

Evidence for dynamic alteration in histone gene clusters of *Caenorhabditis elegans*: a topoisomerase II connection?

SUZAN J. HOLT¹, WILLIAM A. CRESS² AND JOHANNES VAN STADEN^{3*}

¹Division of Biological Sciences, University of Missouri, 312 Tucker Hall, Columbia, Missouri, USA, 65211-7400

²Instituto Potosino De Investigacion Cientifica Y Tecnologica, A.C., Av. Venustiano Carranza 2425-A Col. Bellas Lomas C.P. 78210, San Luis Potosi, Mexico

³Research Centre for Plant Growth and Development, School of Botany and Zoology, University of Natal Pietermaritzburg, Private Bag X01, Scottsville 3209, South Africa

(Received 5 January 2001 and in revised form 13 August 2001)

Summary

Chromatin integrity is maintained throughout the cell cycle through repair mechanisms and intrinsically by the ordered packaging of DNA in association with histone proteins; however, aberrant rearrangements within and between chromosomes do occur. The role of the nuclear matrix protein topoisomerase II (TopoII) in generating chromosome breakpoints has been a focus of recent investigations. TopoII preferentially binds *in vitro* to scaffold-associated regions (SARs) and is involved in many DNA processing activities that require chromosome untangling. SARs, biochemically defined DNA elements rich in A+T, have been proposed to serve as structural boundaries for chromatin loops and to delineate functional domains. In our investigation of gene compartmentalization in a eukaryotic genome, SAR-associated nucleotide motifs from *Drosophila* were mapped in the regions of three histone gene clusters in an *in silico* analysis of the genome of *Caenorhabditis elegans*. Sites with similarity to the 15 bp consensus for TopoII cleavage were found predominantly in A+T enriched intergenic regions. Reiteration of sites matching the TopoII core consensus led to the identification of a novel core histone gene on chromosome IV and provided evidence for duplication and inversion in each of the three histone gene clusters. Breakpoint analysis of DNA flanking reiterated regions revealed potential sites for TopoII cleavage and a base composition phenomenon suggestive of a trigger for inversion events.

1. Introduction

The integrity of chromatin in eukaryotic genomes is maintained from one cell cycle to the next not only by DNA repair processes but also by intrinsic associations with histone proteins and an ordered hierarchy of packaging, which is used fully in the condensed state at metaphase. Aberrant chromosome rearrangements do occur, however, and are attributed to breakage–rejoining events that lead to discontinuities or possibly to exchanges between local regions of instability (Wagner *et al.*, 1993). We need to elucidate how interphase events during replication and repair might be involved in the occurrence of spontaneous breaks (Wagner *et al.*, 1993).

The nuclear matrix protein topoisomerase II (TopoII) has been a focus in recent studies of

chromosome breakpoints (Kingma *et al.*, 1997; Zhou *et al.*, 1997; Fajkus *et al.*, 1998; Kanoe *et al.*, 1999) and has been shown to mediate illegitimate recombination in eukaryotic DNA (Bae *et al.*, 1988). TopoII is a major non-histone protein in the metaphase scaffold (Gasser *et al.*, 1986) and the nuclear matrix (Berrios *et al.*, 1985) and binds preferentially *in vitro* to scaffold-associated regions or SARs (Adachi *et al.*, 1989). SARs are biochemically defined DNA elements rich in A+T and are implicated as structural boundaries for chromatin loops and as determinants of functional domains (Laemmler *et al.*, 1992). The ability of TopoII to make double strand breaks in DNA (Sander & Hsieh, 1983) and to pass another double strand of DNA through the break before religation is useful for resolving topological tangles, as in sister chromatid separation during mitosis (Holm, 1994) and daughter DNA molecule separation at the end of replication (Di Nardo *et al.*, 1984). Among

* Corresponding author. e-mail: vanstadenj@nu.ac.za

other functions of TopoII is the control of torsional stress during transcription (Wang *et al.*, 1990) and possibly in chromatin loops (Breyne *et al.*, 1994; Pikaard, 1998).

To investigate the structural compartmentalization of genic regions in a eukaryotic genome, we used features gleaned from SAR studies of the histone gene repeat in *Drosophila* (Gasser & Laemmli, 1986*a*; Gasser & Laemmli, 1986*b*) in conjunction with the genomic sequence of *Caenorhabditis elegans*, available in public access databases (The *C. elegans* Sequencing Consortium, 1998). In the *C. elegans* genome, the nucleosomal core histones (H2A, H2B, H3 and H4) occur in gene clusters of variable composition but without the nucleosome linker-associated H1 (Roberts *et al.*, 1987). Here, we report evidence for duplication and inversion rearrangements as shown by fingerprinting with nucleotide motifs, including the core consensus for TopoII cleavage from *Drosophila* (Sander & Hsieh, 1985), in three histone gene clusters. Analysis at breakpoints revealed potential sites for TopoII cleavage and a base composition phenomenon associated with apparent chromosome inversions.

2. Materials and methods

(i) Computer systems and software

A *C. elegans* Database (ACeDB), developed by Richard Durbin and Jean Thierry-Mieg (1991–), was operated via Unix systems with SUN workstations and R2000 X terminal. Documentation, code and data were accessed from <ftp://ncbi.nlm.nih.gov/repository/acedb/> and/or <ftp://ftp.sanger.ac.uk/pub/acedb/> for ACeDB 4.5 and ACeDB 4.7, data series WS3 (9/97) and WS6 (12/98). Basic Local Alignment Search Tool (BLAST) analysis was performed as BLASTN 1.4.11 (Altschul *et al.*, 1990) from <http://www.ncbi.nlm.nih.gov/BLAST/> with non-redundant nucleic acid sequences in GenBank, EMBL, DDBJ and PDB databases in December 1997. DNA Sequence Analysis System (DNASIS[™]) (1991; Hitachi Software Engineering) was used for DNA primary structure analysis including HARR plot analysis (Needleman & Wunsch, 1970), sequence comparisons and editing. Automated amino acid alignments were performed with ClustalW 1.7 (revised from Higgins *et al.*, 1991) and/or Multiple Alignment Program (Huang, 1994) from the Baylor College of Medicine Search Launcher (refer to <http://searchlauncher.bcm.tmc.edu>).

(ii) Sequences

C. elegans genomic sequences analysed were (with their accession numbers) cosmids F35H10 (U40934), F17E9 (AFO47656), ZK131 (Z83245), F08G2

(Z81495) and K06C4 (U64843). Preliminary analyses of F35H10, F17E9, ZK131 and K06C4 were made with the ACeDB WS3 data release; F08G2 was not available in the WS3 release. Later analyses were performed with the WS6 data release, in which ZK131 and K06C4 were in an inverted orientation compared with the WS3 release. The WS6 data version for ZK131, F08G2 and K06C4 is graphically illustrated in the present report. The updated version of ACeDB, as well as the WS3 version in the case of F17E9 and F35H10, correspond to the following Genbank nucleotide sequences (verified by J. Spieth, pers. commun.): F35H10 corresponds to gi1072149; F17E9 corresponds to gi2935370; ZK131 corresponds to gi3647045; F08G2 corresponds to gi3642236; K06C4 corresponds to gi9755504. Base 29491 (minus strand) in ZK131 (gi3647045) should be 'C' rather than 'G' (verified by R. Durbin, pers. commun.), which gives 'CTT' rather than 'GTT' at 29491 to 29489 and leucine rather than valine at position 11 (methionine = 1) of the deduced protein product of ZK131.4.

(iii) Nucleotide motif localization

The occurrence and distribution of the consensus for TopoII cleavage in *Drosophila* (Sander & Hsieh, 1985) was determined by searches of entire clones and/or selected regions encompassing core histone gene clusters with the ACeDB DNA Analysis tool. A perfect match was required for the hexamer core (WAYATT) and lower stringencies (maximum of three mismatches and other criteria as presented in Results) were allowed for the 15-bp consensus (GTNWAYATTNATNNG). Other nucleotide sequence motifs associated with SARs of *Drosophila* (A boxes and T boxes) (Gasser & Laemmli, 1986*b*) used in the study are reported only for the F35H10/F17E9 gene cluster. The A box motifs were either 10-mers or 11-mers as follows: AATAAATAAT, AATAAATAAA, AAAAAATAAA, AAGAAACAAT, TATATATAAA, AATAATATAAA, AATAAATA-TA, AATAATATATA, TATAAATAAA and AATAAACAAA. The T box motifs used were the following 10-mers: TTATTTTTTT, TTATATTATT, TTTT-ATTATT, TTATTTTATA, TTATGTTTAT and TAATTTTTTT. From distribution analysis of overlapping thymine runs (T tracts) in ZK131, reiterated spacing patterns of T₄ tracts were noted. Therefore, the ZK131/F08G2 histone gene cluster region was analysed for T₄ tract distribution.

(iv) Amino acid alignments

Identities of nucleosomal core histone genes reported here were confirmed through comparison of deduced amino acid sequences by automated and/or visual

alignments with the derived amino acid sequences of *his-9* to *his-12*, experimentally determined core histone genes (Roberts *et al.*, 1989). Histone gene locations in K06C4 (WS6) were determined relative to the 34 bp 3' conserved element (see Results).

3. Results and Discussion

(i) Fingerprints implicate alterations

F35H10/F17E9 gene cluster

To study histone gene compartmentalization within chromosomal regions of the *C. elegans* genome, we began on chromosome IV, genetic map locus 3.83 to 3.85, where several histone genes were predicted in ACeDB in a region of overlap between cosmids F35H10 and F17E9. Translations of five predicted coding regions in the F35H10–F17E9 overlap were found to have 100% identity to the corresponding deduced amino acid sequences of nucleosomal core histones reported by Roberts *et al.* (1989) (i.e. *his-9* (H3), *his-10* (H4), *his-11* (H2B), or *his-12* (H2A)). The histone gene cluster shown in Fig. 1*a* at the F35H10–F17E9 overlap includes *his-31* to *his-34* (F17E9.12, F17E9.10, F17E9.13 and F17E9.9, respectively), annotated as members of the HIS5 cluster. F35H10.1 (an H2A) is proposed to be *his-30* of the HIS5 cluster. The histone gene repeat in *Drosophila*, composed of genes for histones H3–H4–H2A–H2B–H1, has been found to be delineated by regions that attach to nuclear scaffolds or SARs in the intergenic spacers between histone gene repeats (Mirkovitch *et al.*, 1984). Both F35H10 and F17E9 were searched for the presence of nucleotide motifs characteristic of SARs from *Drosophila* as a strategy for localization of potential SAR domains (see Materials and methods). Three types of SAR-associated motifs (A boxes, T boxes and core consensus sites for TopoII cleavage (WAYATT)) were found with perfect identity in F35H10 and F17E9 (Fig. 1*a, c*). No identical 15 bp matches to the *Drosophila* consensus for TopoII cleavage were found but similar sites were present (discussed below).

A recurrent positional pattern of WAYATT sites was noted, encompassing an H2A in F35H10 and flanking the F17E9 H2A–H2B (Fig. 1*a*). In the vicinity of the F35H10 H2A, a 122-amino acid sequence (Fig. 1*b*) was found to have 100% identity to the deduced amino acid sequence of *his-11* (Roberts *et al.*, 1989). The H2B in F35H10 (35616–35248) completes an H2B–H2A pair, which is in reverse orientation compared with the H2A–H2B in F17E9. The mirrored patterns of WAYATT sites in the F35H10–F17E9 gene cluster (Fig. 1*a, c*) imply a non-tandem duplication with inversion of an H2A–H2B pair. A six-member core histone gene cluster is

proposed at the F35H10–F17E9 region, which differs in organization and composition from the reported clusters HIS1 through HIS4 (Roberts *et al.*, 1987), as shown in Table 1. The putative H2B at the F35H10 locus had not been predicted previously and is proposed to be a new member of HIS5 on IV, as locus *his-29* (F35H10.11). Also mapped in Fig. 1*c* are sites with 12 or 13 matches to the 15-nucleotide consensus for TopoII cleavage, where the core WAYATT is at the predicted consensus position on the same strand or in the relative position on the opposite strand.

The 1.6 kb segment between the F35H10 H2A (F35H10.1) and the H4 (F17E9.12) contained 16 core consensus sites for TopoII cleavage, two T boxes and one A box (star bordered region in Fig. 1*c*). A BLASTN analysis (Altschul *et al.*, 1990) of the 1.6 kb (plus strand) segment, run in December of 1997, resulted in sequence matches to 33 nucleotide sources ($P \leq 0.10$), 20 of which were from *C. elegans*. Analysis of the BLASTN output revealed sequences of similarity (i.e. high-scoring segment pairs or HSPs) between several *C. elegans* clones and regions near the ends of the 1.6 kb segment, including regions similar to sequences in cosmids ZK131 and K06C4.

The HSPs with ZK131 had high similarity (33 or 34 matches) to a 34-bp sequence (5'-ACCGAACCCA-ACGGCCCTCTTTAGGGCCACAAAT-3') known to occur 3' to core histone genes in *C. elegans* (Roberts *et al.*, 1989). A portion of the 34-bp sequence has close similarity to a 23-bp homology block found 3' to histone genes of other organisms (Hentschel & Birnstiel, 1981). The conserved 3' sequence contains dyad symmetry and has been associated with efficient 3' processing of histone transcripts (Birchmeier *et al.*, 1984), possibly by stabilization of the processing complex via a hairpin binding factor (Streit *et al.*, 1993). Highly similar 34-bp sequences (32–34 matches to the 34-nucleotide sequence reported by Roberts *et al.* (1989)) were also found 3' to the six core histone genes in the F35H10–F17E9 overlap region (Fig. 1*c*).

ZK131/F08G2 and K06C4 gene clusters

The 3' conserved element mapped 3' to ten coding sequences in ZK131 (Fig. 2 shows the WS6 orientation of ZK131). The ten coding sequences translate to known core histones (Roberts *et al.*, 1989) as shown in Fig. 2 and predicted in ACeDB WS6, with perfect amino acid identity (after correction of a base error at ZK131 29491, see Materials and methods). The organization of the ZK131 histone genes, exact identities of the predicted amino acids and *in silico* restriction analyses (data not shown) support the designation of the location of HIS3 (Roberts *et al.*, 1987; Roberts *et al.*, 1989) in ZK131 on chromosome II, with ZK131.3 to ZK131.10 corresponding to *his-9*

Table 1. Comparison of histone gene clusters in *Caenorhabditis elegans*

Histone gene cluster ^a	Histone gene order
HIS1	H4-H3-H2A-H2B
HIS2	H4-H3-H2A-H2B
HIS3	H3-H4-H2B-H2A-H3-H4-H2B-H2A
HIS4	H3-H4-H2A-H2B-H2A-H2B
F35H10-F17E9 cluster	H2B-H2A-H4-H3-H2A-H2B ^b
	H2B-H2A-H3-H4-H2A-H2B ^c

^a Histone gene clusters refer to closely linked sets of histone genes named HIS1 to HISn (Roberts *et al.*, 1987). The histone gene cluster located in the overlap region of cosmids F35H10 and F17E9 is described in the present paper.

^b The left-to-right orientation of the F35H10-F17E9 cluster follows the base sequence order from low to high base position as oriented on chromosome IV in ACeDB.

^c The F35H10-F17E9 cluster is shown in reverse orientation for comparison to the HISn clusters (Roberts *et al.*, 1987).

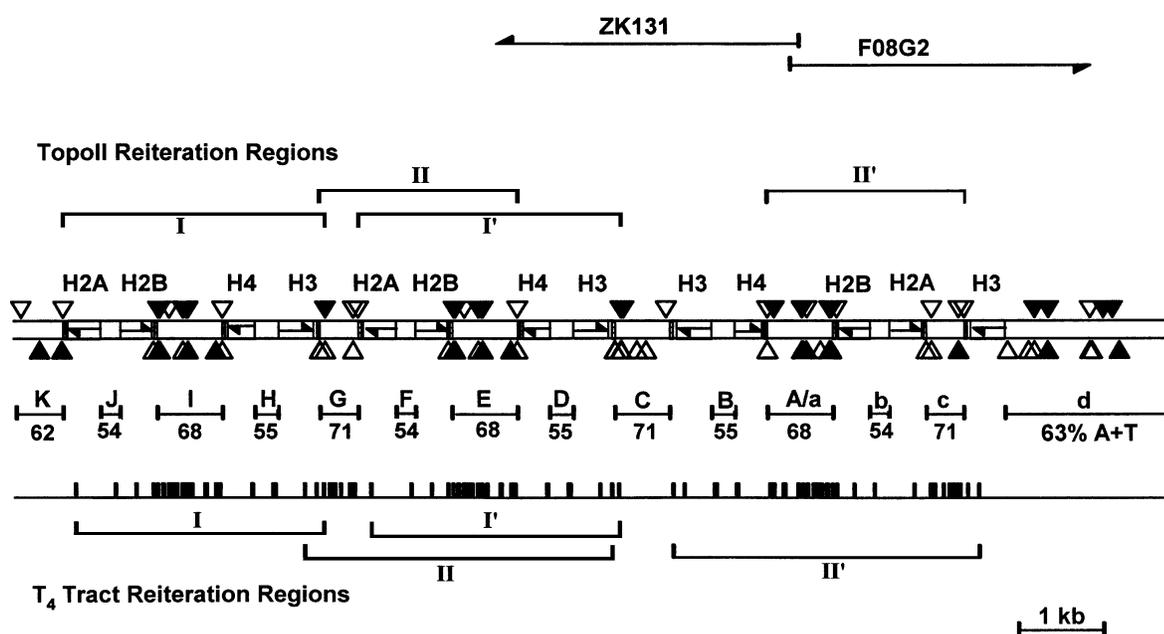


Fig. 2. Evidence for sequential duplication and inversion in the nucleosomal-core histone gene cluster at the ZK131-F08G2 overlap on chromosome II. A 13-member histone gene cluster (ZK131.10 to ZK131.1 and F08G2.1 to F08G2.3 consecutively from left to right) is shown fingerprinted by TopoII core consensus sites (WAYATT) and thymine tracts. Gene orientation, TopoII consensus cleavage sites (∇/Δ or $\blacktriangledown/\blacktriangle$) and conserved 34-bp regions 3' to each histone gene (etched bars) are as described in Fig. 1. The percentage A+T in the intergenic zones is indicated; these zones are bounded by either 34 bp elements 3' to core histone genes or by 5' ends of histone genes. Zones K and d are each bounded on one side by non-histone coding regions (not shown). T_4 tracts that were distributed in reiterated patterns are shown as vertical bars. The first set of regions with reiteration of the distribution pattern for TopoII core consensus sites (or for T_4 tracts) are designated Regions I and I', and occur only in ZK131. In the second set of reiteration regions (Regions II and II'), Region II' contains motif distribution patterns that are inversely oriented compared with the patterns in Region II, and crosses the ZK131-F08G2 overlap.

consensus for TopoII cleavage in *Drosophila* (Sander & Hsieh, 1985). BLASTN analysis of a 1.6 kb inter-histone segment (between the stars) led to the localization of 34-bp conserved elements 3' to each histone gene (etched bars). The percentage A+T content in eight inter-histone zones (A-H) are shown. Zone boundaries are delimited by a non-histone coding region (F-2 = F35H10.2), 34-bp sequences 3' to core histone genes and 5' ends of core histone genes. Zone C is bounded on one side by zone D at the start of a region with similarity to *C. elegans* clone F58B3, and zone H ends 2 kb from a conserved 34-bp sequence 3' to F17E9.9 (H2B).

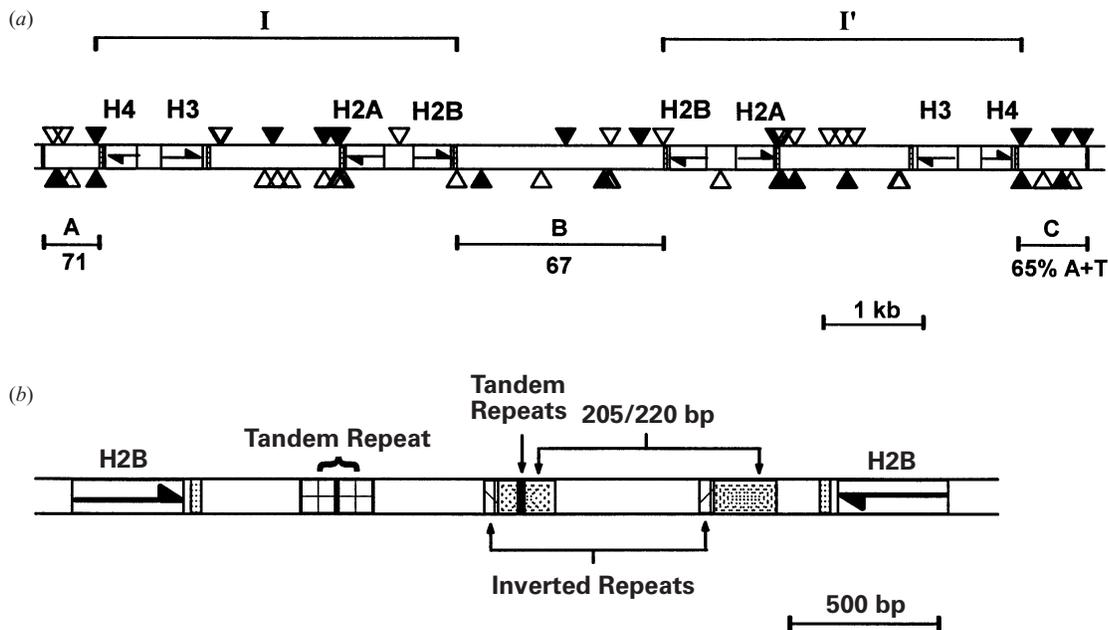


Fig. 3. Evidence for tandem duplication with inversion in the nucleosomal-core histone gene cluster in K06C4 on chromosome V. (a) An eight-member cluster of core histone genes is shown that corresponds to K06C4.2 (*his-28*), K06C4.13 (*his-27*), K06C4.3 (*his-21*), K06C4.12 (*his-22*), K06C4.4 (*his-20*), K06C4.11 (*his-19*), K06C4.5 (*his-17*) and K06C4.10 (*his-18*) consecutively from left to right. Gene orientation, TopoII consensus cleavage sites (∇/Δ or $\blacktriangledown/\blacktriangle$) and conserved 34-bp elements (etched bars) are as described in Fig. 1. Regions of reiterated and inverted positional patterns of core consensus sites for TopoII cleavage are marked as I and I'. The percentage A + T in zones 3' to putative duplicated regions is indicated; these regions are bounded by the 34-bp conserved element 3' to core histone genes or the start of predicted non-histone coding regions, which flank the histone gene cluster. (b) Battery of repetitive DNA in the 3' spacer between inverted H2B genes in K06C4. Two inverted H2B genes (K06C4.12 and K06C4.4, left to right) are shown with 3', 34-bp conserved elements (etched bars) and a series of repetitive DNA sequences: 120/121-bp tandem repeats (99% identity), 48/49-bp inverted repeats that overlap 205/220 bp regions that are 76% identical and six tandem repeats of GCAT.

to *his-16*, respectively. The histone gene cluster in ZK131 is next to core histone genes in cosmid F08G2, (F08G2.1 to F08G2.3), which we found to be identical at the amino acid level to *his-11* (H2B), *his-12* (H2A) and *his-9* (H3), respectively.

Based upon the frequency of WAYATT sites in the F35H10–F17E9 gene cluster and the usefulness of the six-base motif in characterization of the genomic sequence history, the same approach was used to fingerprint the histone gene cluster in ZK131. Early analysis of WAYATT site distribution in ZK131, before the F08G2 sequence was available, revealed reiteration of positional patterning (TopoII Regions I and I' in Fig. 2) and provided evidence for a tandem duplication of a full set of core histone genes. In addition to SAR motif analysis in our study, scans were made in several clones for thymine runs (T tracts), because T tracts were noted near certain WAYATT sites and were present in sites used to develop the *Drosophila* consensus for TopoII cleavage by Sander and Hsieh (1985). From an analysis of T₄ tracts in ZK131 a reiteration pattern in tract sites was noted that implied an additional chromosome rearrangement. Analysis of the expanded sequence of the ZK131–F08G2 overlap revealed two sets of

regions with reiterated patterns of TopoII core cleavage sites (or T₄ tracts): Regions I and I', found only in ZK131, and Regions II and II', mirrored patterns that involved both ZK131 and the overlap of ZK131–F08G2 (Fig. 2). Combined with the TopoII fingerprint data, the T₄ tract fingerprints indicate a second duplication, but in this instance with inversion, of an entire core histone gene set and at least a portion of an H3 (ZK131.7). Thus, the 13-member core histone gene cluster on chromosome II contains apparent sequential duplications with an inversion of the second duplication.

C. elegans cosmid K06C4, noted through the BLASTN analysis of the 1.6 kb segment in the F35H10–F17E9 cluster (Fig. 1c), contains an eight-gene cluster of predicted core histone genes annotated as HIS4 (H4–H3–H2A–H2B–H2B–H2A–H3–H4) (Fig. 3a). K06C4 coding regions, located relative to the 34 bp 3' conserved element (Fig. 3a), were translated and compared to known core histones (Roberts *et al.*, 1989). Perfect identity was found for the H4, H3 and H2A classes of histones compared to the deduced products of *his-10*, *his-9* and *his-12*, respectively. Each H2B translation from the K06C4 gene cluster had alanine rather than proline after the

start methionine but otherwise was identical to the deduced amino acid sequence of *his-11*. Vanfleteren *et al.* (1986) found ~90% of H2B subtypes in *C. elegans* with alanine as the N-terminal residue and the remaining 10% started with proline. The arrangement of the histone genes within the K06C4 cluster suggested a duplication with inversion. The distribution patterns of WAYATT sites clearly indicated two reverse-oriented regions, each containing a complete set of core histone genes and thus support the prediction of a duplication with inversion of a full set of core histone genes in K06C4 on V (Fig. 3a).

(ii) Breakpoints flank expected rearrangements

Intergenic regions in the three histone gene clusters (F35H10–F17E9, ZK131–F08G2 and K06C4) were analysed for base content (Figs 1 to 3). The A+T composition was enriched (62–71%) in zones 3' to histone genes and at moderate levels in 5' spacers (54–57%). The coding regions of one set of the four classes of nucleosomal core histones (ZK131.7 to ZK131.10) were 51–53% G+C. Portmann *et al.* (1976) illustrated the enriched A+T spacers between histone genes in the sea urchin histone repeat by partial denaturation mapping and electron microscopy. Unlike the sea urchin histone repeat organization, the core histone genes in *C. elegans* are organized as divergently transcribed pairs with shared 5' spacers of moderate A+T content in addition to the 3' spacer regions of A+T enrichment.

In the F35H10–F17E9 gene cluster, zones H, A, F and C, which flank the mirrored H2A–H2B and H2B–H2A pairs (Fig. 1c), were analysed individually by HARR plot analysis for regions of repetitive DNA (data not shown). A 20-bp sequence (5'-AGTTGT-TCCAATAAAATTTT-3') was found at two locations in zone H, proximal (76 bp 3' to H2B, F17E9.9) and distal (404 bp 3' to H2B, F17E9.9). The proximal 20-bp repeat was also found at the comparable location (76 bp 3' to H2B) in zone A, but the distal repeat was missing. When the first 400 bp of zone H and the last 400 bp of zone A (reverse complement (r-c)) were aligned, the distal 20-bp repeat in zone H was found to be 28 bp downstream from a break in identity between the two zones (Fig. 4a). Comparison of the first 330 bp of zones F_{r-c} and C revealed a similar break in identity. The remaining 84 bp of the H and A_{r-c} zones were tallied for A+T content and, surprisingly, even though they are in regions with low sequence similarity base for base, they had identical A+T content (75%). The same approach was taken for the 84 bp past the break in identity between zones F_{r-c} and C, and, again, identical A+T content (70%) was found in each 84 bp block (Fig. 4a).

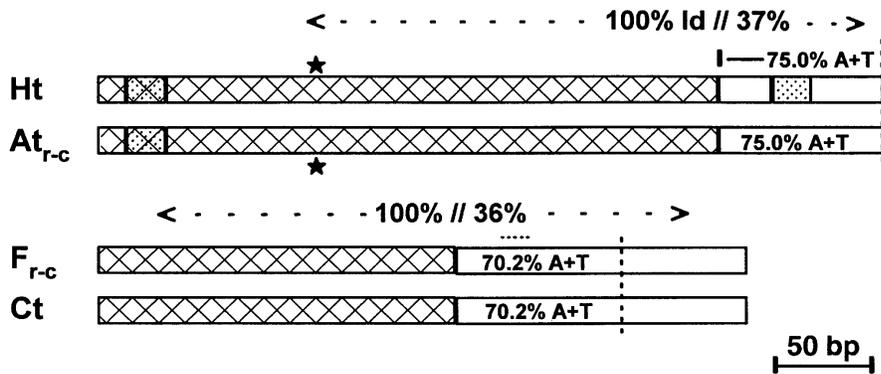
The same general approaches used with the F35H10–F17E9 gene cluster were taken for com-

parisons of zones (entire or truncated) that flanked putatively reiterated gene segments in ZK131–F08G2 and K06C4 to locate breakpoints in sequence similarity and repetitive DNA. Features of the zone comparisons are shown in Fig. 4b, c. Sharp breaks in sequence identity were also evident in zone comparisons from the K06C4 cluster (Fig. 4c) as were seen in the F35H10–F17E9 cluster. The comparison of zones H_{r-c} and d in ZK131 (Fig. 4b) by gapped alignment revealed a drop in similarity roughly midway into the regions compared and 1 bp past a conserved 16-bp sequence element, which was reported by Roberts *et al.* (1989) to occur in the 5' spacer of core histone genes in *C. elegans*. Other conserved elements reported by Roberts *et al.* (1989) were also noted in zones H and d, with perfect identity being present in zone H (the 5' H3–H4 spacer in ZK131) and reduced identities in zone d (5' to H3 in F08G2).

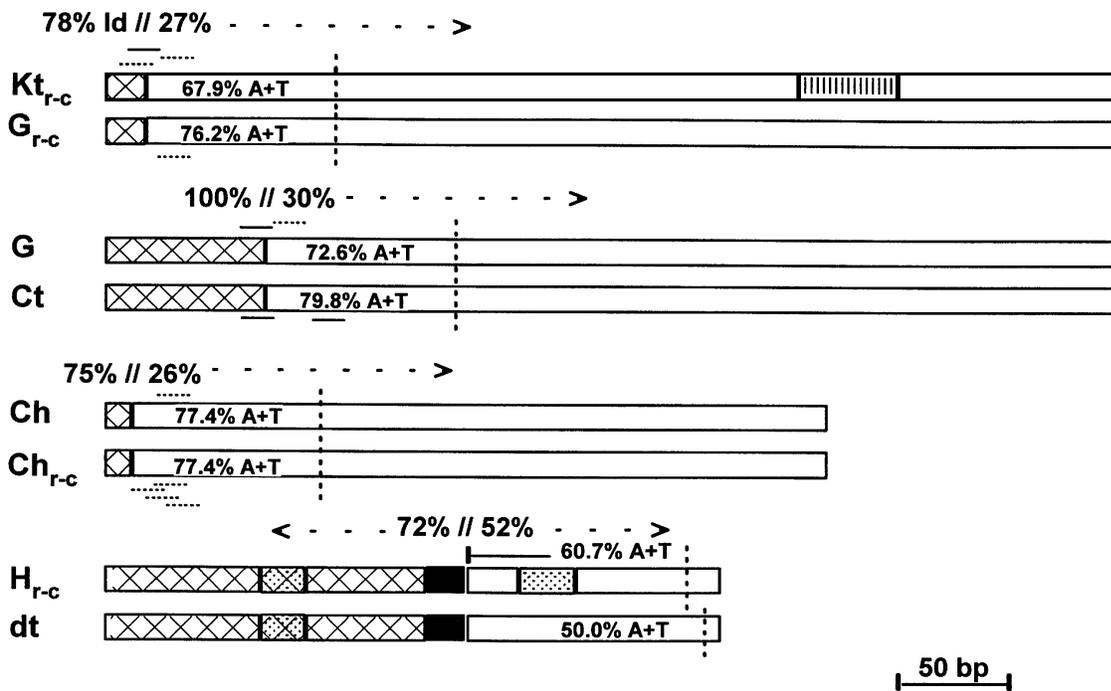
The first 84 bp past the region of similarity between zones C and C_{r-c} in ZK131 also contained identical A+T content, at 77% (Fig. 4b). Nearly identical A+T content (within 1 base) was found in 84 bp blocks flanking the breakpoint between K06C4 zones A_{r-c} and C, and the similarity in A+T content contrasted with wider differences in the second and third 84 bp blocks more distal to the breakpoint (Fig. 4c). Variable A+T content in 84 bp regions flanked breakpoints between zones K_{r-c} and G_{r-c}, G and C, and H_{r-c} and d in the ZK131–F08G2 cluster, and between zones B and B_{r-c} in K06C4. In total, for all three core histone gene clusters, 84 bp blocks flanking putatively inverted DNA segments were found to have identical or nearly identical A+T content in four out of six paired comparisons, even though sequence identity in the 84 bp blocks was low (29–37%).

Two comparisons of zones flanking putative inversions that did not have matching A+T content immediately beyond the break in similarity were K06C4 zones B and B_{r-c}, and ZK131 zone H and F08G2 zone d. In the K06C4 B–B_{r-c} comparison, the 84 bp block falls very near a region of repetitive DNA (Fig. 4c). Zone B in K06C4 had several types of repetitive DNA, including inverted repeats and tandem repeats (Fig. 3b). The repeats were not found in zones A or C, however, and might have had an influence on the insertion site for a duplicated and inverted segment or on the duplication/inversion event itself. In the ZK131–F08G2 gene cluster, zone H is of moderate A+T content and is a shared promoter region between H3 and H4 (Robert *et al.*, 1989). Zone H is thus dissimilar in sequence environment to the enriched A+T regions where breakpoints occurred in all of the other zones analysed. Also, the breakpoint between zones H_{r-c} and d (ZK131–F08G2) is adjacent to a conserved 16-bp element, which might imply some association between the 16-bp element and the break in this region of the promoter. ZK131

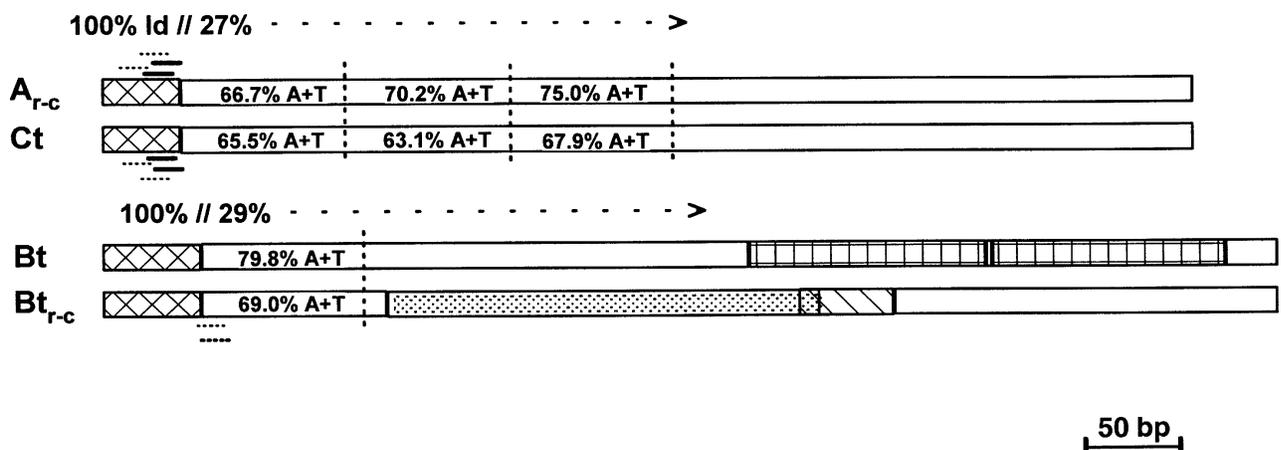
(a) F35H10/F17E9



(b) ZK131/F08G2



(c) K06C4



zones K_{r-c} and G_{r-c} , and zones G and C flank DNA segments that appear to be duplications without inversion. To summarize, the data from most inverted duplications in the core histone gene clusters analysed show an association between an inverted orientation of duplicated DNA and A+T content rather than base sequence *per se* in short blocks of DNA adjacent to breakpoints in similarity.

(iii) *Similarity to TopoII consensus sites in breakpoint regions*

The core consensus for TopoII cleavage was found predominantly in enriched A+T intergenic regions of the histone gene clusters on chromosomes IV and II (Figs 1, 2) and in three intergenic zones analysed in K06C4 (Fig. 3). Because the core consensus WAYATT is A+T rich, the motif might be expected to occur in A+T rich regions. Nevertheless, for 169 kb searched in five clones, one core consensus TopoII cleavage site was found on average every 202–203 bp, a distribution in good agreement with the average number of binding sites at saturation for TopoII in *Drosophila* (200) (Adachi *et al.*, 1989) and an even denser distribution than expected in DNA of random composition (one WAYATT site every 512 bp). Although no sites with perfect identity to the 15 bp consensus from *Drosophila* were found, a subset of core consensus sites were located within or opposite sites with 80 or 87% identity to the 15-bp consensus for TopoII cleavage (Figs 1 to 3). Fig. 4 shows the locations of sites with similarity to the TopoII 15-bp consensus and occurring near breakpoints in sequence similarity. Of eight breakpoints between regions flanking duplicated segments in the histone gene clusters examined, six revealed sites with 12 or 13 matches to the 15-bp consensus for TopoII cleavage either near or crossing breakpoints (Fig. 4). In addition, zone A of the

F35H10–F17E9 gene cluster contained an eight-base sequence (TCTTAATC) across the breakpoint with zone H. The eight-base motif TCTTAATC was also found in zones H and A within sites with 12 matches to the 15-bp *Drosophila* consensus for TopoII cleavage and located ~200bp from the breakpoint in identity between the two zones (Figs 4a, 5). The occurrence of the eight-base sequence at two sites within 205 bp in zone A (F35H10) is improbable under random conditions because only one such eight-base motif would be expected to occur in roughly 18700 bp of randomly ordered DNA with base composition as in the analysed 400-base region of zone A. In Fig. 5 12 sequences that cross breakpoints are compared with the *Drosophila* TopoII consensus and the *C. elegans* site that has 80% identity to the *Drosophila* TopoII consensus and contains the motif TCTTAATC.

The sequence TCTT was found six times at breakpoints and other four-base T-rich sequences were found in four additional instances (Fig. 5). Of 16 sequences analysed for breakpoint sites, 13 were in zones of $\geq 65\%$ A+T (Figs 1c, 2, 3a) and eight of the 13 were marked by TCTT or TTCT prior to a break in similarity with the compared sequence. Given the T and C content in each of the 13 A+T rich zones, less than one TCTT or TTCT site would be expected to occur in any 52 bases of any of the zones. Yet the observed eight TCTT/TCTT motifs constituted 32 bases of the particular 52 bases present prior to breaks in similarity, with a four-base site frequency 15–37 times the expected frequencies in randomly ordered DNA. Therefore, the observed recurrence of the TCTT/TTCT motifs at breakpoints is unlikely to be due to trivial incidence in A+T rich regions.

Negrini *et al.* (1993) found TopoII sites near (< 30 bp from) breakpoints of a translocation between chromosomes 11 and 9 in cells of a leukaemic infant. In an amsacrine-induced rearrangement involving the

Fig. 4. Breakpoint analysis in core histone gene clusters on chromosomes II, IV and V. Portions or entire zones analysed for base content and flanking putative duplicated regions (Figs 1c, 2, 3a) were compared for similarity. High similarity between compared sequences is shown by cross hatching; double forward slashes (//) mark an abrupt change in percent identities and thus the apparent breakpoint sites. Sites with similarity to the 15 bp consensus for TopoII cleavage in *Drosophila* (Sander & Hsieh, 1985) and located (entirely or partly) within ± 30 bp from the breakpoints are shown as solid (13 matches out of 15 bases) or dashed (12 matches out of 15 bases) lines above or below the zones (the strand is not specified). TopoII consensus cleavage sites with a core hexamer (WAYATT) on the same or opposite strand are shown in bold. Vertical dashed lines mark 84 bp sections immediately past regions of similarity between compared zones. Abbreviations: t, truncated zone length; h, half zone length; r-c, reverse complement; Id, identity. (a) Comparisons of zones 3' to putative duplications in the F35H10–F17E9 overlap (Fig. 1c). Dotted regions in zones A and H are 20-bp repeats. Sites at 205 bp from the breakpoint (stars) have 12 bases out of 15 that agree with the TopoII consensus for cleavage and contain the sequence TCTTAATC, which is also present in zone A at the breakpoint in identity between zones H and A (Fig. 5). (b) Comparisons of zones 3' or 5' to putative duplications in the ZK131–F08G2 overlap (Fig. 2). In zones H and d, which were compared by gapped DNA alignment, dotted regions (20 bp elements) and solid regions (16 bp conserved element) are conserved elements in the 5' spacer between core histone genes of *C. elegans* (Roberts *et al.*, 1989). The region marked with vertical lines in zone K_{t-r-c} represents repetitive AG ... AG. (c) Comparisons of zones 3' to putative duplications in K06C4 (Fig. 3a). Repetitive regions in zone B are tandem repeats (square hatching) and an inverted repeat (diagonal hatching) that overlaps a 205 bp region (dotted) with 76% identity to 220 bp 3' to H2B at K06C4.4 (Fig. 3b).

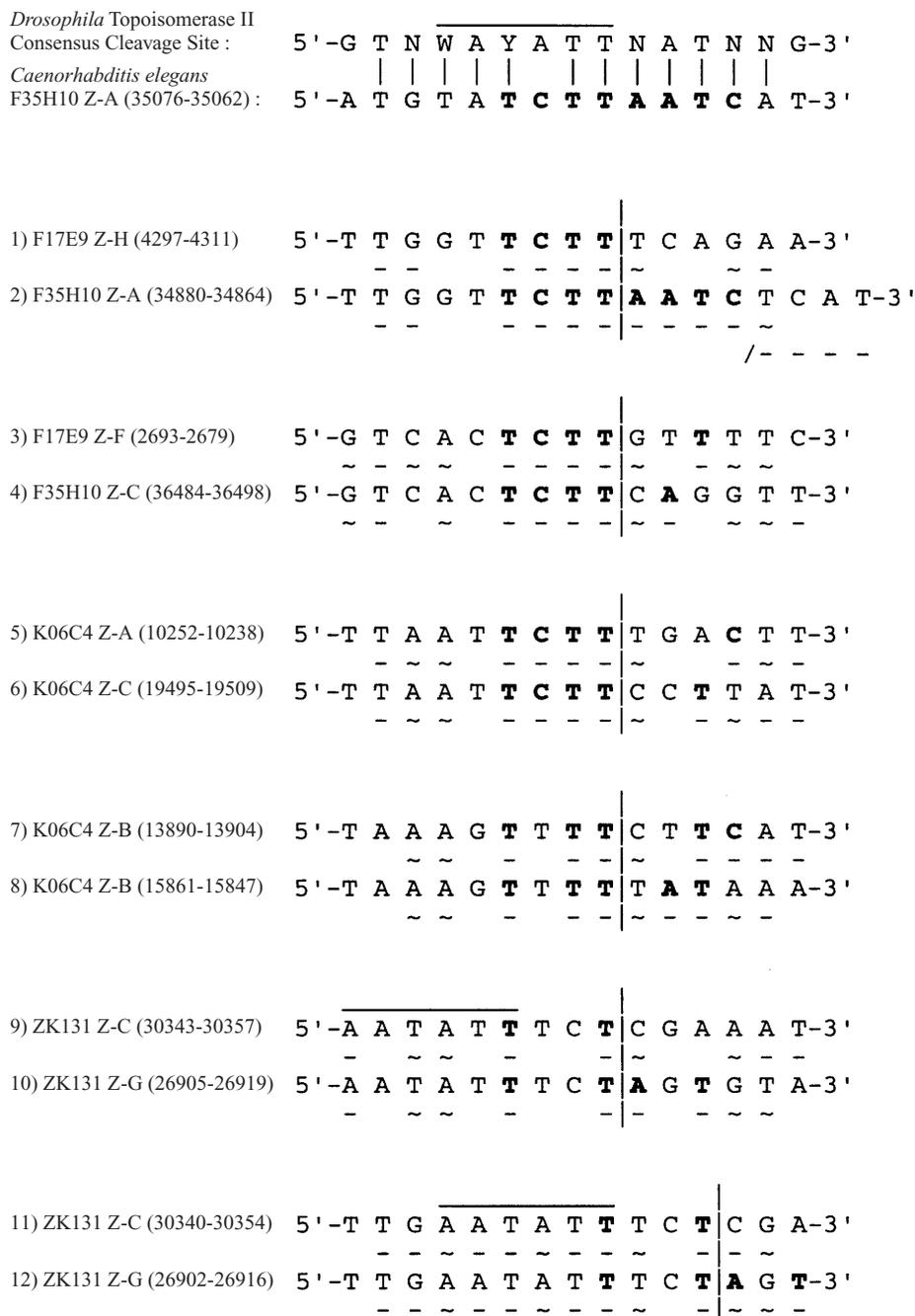


Fig. 5. Similarity of nucleotide sequences in breakpoint regions to the consensus for cleavage by TopoII in *Drosophila* (Sander & Hsieh, 1985). A 15-bp site in F35H10 zone A (F35H10 35076–35062) is shown compared to the *Drosophila* consensus for TopoII cleavage. 12 sites at or near breakpoints (vertical dashed lines) in histone gene clusters in F35H10–F17E9, K06C4 and ZK131 are compared to the *C. elegans* site for reference and to the *Drosophila* consensus. Straight lines under a base indicate a match to the *C. elegans* reference site. Additional agreement with the *Drosophila* consensus is shown by tildas. The *Drosophila* core consensus at the TopoII cleavage site is overlined. Matches to the eight-base sequence TCTTAATC within the *C. elegans* reference site are shown in bold \pm four bases from breakpoints. The eight-base sequence TCTTAATC lies symmetrically across a breakpoint between zones H and A in the F35H10–F17E9 histone gene cluster (Fig. 4a). The 3' half of sequence 2 differs from the 3' half of the *C. elegans* reference site by the presence of one TC repeat. Sequences 11 and 12 begin three bases upstream from sequences 9 and 10.

aprt gene and an unrelated sequence, Zhou *et al.* (1997) found that the breakpoints in both parental sequences contained strong sites for amsacrine-stimulated cleavage by TopoII. TopoII inhibitors

stabilize the TopoII–DNA complex and prevent re-ligation, and are used as anticancer drugs. However, on occasion, secondary malignancies with chromosome rearrangements develop following treatment

with TopoII inhibitors (Zhou *et al.*, 1997). Although Kudo *et al.* (1998) did not find TopoII sites near breakpoints associated with a translocation in cells from etoposide (TopoII inhibitor) related leukaemia, they proposed that chromatin structure might be more important than specific consensus sequences in the distribution of breakpoints. The sequence makeup of most zones flanking inversions in the *C. elegans* histone gene clusters is dissimilar in base order but shows perfect or near perfect identity in A + T content. The exact effect of this base composition phenomenon is not known. Because A tracts are capable of strong base stacking and confer rigidity to the DNA strand (Nelson *et al.*, 1987), DNA conformation near the sites of duplicate insertions might have been influenced by the particular 'moulding' effect of the A + T makeup in the local neighbourhood.

Sperry *et al.* (1989) found MARs (matrix association regions) to be at genomic sites of insertion, deletion and translocation. They also found that a strong *in vitro* cleavage site of TopoII corresponded with a translocation breakpoint in an MAR of the mouse immunoglobulin κ -chain gene. With the exception of the promoter regions (zones H and dt) in the ZK131–F08G2 gene cluster, the breakpoints found in the three histone gene clusters occurred in A + T enriched intergenic regions that contained sites with similarity to the *Drosophila* TopoII consensus for cleavage. Sites with similarity to the *Drosophila* consensus for TopoII cleavage (Sander & Hsieh, 1985) have been located in SARs or MARs in numerous species, including human (Jarman & Higgs, 1988), mouse (Cockerill & Garrard, 1988), plant (Slatter *et al.*, 1992) and *Drosophila* (Gasser & Laemmli, 1986a). From our results with the *Drosophila* consensus, evidence for possible involvement of TopoII in chromosome rearrangements in *C. elegans* is apparent. In addition, intergenic A + T rich regions with sites similar to the TopoII consensus, along with other SAR-associated motifs, might contain candidate SARs that occur as functional or boundary elements between sets of histone gene clusters (e.g. zones C and H in the F35H10–F17E9 overlap (Fig. 1c)).

In conclusion, evidence from nucleotide motif fingerprinting of three core histone gene clusters in the genome of *C. elegans* supports the occurrence of multiple alterations within the gene clusters; that is, there has been non-tandem duplication with inversion (F35H10–F17E9), tandem duplication without inversion (ZK131) and tandem duplication with inversion (K06C4 and ZK131–F08G2). The sequence makeup in proposed breakpoint regions indicates the potential involvement of TopoII in the rearrangements and a correlation between the A + T base composition at insertion sites and the orientation of duplicated segments.

We acknowledge funding from the Foundation for Research and Development, Pretoria, South Africa and the University of Natal Research Fund, Pietermaritzburg, South Africa. We extend thanks to C. Reid and R. Durbin for assistance with ACeDB installation and operation. Access to general resources at the University of Arkansas at Fayetteville and the University of Missouri at Columbia, USA, during the manuscript preparation is appreciated.

References

- Adachi, Y., Käs, E. & Laemmli, U. K. (1989). Preferential, cooperative binding of DNA topoisomerase II to scaffold-associated regions. *The EMBO Journal* **8**, 3997–4006.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- Bae, Y.-S., Kawasaki, I., Ikeda, H. & Liu, L. F. (1988). Illegitimate recombination mediated by calf thymus DNA topoisomerase II *in vitro*. *Proceedings of the National Academy of Sciences of the USA* **85**, 2076–2080.
- Berrios, M., Osheroff, N. & Fisher, P. A. (1985). *In situ* localization of DNA topoisomerase II, a major polypeptide component of the *Drosophila* nuclear matrix fraction. *Proceedings of the National Academy of Sciences of the USA* **82**, 4142–4146.
- Birchmeier, C., Schümperli, D., Sconzo, G. & Birnstiel, M. L. (1984). 3' editing of mRNAs: Sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. *Proceedings of the National Academy of Sciences of the USA* **81**, 1057–1061.
- Breyne, P., Van Montagu, M. & Gheysen, G. (1994). The role of scaffold attachment regions in the structural and functional organization of plant chromatin. *Transgenic Research* **3**, 195–202.
- Cockerill, P. N. & Garrard, W. T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* **44**, 273–282.
- DiNardo, S., Voelkel, K. & Sternglanz, R. (1984). DNA topoisomerase II mutant of *Saccharomyces cerevisiae*: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proceedings of the National Academy of Sciences of the USA* **81**, 2616–2620.
- Fajkus, J., Nicklas, J. A. & Hancock, R. (1998). DNA loop domains in a 1.4-Mb region around the human *hprt* gene mapped by cleavage mediated by nuclear matrix-associated topoisomerase II. *Molecular and General Genetics* **260**, 410–416.
- Gasser, S. M. & Laemmli, U. K. (1986a). The organisation of chromatin loops: characterization of a scaffold attachment site. *The EMBO Journal* **5**, 511–518.
- Gasser, S. M. & Laemmli, U. K. (1986b). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. *Cell* **46**, 521–530.
- Gasser, S. M., Laroche, T., Falquet, J., Boy de la Tour, E. & Laemmli, U. K. (1986). Metaphase chromosome structure. Involvement of topoisomerase II. *Journal of Molecular Biology* **188**, 613–629.
- Hentschel, C. C. & Birnstiel, M. L. (1981). The organization and expression of histone gene families. *Cell* **25**, 301–313.
- Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* **8**, 189–191.

- Holm, C. (1994). Coming undone: how to untangle a chromosome. *Cell* **77**, 955–957.
- Huang, X. (1994). On global sequence alignment. *CABIOS* **10**, 227–235.
- Jarman, A. P. & Higgs, D. R. (1988). Nuclear scaffold attachment sites in the human globin gene complexes. *The EMBO Journal* **7**, 3337–3344.
- Kanoe, H., Nakayama, T., Hosaka, T., Murakami, H., Yamamoto, H., Nakashima, Y., Tsuboyama, T., Nakamura, T., Ron, D., Sasaki, M. S. & Toguchida, J. (1999). Characteristics of genomic breakpoints in TLS-CHOP translocations in liposarcomas suggest the involvement of Translin and topoisomerase II in the process of translocation. *Oncogene* **18**, 721–729.
- Kingma, P. S., Greider, C. A. & Osheroff, N. (1997). Spontaneous DNA lesions poison human topoisomerase II α and stimulate cleavage proximal to leukemic 11q23 chromosomal breakpoints. *Biochemistry* **36**, 5934–5939.
- Kudo, K., Yoshida, H., Kiyoi, H., Numata, S., Horibe, K. & Naoe, T. (1998). Etoposide-related acute promyelocytic leukemia. *Leukemia* **12**, 1171–1175.
- Laemmli, U. K., Käs, E., Poljak, L. & Adachi, Y. (1992). Scaffold-associated regions: *cis*-acting determinants of chromatin structural loops and functional domains. *Current Opinion in Genetics and Development* **2**, 275–285.
- Mirkovitch, J., Mirault, M. E. & Laemmli, U. K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39**, 223–232.
- Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**, 443–453.
- Negrini, M., Felix, C. A., Martin, C., Lange, B. J., Nakamura, T., Canaani, E. & Croce, C. M. (1993). Potential topoisomerase II DNA-binding sites at the breakpoints of a t(9;11) chromosome translocation in acute myeloid leukemia. *Cancer Research* **53**, 4489–4492.
- Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). The structure of an oligo(dA)·oligo(dT) tract and its biological implications. *Nature* **330**, 221–226.
- Pikaard, C. S. (1998). Chromosome topology – organizing genes by loops and bounds. *Plant Cell* **10**, 1229–1232.
- Portmann, R., Schaffner, W. & Birnstiel, M. (1976). Partial denaturation mapping of cloned histone DNA from the sea urchin *Psammechinus miliaris*. *Nature* **264**, 31–34.
- Roberts, S. B., Sanicola, M., Emmons, S. W. & Childs, G. (1987). Molecular characterization of the histone gene family of *Caenorhabditis elegans*. *Journal of Molecular Biology* **196**, 27–38.
- Roberts, S. B., Emmons, S. W. & Childs, G. (1989). Nucleotide sequences of *Caenorhabditis elegans* core histone genes, genes for different histone classes share common flanking sequence elements. *Journal of Molecular Biology* **206**, 567–577.
- Sander, M. & Hsieh, T. (1983). Double strand DNA cleavage by type II DNA topoisomerase from *Drosophila melanogaster*. *Journal of Biological Chemistry* **258**, 8421–8428.
- Sander, M. & Hsieh, T. (1985). *Drosophila* topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. *Nucleic Acids Research* **13**, 1057–1072.
- Slatter, R. E., Dupree, P. & Gray, J. C. (1991). A scaffold-associated DNA region is located downstream of the pea plastocyanin gene. *The Plant Cell* **3**, 1239–1250.
- Sperry, A. O., Blasquez, V. C. & Garrard, W. T. (1989). Dysfunction of chromosomal loop attachment sites: illegitimate recombination linked to matrix association regions and topoisomerase II. *Proceedings of the National Academy of Sciences of the USA* **86**, 5497–5501.
- Streit, A., Koning, T. W., Soldati, D., Melin, L. & Schümperli, D. (1993). Variable effects of the conserved RNA hairpin element upon 3' end processing of histone pre-mRNA *in vitro*. *Nucleic Acids Research* **21**, 1569–1575.
- The *C. elegans* Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018.
- Vanfleteren, J. R., Van Bun, S. M., Delcambe, L. L. & Van Beeumen, J. J. (1986). Multiple forms of histone H2B from the nematode *Caenorhabditis elegans*. *Biochemical Journal* **235**, 769–773.
- Wagner, R. P., Maguire, M. P. & Stallings, R. L. (1993). Variations in chromatin organization and amount. In *Chromosomes, A Synthesis*, pp. 183–287. New York: Wiley-Liss.
- Wang, J. C., Caron, P. R. & Kim, R. A. (1990). The role of DNA topoisomerases in recombination and genome stability: a double-edged sword? *Cell* **62**, 403–406.
- Zhou, R.-H., Wang, P., Zou, Y., Jackson-Cook, C. K. & Povirk, L. F. (1997). A precise interchromosomal reciprocal exchange between hot spots for cleavable complex formation by topoisomerase II in amsacrine-treated Chinese hamster ovary cells. *Cancer Research* **57**, 4699–4702.