An inverted sea urchin histone gene sequence with breakpoints between TATA boxes and mRNA cap sites

Luigi Vitelli* and Eric S. Weinberg

Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 11 November 1982; Revised and Accepted 2 March 1983

ABSTRACT

A sea urchin histone gene fragment containing inverted regions of the normal repeat has been cloned in pBR322. Restriction enzyme mapping and homoduplex analysis of this fragment indicate that the H1-H4 spacer of one repeat is situated alongside the inverted H2A-H1 spacer of another repeat. The site of the breakpoint has been sequenced and compared with homologous stretches of the normal repeat. The breakpoints in the original duplexes were found to be within 4-6 bp of the H4 mRNA cap site and 8-10 bp of the H1 mRNA cap site in the standard repeat. The breakpoints in both original duplexes contain short direct repeats and an overlap of 3 bases. As this is the first breakpoint resulting in the apposition of inverted sequence to be analyzed at the level of DNA sequence, we speculate whether structural features described here are typical of such rearrangements. The structure observed is consistent with, but does not prove, that the sequence is the endpoint of a true inversion since only one junction has been isolated and characterized.

INTRODUCTION

Gene rearrangements in eukaryotic genomes have been known to occur ever since the initial recognition of chromosomal inversions and translocations. Early analysis of chromosomal rearrangements showed that these events involve non-homologous (illegitimate) recombination, at least when observed at the chromosomal level (1). It is of interest to find out whether naturally occurring inversions, translocations, and deletions generally involve localized sequence homology, contain transposon-like elements, or have some other common feature. No information is available at this time about base sequence at the junctions of naturally occurring inversions. We have found a sea urchin histone gene fragment which contains an oppositely oriented sequence and report here the structure and sequence of the inversion site.

The genes coding for the early embryonic histones of the sea urchin, *Strongylocentrotus purpuratus*, are contained in a 6-7 kb structure which is tandemly repeated several hundred times per genome (2-4). Within any one individual sea urchin of this species, there are a limited number

of major repeat types (most often 2) which differ slightly in the size or sequence of the spacer regions that occur between the histone coding sequences (5). Most of the major repeats are organized in homologous clusters since partial restriction enzyme digestions reveal little or no interspersions of the major repeats of different size (6). The histone gene fragment which we will describe here is not a typical sea urchin histone gene repeat. In addition to the major repeats revealed by blot hybridizations, many other histone DNA fragments are seen. Some of these may be parts of the repeat which have become situated away from the clustered units (7), some may be cross-reacting sequences coding for histones expressed at different developmental stages (8-10), and some may represent the ends of a cluster of repeats or some other organization. The variety of these fragments is only now just being explored. We present here, an analysis of one such fragment which consists of oppositely oriented histone repeat sequences. Comparison of this inverted sequence with analogous sequences in a normal repeat reveals several structural features which might be typical for points of inversion.

MATERIALS AND METHODS

Preparation, restriction enzyme digestion, and electrophoresis of sea urchin DNA

DNA from the sperm of a single sea urchin individual, which we denoted as 7N, (obtained from Peninsula Marine Biologicals, Sand City, CA) was prepared by slight modifications of the methods of Stafford and Bieber (11) and Joseph and Stafford (12). Restriction enzymes, purchased from New England Biolabs or Bethesda Research Laboratories, were used as recommended by the supplier. For sequential enzyme digestions, DNA fragments, electrophoresed on low melting agarose (BRL), were eluted and digested with a second restriction enzyme. Agarose gel electrophoresis, filter blotting, and hybridization were done essentially as previously described (5).

Cloning of restriction enzyme fragments, and preparation of hybridization probes

DNA from the sea urchin, 7N, was digested with Bam HI and ligated to Bam HI - digested pBR322. Cloning with pBR322 in an E. coli HB101 host followed standard procedures (13) and histone DNA-containing colonies were identified by hybridization to filter blots (14) with a full length nick-translated (15) S. purpuratus histone repeat fragment

(prepared from a lambda phage recombinant, λ Sp his-1, available in our laboratory). Preparations of small quantities of DNA from the positive clones were made by the method of Birnboim and Doly (16). In addition to the full-length histone repeat hybridization probes prepared from the lambda phage recombinant or from pC02A (the 6.3kb Hind III full length repeat fragment from pC02 (5) recloned into pBR322), several subcloned DNAs, diagrammed in Figure 1, were used. The H4 and H1 subclones were respectively, pHR15 and pHR17, the 1.0 kb Hha I and 1.7 kb Hha I + Eco RI fragments of pC02 cloned into the Pst I site of pBR322 by G-C tailing. These subclones contain the genes themselves and flanking 5' and 3' spacer sequences. The H2A probe was pRC9, a gift of Ron Cohn and Larry Kedes, which has the Eco RI-Hind III fragment containing the 3' half of the H2A gene and some spacer 3' to the gene (4). The H2B probe, C02M2B1, was the 1.20 kb Bam HI fragment of pC02 cloned in the Bam HI site of M13. The H3-containing subclone, C02M31, was a blunt-ended 0.50 kb Hha I fragment of pC02 cloned into the Hinc II site of M13. The H2B fragment contained some H2B-H3 spacer in addition to the H2B gene. The H3 subclone contained almost the complete coding sequence and a small amount of 5' spacer. The relationship of these fragments to the map of the repeat can be deduced from the map shown in Figure 1. Labeled probes from the plasmid clones were prepared by nick-translation (15), and single stranded probes were prepared from M13 clones by synthesis on primed single stranded phage.

Electron microscopy

7N32L DNA, digested with Sal I, was denatured in 0.1 M NaOH, neutralized in 0.2 M Tris HCl (pH 7.2), the solution adjusted to be 50% formamide. Spreading in a 10% formamide hypophase, staining, and preparation of the sample followed the procedures of Davis et al. (17) except that EDTA was omitted from the spreading solution in the hypophase. Single stranded and replicative form ϕ X174 DNA were added as standards for measuring single and double stranded DNA.

DNA Sequencing

DNA fragments, isolated from polyacrylamide gels by electroelution were purified on small columns of Whatman DE53 cellulose. The ends of fragments were either labeled using the Klenow fragment of DNA polymerase or with polynucleotide kinase. Slight modifications of the methods of Maxam and Gilbert (18) were used. The strategy used to sequence 7N32L (containing the inverted region) can be followed on Figure 1. The plasmid was cut at the single Xho I site (with Ava I), labeled with [32 P]-dCTP

Figure 1. Sequence organization of histone gene fragments. Top: The restriction map of pC02, a full length repeat cloned into the Hind III site of pBR322 (5). Middle: restriction map of 7N44, a 5.3 kb Bam HI fragment of the larger standard repeat of individual 7N, inserted into the Bam HI site of pBR322. Bottom: Restriction map of 7N32L, a 5.8 kb Bam HI fragment containing the inverted sequence, inserted into the Bam HI site of pBR322. Subclones used in the hybridizations are indicated directly below the pC02 map. The asterisk indicates the position of the 200 bp which are absent in one of the 7N major repeat types. The regions sequenced are indicated by horizontal arrows below the 7N44 and 7N32L maps (A,B,C,D refers to the strands of Figure 6). Cross hatched areas are pBR322 sequences and solid dark areas are histone mRNA coding regions. Restriction sites are as indicated by the key.

using the Klenow fragment, and digested with Eco RI. The strand sequenced was 3'→5', away from the Xho I site towards the right side of the structure as drawn in Figure 1. The H1-H4 spacer region of 7N44 was sequenced by cutting the plasmid at the single Xho I site (with Ava I), labeling with [³²P]-dCTP with the Klenow fragment, cutting with Sal I, and isolating the fragment containing the 5' end of the H4 gene. The strand sequenced was 3'→5', away from the Xho I site towards the H4 gene. The other strand of this same fragment was also sequenced after labeling with γ-³²P-ATP using T4 kinase. The H2A-H1 spacer region of 7N44 was sequenced by labeling the Sma I - Xho I fragment containing the H1 gene with γ-[³²P]-ATP and cutting with Pst I. The fragment was sequenced 5'→3' away from the Sma I site towards the Pst I site as shown in Figure 1.

RESULTS

Initial characterization of cloned DNA of individual 7N

The two major histone repeat types in the individual sea urchin, 7N, are approximately 6.2 and 6.5 kb. As seen in Figure 2, although these are the sizes of the major fragments when the repeats are cut once with Hind III, many other histone DNA sequences can be detected in the genome. Bam HI cuts the major repeats into two fragments: the smaller 1.2 kb fragment contains the H2B gene; the larger fragment, containing the other four histone genes, is either 5.0 or 5.3 kb depending on the major repeat type. The difference in size of the two major repeats is mainly found within the H4-H2B spacer region. Although the smaller repeat is more typical of the species, the larger repeat is a fairly common polymorphism in the sea urchin population (5,6,19). In the course of isolating the two large Bam HI fragments, several pieces of unexpected size were cloned in pBR322. They correspond in size to some of the faint bands seen in the genomic blot of Figure 2B, but most have not been characterized. One of these cloned DNA segments, 7N32L, however, was quite unusual and will be described here.

The initial restriction analysis of the 7N32L DNA and the more common major repeat fragments is presented in Figure 3. The Bam HI products, hybridized with a complete histone gene repeat, are shown in Figure 3A. Clone 7N44 contains a 5.3 kb Bam HI fragment and clone 7N65 contains a 5.0 kb fragment, the sizes expected from the major repeat classes (Figure 2A). However, the histone DNA product of Bam HI digestion of 7N32L was a larger 5.8 kb in size. Figure 3B presents the restriction analysis of Bam HI + Hind III and Bam HI + Hha I digestions. Clones 7N44 and 7N16 gave the fragments expected for a 6.5 kb major repeat and clones 7N65 and 7N24 yielded fragments expected for a 6.2 kb major repeat. The fragment sizes agree with the restriction map of pC02 (5) as presented in Figure 1, except that one repeat type contains the extra 200 nucleotides in the H4-H2B spacer region. (When comparing the Bam HI + Hha I digests of the two major repeat types, the position of the extra 200 nucleotides must be considered. The 0.73 kb fragment of the large repeat type is analogous to the 0.52 fragment of the smaller repeat type. This fragment is not seen since it comigrates with the 0.50 kb H3-containing fragment present in all clones.) The digestions of clone 7N32L, however, yielded somewhat different patterns. In the Bam HI + Hind III digest, a 2.9 kb fragment (rather than the usual 3.6 or 3.9 kb) was seen in addition to the expected

Figure 2. Southern blot of genomic DNA from a single sea urchin, 7N, hybridized with a histone gene probe. DNA was digested with Bam HI, Bam HI and Sal I, or Hind III and electrophoresed on a 1.2% agarose gel. After transfer to nitrocellulose paper, the DNA was hybridized with λ Sp his-1 DNA, (a complete *S. purpuratus* histone gene repeat cloned in lambda phage). A) A short exposure visualizing the fragments of the major repeats. B) A long exposure of the same filter to show the rarer fragments. *—fragments which hybridized to an H1 probe but not an H4 probe. • fragments which hybridized to the H4 probe but not the H1 probe.

1.5 kb band. In the Bam HI + Hha I digest, most of the fragments migrated as in the usual repeat, but one fragment was a slightly larger 1.1 kb instead of 1.0 kb. In both the Bam HI + Hha I and Bam HI + Hind III digests of 7N32L, the fragment sizes totaled up to 4.4 kb instead of the expected 5.8 kb.

The Bam HI + Hha I fragments of clone 7N32L were then further analyzed by hybridizing blots with labeled DNA from subclones of the repeat, specific for each of the 5 genes (descriptions of the subclones are given in Materials and Methods). As expected, the H3 subclone hybridized to the 0.50 kb fragment, the H1 subclone hybridized to the 2.1 and 1.1 kb fragments (this subclone contains flanking H1-H2A and H4-H1 spacers), and the H2B subclone did not hybridize at all (the H2B gene is within the 1.2 kb Bam fragment not contained within the 7N32L clone). Surprisingly, however, the H4 subclone hybridized only very weakly to the

Figure 3. Restriction enzyme digests of cloned histone DNA. DNA from individual 7N was digested with Bam HI and cloned into the pBR322 site by procedures cited in Materials and Methods. Small plasmid DNA preparations were made from 5 positive clones and digested with Bam HI to release the insert. Panel A shows the Southern blot of 3 clones after Bam HI digestion. In panel B, these 3 cloned DNAs and 2 additional ones were cut with either Hha I (a) or Hind III (b) after Bam HI digestion. The filters were hybridized with the full length repeat (λ Sp his-1). The samples marked 44 and 16 are cloned from the larger major repeat type, 24 and 65 are from the smaller major repeat, and 32 is the unusual clone.

1.1 kb fragment in contrast to the usual strong hybridization with the 1.0 kb fragment of the more conventional repeats. The 1.1 kb fragment did, however, hybridize with a subclone containing the H2A sequence. We therefore decided to study the 7N32L fragment in more detail since at least part of the sequence organization was quite different from the usual repeat.

Restriction enzyme mapping shows symmetry

A detailed restriction map of the cloned fragment was obtained from the results of double and triple restriction enzyme digestions and by sequential digestions of fragments prepared from low-melting agarose gels. As shown in the map of 7N32L in Figure 4, a large area containing symmetrical enzyme sites was found. At first, this was not apparent from the hybridized blots since each of many of the digestions yielded two

Figure 4. Restriction enzyme map of a typical histone gene repeat oriented in both directions compared with a restriction map of 7N32L. The enzyme sites of the repeat in pC02 were used for the map of the normal gene (6). Short arrows indicate the direction of transcription. Long arrows under the map of 7N32L indicate the regions of symmetry in this clone. The dashed line shows the point of crossover. Restriction enzymes used are represented by the key at the bottom of the figure. Thickened lines map the transcribed genes.

identical fragments from the cloned DNA. The first indication of the actual map was provided by comparing the Hha I digest of 7N32L with the Hha I + Bam HI digest. In the latter case, a 0.50 kb fragment hybridized to the H3 subclone; in the former case, two bands were seen at 0.57 kb and 0.74 kb. The asymmetry of Hha I sites around the Bam HI site of pBR322 allowed the discrimination of the two end fragments of the insert, both of which contained the H3 gene. Fragments from a Bam HI + Xho I digest and a Bam HI + Hind III digest of 7N32L were obtained from low-melting agarose gels, cut with the enzymes shown in Figure 4, resolved by electrophoresis, and then hybridized with the various subclones. The unambiguous map shown in Figure 4 was obtained. The very faint hybridization of the H4 subclone to 7N32L DNA mentioned above was apparently due to the small portion of the H4-H1 spacer contained within the H4 subclone used (see Figure 1). From the observation that the restriction enzyme sites of 7N32L correspond exactly to two regions of the map of the standard repeat, pC02, we tentatively concluded that the clone contained an oppositely oriented histone repeat sequence. The H1-H4 spacer of one repeat was now apposed to a region of the H2A-H1 spacer of

an inverted repeat.

Direct proof of an inverted sequence

If the model implied by the restriction enzyme map were correct, histone coding regions should be found on both DNA strands. A single stranded probe for the H3 region was prepared from a fragment subcloned in M13. The Bam HI-digested 7N32L DNA was strand-separated by treatment with NaOH and electrophoresed in an agarose gel. Although the single strands of the histone gene fragment of the clone were difficult to resolve on the gel, two closely migrating bands could be seen (not shown). Both of the bands hybridized with the single strand M13 probe, indicating that two H3 regions oriented in opposite directions, were present in the cloned fragment.

The DNA was then examined by electron microscopy, which provided a much more convincing corroboration of our map. The inverted regions were expected to snap back after denaturation of the cloned DNA. The schematic description of the protocol, and the results obtained, are shown in Figure 5. The 7N32L DNA was cut with Sal I, which cuts once in pBR322 275 bp from the Bam HI site, but does not cut within the inserted histone DNA sequence. As illustrated, a homoduplex should yield unpaired single strands of pBR322 of 0.27 kb and 4.1 kb, a single strand loop of approximately 1.3 kb and a duplex region corresponding to the area of symmetry of 2.2 kb. This is exactly what is seen in the micrographs of Figure 5, confirming the organization predicted by the restriction enzyme analysis.

Site and sequence of the breakpoint

The restriction enzyme analysis of the insert indicates a normal repeat organization, reading from left to right in Figure 4, until the Hha I site just 5' to the H4 gene is reached. The next site found in 7N32L is a Sma I site, 165 bp to the right of the Hha I site. This site does not align with the standard map in the same orientation but does align with the Sma I site just 5' to the H1 gene of the normal repeat if the normal repeat is compared in an inverted orientation. The breakpoint is therefore between the Hha I and Sma I sites of 7N32L, the H1-H4 spacer now situated alongside an inverted H2A-H1 spacer sequence.

The exact regions of homology were determined by sequencing parts of 7N32L and the more usual repeat, 7N44 DNA. Regions of sequence just 5' to the H4 and H1 genes and within the H1 gene from another standard clone, pSp2, were also available for comparison (20-22). Figure 6 shows the

Figure 5. Homoduplex analyses of 7N32L DNA. The intact plasmid DNA was digested with Sal I and then denatured in 0.1 N NaOH. After 10 min at room temperature, the sample was neutralized with a small volume of 2 M Tris buffer (pH 7.2), made 50% in formamide and then spread for electron microscopy. The structures seen are illustrated in the left panels; and the scheme of snapback is shown in the right panel. Restriction enzyme sites are indicated for reference to the organization shown in Figure 4.

sequence of 7N32L around the point of inversion compared to two regions of the normal repeat, 7N44. The sequence of both strands of 7N32L is shown; these strands are marked b and c in the figure. Directly above and below are sequences of two regions of the sense strand of 7N44. The sense strand of the sequence 5' to the H4 gene is marked a and that 5' to the H1 gene is marked d. These two sequences are contiguous in the original repeat since all 5 histones are coded for on the same strand. The sense strand in the 7N44 H4 region (H1-H4 spacer), is aligned 5'→3', whereas the sense strand in the 7N44 H2A-H1 spacer is aligned in the opposite 3'←5' direction. As seen, the 5'→3' strand of 7N32L matches exactly with the 5'→3' sense strand of the H1-H4 spacer of 7N44 for 180 bp until homology is abruptly lost. The other strand of 7N32L is homologous to the 7N44 sense strand of the H2A-H1 spacer when aligned in the opposite direction (3'←5'). The homology extends from the H2A spacer into the H1 region until homology is lost at position 0 in Figure 6. There is only a 3 nucleotide overlap of sequence in these two regions:

Figure 6. Sequence of 7N32L compared to that of normal repeats. The top line, is the sequence of the sense strand of 7N44 just proximal to the H4 gene (strand a). The middle two lines are the sequence of the 7N32L duplex (strands b and c). The bottom line is the sequence of the sense strand of 7N44 just proximal to the H1 gene and including part of the H1 coding region (strand d). The orientations of the two sense strands, a and d, are reversed. Homology is indicated by the vertical lines between the sequences. Bases 0-3 are boxed and comprise the overlap sequence. The cap site is denoted by * and direction of transcription of the mRNAs is indicated. TATA boxes are boldly underlined. The direct repeats at the overlap region are lightly underlined. Indicated directly above strand a and below strand d are the only differences in sequence between pSp2 and 7N44 in this region. A base next to an arrowhead indicates a single base insertion. Hyphens indicate deletions, and the other changes are base substitutions.

Figure 7. Summary of features of the inverted sequence of 7N32L. The symbols of restriction enzyme sites are as indicated in Figure 4.

at the 5'-GGC-3' of the H4 region and the 3'-CCG-5' of the H1 region.

The crossover point in each region is reached at a surprising location between the TATA box and cap site of the adjacent downstream gene. In 7N44 the TATA box of the H4 gene is reached at positions -27 to -20, the 5' end of the H4 mRNA at +6 and the ATG initiation codon would be at +73 to +75 (not shown). Homology is lost at position 3, 3 nucleotides before the H4 cap site. Reading in the opposite direction, the H1 TATA box is found at +22 to +17, the cap site at -8 and the ATG initiation codon at -46 to -48. Homology is lost at position -1, 7 nucleotides from the H1 cap site. The resulting structure of 7N32L is diagrammed in Figure 7; two oppositely situated different spacers are connected several nucleotides downstream from the TATA boxes of what would have been the next genes.

Since inverted segments have not been previously analyzed at the level of DNA sequence, it is interesting to note if there are any distinctive features in the original or rearranged sequences. The most impressive similarity between the site of interchange is that it involves the region between the TATA box and mRNA cap site in both original sequences. As drawn, if the crossover point occurred between -1 and 0, the junction would be 7 or 8 bp from the two cap sites, and about 19 bp from the TATA boxes, depending on what site is chosen as the TATA box. The identity of position in the two regions is very striking.

Another feature of the crossover site is that it contains direct repeats in both original sequences. As seen in Figure 6, the C at position 2 in the H4 region (strand a) is part of the direct repeat 5-CATCAT-3'. The CCG in the H1 region (strand d) is part of the direct repeat 3'-AGCCGAGCCG-5'. This position is also distinctive because it contains the symmetrical sequences AGCCGA and GCCGAGCCG.

Although homology in sites of duplication and deletion is usually between strands of similar polarity when the original sequences are aligned in the same direction, inversions might involve homology between

2147

right of the breakpoint. In addition, there are 4 other 4 base homologies which align: TGGG (#1), AGTC (#2), TCTC (#3), and CGCC (#5). Whether this level of homology is significant for the duplex interaction will probably only be evident when more sequences involved in inversions are obtained.

Short misalignments and lack of pairing such as found in sequences #4 and #7 are probably not due to polymorphisms since the sequences in 7N32L are the same in this region as in 7N44. There is polymorphic variation, however, between sequences from different sea urchins as shown in Figure 6. The differences in sequence between 7N44 and pSp2 in the two regions we compared are shown by the extra bases just above strand a and below strand d in the figure. The 11 changes seen in the 339 bases compared are typical of the level of polymorphism between spacers in repeats from different individuals (19). We do not think that the A at position 3 of the H4 sense strand of 7N44 is a polymorphic variation found only in 7N44 since it is also found in pSp2. The overlap of sequences is therefore truly only 3 bases and not a 5 bp sequence which would result from the absence of A in the H4 region of the parental duplex.

The two original sequences at the sites of interchange, therefore, share very similar features: they contain a small overlap region which can be drawn as GGC or GGC-TC, they contain additional small regions of homology, both sites involve small direct repeats, and the event occurs in the region situated between the TATA box and cap site in each region.

DISCUSSION

Direct repeats and homologies in the rearrangement

The nature of sequences involved in rearrangements such as inversions and translocations is still very much unknown. The original duplexes might recombine completely at random, or there might be some particular feature common to these events. Although studies of inversions at the level of chromosome pairing reveal little direct homology in the interacting chromatids, regions of homology on the order of thousands of kb would not have been evident. Inversions might be site specific in that they involve small regions of homology or they might be related to some other feature such as proximity to a transposon or to some functional state of the chromatin. Except for events mediated by transposons, there is no information on the DNA sequences involved at the breakpoints of inversions. The results described above for the 7N32L sequence indicate several interesting features at the site of apposition. These

characteristics are similar to those involved in certain previously analyzed events of rearrangements in prokaryotes.

Several studies of prokaryotic insertions and deletions implicate the role of direct repeats in rearrangements. The sequence at the hotspot of mutation in the lac I gene is the repeated sequence (CTGG)₃ (23). The points of crossover in 7N44 also contain small direct repeats. These sequences might create an easily perturbed site as does the short repeat in the lac I gene.

A careful study of the DNA sequences at the border of deletions of the lac I gene shows that in virtually every case, the deleted segment is situated between what were direct repeats in the original sequence (24). The most common site of deletion is between sequences of 17 bp which share 14 bp of homology, while occasional deletions were found between direct repeats sharing as few as 4 of 5 or 5 of 6 bp (24). The finding of a very small sequence overlap between the two interacting regions in the normal sea urchin histone gene repeats is very much like the site specific nature of the lac I deletions, except that homology is not on the same DNA strand. The short region of homology is found on the 5'→3' sense strand of one duplex and the 5'→3' anti-sense strand of the other duplex when both are read from left to right as shown in Figure 8. In this sense, the two duplexes would interact in the same way as they might in a deletion, except that they would have been arranged alongside each other with histone coding regions oriented in opposite directions. In addition to the small homology at the crossover site, there is additional complementarity in sequence along the interacting duplexes. These sequences are in fact longer than the 3-5 base region at the crossover site and might be involved in some initial recognition of the interacting duplexes. What is certain from our result, however, is that a long region of extensive homology in proximity to the crossover site is not required for the inversion to occur. Both the short homologies and the repeated sequences at each site would make an interaction more probable, by analogy to mutations in the lac I gene (23, 24).

Position of rearrangement between TATA and cap sites

The site of rearrangement we describe occurs several bases proximal to the cap sites of two different histone genes. It is perhaps noteworthy to mention that IS1-related deletions have breakpoints in the host DNA which are identical or similar in sequence to promoter regions (25). Ohtsubo et al. have postulated that RNA polymerase could play an important

role in site-specific recombination phenomena (25). The specific recombination site within the $\gamma\delta$ transposable element is also at a promoter region which is specifically recognized by a resolvase protein (26). This 19 bp region contains promoters for two genes oriented in opposite directions in this case. We might speculate that, in general, proteins such as RNA polymerase provide a DNA structure susceptible to recombination events in eukaryotes as well, and that these events result in rare gene rearrangements. The fact that the breakpoints in both original sequences of 7N44 are found virtually in the same position relative to the TATA sequences is probably not coincidental. Random interactions between short homologous sequences would occur at many places along the oppositely oriented duplexes and it is highly unlikely that the two promoter areas would both be involved if homology alone determined the breakpoint. The promoter regions of the two duplexes might be unwound so that random base pairing could occur between the strands. Although there is evidence that transposon-like elements are involved in deletions and insertions in Drosophila melanogaster (27-32), the basis of many eukaryotic rearrangements may, in fact, be the random apposition of regions of small homology which have particular secondary structure due to interaction with protein.

Origin of the inverted sequence

We must, of course, consider the possibility that the rearrangement we postulate occurred in E. coli during or after cloning, and not within the sea urchin genome. Several lines of evidence support, but do not prove, that this structure occurs in the sea urchin DNA. First, the Southern blots of DNA from individual 7N (Figure 2) do reveal bands in the regions expected for restriction enzyme digestion of the 7N32L sequence. The putative fragments hybridize with the full length histone repeat probe or with an H1 subclone but, as expected, not with a probe containing the H4 gene region. Second, preliminary results indicate that there may be other specific histone fragments which can be isolated from the DNA of individual sea urchins by first purifying a snap-back DNA fraction after denaturation, although the fragment studied here is unique to one individual. Third, of over 5000 colonies containing plasmids with inserts that were screened in the original cloning procedure, only 7 contained histone DNA sequence. It is very unlikely that one of these positive clones contained two independently derived histone DNA Bam HI fragments. Last, when comparing the sequence of the two H2A-H1 spacers of 7N32L (data

not shown), a number of base differences were seen. The two parts of 7N32L, therefore, probably did not arise by recombination of amplified copies of a single repeat. Although the above reasons make it unlikely that the 7N32L structure was generated during cloning, a completely unambiguous determination would be possible only by independently recloning the sequence from 7N genomic DNA.

The sequences of 7N32L at the point of crossover match exactly with the sequence of the normal clone, 7N44. The 7N44 repeat is the larger type of the two common repeats of 7N44. Although we do not have sequence information from the smaller common repeat from individual 7N, we can compare our sequence with the previously published sequence of the same regions of another repeat, in the clone pSp2 (20-21). In the H4 region shown in Figure 6, pSp2 differs at 9 positions from 7N44 (5 base substitutions and 4 base additions). In the H1 region there is only 1 base change and an addition/deletion of 2 bp. This degree of divergence (3%) is typical of other spacer regions when repeats from different individuals are compared (19). Since 7N44 and 7N32L exactly match in this region it is likely that at least part of 7N32L is derived from the 7N cluster containing the 7N44-type repeat. However, only additional sequence analysis will reveal what divergence actually exists between nearest neighboring repeats within a cluster.

There is no evidence that the 7N32L sequence is an "orphon" or dispersed repeat. Whereas the sequence of an orphon is found to differ from the normal repeat (7), the sequence of 7N32L in the region of the inversion matches that of 7N44 exactly. The boundary analyzed in 7N32L involves only sequence from the typical histone repeat and not from some other DNA region. Although a fragment larger than that cloned in 7N32L would have to be analyzed to determine the relationships of the sequence to other clustered histone genes, it is clear that the boundary of inversion in 7N32L is highly conserved histone DNA.

Although the term "inversion" has been used in this report, we should emphasize that only one boundary has been analyzed and a better term might be "inverted apposition". We have no information to indicate whether a segment actually inverted in the sense of an exact reciprocal reintegration resulting in no loss of genetic material. The results imply that there was at least a single recombination between adjacent and oppositely oriented histone gene repeats, an event which would be sufficient to obtain the translocation of an inverted segment. It is also

important to note that the inverted sequence might not be the result of a true recombination event but rather of a conversion event.

Finally, we would point out that events similar to those involved in forming the 7N32L structure must have also occurred during the evolution of the histone gene set in some organisms. Although the five histone genes are found oriented in the same direction in the sea urchin repeat, they are oriented in opposite orientations in repeats of Drosophila and some other species (33). In Drosophila (34), the H3 and H4 genes and also the H2A and H2B genes are oppositely oriented with closely spaced promoters. An event similar to that which resulted in 7N32L may have occurred; the inverted repeat might have been amplified by unequal crossing over events to yield a newly oriented repeat structure.

Whether the structure we have found is merely a curiosity or is typical of the site of other rearrangements will be determined when breakpoints of other natural inversions have been analyzed. It is suggested from this analysis that promoter regions are particularly susceptible to rearrangements and that the exact site of crossover may be determined by small stretches of localized homology and the presence of short direct repeats.

ACKNOWLEDGEMENTS

We thank Dr. Lawrence Yager for his help and advice on DNA sequencing procedures and Dr. Gary Kettner for helping us do the homoduplex mapping. Histone gene regions were subcloned by Lawrence Yager and Patricia Kuwabara in our laboratory. This work was supported by NIH grant 22155 to E.S.W.

*Present address of L.V.: Istituto di Istologia e Embriologia, via A. Volta, 4, 56100 Pisa, Italy

REFERENCES

1. Stutevant, A.H. and Beadle, G.W. (1940). An Introduction to Genetics (W.B. Saunders Company, Philadelphia and London), pp 127-191.
2. Kedes, L.H., Cohn, R.H., Chang, A.C.Y. and Cohen, S.N. (1975). Cell 6, 359-369.
3. Weinberg, E.S., Overton, G.C., Shutt, R.H. and Reeder, R.H. (1975). Proc. Natl. Acad. Sci. USA 72, 4815-4819.
4. Cohn, R.H., Lowry, J.C., and Kedes, L.H. (1976). Cell 9, 147-161.
5. Overton, G.C. and Weinberg, E.S. (1978). Cell 14, 247-257.
6. Overton, G.C. (1978). Thesis: An Analysis of the Structure and Heterogeneity of Histone Genes in the Sea Urchin, S. purpuratus (Johns Hopkins University).
7. Childs, G., Maxson, R., Cohn, R.H. and Kedes, L. (1981). Cell 23, 651-663.

8. Kunkel, N.S. and Weinberg, E.S. (1978). *Cell* 14, 313-326.
9. Grunstein, M. (1978). *Proc. Natl. Acad. Sci. USA* 75, 4135-4139.
10. Childs, G., Maxson, R. and Kedes, L.H. (1979). *Dev. Biol.* 73, 153-173
11. Stafford, D.W. and Bieber, D. (1975). *Biochim. Biophys. Acta* 378, 18-21.
12. Joseph, D.R. and Stafford, D.W. (1976). *Biochim. Biophys. Acta* 418, 167-174.
13. Davis, R.W., Botstein, D. and Roth, J.R. (1980). *A Manual for Genetic Engineering Advanced Bacterial Genetics*. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) pp. 138-141.
14. Grunstein, M. and Hogness, D.S. (1975). *Proc. Natl. Acad. Sci. USA* 72, 3961-3965.
15. Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975). *Proc. Natl. Acad. Sci. USA* 72, 1184-1188.
16. Birnboim, H.C. and Doly, J. (1979). *Nuc. Acids Res.* 7, 1513-1523.
17. Davis, R.W., Simon, M., and Davidson, N. (1971). *Meth. Enzymol.* 21, 413-427.
18. Maxam, A.H. and Gilbert, W. (1980). *Meth. Enzymol.* 65, 499-560.
19. Yager, L.H., Kaumeyer, J., Farrelly, L.L., and Weinberg, E.S. (1982). In preparation.
20. Sures, I., Levy, S., and Kedes, L.H. (1980). *Proc. Natl. Acad. Sci. USA* 77, 1265-1269.
21. Grunstein, M., Diamond, K.E., Knoppel, E. and Grunstein, J.E. (1981). *Biochemistry* 20, 1216-1223.
22. Levy, S., Sures, I., and Kedes, L.H., (1982). *J. Biol. Chem.* 257, 9438-9443.
23. Farabaugh, P.J., Schmeissner, U., Hofer, M., and Miller, J.H. (1978). *J. Mol. Biol.* 126, 847-857.
24. Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982). *Cell* 29, 319-328.
25. Ohtsubo, H. and Ohtsubo, E. (1978). *Proc. Natl. Acad. Sci. USA* 75, 615-619.
26. Grindley, N.D.F., Lauth, M.R., Wells, R.G., Wityk, R.J., Salvo, J.J. and Reed, R.R. (1982). *Cell* 30, 19-27.
27. Dunsmuir, P., Broein, W.J., Jr., Simon, M.A., and Rubin, G.M. (1980). *Cell* 21, 575-579.
28. Spradling, A.C. and Rubin, G.M. (1981). *Ann. Rev. Genet.* 15, 219-264.
29. Collins, M. and Rubin, G.M. (1982). *Cell* 30, 71-79.
30. Zachar, Z. and Bingham, P.M. (1982). *Cell* 30, 529-541.
31. Levis, R. and Rubin, G.M. (1982). *Cell* 30, 543-550.
32. Levis, R., Collins, M. and Rubin, G.M. (1982). *Cell* 30, 551-565.
33. Hentschel, C.C. and Birnstiel, M.L. (1981). *Cell* 25, 301-313.
34. Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). *Cold Spring Harbor Symp. Quant. Biol.* 42, 1047-1051.