Conf-9403172--1

LA-UR- 94-3466

Title: MOLECULAR ARCHITECTURE OF CLASSICAL CYTOLOGICAL LANDMARKS: CENTROMERS AND TELOMERS

Author(s): J. Meyne

Submitted to: US-Japan Joint Seminar: Classical & Molecular Cytogenetic Analysis of Cereal Genomes

Los Alamos
NATIONAL LABORATORY

Form No. 836 R5
ST 2629 10/91

Molecular Architecture of Classical Cytological Landmarks: Centromeres and Telomeres

Julianne Meyne, Genomics and Structural Biology Group, Life Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545

To be published in:
Proceedings of the US-Japan Joint Seminar: Classical and Molecular Cytogenetic Analysis of Cereal Genomes (B. Gill and J. Raup, eds.) In Press (1994).

I am going to make a transition here, as we work almost entirely with mammalian systems. I would like to start with a brief review of some of the repetitive DNA sequence families found in the human genome and their relative locations in metaphase chromosomes. Each arm of each chromosome terminates with a tandemly repeated telomere sequence. In vertebrates this is (TTAGGG)n, rather than (TTTAGGG)n as found at plant telomeres. Moving along each chromosome arm toward the centromere is a region frequently referred to as the subtelomere. This is a very complex area. The organization of this region is just beginning to unfold. There appear to be several repeats in this area, some of which are specific for individual subtelomeric sites and others found in at least one subtelomeric region of several chromosomes. The alpha satellite DNA repeat family is located within the pericentric region. After fluorescent in situ hybridization (FISH) these sequences appear to be at or near the cytological centromere. The classical satellite DNA repeat families are also present in the pericentromeric regions of many chromosomes. In addition, some human chromosomes, most notably 1, 9, 16, and Y, contain large blocks of satellite DNA repeats in heterochromatic regions of the long arm (Moyzis et al. 1987). Interspersed throughout the genome are the ubiquitous repeats such as Alu, L1, GT:AC, and a variety of other sequences (Moyzis et al. 1989). These are just some of the major repetitive DNA families. There are many other less frequently represented repeats within the human genome.

Both the human telomere repeat and the pericentromeric repeat sequence $(GGAAT)_n$ were isolated based on evolutionary conservation. Their isolation was based on the premise that chromosomal features as structurally and functionally important as telomeres and centromeres should be highly conserved. Both sequences were isolated by high stringency screening of a human repetitive DNA library with rodent repetitive DNA (see Moyzis et al. 1988 and Moyzis 1991 for details). The pHuR library (plasmid Human Repeat) used for this project was enriched for repetitive DNA by using a modification of the standard DNA library preparation method. Usually DNA for a library is cut with restriction enzymes, packaged, infected, and the library is screened. A problem with this approach is that many tandem repeats don't have any (or many) common restriction sites. Therefore, many of the repeat sequences will not be represented in the library because they are not restricted to a viable length for the vector used. To prepare the pHuR library, human DNA was mechanically sheared to a small size. These relatively short DNA fragments were denatured and then renatured to C_0t 50. Theoretically only repetitive DNA sequences should renature under C_0t 50 conditions. The single-stranded regions were digested using S1 nuclease, leaving the double-stranded, renatured repeat sequences. In other words, the single copy DNA and low abundance repeat families that do not renature under C_0t 50 conditions were removed by the enzyme digestion. The remaining double-stranded regions were tailed, packaged into the plasmid, and screened. This method can also be used to construct repetitive DNA libraries from other species.

Screening the pHuR library for evolutionarily conserved sequences was done using standard methods. First, the C_0t 50 human repeat library was screened with C_0t 50 human DNA to confirm that the clones were indeed human repeats. This was followed by a fairly moderate stringency hybridization, requiring 70-75% identity, with C_0t 50 hamster DNA. The stringency was then increased to the level where near

perfect identity was required for clone detection. The only clones that should be detected by this screening are sequences that are shared by both hamster and human. Six clones were strongly positive at this stringency level. Of these, four were the ubiquitous repeat GT:AC. All animal genomes have this repeat interspersed along the length of the chromosomes. The two remaining clones, pHuR 93 and pHuR 143 were basically tandem repeats of $(TTAGGG)_n$. Because this sequence was the same as the telomere sequence reported for trypanosomes we assumed that this was the human telomere. A variety of methods were used for molecular characterization and localization of the sequence to confirm this as the human telomere (Moyzis et al. 1988).

As an adjunct to the molecular studies we used FISH analysis of the sequence to provide visual proof that this sequence is at the termini of the chromosomes (Moyzis et al. 1988). We prefer to use synthetic deoxyoligomer probes, rather than the pHuR clones, because they are more efficient for large-scale studies (Meyne and Moyzis 1989). We use $(GGGTTA)_7$ and $(TAACCC)_7$ as probes for telomere FISH. The synthetic DNA strands are tailed with biotin using terminal deoxynucleotidyl-transferase. We get fairly large signals even though our probe is only 42 bp long, probably because of formation of out-of-register concatamers by the deoxyoligomers. In a survey of over 100 species of vertebrates we found that many species have telomere sequences in the pericentric region and/or at interstitial sites (Meyne et al. 1989 and 1990). The non-telomeric sites of $(TTAGGG)_n$ do not seem to have any adverse effect on chromosome function.

Structural studies of the telomere region are also providing important information about telomeres. Telomeres of all the species that have been studied are very similar. Most have a repeating sequence where the G-rich strand contains clusters of 3-4 Gs . Model system studies show the terminal repeats of the telomere sequence can form a structure called a G quartet (reviewed in Williamson 1994). This quadraplex unit may exist in vivo and, if so, probably has structural and functional significance. Some structural features appear to vary with complexity of species. For example, Makarov et al. (1993) have reported the telomeres of rat chromosomes are constructed of closely spaced nucleosomes, while the telomeres of lower eukaryotes show no evidence of nucleosomal structure.

Continuation of the search for additional evolutionarily conserved sequences from the pHuR library led to the isolation of sequences from the human satellite DNA repeat families localized in the pericentromeric region of human chromosomes (see Grady et al. 1994 for details). After screening with hamster repetitive DNA, three clones were isolated. One of these clones, pHuR 98, is located predominantly in the pericentromeric heterochromatin of human chromosome 9 (Moyzis et al. 1987). Screening with mouse, rather than hamster, repetitive DNA yielded five additional clones of interest. The common feature of all eight sequences was the five nucleotide repeat $(GGAAT)_n$. Human chromosomes contain three basic classical satellites (Sat I, Sat II, and Sat III) which were originally isolated using ultracentrifugation and have since been characterized using molecular methods (Prosser et al. 1986). The 5-mer sequence, $(GGAAT)_n$, is a common feature shared by satellites II and III.

Zoo blot analysis of species ranging from human to yeast using either the pHuR 98 clone or synthetic deoxyoligomers of the core consensus sequences from satellite II and III simple sequences showed positive hybridization in all species except yeast. Because this method is optimized to detect repetitive sequences, it is not suprising that yeast DNA did not have a hybridization signal. The positive signals from maize, *Drosophila*, and sea urchin indicate conservation of this sequence.

One of the interesting features of the core sequences of human satellites I, II, and III is a striking similarity to the yeast centromere sequences CDE I, CDE II, and CDE III. For instance, yeast element CDE III has a functionally critical core component (Fitzgerald-Hayes et al. 1982). Mutation of two specific cytosines in the CDE III core significantly inhibits centromere function. Eight of the first nine bases of this core region can be aligned with similar regions from satellites II and III. Only deletion of a central thymidine prevents complete homology of these regions. Even with such a short sequence the probability of this similarity occurring by chance is quite large. Other homologies are also present between these families of sequences (Grady et al. 1992) and were noted after identification of the yeast sequences (Fitzgerald-Hayes et al. 1982).

FISH analysis using a synthetic probe of the 5-mer sequence, $(GGAAT)_4$, showed positive signals at the centromeric regions of a number of human chromosomes, with the largest signals in the heterochromatic regions of chromosomes 9 and Y (see Grady et al. 1992 for FISH data). Because of the evolutionary conservation indicated by the zoo blot, we also hybridized the $(GGAAT)_4$ probe to *Drosophila* polytene chromosomes. The observed fluorescence signal was very weak. We then used the 5-mer sequence as a primer for PCR of *Drosophila* DNA. The biotin labeled probe prepared from the PCR products showed a fairly intense signal in the chromocenter, or centromeric region, of the *Drosophila* polytene preparations. The more intense hybridization signal from the PCR products using the 5-mer as a primer indicates the homologous sequence in *Drosophila* may be more interspersed with other sequences, as opposed to the tandem repeat arrangement observed in human chromosomes.

Thermal hyperchromicity profiles indicated that $(GGAAT)_n$ is also a structurally interesting sequence. The C-strand of the 5-mer sequence shows no evidence of secondary structure. The melting curve of the G-strand, however, is very similar to that of the duplex for this sequence. Modification of the sequence to $(GGATT)_n$, indicates the presence of a similar structure. Data collected from the various possible permutations of the sequence indicate the second guanine in the sequence can be changed without having much effect on hyperchromicity. Changing any other of the bases can result in loss of the ability to form secondary structure (Grady et al. 1992).

DNA mobility shift assays were conducted to determine if nuclear proteins bind specifically to this sequence. Nuclear extracts from HeLa cells were incubated with pHuR 98 to allow the proteins to attach. The mass excess competitor DNA curve resulting from these assays indicated a 10,400-fold greater affinity of the protein for the pHuR 98 sequence than for E. coli or other competitor DNA. Related sequences demonstrate very similar results. Use of other repeats, such as Alu, alpha satellite, or the telomere sequence does not produce this phenomenon. These data indicate

sequence-specific binding of some protein(s) to this sequence family (Grady et al. 1992).

NMR spectroscopy revealed the duplex formed by $(GGAAT)_n$ and its complementary sequence is a normal Watson-Crick duplex (Gupta et al 1993 and Catasti et al. 1994). As indicated by the hyperchromicity studies, the C-rich strand was just a random coil state. The G-rich strand displayed an ordered structure that we call a dumbbell or stem-loop structure. If the oligomer had a length of 2, 3, 4, or 6 repeats this structure was observed. This does not necessarily mean this is the structure existing within the chromosome, but the sequence is capable of forming these structures. Based on the conservation of $(GGAAT)_n$ at the molecular level, as indicated by the zoo blots, the sequence family is greater than one billion years old. To date, it is the most conserved region that's been found in the human centromeric region. In addition, this structure can form with very little free energy cost. If repeats like this are located on the outside of the chromosome surface at mitosis, they could provide recognition sites for kinetochore function. It is interesting to speculate that this type of sequence could form a specific DNA structure that could be recognized by nuclear proteins.

I will conclude by briefly reviewing a strand-specific modification of the FISH method recently developed in our lab. We call it CO-FISH for "Chromosome Orientation-FISH (Goodwin and Meyne 1993 and Meyne et al. 1994). The cells are grown in culture medium containing bromodeoxyuridine (BrdU) for only one cycle. The BrdU partially replaces the thymidine in the newly replicated strand of each chromatid. The method is quite similar to that for differential staining to reveal sister chromatid exchange. After preparation of metaphase chromosomes using standard methods, the slides are stained with Hoechst 33258 and exposed to long wave ultraviolet light. This treatment produces many nicks in the BrdU-substituted strand, while leaving the original unsubstituted strand intact. Exonuclease III is used to digest the fragments of nicked DNA which, in effect, removes one strand of the DNA duplex from each chromatid. We then hybridize the chromosomes prepared by this method with single-stranded synthetic deoxyoligomer probes for repetitive sequences. If the simple repeating units of the tandemly repeated sequence are arranged head-to-tail, the hybridization will only be to one strand. Thus, we can hybridize a single-stranded probe to only one chromatid of a metaphase chromosome. If a sequence has a mixed orientation, both strands will be labeled and the hybridization will have the same pattern as standard FISH.

Because the G-strand of vertebrate telomeres is assumed to overhang the 3' end of the DNA duplex within each chromatid, we can also use this method to determine the direction of specific strands of repetitive sequences. For example, hybridization of the C-strand to the telomeres of chromosomes after the CO-FISH procedure, will produce a hybridization only at the 5' end of the remaining strand in each chromatid. By co-hybridizing the C-strand of the telomere and a single strand of a repetitive sequence, we can determine the direction of the repeats. Future application of this method will require increasing the resolution of the procedure to detect single copy sequences.

References

Catasti P, et al. 1994. Biochem 33: 3819-3830.
Fitzgerald-Hayes M, et al. 1982. Cell 29: 235-244.
Grady DL, et al. 1992. Proc Natl Acad Sci USA 89: 1695-1699.
Goodwin EH, Meyne J. 1993. Cytogenet Cell Genet 63: 126-127.
Gupta G, et al. 1993. In: Structural Biology: The State of the Art. Adenine Press.
Makarov VL, et al. 1993. Cell 73: 775-787.
Meyne J, Moyzis RK. 1989. Genomics 4: 472-478.
Meyne J, et al. 1989. Proc Natl Acad Sci USA 86: 7049-7053.
Meyne J, et al. 1990. Chromosoma 99: 3-10.
Meyne J, et al. 1994. Chromosoma 103: 99-103.
Moyzis RK, et al. 1987. Chromosoma 95: 375-386.
Moyzis RK, et al. 1988. Proc Natl Acad Sci USA 85: 6622-6626.
Moyzis RK, et al. 1989. Genomics 4: 472-478.
Moyzis RK. 1991. Scientific American 265: 48-55.
Prosser J, et al. 1986. J Mol Biol 187: 145-155
Williamson JR. 1994 Ann Rev Biophys Biomolec Struct 23: 703-730.

QUESTIONS

Birchler: Have you examined the relationship of the alpha sequence and the (GGAAT)n repeat at all?

Meyne: We really have not done very much about that particular relationship, but several other groups are working on it. I think it's at the level of our technical expertise for FISH. At the moment it looks like they are probably adjacent to one another. There are some molecular studies from a group in Australia showing there may be small blocks of repeats within other repeats. The blocks are so small that they did not showing up as separate regions after FISH. The region of the junctions was sequenced. Other groups are showing similar organization using extended DNA methods.

BS Gill: We heard Hunt Willard gave us a seminar last year at the Agronomy Society meetings and if I remember correctly, he said that it is the overall three dimensional structure of the centromeric region that is more important rather than any specific sequences. Can you comment on that?

Meyne: I think that statement is very fair and will prove to be true. But there has to be some specificity for sequence types to provide the three-dimensional structure. You can't just take a random sequence and expect a specific structure. I think there may be several repeat sequences associated with the structural centromere. We don't claim $(GGAAT)_n$ is the centromere sequence, particularly in the functional sense. We do propose that it may be part of the centromeric structure, however. The actual centromere itself, that region which when removed from the chromosome results in cessation of mitotic function, is probably (as in yeast) relatively small. But most commonly studied species also have a larger area that encases and protects the centromere from gene activity that also seems to be important. The sequences such as $(GGAAT)_n$ and other repeats may form structures critical for centromere integrity.

Murata: Can you say something about the present situation of artificial chromosomes in human beings?

Meyne: I think mammalian artificial chromosomes are a little ways off yet. Probably the first wave of "artificial" chromosomes will be derived from transgene and transfection methodologies. I think mammalian chromosomes are complex enough that a truly artificial chromosome is not imminent. Certainly a lot of laboratories are thinking about it, but very few labs have the expertise or funding to devote a great deal of time to a project this large. I think there will be many smaller projects directed toward this goal that will yield a great deal of structural information.

Stelly: Thinking back about 15 or 20 years perhaps, I'm not sure that to many of us anticipated the telomere sequence would be conserved. Is the centromere sequence perhaps more variable? Would you comment on that?
Meyne: I think part of it is what I alluded to earlier in that amplifications in this region do not necessarily affect the function of the region. It appears there have been a lot of amplifications of sequences around the functional centromere. There is clear evidence the alpha satellite is primate-specific and has amplified and diverged to the point of chromosomal specificity of some subfamilies. The mouse has its own set of satellite DNAs, and the rat, and so on. Many of these are at the centromere so it is easy to speculate there is some protective effect. If there was no structural or functional requirement, why are there so many pericentromeric repeats. There does not seem to be a requirement for a specific repeat sequence surrounding the functional centromere because the sequences have changed over evolutionary time without any apparent problems. While the centromeric region appears to be more variable than the telomere, I feel the actual, functional centromere will also be highly conserved. The organization of the centromere is just starting to be unraveled. It is still an area of more speculation than revelation.