Physical map of the centromeric region of human chromosome 7: relationship between two distinct alpha satellite arrays

Rachel Wevrick[1,2] and Huntington F. Willard[1]*

[1]Department of Genetics, Stanford University, Stanford, CA 94305, USA and [2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ABSTRACT

A long-range physical map of the centromeric region of human chromosome 7 has been constructed in order to define the region containing sequences with potential involvement in centromere function. The map is centered around alpha satellite DNA, a family of tandemly repeated DNA forming arrays of hundreds to thousands of kilobasepairs at the primary constriction of every human chromosome. Two distinct alpha satellite arrays (the loci D7Z1 and D7Z2) have previously been localized to chromosome 7. Detailed one- and two- locus maps of the chromosome 7 centromere have been constructed. Our data indicate that D7Z1 and D7Z2 arrays are not interspersed with each other but are both present on a common Mlu I restriction fragment estimated to be 3500 kb and 5500 kb on two different chromosome 7's investigated. These long-range maps, combined with previous measurements of the D7Z1 and D7Z2 array lengths, are used to construct a consensus map of the centromere of chromosome 7. The analysis used to construct the map provides, by extension, a framework for analysis of the structure of DNA in the centromeric regions of other human and mammalian chromosomes.

INTRODUCTION

The elucidation of the DNA sequences which comprise the centromeres of eukaryotic chromosomes remains a key step in the understanding of how this chromosomal element functions in the attachment of chromosomes to the mitotic and meiotic spindles (1, 2). The minimum requirement for centromeric DNA is that it contain sequences which form a recognition site for kinetochore proteins (3). The region in which such sequences exist is defined functionally only in the lower eukaryotes *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, and the exact sequences are known only in *S. cerevisiae* (4). In mammalian chromosomes, the centromere is defined cytogenetically by the primary constriction of a metaphase chromosome. Given current cytogenetic resolution, this region may encompass many millions of basepairs in each chromosome.

At present, the best candidate for a sequence important to centromere structure in human chromosomes is alpha satellite DNA, a family of tandemly repeated DNA present at every human centromere (5, 6, 7). A specific sequence from some alpha satellite repeats (8) has been shown to bind a centromere-specific protein, CENP-B (9), which has been independently localized to the heterochromatin subjacent to the kinetochore (10). The demonstrated relationship between alpha satellite DNA and CENP-B (8, 10, 11), and the localization of alpha satellite at every human centromere suggests that alpha satellite DNA may play a role in the structure or function of the human centromere (5, 11). Other candidate centromere sequences may be difficult to recognize in the absence of a functional assay. In addition, a better understanding of the physical organization of the sequences which are present in the region of the primary constriction is required.

Alpha satellite DNA is a primate-specific satellite DNA family composed of ~171 bp monomer units, which are tandemly arranged in a largely chromosome-specific manner to form higher-order repeat units (6). Both the length of alpha satellite arrays and the distribution of particular restriction sites in and flanking the arrays are highly variable between homologous human chromosomes (12, 13, 14). Despite this extensive polymorphism, it has been possible to obtain an estimate of the size of arrays on different homologous and non-homologous chromosomes (12, 14, 15, 16).

The centromeric region of chromosome 7 contains two distinct arrays of alpha satellite (17, 18). D7Z1 is composed of multiple six monomer higher-order repeat units, tandemly repeated to form an array of average length ~2580 kb, with a range on different copies of chromosome 7 of 1530–3810 kb, estimated using pulsed-field gel electrophoresis (12). D7Z2 is one of the smallest arrays measured, at ~265 kb (range 100–550 kb). In this work, we present a long-range physical map of the centromeric region of chromosome 7, including both alpha satellite arrays, for two different chromosome 7's isolated in rodent-human somatic cell hybrids.

---

* To whom correspondence should be addressed

MATERIALS AND METHODS

Somatic cell hybrid DNA

A50-1Acl3A is a mouse-human somatic cell hybrid containing as its only human component a translocation between the X chromosome and chromosome 7, which carries one chromosome 7 centromere and no other human centromere (18). KO15 is a hamster-human somatic cell hybrid containing a single chromosome 7 as its only human component (19). Cytogenetically the chromosome 7 appears to have a short arm rearrangement not involving the centromere (unpublished data). DNA was isolated for pulsed-field gel electrophoresis as previously described (11).

In situ hybridization

Metaphase chromosomes from A50-1Acl3A and a normal male were prepared and hybridized to biotinylated probes pMGB7 (detecting the locus D7Z2) or pα7t1 (detecting the locus D7Z1) essentially as previously described (16). Biotinylated probes were obtained from Oncor, Inc. (Gaithersburg, MD.). High stringency

conditions included hybridization at 37°C in 65% formamide, and washing at 43°C in 65% formamide. Detection of the hybridization signal was performed using a commercially available chromosome *in situ* kit (Oncor, Inc.). Slides were counterstained in propidium iodide and destained before viewing.

Pulsed-field gel electrophoresis and Southern hybridization

Restriction endonuclease digestion, electrophoresis, transfer to nylon membrane, prehybridization and hybridization were performed essentially as described (20). Hybridization of Southern blots with each probe was performed under conditions of high stringency (53°C hybridization in 50% formamide, 3×SSC, final wash in 0.1×SSC at 65°C). Contour-clamped, homogeneous electric field electrophoresis (CHEF) was performed using the CHEF-DRII system (BioRad). Conditions for the CHEF gel electrophoresis are given in the figure legends. Yeast chromosomes (strain YNN295, Bio-Rad), lambda DNA concatemers (Pharmacia) and lambda DNA digested with Hind III were used as size markers.
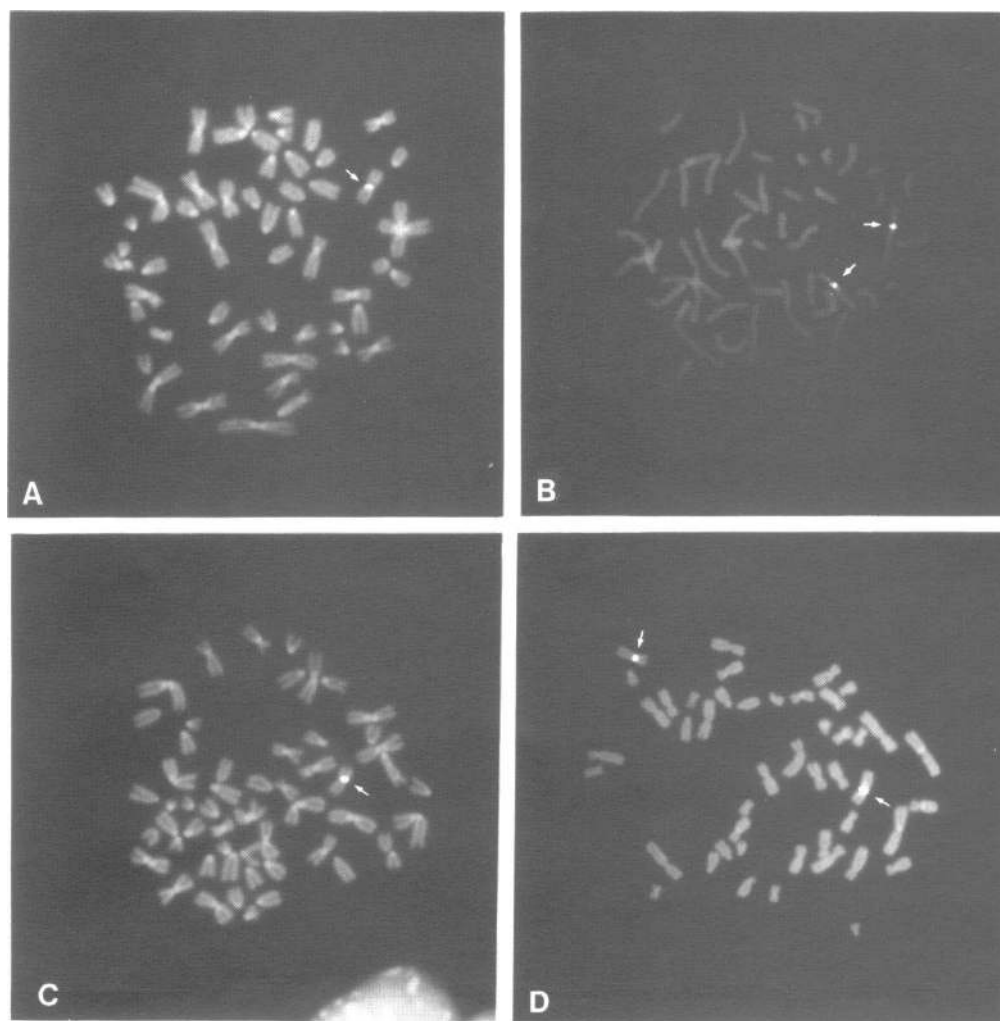

Fig. 1 *In situ* hybridization of the biotin-labeled probes pMGB7 (locus D7Z2) and pα7t1 (locus D7Z1) to metaphase chromosomes from a somatic cell hybrid and a normal human male. A) probe pMGB7, cell line A50-1Acl3A, B) pMGB7, normal male, C) pα7t1, A50-1Acl3A, D) pα7t1, normal male. In each case, the signal is localized to the primary constriction of the chromosome 7 (KO15) or both chromosome 7 homologues (normal male).

RESULTS

Localization of D7Z1 and D7Z2 by *in situ* hybridization

The loci D7Z1 and D7Z2 were originally assigned to human chromosome 7 by Southern blot hybridization of DNA from a hybrid panel (18). In order to establish the presumed centromeric localization of the two arrays, *in situ* hybridization of the probes pMGB7 and pα7t1 to metaphase chromosomes was performed. Using conditions of high stringency under which the two probes do not cross-hybridize (12, 18), biotinylated pMGB7 hybridizes to the centromere of the 7/X translocation chromosome in A50-1Acl3A (Fig. 1, panel A) and to the centromere of the two chromosomes 7 in a normal male (Fig. 1, panel B). Similarly, using biotinylated pα7t1 as a probe under high stringency conditions, the chromosome 7 centromeres in both A50-1Acl3A and the human cells are specifically labeled (Fig. 1, panels C and D). The more intense signal from the pα7t1 probe reflects the larger target size of the D7Z1 array, which is on average ten times larger than the D7Z2 array (12). The signals from the two probes appeared to be coincident at the primary constriction at this level of resolution.

Pulsed-field mapping of two alpha satellite arrays: single-locus maps

We turned to pulsed field gel electrophoresis in order to define the physical relationships between the two alpha satellite arrays. In previous experiments using the two alpha satellite probes pMGB7 and pα7t1 to detect large alpha satellite DNA-containing restriction fragments on CHEF blots of DNA from individuals in a large three-generation family, it had been demonstrated that both the array sizes and the placement of individual restriction sites within and outside of the arrays varies between copies of chromosome 7 (12). Thus the analysis of single chromosome 7's isolated in two different somatic cell hybrids was necessary.

Knowledge of the restriction map of single higher-order repeats belonging to each chromosome 7 array (18) allowed a choice of useful restriction enzymes for long-range mapping(12). DNA from each of two somatic-cell hybrids containing a chromosome 7 centromeric region was digested with a set of 17 restriction enzymes, expected to have few or no sites within each array, and analysed by CHEF gel electrophoresis and Southern blotting. If restriction sites for the particular enzyme existed within the tandem array, several fragments hybridized to the alpha satellite probe since each probe detects a locus spanning hundreds to thousands of kb of DNA. Using single and double digests for this class of enzyme (e.g. BamH I and Bgl II for D7Z2 in A50-1Acl3A, lanes 2 and 3, Fig. 2A) a restriction map was developed for the internal sites of each of the two arrays in each chromosome 7. These fragment sizes were tallied and their sum gives a maximum estimate of the array size. This sum was consistent between these enzymes, and these enzyme sites are virtually coincident at the edges of the arrays (Fig. 3). For example, using this approach the locus D7Z2 was estimated to span 170 kb in A50-1Acl3A, determined with restriction enzymes such as BamH I, Bgl II, and Xba I.

A second class of enzymes revealed only a single fragment on a Southern blot of a CHEF gel, either of the same size as estimated above (e.g. Eco RV for the D7Z2 in A50-1Acl3A, lane 5, Fig. 2A) or a larger size (e.g. Apa I and BstE II, lanes 1 and 4, Fig. 2A). These enzymes were used to develop a map of the sites flanking each array. If double digests with these enzymes failed to yield a smaller fragment than the smaller of

the two fragments detected in either of the single enzyme digests, then it was possible to determine the order of these flanking sites, but not their distances from the ends of the array. The blot used for pMGB7 in 2A, lanes 1−5 was stripped of this probe and rehybridized under high stringency to pα7t1, revealing a signal in the compression zone of the gel in the Apa I and BamH I digests (lanes 6 and 7). No cross-hybridization to the fragments from lanes 1 and 2 was detected.

Each probe was used in this fashion to develop a map in and around the array to which the probe hybridizes, independent of the second probe (Fig. 3A, D7Z1 and 3B, D7Z2). The maps
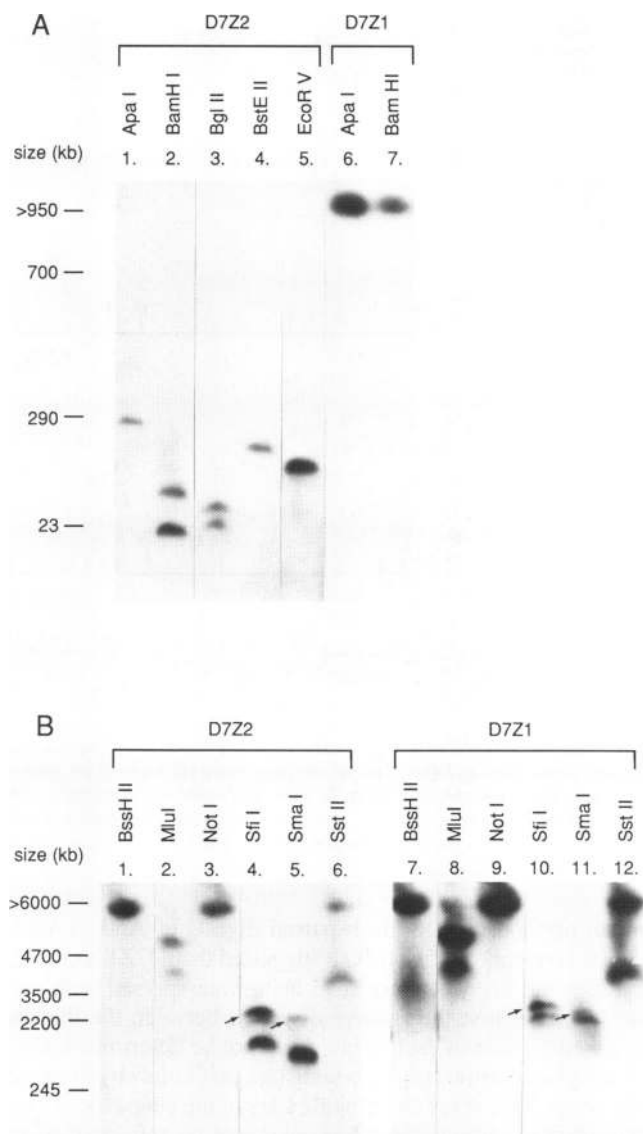

Fig. 2 CHEF gel restriction analysis of the alpha satellite arrays on chromosome 7 (A, hybrid A50-1Acl3A, B, hybrid KO15). The restriction enzymes used for DNA digestion are marked above each lane, and size markers, measured by YNN295 yeast chromosomes (Bio-Rad) and lambda DNA digested with Hind III, are indicated on the left. In addition, lambda ladders were used as markers to size accurately fragments between 23 and 290 kb in size. The locus detected in each case is noted by the D number above the lanes. Conditions for the gel in 2A were 1% agarose in 0.5×TBE, 200V for 24 hours, with a pulse time ramped from 20−80s. Conditions for the gel in 2B were 0.7% Megarose (Clontech) in 0.5×TBE, 50V, 120h at a pulse time of one hour followed by 20h at a pulse time of 90s.
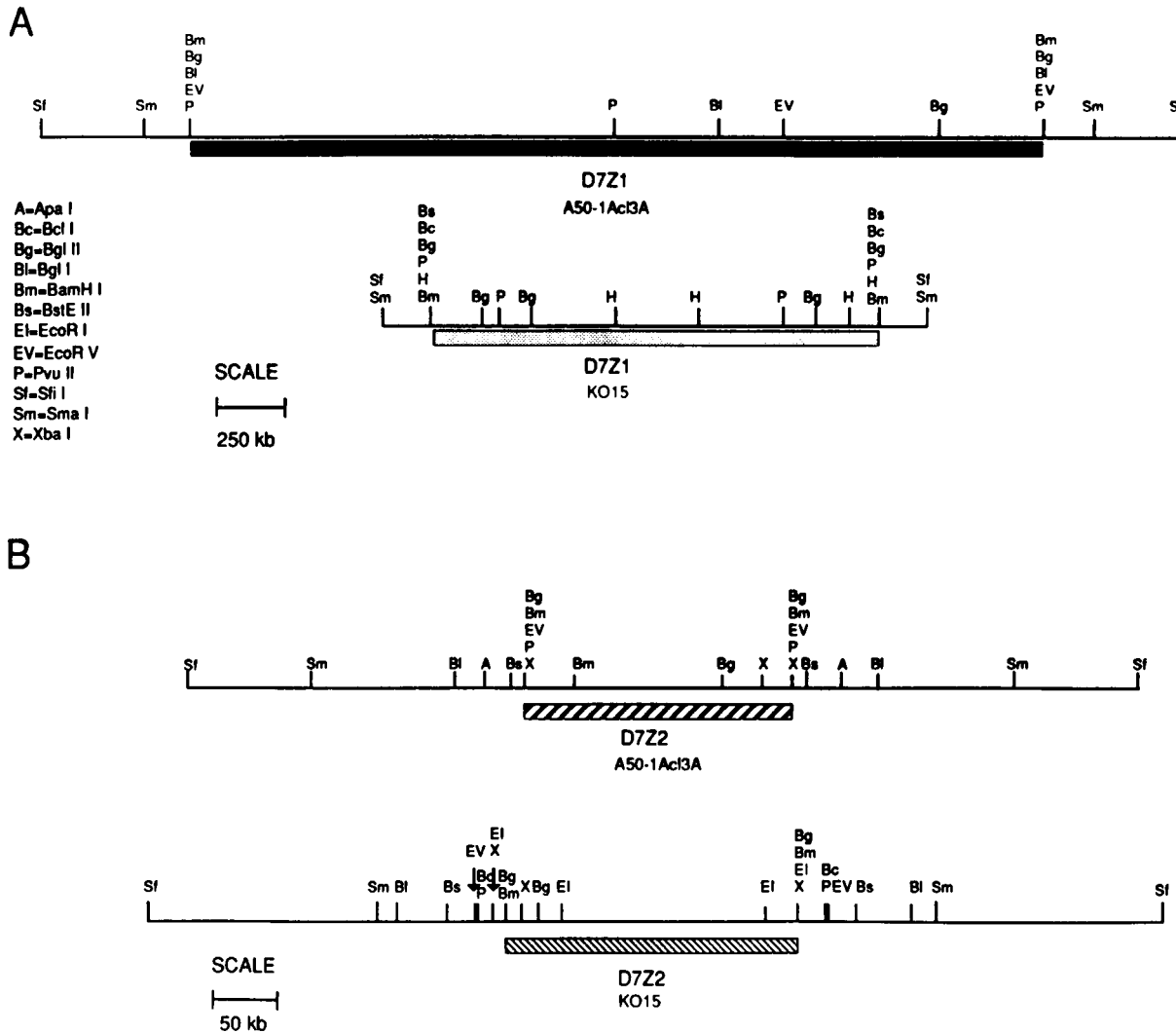
Fig. 3 Single-locus restriction maps were derived for each locus separately, using the restriction enzymes listed in the key. The rectangles represent the regions containing alpha satellite DNA. Not all enzymes were appropriate for both maps. Note the difference in scale between the maps in A (probe pα7t1, locus D7Z1) and those in B (probe pMGB7, locus D7Z2). The name of the somatic cell hybrid carrying the chromosome 7 from which the map was derived is listed below the name of the locus. In some cases, the restriction sites were placed in a symmetric arrangement because it was not possible to derive the distances from each edge of the array, only the total distance between sites. See the text for further information concerning the placement of specific restriction sites.

confirm previous results from partial digests of A50−1Acl3A DNA on conventional gels which suggested that D7Z1 and D7Z2 were separate arrays, rather than being interspersed with each other (18). Because the relative distances between the flanking sites on either side of each array could not be determined, these sites are placed arbitrarily in equidistant positions on either side of the array. This gives the flanking sites of the single locus maps a symmetrical appearance which may not be reflective of the situation on the chromosome, but indicates only the order of these sites (Fig. 3). In addition, because of the much larger size of the D7Z1 array and the decreased resolution of CHEF gels in this size range, the potential error in the position of the flanking sites for this array is larger than that for the smaller D7Z2 array.

Pulsed-field mapping of two alpha satellite arrays: two-locus maps

Standard rare-cutting enzymes (i.e., those with CpG-rich recognition sites) were used next to define the position of the arrays with respect to each other. Restriction sites for some such

enzymes exist between the two arrays. For example, D7Z2 is present on a 450 kb Sma I fragment, and D7Z1 on a 1500 kb Sma I fragment in KO15 (Fig. 2B, lanes 5 and 11). Partial digests using either Sfi I or Sma I gave a result consistent with the D7Z2 and D7Z1 arrays residing on adjacent Sfi I restriction fragments, and adjacent Sma I fragments (upper bands, in lanes 4,5,10 and 11, Fig. 2B). Finally, a limited number of rare-cutting enzymes gave the same size fragment with either probe, indicating that the two arrays could be co-localized on the same large restriction fragment (e.g. Mlu I and Sst II digest of KO15, Fig. 2B, lanes 2,6,8, and 12). Both the Mlu I and Sst II digests are partial, as seen by the presence of the upper band in each lane.

Using combinations of the types of enzymes listed above, physical maps were constructed of the region containing both D7Z1 and D7Z2 for both chromosome 7's examined (Fig. 4A). These two-locus maps could be constructed using the information provided by restriction fragments which harboured both arrays. However, the orientation of the initial one-locus maps with respect to the two-probe map, as well as the orientation of the two-locus
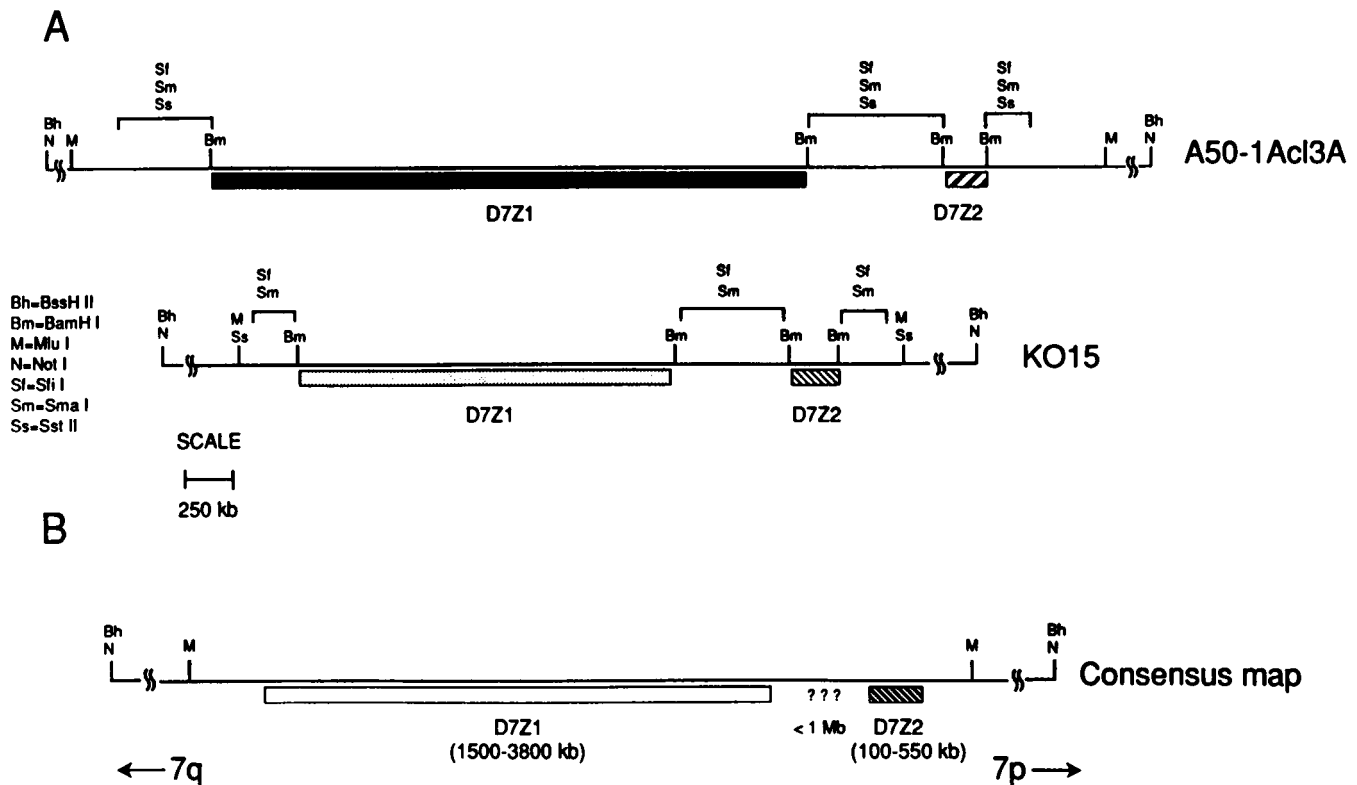
A

B

Fig. 4. A. Using the individual maps presented in Fig. 3 and the additional enzymes listed in the key, two-locus restriction maps were constructed which include both alpha satellite arrays on each chromosome. The wavy lines indicate breaks in the map, for cases in which the restriction fragment produced by digestion at sites outside the breaks were not resolved. As described in the text, the bracket below a set of restriction enzyme sites indicates that the site could not be placed exactly although the size of the restriction fragment produced by that enzyme is known. B. A consensus map of the centromeric region of chromosome 7 (discussed in the text) was derived from the two maps presented in A and previous data on the sizes of chromosome 7 arrays (12), with the shaded rectangles representing the alpha satellite arrays. The range in size of each array is indicated below the locus name. Certain rare-cutting enzymes (e.g. Not I, Mlu I and BssH II) are expected to release a fragment containing both arrays in most (or all) chromosome 7's.

map with respect to the chromosome arms remain to be determined (see Discussion). The exact distance between the two arrays on each copy of chromosome 7 examined cannot be measured accurately. However, it can be estimated to be less than ~1000 kb based on the difference between the size of each Sfi I (or Sma I) fragment and the BamH I fragment (which approximates the array size) within it (Fig. 4A). Two enzymes, Not I and BssH II, gave unresolved bands of greater than 6 Mb upon restriction enzyme analysis in all experiments. These sites were therefore placed at the boundaries of the two-locus maps, with breaks in the maps to indicate that the sizes of these fragments are unknown.

In the two-locus maps, brackets under the restriction enzyme sites between the two arrays reflect uncertainty in the order of the sites, which cannot be ordered unambiguously without the use of a third probe from this region. Thus, although the exact sizes of the restriction fragments produced by these enzymes are known, the placement of each site can be indicated only approximately in Fig. 4A. Although the size of the brackets indicates the region in which these particular sites are located, the single-locus maps (Fig. 3) give more direct information in this regard.

DISCUSSION

The isolation of the DNA sequences required for structural elements of chromosomes, such as centromeres, requires a

fundamental understanding of the nature and organization of sequences in the region of interest. In the centromeric region, alpha satellite DNA makes up the bulk of the DNA so far identified (5) and has thus formed the basis for our long-range restriction maps. In both chromosomes examined, the two arrays are present on a single large rare-cutting fragment of greater than 3.5 Megabases (Fig. 2). Given the localization of alpha satellite DNA at the primary constriction of chromosome 7 by *in situ* hybridization, it is not unreasonable to expect that the maps presented here contain some if not all of the sequences responsible for centromere activity.

Because of the extensive polymorphism in restriction sites and alpha satellite array lengths between chromosome 7 homologues (12), the maps presented are accurate in detail only for the two copies of chromosome 7 described. However, the general features of the maps are representative of the chromosome 7 centromere region and some features may extend to other centromeres as well. In each case, the restriction maps of the D7Z1 and D7Z2 arrays and surrounding DNA could be developed independently, implying that the two arrays are not interspersed and are separated by less than 1 Mb. The sizes of the D7Z1 arrays (3100 kb and 1800 kb) and the D7Z2 arrays (170 kb and 230 kb) are within the range previously measured (12). Although the orientation of the maps with respect to the rest of the chromosome is not known with certainty, the D7Z1 array can be tentatively assigned to the long arm side of the centromeric region, and D7Z2 to the short arm side by two different lines of evidence. An isochromosome

for the short arm of chromosome 7 identified in a glioma cell line lacked D7Z1 sequences as detected by *in situ* hybridization (21),thus assigning D7Z1 to the missing long arm portion of this rearrangement. This orientation has been confirmed in two independent studies by simultaneous two colour *in situ* hybridization of the two probes (D. Pinkel, personal communication, and T. Haaf, personal communication).

It was *a priori* possible that there were multiple copies of each array with identical restriction maps present on each chromosome 7. However, this can now be ruled out since there is insufficient space in the Sma I fragments which contain the individual arrays to accomodate two copies of the array, and there is insufficient space in the Mlu I fragments which contain the pair of arrays to permit a tandem duplication.

A 'consensus' centromere map (Fig. 4B), a generic version of these two centromere maps, was constructed, using the maps in Fig 4A and the previously determined range in size of the two arrays on chromosome 7 (12). In this map, the flexibility in array sizes is indicated by the range in size below the shaded box which represents the average size alpha satellite array. The distance between the two arrays is approximated as up to 1 Mb from the data in Fig. 4A. Although much information is lost in the transfer between the map of this region in a particular chromosome and a consensus map, the consensus map provides information concerning the expected appearance of this region in any chromosome 7, including array size estimates and expected restriction sites. The rare-cutting restriction fragments adjacent to the one containing both alpha satellite arrays are expected to contain either typical single copy sequence DNA from the chromosome arms or a third type of repetitive DNA. If they contain typical sequence DNA, then it can be cloned and mapped using standard methods. This DNA is then likely to lie on rare-cutting restriction fragments which are not as highly polymorphic as are the alpha satellite-containing rare-cutting fragments.

If the adjacent rare-cutting restriction fragment contains DNA of non-typical sequence, then a separate consensus map of this DNA must be constructed and joined to the alpha satellite-based maps presented here. The need for additional consensus maps is predicted for centromeric and pericentromeric physical maps on chromosomes which are known to contain classical satellite DNA (22) (e.g. chromosomes 1, 9, 16 and Y) or beta satellite DNA (23) (e.g. at least chromosomes 1, 9, and the acrocentrics). In either case, the development of consensus maps for regions of repetitive, highly polymorphic DNA is essential to the completion of physical chromosome maps (24).

The data presented in Figures 2−4 demonstrate that the D7Z1 and D7Z2 arrays are separated by DNA of unknown sequence. The appearance of restriction sites between the two arrays for enzymes (such as Sma I and Sfi I) which do not ordinarily have sites in any alpha satellite DNA (25) may indicate at least some of this DNA is not typical alpha satellite DNA. Some of this DNA may consist of alpha satellite which is diverged in sequence from either of the two probes, and so is not detected by them at high stringency. Identification of this DNA awaits the cloning of sequences adjacent to the arrays. In addition, the presence within the arrays of restriction sites for enzymes which do not ordinarily cut within the higher-order repeat unit was noted. However, the sites are not clustered as might be expected if a foreign sequence were embedded in the array, implying that the presence of these sites may merely reflect the ~1−2% differences in sequence between higher-order repeat units (6, 26).

Experience in mapping in centromeric regions indicates that cloning of the DNA across the centromeric region may also require a different approach from that used for the chromosome arms. The advent of yeast artificial chromosome (YAC) techniques (27) has now made it possible to isolate larger fragments of DNA, with a greater likelihood of detecting the junction between alpha satellite DNA and flanking sequences. However, because YACs containing tandem alpha satellite repeats have been shown to be prone to rearrangement (28), YACs may not represent a faithful copy of the original genomic structure. As well, restriction site biases intrinsic to each alpha satellite array may preclude the use of some YAC libraries with small insert sizes for arrays which do not contain certain restriction sites for tens or hundreds of kb of DNA. Despite these drawbacks, YAC technology can still be used for mapping centromeric DNA (28). Single- or low-copy probes isolated from YACs or cosmids containing repetitive DNA can be added to the consensus map using pulsed-field gel mapping.

We have developed a physical map of the centromere of chromosome 7 as a guide for the identification and cloning of elements responsible for centromere activity. The construction of consensus maps of regions of the genome, including centromeres, which may be difficult to map and clone due to extensive variation between different copies of the same region, is a necessary prelude to identifying regions of potential significance. Finally, the ability to clone such sequences into YACs or other suitable vectors using a strategy guided by physical maps of the region may eventually allow the development of a functional assay for centromere sequences in mammalian cells.

ACKNOWLEDGEMENTS

REFERENCES

1. Brinkley, B.R. (1990) Cell Motil. Cytoskeleton, 16, 104−109.
2. Rieder, C.L. (1982) Int. Rev. Cytol., 79, 1−58.
3. Nicklas, R.B. (1971) Adv. Cell Biol., 2, 225−298.
4. Clarke, L. (1990) Trends Genet., 6, 150−154.
5. Willard, H.F. (1990) Trends Genet., 6, 410−416.
6. Willard, H.F. and Waye, J.S. (1987) Trends Genet., 3, 192−198.
7. Manuelidis, L. (1978) Chromosoma, 66, 23−32.
8. Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) J. Cell Biol., 109, 1963−1973.
9. Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) J. Cell Biol., 104, 817−829.
10. Cooke, C.A., Bernat, R.L. and Earnshaw, W.C. (1990) J. Cell Biol., 110, 1475−1488.
11. Wevrick, R., Earnshaw, W.C., Howard-Peebles, P.N. and Willard, H.F. (1990) Mol. Cell. Biol., 10, 6374−6380.
12. Wevrick, R. and Willard, H.F. (1989) Proc. Natl. Acad. Sci. (USA), 86, 9394−9398.
13. Jabs, E.W., Goble, C.A. and Cutting, G.R. (1989) Proc. Natl. Acad. Sci. USA, 86, 202−206.
14. Oakey, R. and Tyler-Smith, C. (1990) Genomics, 7, 325−330.
15. Mahtani, M.M. and Willard, H.F. (1990) Genomics, 7, 607−613.
16. Greig, G.M., England, S.B., Bedford, H.M. and Willard, H.F. (1989) Am. J. Hum. Genet, 45, 862−872.
17. Jørgensen, A.L., Bostock, C.J. and Bak, A.L. (1986) J. Mol. Biol., 187, 185−196.

18. Waye, J.S., England, S.B. and Willard, H.F. (1987) Mol. Cell. Biol., **7**, 349−356.
19. Rommens, J.M., Zengerling, S., Burns, J., Melmer, G., Kerem, B., Plavsic, N., Zsiga, M., Kennedy, D., Markiewicz, D., Rozmahel, R., Riordan, J.R., Buchwald, M. and Tsui, L.-C. (1988) Am. J. Hum. Genet., **43**, 645−663.
20. Willard, H.F., Smith, K.D. and Sutherland, J. (1983) Nuc. Acids Res., **11**, 2017−2033.
21. Cremer, T., Lichter, P., Borden, J., Ward, D.C. and Manuelidis, L. (1988) Hum. Genet., **80**, 235−246.
22. Prosser, J., Frommer, M., Paul, C. and Vincent, P.C. (1985) J. Mol. Biol., **187**, 145−155.
23. Waye, J.S. and Willard, H.F. (1989) Proc. Natl. Acad. Sci. (USA), **86**, 6250−6254.
24. Tyler-Smith, C. (1987) Development, **101**, 93−100.
25. Waye, J.S. and Willard, H.F. (1987) Nucleic Acids Res., **15**, 7549−7569.
26. Durfy, S.J. and Willard, H.F. (1989) Genomics, **5**, 810−821.
27. Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Science, **236**, 806−812.
28. Neil, D.L., Villasante, A., Fisher, R.B., Vetrie, D., Cox, B. and Tyler-Smith, C. (1990) Nucleic Acids Res., **18**, 1421−1428.