hybridization protocol, and adds to the specificity of the overlap data. Cosmids are grown and DNA is prepared by an Autogen robot. DNAs are digested and analyzed by the AB362 GeneScanner for collection of fingerprint data. The use of novel fluorescent dyes (BODIPY) in this application has increased signal strength markedly. End fragment detection is currently carried out with traditional Southern hybridization, however additional dyes will permit detection without hybridization in the GeneScanner protocol. Data are transferred to a Sybase database and analyzed with ODS (J. Arnold, U. Georgia) software for overlap. ODS output is ported to GRAM (LANL) for map construction. A fully automated approach has yet to be achieved, but this goal is increasingly in reach.

Sequencing. An independently funded project awarded to RAG seeks to develop long-range genomic sequence for ~2 Mb of the human X chromosome. In support of this project, cosmids have been constructed and isolated for the 1.6 Mb region between FRAXA and FRAXF in Xq27.3-Xq28. To date, the complete sequences of the regions surrounding the FMR1 and IDS genes have been determined (180 and 130 kb, respectively), along with an additional ~700 kb of the interval. This sequence has led to identification of the gene involved in FRAXE mental retardation. Additional sequence in Xq28 has been determined, including that of a cosmid containing the two genes, DXS1357E and a creatine transporter. This sequence has been duplicated to chromosome 16pl 1 in recent evolutionary history. Comparative sequence analysis reveals 94% sequence identity over 25 kb, and the presence of pentameric repeats which are likely to have mediated the duplication event. A number of technical advances in sequencing have been developed, including the use of BODIPY dyes in AB373 sequencing protocols, which has offered enhanced base calling due to reduced mobility shifting, improved single strand template protocols for much reduced cost, and streamlined informatics processes for assembly and annotation.

DOE Grant Nos. DE-FG05-92ER61401 and DE-FG03-94ER61830 and NIH Grant No. 5P30 HG00210.

*Sequence-Specific Proteins Binding to the Repetitive Sequences of High Eukaryotic Genome

Olga Podgornaya, Ivan Lobov, Ivan Matveev, Dmitry Lukjanov, Natella Enukashvily, and Elena Bugaeva
Institute of Cytology; Russian Academy of Sciences; St. Petersburg 194064, Russia
Telephone and Fax: +7-812/520-9703
podg@ivm.stud.pu.ru

Repetitive sequences occupy the most part of the whole eukaryotic genome but up to the last few years there has not been much interest in their role. The situation changed when alpha-satellites in human and minor satellites in mouse became candidates for centromere function responsibility. A number of centromere-specific proteins are under investigation but none seems to distinguish centromeric functions of exact sequences among long arrays of tandemly repeated satellites. The proteins associated with that array are poorly known. We are trying to find out what proteins are involved in maintaining the heterochromatin structure of different types of repetitive sequences.

The major proportion of total genomic satellite DNA remains attached to the nuclear matrix (NM) after DNase1 and high salt treatment. We followed this association in various steps during NM preparation by in situ hybridization with the mouse satellite probe. Two mouse species were used -*M. musculus* and *M. spretus*. Both contain the same repertoire of satellite DNAs but in different amounts. In *M. musculus* the centromeric heterochromatin contains major satellite (MA) as the principal component. In *M. spretus* the minor satellite (MI) is predominant. To test DNA-binding activity of the proteins after chromatography of the soluble NM proteins on cationic and anionic ion-exchange columns, gel shift assays were performed with cloned dimer of MA and a trimer of MI. To produce antibodies, the DNA-protein complexes obtained from large-scale gel-shift assays were isolated and injected into a guinea pig.

The gel shift assay with column fractions from *M. musculus* NM and MA shows a ladder of complexes. The complexes could be competed out with an excess of MA DNA but not with the same amount of *E. coli* DNA. Antibodies from the immune serum caused a hypershift of the MA/NM protein complexes. Preimmune serum at the same dilution did not alter the mobility of the complexes. A combination of western and Southern blots allows us to conclude that a protein with a molecular weight of about 80 kD and some similarity to the intermediate filaments is responsible for the MA/NM interaction.

Specific DNA-binding activity to the MI has been tested after column fractionation of the *M. spretus* NM extract. A ladder of complexes can be competed out with an excess of unlabeled MI but not *E. coli* or MA DNA. MI contains the CENPB-box sequence, which is the binding site for the protein CENPB, one of the centromeric proteins. Fractions from the NM extract with MI-specific binding activity do not contain CENPB, as shown by western blotting with anti-CENPB antibodies.

The same kind of work is going on with human analogs of MA and MI sequences, using large clones of satellite and alpha-satellite DNA and nuclear matrices.