of different heat-shock genes are the same as those conserved in the repetition of C9 sequence suggesting that functional constraints have operated at least during the initial phases of the amplification process. However, this conclusion, based on the computer analysis of the sequence alone, may be misleading. We therefore decided to test by other means any residual functional activity of the amplified HSE elements. We used the entire 7kb long HindIII fragment cloned in prC9 as a probe in an experiment of Northern transfer of total RNA extracted before and after a heat shock treatment of two hours at 37°C (see fig.6). As a control of the heat-shock condition we hybridized two lanes of the same gel with a *Drosophila melanogaster* hsp-70 probe (18). We were unable to detect transcription in the 7kb surrounding the element in prC9 even exposing the filter several days longer than the hsp-70 control. An enhanced background in heat-shocked RNA lanes is present especially around the 18S region. We therefore performed an RNase protection assay (not shown) using as probes single strand RNA complementary to the 2.7kb fragment on the left and the 3kb fragment on the right of the C9 element in prC9 to further investigate if in any condition (heat-shocked and not) we could detect any transcript, but no bands were identified in conditions in which we could easily see a *C. elegans* control gene.

Analysis of the behaviour of prC9 sequence in a heterologous system

The transcription studies described above exclude a conclusion that the element of the C9 family we studied works as a bonafide heat-shock promoter in *C. elegans*. We thought that the peculiar organization in tandem of the C9 sequence and the alternative configurations that the DNA helices can assume might be a handicap in the binding of the heat-shock regulatory factors. To test this hypothesis, we decided to use a heterologous system: *Saccharomyces*

Fig.6: Northern transfer of total *C. elegans* RNA from a mixed population: odd lanes contain RNA from worms grown at standard temperatures, even lanes RNA after heat-shock. The four lanes on the left were hybridized with the entire 7kb HindIII fragment shown below. The lanes on the right were hybridized with a *Drosophila melanogaster* 70 kd heat-shock protein gene probe.

cerevisiae. We were encouraged to use this system because, although the fine regulation of the heat shock response may follow different strategies in different species (26), it is known that *Drosophila* heat-shock genes are inducible in *S. cerevisiae*, as they are in *C. elegans* (27,28) and both *Drosophila* and *C. elegans* heat-shock genes are inducible in mammalian cells (3,4,6).

We cloned the 242bp insert of pC9N40 into a multicopy test-plasmid upstream of a deletion-substitution derivative of the HIS-3 gene (his-del4, see Materials and Methods). This deletion-substitution sequence behaves as a pseudopromoter that is activatable by the proximity of yeast (29) or heterologous (30) cis acting control elements. This test-plasmid (shown in fig.7) is unable to restore histidine prototrophy to a his3- *S. cerevisiae* strain at any temperature. If we insert our sequence in the polylinker situated 260 nucleotides upstream of the his-3 gene and we transform his3- yeast cells we observe recovery of the

Fig.7: test-plasmid pGM8 used for expression in *S. cerevisiae* (see Materials and Methods for description). Solids blocks indicate TATA-boxes and TATA-like sequences, shaded blocks represent UAS (Upstream Activating Site; this element is deleted upstream his-del4), empty block shows the position of the ARS (Autonomous Replicating Site). The arrow indicates the site of the insertion of the DNA element to be tested, in our case the 242bp fragment.

HIS⁺ phenotype at all temperatures above 30°C. The RNA analysis is shown in fig.8. In the control plasmid the HIS-3 sequences are not significantly transcribed either at 21°C or after heat-shock at 39°C. HIS-3 RNA is also absent when the yeast cells transformed with the plasmid harboring the 242bp insert upstream of the HIS-3 gene are grown at 21°C. In contrast, a strong signal appears when these cells are heat-shocked at 39°C. In the same figure we show the result obtained by hybridizing the same filter with the Pst-BglII fragment from the yeast hsp-26 gene that we used as a control of the heat-shock condition. The signal from the endogenous heat-shock gene is reduced at 39°C in the cells harboring the test-plasmid with the 242bp insert compared to those carrying the test-vector without insert. Perhaps there is titration of the heat-shock factor by abnormally high level of HSEs in the cell. It has been observed in *Drosophila* that the introduction of an artificial poly-HSE sequence in the cell results in a reduction of the heat-induced transcripts from co-transformed heat-shock genes due to competition for heat-shock factor (31). We were unable to test this hypothesis rigorously because the ARS plasmid we use is unstable and the percent of cells carrying the plasmid varies extensively between 20% and 50% (in our hands about a third of the cells in any one colony contains the plasmid even under selective conditions).

A different pattern is observed (see fig.9) when the cells are grown at 30°C and then heat-induced. Although most laboratory yeast strains are normally grown at this temperature, we observed that both the hsp-26 gene in the genome and the his-del4 gene downstream from our sequence on the plasmid are already induced at 30°C and constitutively expressed. While the hsp-26 gene transcript

Fig.8: Northern transfer of yeast RNA: odd lanes contain RNA from yeast cells grown at 21°C, even lanes contain RNA from heat-shocked cells. First and second lanes contain RNA from yeast cells transformed with the test-plasmid. Third and fourth lanes contain RNA from yeast cells transformed with the test-plasmid harbouring the 242bp fragment 260nt upstream the HIS-3 gene. The filter hybridized with the HIS-3 internal HindIII fragment is shown on the left while the same filter hybridized with the yeast heat-shock protein 26 gene probe is shown on the right.

is enhanced by an increase in temperature, the HIS-3 gene transcript is not further induced. We think that the 242 bp sequence has a high affinity for the regulatory active factors and is already saturated at 30°C.

DISCUSSION

Two important features are apparent in our repetitive DNA family : a satellite-like structure (composed of a short, 10bp, tandemly repeated consensus sequence) and a strong resemblance of the consensus to a functional regulatory DNA element (HSE). Satellite DNA in most eukaryotes is composed of short DNA sequences repeated in tandem. This type of repetitive sequences is usually associated to pericentromeric heterochromatin (32). In *C. elegans* centromeres are not discrete structures but both genetic and cytological evidences indicate that they are diffuse along the chromosomes (33). Neither pericentromeric heterochromatin nor satellite-like repetitive DNA sequences have

Fig.9: Northern transfer of yeast RNA: even lanes contain RNA from yeast cells grown at 30°C, odd lanes RNA after heat-shock from 30°C to 37°C. First and second lanes contain RNA from yeast cells transformed with the test-plasmid harbouring the 242bp fragment 260nt upstream the HIS-3 gene; third and fourth lanes contain RNA from yeast cells transformed with the test plasmid. The filter hybridized with the HIS-3 internal HindIII fragment is on the left and the filter hybridized with the yeast heat-shock protein 26 gene probe is on the right.

been described so far, although clusters of short palindromes were identified (1,2). From our Southern transfer analysis and from the screening of the genomic library we calculate that the members of the C9 family are interspersed in the genome at 20/30 loci. The dispersion of this satellite like sequence might be a direct consequence of the diffusion of centromeric structures. C9 is the first example of a repetitive family which is strictly related to a non-coding regulatory sequence with a known function (binding the heat-shock transcription factor, HSTF). This resemblance of the C9 consensus and the HSE consensus is probably not a mere coincidence. In fact, even if a 10bp sequence is a relatively simple unit, only the nucleotides that are also conserved among the HSEs, and not the entire unit, are found to be conserved during the amplification process. Furthermore, we showed that our sequence is capable of mimicking a heat-shock promoter in a heterologous system. Of course we do not have any direct evidence for a productive interaction between our repetitive DNA family and heat-shock response in *C. elegans*. Since it has been observed in *D. melanogaster* cells that the introduction of an artificial poly-HSEs sequence results in a reduction of the heat-induced

transcripts from coinjected heat-shock genes, due to competition for the heat-shock factor (31), one might expect that *C. elegans* cells, which naturally contain a large number of HSE-like sequences, should be unable to respond to stress. Since this is not the case, this sequence in the C9 family must have evolved either a direct function, specific for *C. elegans*, in the regulation of the phenomenon (such as, for example, storage of inactive factors in absence of induction) or a mechanism to neutralize the competition effect. This in turn could be achieved either through the complex pattern of repetition itself, with the different configurations it implies, or through a particular location of these sequences in the nucleus.

The event of amplification and redistribution of the HSE-like sequences, which led to the formation of the C9 family, seems relatively recent because the original unit is still identifiable and we could not detect related DNA families in other eukaryotes or nematodes with the exception of *C. briggsae* where more work is needed on a XbaI family with, at least superficially, similar features.

We propose a possible model for the origin of this repetitive family in particular and of satellite-like repetitive DNA in eukaryotic genomes in general: initial cycles of duplication of the 10bp module (and consequentially of HSE palindromes) occurred, conferring a selectable advantage to genomes by a better response to stress. We know in fact that often more than one HSE is present in front of heat-inducible genes and that, at least in *Drosophila*, the binding of the HST-factor to adjacent HSEs occurs in a cooperative fashion (34,35). Once the first round of amplification occurs, then other phenomena such as non-reciprocal recombination between non-homologous chromosomes, unequal crossover, saltatory replication could be responsible for further amplification, spreading throughout the genome and fixation in the population (36,37,38). After several generations, because of the accumulation of new mutations or because of the structure generated by the multiple repetitions, the original structure and possibly function of the initial sequence might be lost and new features be gained. An appealing hypothesis is that just as proteins with affinity to nucleic acids (DNA binding proteins) are known to have a variety of functions in the nucleus, conversely some DNA sequences might exist (Protein binding DNAs) with a high affinity for proteins which are used in the nucleus to assemble different parts of the genome in complex superstructures. These sequences may evolve from shorter DNA stretches that bind specific factors and after amplification lose in part their specificity but

gain a broader spectrum of affinities. We may imagine that other satellite-like sequences might have originated from functional DNA elements but either the sequence unit is too degenerated to be recognized as related to the original one, or this latter may still be unknown to us. We think that in this specific case we have managed to entrap such a sequence during an intermediate stage in this process.

Acknowledgments

We are grateful to Dr. Silvana Gargano for providing the *C. elegans* cosmid library and to Dr. G. Cancrini for the microfilariiae. We thank Dr. John F. Pulitzer for comments on the manuscript and Dr. John F. Pulitzer and Dr. John E. Sulston for stimulating discussions and suggestions. This work was supported in part by Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie di CNR, Roma. Some nematode strains used in this study were provided by the *C. elegans* Genetics Center, which is supported by contract number N01-AG-2113 between the National Institute of Health and the Curator of the University of Missouri.

*To whom correspondence should be addressed

REFERENCES

1. Sulston, J.E. and Brenner, S. (1974) *Genetics* **77**: 95-104
2. Emmons, S.W., Rosenzweig, B. and Hirsh, D. (1980) *J. Mol. Biol.* **144**: 481-500.
3. Pelham, H.R.B. (1982) *Cell* **30**: 517-528
4. Pelham, H.R.B. (1982) *EMBO J.* **1**: 1473-1477
5. Wu, C. (1980) *Nature*, **286**: 854-860
6. Kay, R.J., Bolssy, R.J., Russnak, R.H., Candido, P.M., (1986) *Mol. Cell. Biol.*, **6**: 3134, 3143,
7. Snutch, T.P., Hechl, M.F.P., Ballille, D.L. (1988) *Gene*, **64**: 241-255.
8. Brenner, S. (1974). *Genetics* **77**: 71-94.
9. Tschumper, G. and J. Carbon (1980) *Gene* **10**: 157-166.
10. Struhl, K. (1985) *Proc. Natl. Acad. Sci. USA* **82**: 8419-8423.
11. Hanahan, D. and Meselson, M. (1980). *Gene* **10**: 63-67.
12. Dente, L., Cesareni, G. and R. Cortese (1983) *Nucl Acids Res.* **11**: 1645-1655
13. La Volpe, A., Simeone, A., D'Esposito, M., Scotto, L., Fidanza, V., de Falco, A. and E. Boncinelli. (1985) *J. Mol. Biol.* **183**: 213-223.
14. Struhl, K. (1981) *J. Mol. Biol.* **152**: 517-533.
15. Sherman, F., Fink, G. R., Lawrence, C. W., (1974) *Methods in Yeast Genetics*. (Cold Spring Harbor, New York; Cold Spring Harbor Laboratory)
16. Southern, E. M. (1975). *J. Mol. Biol.* **98**: 503-517.
17. Struhl, K. and r. W. Davis. (1981) *J. Mol. Biol.* **152**: 535-552.
18. Corces, V., Pellicer, A., Axel, R., Meselson, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**: 7038-7042
19. Petko, L. and Lindquist, S. (1986) *Cell* **45**: 885-894.
20. Ito H., Fukuda, Y., Murata, K. and Kimura, A. (1983) *J. Bacteriol* **153**: 163-168.
21. Maniatis, T., Fritsch, E. F. and J. Sambrook. (1982) *Molecular Cloning, a Laboratory Manual*. (Cold Spring Harbor, New York; Cold Spring Harbor

- Laboratory).
22. Melton, D.A., Krieg, P.A., Rebagliati, M. R., Maniatis, T., Zinn, K. & M. R. Green. 1983. Nucl. Acids Res. 12: 7035-7056.
 23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467.
 24. Lindquist, S. (1986) Ann. Rev. Biochem. 55: 1151-1191
 25. Emmons, S.W., Yesner, L., Ruan, K., Katzenberg, D. (1983) Cell Cell: 55-65.
 26. Sorger, P.K., Lewis, M.J., Pelham, H.R.B. (1987) Nature, 329: 81-84.
 27. Lis, J., Costlow, N., de Banzie, J., Knipple, D., O'Connor, D., Sinclair, L. (1982). in Schlesinger, M.J., Ashburner, M. and Tissieres, A. (Eds.) Heat Shock from Bacteria to Man, Cold Spring Harbor Laboratory p.57
 28. Fire, A. (1986) EMBO J. 5: 2673-2680.
 29. Boeke, J.D., Garfinkel, D.J., Styles, C.A., Fink, G.R. (1985) Cell 40: 491-500.
 30. Pulitzer, J.F., Ciaramella, M. (1986) J. Cell. Biochem., suppl. 10D: 323
 31. Xiao, H., Lis, J. (1986) Mol. Cell. Biol. 6: 3200-3206.
 32. Brutlag, D.L. (1980) Ann. Rev. Genet. 14: 121-144
 33. Albertson, D.G., Nichol Thomson, J. (1982) Chromosoma 86: 409-428
 34. Zimarino, V., Wu, C. (1987) Nature 327: 727-730
 35. Topol, J., Ruden, D.M., Parker, C.S. (1985) Cell 42: 527-537.
 36. Smith, G.P. (1973) Cold Spring Harb. Symp. Quant. Biol. 38: 507-513
 37. Dover, G.A. (1982) Nature 299: 111-117
 38. Dover, G.A., Flavell, R.B., (1984) Cell 38: 621-624.
 39. La Volpe, A., Taggart, M., MacLeod, D., Bird, A. (1983) Cold Spring Harb. Symp. Quant. Biol. 47: 585-592