Two human repetitive DNA elements: a new interspersed repeat found in the factor IX gene, and a satellite 11 tandem repeat sequence

Roger D.Drinkwater, Leigh A.Burgoyne and Jennifer D.Skinner

School of Biological Sciences, Flinders University, Adelaide, Australia

Submitted 30 September 1986

A human sperm library enriched in repetitive sequences has yielded two sequences (35s and 227s) which have not been classified as repeat elements by the major sequence data banks.

35s. A Genbank search for the 107 bp sequence of 35s (see below) found a match along its full length with a section of the human Factor IX gene (intron-A 5192-5296: 79% homology) (1). However, restriction fragment/southern blot/35s probe analysis did not show a pattern that would be expected of the 35s sequence were it to reside solely in the Factor IX gene. The restriction pattern for eight enzymes is shown below. The banding pattern would imply the existence of many unit length variations of the repeat, inserted into the genome in an interspersed way. The size of the repeat unit in the factor IX gene is probably less than 1700 bp, based on the position of known flanking sequences surrounding the 35s sequence. The position of homology of 35s in factor IX is close to (208 bp) the site of a polymorphism (50 bp insert) in this gene (1,2). The copy number of the 35s sequence in the diploid genome is around 250-500, as estimated by dot blot intensity comparisons with the ALU interspersed repeat.

227s. The 227s sequence belongs to a tandemly repetitive class of DNA. A comparison of the sequence to major data bases has only found a poor homology with a single "human satellite 111 related sequence". However, 227s is almost certainly derived from the satellite 11 class of tandem repeats. The satellite 11 distinguishing features of 227s are the presence of a Taq-I site (3), a fragmented and degenerated ATTC internal tandem repeat (3), and a characteristic restriction pattern with Eco-RI (4). A 227s probed blot of Eco-RI or Hae-III digested human sperm DNA yielded an extensive ladder of bands based on multiples of 220 bp, clearly indicating a common unit length of this size for this type of satellite 11 DNA (see below). The unit multiples for Eco-RI are 6, 8, 10, 11, 12, 14, 15 and 17, and for Hae-III they are 5, 7, 9, and 11. The 227s sequence has a copy number of about 2500.

35s GTTTTCAATGTTTAGATATGCTTAGATTGACAAATCTTACCACTGTGTGACAGTTGCCACAGTATTGAGTACAATACTGTGCTGTACAGGTTTGTAGCCTAGGAG

227s CACAGCTCTCCATGGAGGGATCTATTTCGAATGATCCATTGGATAATTCCGTTTTCGTTTGGTGTGATACCAT

Restriction pattern	35s (band size-bp)	227s (band size-bp)
Alu-I	380, 590, 670, 1000, 1090, 1210	(no bands)
Hae-III	370, 620, 840, 950, 1120, 1360, 2100, 2300, 2680	1110, 1510, <u>1910</u> , 2410
Pst-I	670, 790, 950, 1160, 1570, 1700, 1860, 1990, <u>2430</u> , 3630	(no bands)
Hpa-II	540, 1670	(-)
Xba-I	1610, 1750, 1970, 2220, 2690, <u>3240</u> , <u>3850</u>	(-)
Eco-RI	1200, 1610, <u>1940</u> , 2450	1340, <u>1770</u> , <u>2190</u> , 2410, 2740, 3070, 3370, 3720
Cfo-I	710	(no bands)
Bam HI	2600 (major bands are underlined)	(-)

References: (1) Yoshitake, S., et al., (1985) *Biochemistry* 24(3736), (2) Winship, P. et al., (1984) *Nuc. Acids Res.* 12(8861). (3) Prosser, J. et al. (1986) *J. Mol. Biol.* 187(145). (4) Mitchell, A.R. (1979) *J. Mol. Biol.* 135(127).