Dr. Vladimir Larionov

Interagency Agreement No. DE-AI02-0163079
The project period was from 1/15/01 to 1/14/05.

Title of the project: "**Physical Characterization of human centromeric regions using transformation-associated recombination cloning technology**".

Introduction

Despite their importance in the study of mechanisms that underlie proper chromosome segregation, structural analysis of centromeric regions has proceeded slowly, mainly due to the difficulty of cloning highly repetitive DNA fragments. A special interest in the organization of human centromeric DNA was stimulated a few years ago when two independent groups succeeded in reconstituting a functional human centromere, using constructs carrying centromere-specific alphoid DNA arrays (**4, 5, 7, 10, 1, 3, 9**). This work demonstrated the importance of DNA components in mammalian centromeres and opened a way for studying the structural requirements for *de novo* kinetochore formation and for construction of human artificial chromosomes (HACs) with therapeutic potential. Until recently, HAC formation was investigated with only a few alphoid DNA arrays identified in BAC and YAC libraries. The results of these studies suggest that higher-order repeat alphoid DNAs (HORs), which have frequent CENP-B boxes (17 bp binding sites for CENP-B protein), are competent HAC substrates. The examined arrays represent only a small fraction of alphoid DNA in the centromeric regions and, therefore, their analysis did not allow to make a final conclusion on structural requirements for efficient *de novo* assembly of active kinetochore. Obviously, an additional analysis of alphoid DNA that will involve its classical mutational analysis is required to elucidate which alphoid DNA monomers are capable to seed a centromere-specific chromatin on the transforming alphoid DNA arrays. It is also important to develop a system allowing a regulation of kinetochore assembly in HAC. To address these needs, a new strategy for the construction of alphoid DNA arrays was developed in our lab (**2, 8**). The strategy involves the construction of uniform or hybrid synthetic alphoid DNA arrays by the RCA-TAR technique. This technique comprises two steps: rolling circle amplification of an alphoid DNA dimer and subsequent assembling of the amplified fragments by *in vivo* homologous recombination in yeast (Figure 1). Using this system, we constructed a set of different synthetic alphoid DNA arrays with a predetermined sequence varying in size from 30 to 140 kb and demonstrated that some of the arrays are competent in HAC formation. Because any nucleotide can be changed in a dimer before its amplification, this new technique is optimal for identifying the structural requirements for *de novo* kinetochore formation in HACs. Moreover, the technique makes possible to introduce into alphoid DNA arrays recognition sites for DNA-binding proteins. Recently we also initiated exploiting of this strategy to develop a new generation of HACs with a conditional centromere (**11**).

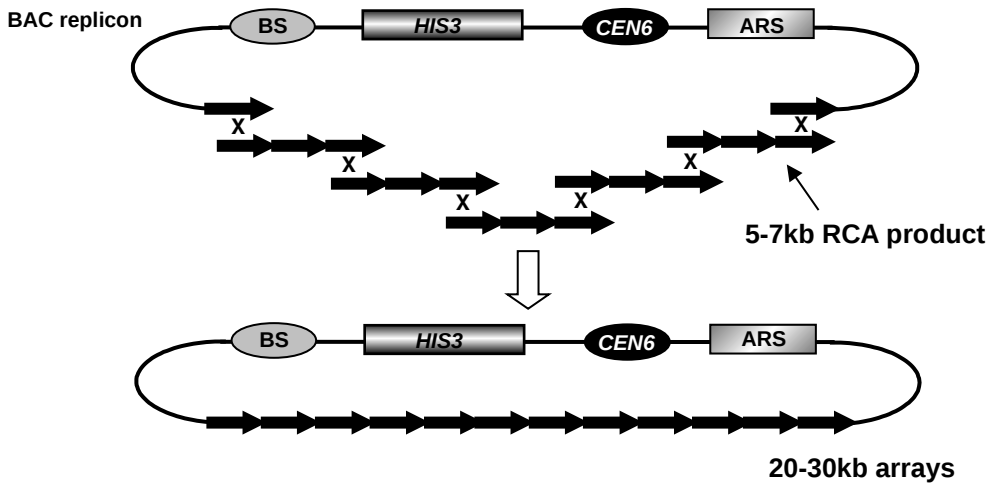

Fig. 1. Assembling of alphoid DNA arrays by *in vivo* recombination in yeast (from reference **2**).

ACCOMPLISHMENTS:

We have made the following progress on the studying of human centromeric regions using transformation-associated recombination cloning technology:

1. Optimal conditions for direct isolation of large segments of centromeric DNA for generation of Human Artificial Chromosomes (HACs) are developed. Despite their importance in the development of a human artificial chromosome (HAC) vector system, centromeric regions remain poorly characterized. In part, this is due to the difficulty of cloning highly repetitive DNA fragments and distinguishing chromosome-specific clones in a genomic library. In our work, we demonstrated the highly selective isolation of human centromeric DNA, using TAR cloning protocols (**7**). A TAR vector with alphoid DNA monomers as targeting sequences was used to isolate large (up to ~400 kb in length) centromeric regions of human chromosomes 2, 5, 8, 11, 15, 19, 21, and 22 from human cells, as well as from monochromosomal hybrid cells. The alphoid DNA array was also isolated from the 12-Mb human mini-chromosome DYq74, which contains the minimum amount of alphoid DNA required for proper chromosome segregation. The DNA contained in most of the centromeric clones was exclusively alphoid, in some cases organized in higher-order repeat arrays. Approximately 15% of the isolates contained alphoid DNA arrays interrupted by nonalphoid DNA sequences, including *Alu* and *LINE* repeats. TAR clones carrying alphoid DNA are reasonably stable during propagation in yeast and *E. coli*, indicating that these clones can be used as a starting material to

construct physical maps of centromeric regions. The ability of the cloned human centromeric regions to form HACs was assessed by transfection into human HT1080 cells. Centromeric clones from DYq74 containing higher-order repeat arrays (HOR) did not support the formation of HACs, indicating that the requirements for the existence of a functional centromere on an endogenous chromosome and those for forming a *de novo* centromere may be distinct. In contrast, constructs with HOR arrays from chromosome 22 formed HACs without detectable acquisition of host DNAs and were mitotically stable in the absence of drug selection. In summary, our results demonstrate that investigation of human centromere organization and the structural requirements for *de novo* kinetochore formation is facilitated by the use of recombinational capture in yeast. Results of this study were published in Nucleic Acids Research (**7**).

2. A novel approach for construction of alphoid DNA arrays with a predefined structure for studying of the human centromere was developed. Human Artificial Chromosomes (HACs) provide a unique opportunity to study kinetochore formation and to develop a new generation of vectors with potential in gene therapy. An investigation into the structure/function relationship in centromeric tandem repeats in HACs requires the ability to manipulate repeat substructure efficiently. We describe here a new method to rapidly amplify human alphoid tandem repeats of a few hundred bp into long DNA arrays up to 120 kb. The method includes rolling circle amplification (RCA) of repeats *in vitro* and assembly of the RCA products by *in vivo* recombination in yeast. The synthetic arrays are competent in HAC formation when transformed into human cells. Because short multimers can be easily modified before amplification, this new technique can identify repeat monomer regions critical for kinetochore seeding. The method may have more general application in elucidating the role of other tandem repeats in chromosome organization and dynamics. Results of this study were published in Nucleic Acids Research (**2, 8**).

3. Analysis of the structural requirements for de novo kinetochore formation using synthetic alphoid DNA. During the past three years we concentrated on elucidation of the structural requirements for de novo kinetochore formation using a set of different HAC constructs with a predefined structure. This work was significantly simplified by developing a novel method for construction of synthetic alphoid DNA arrays (**2**). This method exploiting in vivo recombination in yeast provides a tool for fast generation of long synthetic DNA arrays (up to 120 kb) from oligomers. Specific questions addressed in our study included: a) elucidation of the role of vector sequences in the formation of a HAC from alphoid DNA constructs, b) clarification of a minimal size of CENP-A chromatin assembled on alphoid DNA that is required for a stable maintenance of a functional kinetochore, and c) elucidation of a role of the CENP-B binding site in alphoid DNA for organization of centromeric chromatin. The alpha-satellite (alphoid) DNA constructs competent for HAC formation are built on a vector containing a mammalian selectable marker. While a role of specific alphoid DNA sequences in HAC formation is quite well understood, the contribution of a selectable marker sequence into a kinetochore chromatin domains assembly was not previously investigated. Our recent results obtained

by chromaitn immunoprecipitation (ChIP) analysis have shown that while alphoid DNA induces CENP-A and H3K4me2 assemblies, vector DNA sequences corresponding to a selectable marker (Bsr transcriptional cassette) induce histone H3 methylation, not only euchromatic H3K4me2 and H3K4me3 but also heterochromatic H3K9me3 (**12**). Based on these observations, we suggest that a transcriptional unit may be an essential component of the HAC substrate in forming two different chromatin domains, both of which are critical for the kinetochore function.

We were also interested to know what is a minimal size of CENP-A core that is required for nucleation and maintenance of a functional human centromere. For this purpose, we examined the ability of alpha-satellite DNA arrays in different lengths to nucleate CENP-A chromatin and form a functional kinetochore *de novo* (**14**). Kinetochore assembly was followed by measuring a HAC formation in cultured human cells and by chromatin immunoprecipitation analysis. The results have shown that both the length of alphoid DNA arrays and the density of CENP-B boxes had a strong impact on nucleation, spreading and/or maintenance of CENP-A chromatin, and formation of a functional kinetochore. These effects are attributed to a change in the dynamic balance between assembly of chromatin containing H3K9me3 and chromatin containing CENP-A and CENP-C (**14**). The data obtained suggest that a functional minimum core stably maintained on 30-70 kb alphoid DNA arrays represents an epigenetic memory of centromeric chromatin.

In a separate study, a role of CENP-B in *de novo* centromere and CENP-A chromatin assembly was investigated (**13**). Wild type mouse embryo fibroblasts (MEFs) and CENP-B-deficient MEFs were transfected with BACs carrying human alpha-satellite DNA, and the stability and initial/stable chromatin structure of input alphoid DNA were assessed by immuno-FISH and ChIP assays. The results show that CENP-A chromatin assembly and stable artificial chromosome formation requires binding of endogenous or exogenous CENP-B to CENP-B boxes in input alphoid DNA. However, this process is suppressed if the alphoid DNA is integrated into a mouse chromosome. In CENP-B+/+ MEFs integrated alphoid DNA undergoes H3K9me3 hypermodification and CpG hypermethylation. Although H3K9me3 hypermodification in alphoid DNA is not observed in CENP-B-/- MEFs, it is restored by CENP-B expression. These data suggest a chromatin structural model for the centromeric repetitive DNA in which two alternative chromatin structural pathways are modulated by the presence of CENP-B. Results of these studies have been submitted for publication (**12, 13**).

4. Construction of a new generation of Human Artificial Chromosome with a conditional centromere to study an epigenetic control of the kinetochore function. Using a RCA-TAR method (**2**) we have constructed a novel human artificial chromosome (HAC) to manipulate the epigenetic state of chromatin within an active kinetochore. The HAC has a dimeric alpha-satellite repeat containing one natural monomer with a CENP-B binding site, and one completely artificial synthetic monomer with the CENP-B box replaced by a tetracycline operator (tetO). This HAC exhibits normal kinetochore protein composition and mitotic stability (**11**). Targeting of several tet-repressor (tetR) fusions into the centromere had no effect on kinetochore function. However, altering the chromatin state to a more open configuration with the tTA transcriptional activator or to a

more closed state with the tTS transcription silencer caused mis-segregation and loss of the HAC. tTS binding caused the loss of CENP-A, CENP-B, CENP-C and H3K4me2 from the centromere accompanied by an accumulation of histone H3K9me3. In addition to providing the first clear demonstration that heterochromatin within the centromere is incompatible with kinetochore activity, the conditional centromere of the HAC opens a new spectrum of opportunities for the systematic manipulation of the "histone code" within the kinetochore, and definition of the full epigenetic signature of centromeric chromatin. The new HAC with a conditional centromere has also a great potential as a system for gene delivery and regulated gene expression in mammalian cells. Results of these studies have been submitted for publication (**11**).

5. Use of TAR cloning technology to clarify a role of pericentromeric repeats in organization and function of the human kinetochore. Centromeric and pericentromeric regions of mammalian chromosomes contain repetitive DNA sequences that exhibit a high rate of evolutionary changes. The exact role of these sequences with respect to kinetochore/heterochromatin structure and function is not understood yet. Our recent study investigates the effect of several repetitive elements on expression of a transgene cassette stably integrated into a mouse chromosome in mouse erythroleukemia cells (**6**). Our results show that human gamma-satellite DNA derived from the pericentromeric region of human chromosome 8 prevents vector DNA-induced epigenetic silencing of an eGFP reporter gene. Electrophoretic mobility shift assay showed that gamma-satellite arrays from human chromosomes 8, X and Y contain clusters of CTCF binding sites, and that these sites are protein-bound independent of their methylation status at CpG sites. Chromatin immunoprecipitation experiments confirmed that CTCF binds these sites *in vivo*. Given the discovery of gamma-satellite DNA in most human chromosomes (23 appended), these data suggest that pericentromeric gamma-satellite DNA plays a role in maintaining a mosaic structure in human centromeric-associated chromatin and may prevent heterochromatin spreading into and/or beyond the pericentromeric region. Also, a strong anti-silencing activity of gamma-satellite DNA suggests that these sequences may represent optimal insulators to provide a stable transgene expression in different locations. Results of these studies have been submitted for publication (**6**).

Significance. Long-term objectives of this research include determination of the structural requirements for *de novo* kinetochore formation. This would be a step to develop a novel gene delivery system based on Human Artificial Chromosome constructs. These studies may thus provide both a fundamental knowledge and new points of intervention for therapy.

References

1. Basu, J., Stromberg, G., Compitello, G., Willard, H.F., Van, Bokkelen G. (2005). Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. Nucleic Acids Res. *33*, 587-596.

2. Ebersole, T., Okamoto, Y., Noskov, V.N., Kouprina, N., Kim, J.-H., Leem, S.-H., Barrett, J.C., Masumoto, H., Larionov, V. (2005). Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. Nucleic Acids Res. *33*, e130.

3. Grimes, B.R., Monaco, Z.L. (2005). Artificial and engineered chromosomes: developments and prospects for gene therapy. Chromosoma *41*, 230-241.

4. Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K., Willard, H.F. (1997). Formation of de novo centromeres and construction of first- generation human artificial microchromosomes. Nat. Genet. *15*, 345- 355.

5. Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., Masumoto, H. (1998). Construction of YAC-based mammalian artificial chromosomes. Nat. Biotechnol. *16*, 431- 439.

6. Kim, J. -H., Ebersole, T., Kouprina, N., Noskov, V.N., Ohzeki, J.-I., Masumoto, H., Pavlicek, A., Pack, S.D., Kwon, Y.W., Flanagan, P. T., Loukinov, D., Lobanenkov, V., Larionov, V. (2007). Human pericentromeric gamma-satellite DNA binds CTCF and protects a transgene from vector DNA-induced epigenetic silencing. Submitted.

7. Kouprina, N., Ebersole, T., Koriabine, M., Pak, E., Rogozin, I.B., Katoh, M, Oshimura, M., Ogi, K., Peredelchuk, M., Solomon, G., Brown, W., Barrett, J.C., Larionov, V. (2003). Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. Nucleic Acids Res. *31*, 922-934.

8**.** Kouprina, N., Larionov, V. (2006). TAR cloning: insights into gene function, long-range haplotypes, and genome structure and evolution. Nature Rev. Genetics *7*, 805-812.

9. Monaco, Z.L., Moralli, D. (2006). Progress in artificial chromosome technology. Biochem. Soc. Trans. *34*, 324-327.

10. Moralli, D., Simpson, K.M., Wade-Martins, R., Monaco, Z.L. (2006). A novel human artificial chromosome gene expression system using herpes simplex virus type 1 vectors. EMBO Rep. *7*, 911-918.

11. Nakano, M., Cardinale, S., Noskov, V.N., Gassman, R., Vagnarelli, P., Kandels-Lewis, S., Earnshaw, W.C., Larionov, V., Masumoto, H. (2007). Inactivation of a human kinetochore by a specific targeting of chromatin modifiers. Submitted.

12. Ohzeki, J.-I., Larionov, V., Masumoto, H. (2007). The role of transcriptional cassette in the formation of Human Artificial Chromosome. Submitted.

13. Okada, T., Ohzeki, J.-I., Nakano, M., Yoda, K., Brinkley, W.R., Larionov, V., Masumoto, H. (2007). CENP-B mediates two distinctive chromatin assemblies on satellite DNA, CENP-A or tri-methyl histone H3K9. Submitted.

14. Okamoto, Y., Nakano, M., Ohzeki, J.-I., Larionov, V., Masumoto, H. (2007). A minimal core of CENP-A chromatin is required for nucleation and maintenance of a functional human centromere. EMBO J. *26*, 1279-1291.