Cell, Vol. 34, 25–35, August 1983, Copyright © 1983 by MIT 0092-8674/83/080025-11 $02.00/0

Structures of P Transposable Elements and Their Sites of Insertion and Excision in the Drosophila melanogaster Genome

Kevin O'Hare* and Gerald M. Rubin†
Department of Embryology
Carnegie Institution of Washington
115 West University Parkway
Baltimore, Maryland 21210

Summary

We have isolated and characterized several members of the P transposable element family from a Drosophila melanogaster P strain. Large 2.9 kb elements are present as multiple highly conserved copies together with smaller (0.5–1.6 kb), heterogeneous elements. The complete DNA sequences of the 2.9 kb element and four small elements (previously isolated from hybrid-dysgenesis-induced mutations of the *white* locus) have been determined. Each small element appears to have arisen from the 2.9 kb element by a different internal deletion. P elements have 31 bp perfect inverse terminal repeats and upon insertion duplicate an 8 bp sequence found only once at the site of insertion. Three of the insertions into the *white* locus occurred at the same nucleotide, indicating a high degree of local site specificity for insertion. The basis of this specificity has been investigated by DNA sequence analysis of the sites where 18 P elements are found. A revertant of one of the *white* locus mutants has been found to result from precise excision of the P element, restoring the wild-type DNA sequence.

Introduction

The phenomena of hybrid dysgenesis occur in certain interstrain crosses of Drosophila melanogaster. The germline cells of the progeny of such crosses develop aberrantly, resulting in reduced fertility as well as elevated rates of mutation, male recombination, and chromosome rearrangement (Kidwell, Kidwell, and Sved, 1977). Two independently acting systems of hybrid dysgenesis have been described, called I-R and P-M (Kidwell, 1979; Bregliano and Kidwell, 1983). In the P-M system, dysgenesis occurs when males of a P strain (paternally contributing) are mated with females of an M strain (maternally contributing) but not in the reciprocal cross (M♂ × P♀) nor in P♂ × P♀ or M♂ × M♀ crosses.

P strains have multiple copies of the genetic factors that lead to hybrid dysgenesis. These P factors are dispersed over all major chromosome arms (Engels, 1979a; Engels and Preston, 1980) and are absent from M strains. Studies of mutations of the *white* locus (Rubin et al., 1982; Bingham et al., 1982) support the hypothesis that P factors are transposable elements whose transposition is repressed in P strains but derepressed during hybrid dysgenesis. (The control of transposition is complex and depends on the interaction between nuclear and cytoplasmic determinants—see Discussion.) The sequences inserted in several hybrid-dysgenesis-induced mutations defined a family of cross-homologous DNA sequences, the P elements, which are present as 30–50 dispersed copies in P strains and (with one exception) absent in M strains.

The P elements isolated from these mutations were heterogeneous in size, ranging from 0.5 to 1.6 kb (Rubin et al., 1982). The absolute equivalence of such variable P elements with the genetically defined P factors is unlikely, and at least one M strain (hence lacking P factors by definition) has P elements (Bingham et al., 1982). We have isolated a larger conserved P element, which we believe to be the P factor, and have determined the relationship between it and the smaller P elements. In the experiments of Spradling and Rubin (1982), the biological activity of this large P element was tested by injection into developing M strain embryos. P factor functions were expressed, leading to transposition of the injected P element into germ-line cells of the embryo.

Results

Isolation of a Conserved 2.9 kb P Element

The construction of a library of random genomic DNA fragments from the P strain π_2 in the bacteriophage λ vector Charon 28 has been described (Rubin et al., 1982). This library was screened by using as probe a mixture of the fragments that contain P elements from the *white* locus mutants $w^{\#6}$ and $w^{\#12}$ (Rubin et al., 1982). Positively hybridizing phages were counterscreened with the same fragments from the wild-type *white* locus (Levis et al., 1982a). Phages containing P elements appeared to be present in the library at about 50 times the frequency of phages carrying nonrepetitive *white* locus sequences, consistent with P elements being present at about 50 copies per haploid genome (Bingham et al., 1982). Forty phages containing sequences homologous to P elements were purified. When tested for hybridization individually to the fragments from mutants $w^{\#6}$ and $w^{\#12}$ they all hybridized to both probes, although the intensity of hybridization varied considerably.

Twelve phages were selected at random and characterized by restriction enzyme mapping and DNA blotting by using the P elements from mutants $w^{\#6}$ and $w^{\#12}$ as probes. Four of the 12 phages (λπ12, λπ25, λπ30, and λπ44) carried both a 0.9 kb Hind III fragment and a 1.6 kb Hind III–Sal I fragment, which were contiguous and homologous to the probes (data not shown). The P elements found in the *white* locus of mutant $w^{\#6}$ and one of the 12 phages (λπ23) both appear (by the criteria of comigration and hybridization) to contain the 0.9 kb Hind III fragment but not the 1.6 kb Hind III–Sal I fragment.

Bam HI fragments covering these two fragments were subcloned from λπ12 and λπ25 into pBR322 (plasmids pπ12.20 and pπ25.1), and more detailed restriction en-

* Present address: Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, England.
† Present address: Department of Biochemistry, University of California at Berkeley, Berkeley, California 94720.

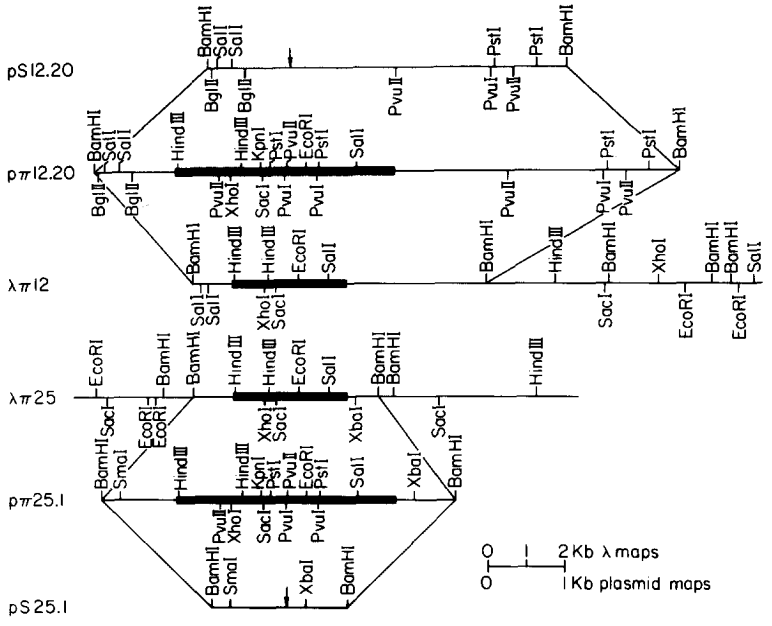

Figure 1. Restriction Site Maps of Cloned 2.9 kb P Elements and Their Sites of Insertion

Maps of two λ clones containing 2.9 kb P elements isolated from the strain π_2, λπ12 and λπ25, are shown. Heavy lines indicate P element sequences. More detailed maps of the Bam HI fragment subclones (plasmids pπ12.20 and pπ25.1) and the corresponding Bam HI fragments cloned from the M strain Canton S (pS12.20 and pS25.1) are given. Arrows indicate sites within the M strain DNA fragments where P elements are inserted in the P strain.

zyme maps were constructed (Figure 1). It is clear that these two DNA fragments have a conserved arrangement of restriction enzyme sites (see Figure 1) where the P element probes hybridized strongly (data not shown). When probes were prepared from the plasmids pπ12.20 and pπ25.1 and hybridized to Bam HI digests of DNA from an M strain (Canton S) in DNA blotting experiments, fragments were detected that were 2.9 kb smaller than the corresponding P strain fragments. Phages homologous to these probes were isolated from a library of this M strain, and the Bam HI fragments were subcloned into pBR322 (plasmids pS12.20 and pS25.1). Comparison of these fragments (Figure 1) with those of the P strain fragments shows that the differences in size and restriction enzyme maps can be explained by the insertion of a 2.9 kb P element. We conclude that this P strain contains multiple copies of a 2.9 kb P element whose structure (at the level of restriction enzyme mapping) is highly conserved between copies.

DNA Sequence of a 2.9 kb P Element

The sequence of the P element from pπ25.1 was determined (see Experimental Procedures) and is shown in Figure 2. The ends of the element were defined in two ways: first by the divergence of sequences between the P elements in pπ12.20 and pπ25.1 and second by a comparison of the sequence of pπ25.1 and that determined for pS25.1 at the site of insertion (see below). The P element from pπ25.1 is 2907 nucleotides long and is bounded by perfect inverse repeats of 31 bases. No differences were detected between this sequence and the sequence determined for 2.6 kb of the P element in pπ12.20 (see Experimental Procedures and Figure 4), illustrating that the sequence of these large P elements is highly conserved.

The protein coding capacity of the 2907 bp P element is illustrated in Figure 3, where the positions of stop codons are shown for each phase of translation for both strands. Three long open reading frames of 714, 792, and 654 bp are found for the upper strand, and several smaller ones of less than 400 bases can be found on both strands.

DNA Sequences of Four P Elements That Inserted into the *white* Locus

The DNA sequences of the P element insertions responsible for the *white* mutations $w^{\#6}$, w^{hd81b9}, $w^{\#12}$, and $w^{hd80k17}$ (Rubin et al., 1982) were determined (see Experimental Procedures). These P elements range in size from 0.5 to 1.6 kb, and their sequences suggest that all of them were derived from a 2.9 kb element by deletion. The elements from $w^{\#12}$ and $w^{hd80k17}$ differ from the 2.9 kb element only by deletion, whereas the element from $w^{\#6}$ has (in addition to a deletion) a single T to A substitution at what corresponds to position 33 of the 2.9 kb element, and that from w^{hd81b9} has (in addition to a deletion) a single T inserted at what corresponds to position 2699 or 2700. Mutants $w^{\#6}$ and $w^{\#12}$ were generated using the strain π_2 as a source of P elements (Simmons and Lim, 1980), so the comparison of the DNA sequences of the P elements from these mutants with the 2.9 kb P element isolated from the same strain demonstrates that, except for the deletions, P element sequences are highly conserved within a strain. Mutants $w^{hd80k17}$ and w^{hd81b9} were made with a different P strain, Harwich (Rubin et al., 1982). Like π_2, Harwich was collected from the wild, but at a different time and from a completely different location. The conservation of DNA sequence between the P elements from the two strains suggests that 2.9 kb P elements from both these strains are extremely similar, if not identical, in sequence.

The positions of the deleted sequences are shown in

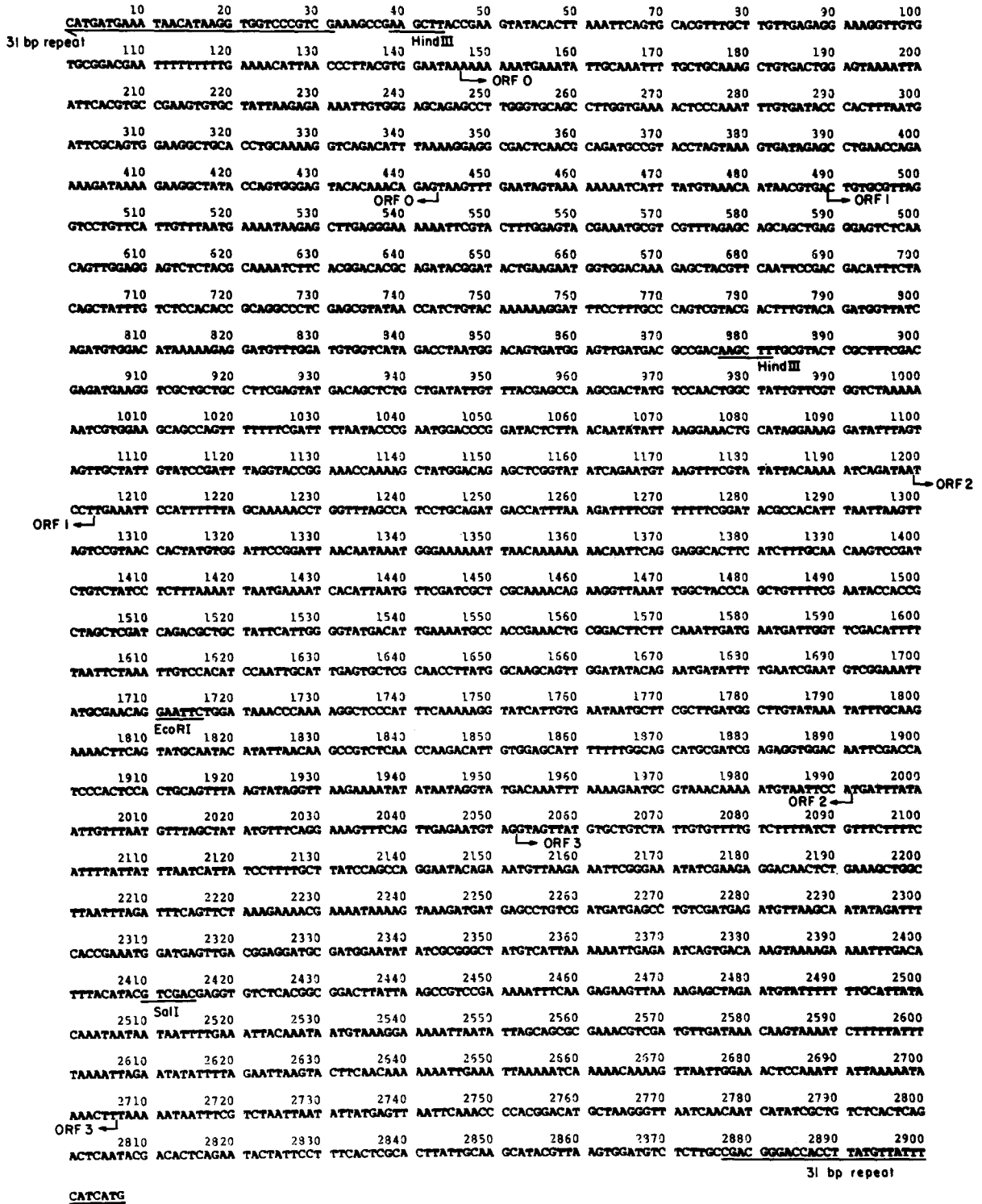

Figure 2. DNA Sequence of the 2.9 kb P Element from pπ25.1

The DNA sequence of only the upper (5′–3′) strand is given. The 31 bp inverse terminal repeats and several restriction enzyme sites are underlined. The limits of the open reading frames (ORFs) discussed in the text—ORF0, ORF1, ORF2, and ORF3—are indicated. Their extents and the position of the first in-phase methionine codon (ATG) are as follows: ORF0, 147–443, ATG at 153; ORF1, 490–1203, ATG at 518; ORF2, 1200–1991, ATG at 1578; ORF3, 2053–2706, ATG at 2059.

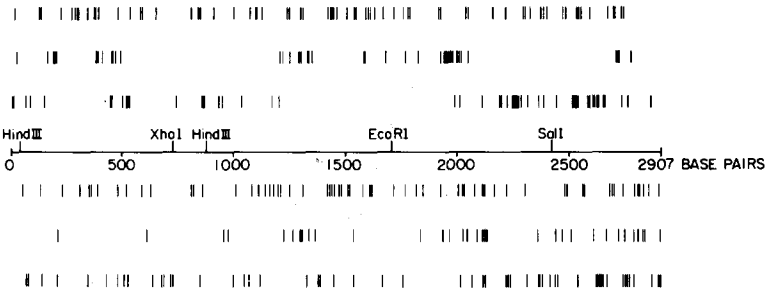

Figure 3. Open Reading Frames in the 2.9 kb P Element

Stop codons in each phase of both strands are shown as vertical lines, above the map for the upper (5′–3′) strand and below the map for the lower (3′–5′) strand.

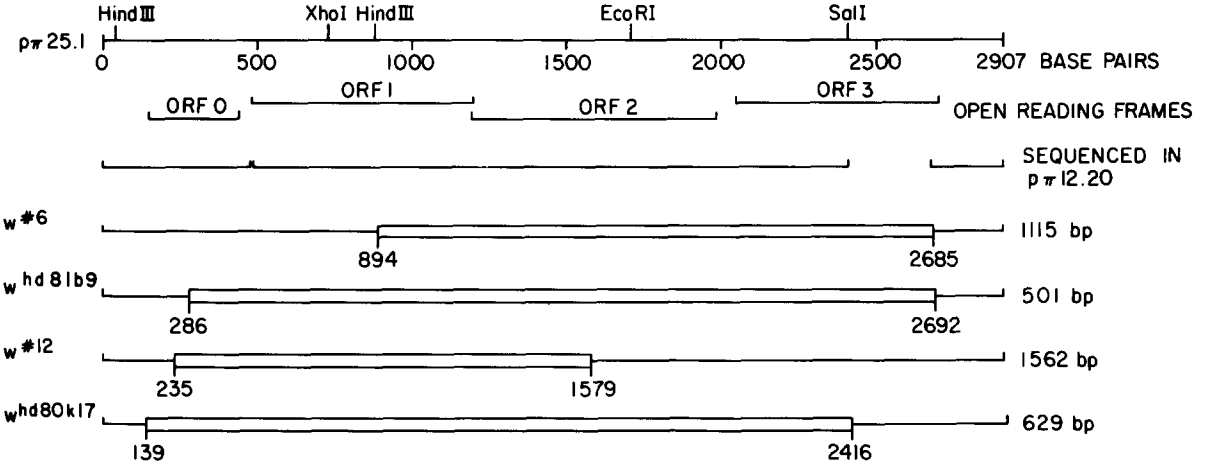

Figure 4. Comparison of the Structure of the 2.9 kb P Element with Those of P Elements Found Inserted in *white* Locus Mutations

Below the map of the 2.9 kb P element from pπ25.1 are shown the positions of the four large open reading frames (ORFs); the region of the 2.9 kb P element from pπ12.20 that has been sequenced; and the sequences deleted (open bars) in the P elements associated with the $w^{\#6}$, w^{hd81b9}, $w^{\#12}$, and $w^{hd80k17}$ mutations. The numbers indicate the positions of the deletion endpoints relative to the 2.9 kb P element sequence shown in Figure 2.

Figure 4, and the DNA sequences surrounding the endpoints of the putative deletions are shown in Figure 5. Included in Figure 5 are DNA sequences at deletion endpoints within some additional small P elements isolated from the π_2 library in the screen for P elements described above. Each deletion endpoint is unique in the sense that none is shared between different elements. This indicates that there is no hotspot for deletion and suggests that each deletion is of independent origin.

Insertion of P Elements Leads to an 8 bp Duplication of the Target DNA

Immediately adjacent to the inverted repeats of 31 bp that define the ends of P elements, 8 bp were found as direct repeat for the four P elements inserted into the *white* locus and for the 2.9 kb elements in pπ12.20 and pπ25.1 (Figure 6). Direct duplications of a few base pairs of the target DNA at the site of insertion have been found after insertion of all previously characterized eucaryotic and procaryotic transposable elements and are thought to be a consequence of the mechanism of transposition (for reviews see Kleckner, 1981; Shapiro, 1983). A fragment containing the site of insertion of the 2.9 kb P element of pπ25.1 was isolated from pS25.1, and its sequence was determined (see Experimental Procedures). The DNA sequences of regions of the *white* locus of the strain π_2, into which P elements had inserted to generate the mutations $w^{\#6}$ and $w^{\#12}$, were similarly determined (see Experimental Procedures). Comparison of these sequences with those flanking the P elements (Figure 6) shows that the 8 bp sequence found directly repeated at both ends of the elements is, in each case, present only once in the target DNA, at the site of insertion.

P Elements Can Precisely Excise

The P-element-induced *white* locus mutations revert at a frequency greater than 10^{-3} during dysgenesis (Rubin et al., 1982). The *white* locus region of a revertant of $w^{\#6}$ ($w^{\#6r}$) has been cloned and, at the level of restriction enzyme mapping, appears to be the same as the wild-type region (Rubin et al., 1982). We have now cloned into M13 and sequenced a fragment from the revertant that includes the site where the P element in $w^{\#6}$ was inserted (see Experimental Procedures). This sequence (Figure 6) demonstrates that reversion is accompanied by loss of both the element and one copy of the 8 bp duplication. Thus the wild-type sequence at the insertion site is precisely restored. By comparison with the DNA sequence of

Figure 5. DNA Sequences at the Deletion Endpoints in P Elements

The sequence of each small P element may (for the most part) be derived by a single deletion from that of the 2.9 kb P element of plasmid pπ25.1. For each of the small P elements the DNA sequence of the novel junction produced by the putative deletion is shown together with the sequences from the 2.9 kb element that surround each of the putative deletion endpoints. Above the small P element sequence is the 2.9 kb element sequence from the left endpoint, with the sequence from the right endpoint below. The numbers correspond to the 2.9 kb P element DNA sequence shown in Figure 2. The sequences have been aligned to emphasize their similarities, and those sequences of the 2.9 kb element that are retained in the small P elements are shown in boldface. The sequences of the P elements from w^{hd81b9}, λπ49, and λπ51 cannot be derived by a simple deletion from that of the 2.9 kb element. Sequences that are not found in the 2.9 kb element at the corresponding positions (T in w^{hd81b9}, ACT in λπ49, and ATCATTAATCATTA in λπ51) are therefore not in boldface.

the corresponding region of the wild-type π_2 chromosome, no differences were found in over 300 bases sequenced for the revertant (data not shown).

Sequence Specificity of P Element Insertion

The P elements in the *white* locus of mutants $w^{\#6}$, w^{hd81b9}, and $w^{hd80k17}$ are inserted at the same nucleotide (see Figure 6). This observation demonstrates that there can be considerable sequence specificity in the insertion of P elements. It should be noted that these mutations were isolated in different laboratories using different P strains. In all the P elements we have characterized, there is a Hind

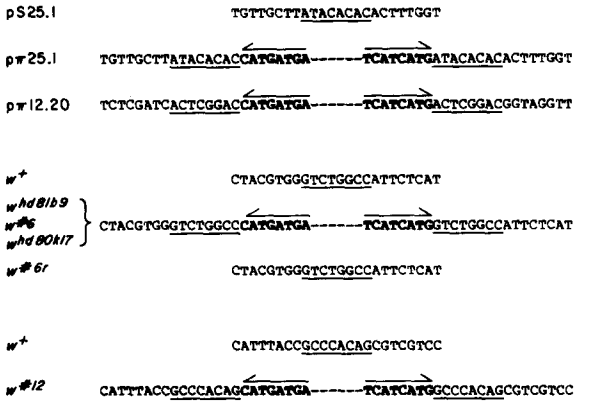

Figure 6. DNA Sequence Changes upon Insertion and Excision of P Elements

The sequences at the ends of the 2.9 kb P elements in pπ12.20 and pπ25.1 are shown together with the sequence from pS25.1 at the position where the P element is inserted in pπ25.1. P element sequences are in boldface, and the 8 bp duplication found at the ends of inserted P elements is underlined. The sequences at the ends of P elements inserted into the *white* locus and the sequences of the wild-type *white* locus and the revertant $w^{\#6r}$ at the positions of insertion are shown in the same manner.

III site 39 bp from one end (which we shall refer to as the left end), and this may be used to orient the P element sequences with respect to their flanking sequences. It may be seen from Figure 10 that the P elements in $w^{\#6}$ and $w^{hd80k17}$ are inserted in one orientation and the element in w^{hd81b9} is inserted in the opposite orientation with respect to the *white* locus sequences. This suggests that during transposition the "target site" as defined by the 8 bp duplication is brought into conjunction with the ends of the P element (which are inverse repeats) and not the internal sequences, so that integration may occur at a given site in either orientation.

To investigate further the sequence specificity of insertion we have made use of this Hind III site to determine the sequence of the ends of 12 additional P elements and the chromosomal sequences adjacent to them. DNAs of phages and plasmids containing P elements were digested with Hind III and Sau 3AI, and the fragments were cloned using M13mp8 cut with Hind III and Bam HI as vector. The recombinants were then screened for P-element-homologous sequences, and the DNA sequences of positively hybridizing phages were determined. The sequences determined, together with those of P elements inserted into the *white* locus, are shown in Figure 7. Although some sequence variation was detected (at positions 32 and 33), the 31 bp inverse repeat was identical in all 18 elements.

We cannot detect any strong homologies in the sequences flanking the inserted P element in the four cases where the sequences on both sides of the elements are known (pπ12.20, pπ25.1, and the two sites where elements have inserted into the *white* locus). Nor were any strong homologies found in the sequences adjacent to the left end for a total of 16 sites. We have, however, observed some similarities among the "target site" 8 bp duplications.

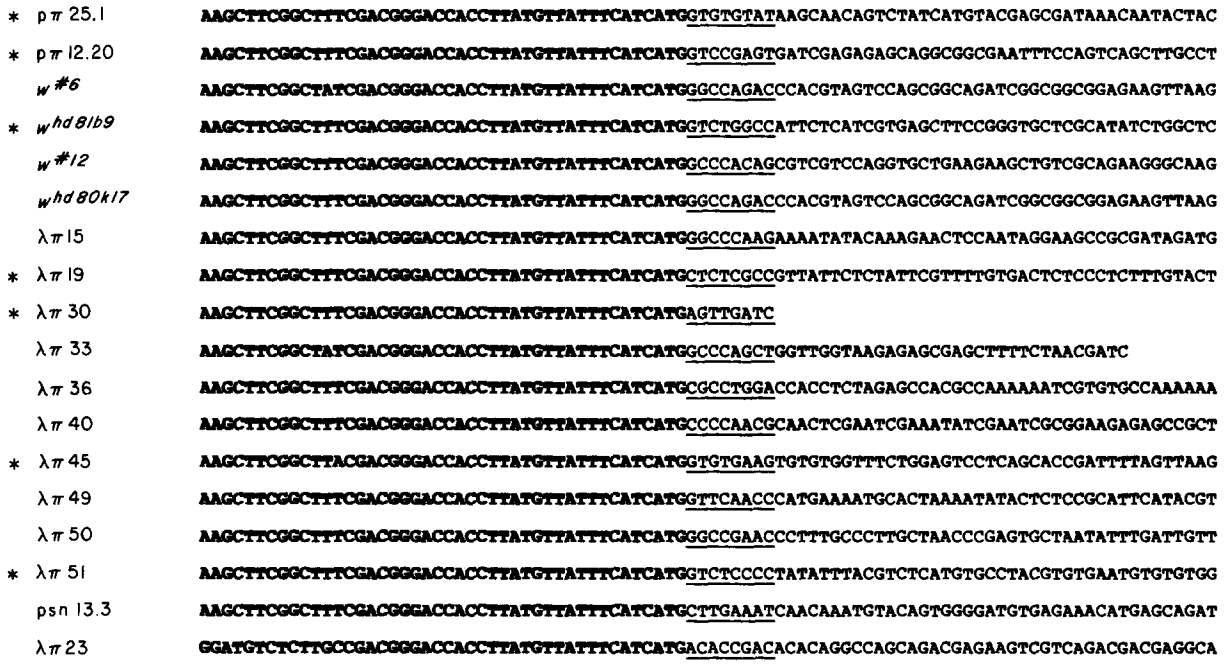

Figure 7. Sequences at the Sites of Residence of P Elements

The DNA sequences at the ends of a total of 18 P elements are given. P element sequences are in boldface, and the 8 bp immediately adjacent to the element are underlined. In all cases except λπ23, the sequence runs from the Hind III site found near the left end of all P elements toward a Sau 3AI site in the flanking sequences. The sequence from λπ23 is that of the other (right) end, as the only clone obtained was from a Hind III site found near the right end of this element to a position in the right flanking sequences. In several cases the same sequence was obtained for different phages and plasmids (λπ30 matched λπ37, and λπ28 matched pπ12.20); therefore, only one example of the sequences is given, as they presumably represent different isolates of the same chromosomal location. The first Sau 3AI site in the flanking sequences for λπ30 and λπ33 is very close to the site of insertion, so only a very short sequence has been determined. By comparison with the P element in pπ25.1, the only sequence differences detected within the P element were in λπ33 and $w^{\#6}$ (both have a T to A substitution at position 33) and in λπ45 (a T to A substitution at position 32). No changes were found within the 31 bp inverse repeats of these 18 elements. For an explanation of the asterisks see the legend to Figure 8.

	1	2	3	4	5	6	7	8
T	0	4	5	0	1	0	0	3
C	4	5	10	15	2	3	4	10
G	11	8	1	3	2	9	1	4
A	3	1	2	0	13	6	13	1
	G	G	C	C	A	G	A	C

Figure 8. A Consensus of the 8 bp Sequences Adjacent to P Elements

The consensus was derived from the sequences underlined in Figure 7. These sequences and their complements were considered because a P element can integrate at a given site in either orientation, so given the cloning strategy used, elements at the same site but in opposite orientations would give complementary 8 bp sequences (see $w^{\#6}$ and w^{hd81b9} as examples). Thirteen of the clones have a C at position 4 in either the 8 bp sequence shown or its complement, while two have a C at this position in both sequences and three do not have a C in this position in either sequence. A partial consensus was made of the 8 bp sequences from the 13 clones with a C at position 4 in only one orientation, and then the sequence from the remaining five cases which most resembled this consensus was included to make the final consensus. The asterisked sequences in Figure 7 are those whose complements were used to construct the consensus sequence. The score of each nucleotide at each position is shown.

In attempting to construct a consensus sequence for these 8 bp, both the sequence shown in Figure 7 and its complement should be considered. This is because a P element can integrate at a given site in either orientation, and so, given the cloning strategy used, elements at the same site but in opposite orientations would give complementary 8 bp sequences (see $w^{\#6}$ and w^{hd81b9} as examples). That we have been able to derive a consensus sequence (Figure 8) shows that insertion is not random with respect to the DNA sequence of the 8 bp target site even though particular target sites may differ from the consensus at up to six of eight positions. The 8 bp duplications were 64% GC, whereas D. melanogaster DNA is only 40% GC overall, and this may play some part in the specificity of insertion of P elements.

Discussion

A New Class of Transposable Elements

The P family represents a new class of eucaryotic transposable elements. They differ from other transposable elements both in their structure and in the fact that their

rate of transposition can be experimentally manipulated. We have characterized a number of P elements ranging in size from 501 to 2907 bp and have shown that while there is considerable variability among the small elements, the large 2907 bp elements are a highly conserved subset. The DNA sequences of the P elements suggest that the small elements have arisen by deletion of internal sequences of the 2.9 kb element, retaining the sequences from its ends. Both the 2.9 kb and the small elements have 31 bp perfect inverse terminal repeats and after insertion are bounded by 8 bp sequences found once at the site of insertion. P elements can excise precisely to restore the original DNA sequence. Independent events resulting in the insertion of a P element in either orientation at the same nucleotide can occur, and the 8 bp adjacent to many P elements show some similarities.

Unlike the *copia*-like transposable elements (reviewed in Spradling and Rubin, 1981), P elements do not have long direct terminal repeats. Considered together, P elements are more variable than the relatively homogeneous *copia*-like elements, but the 2.9 kb P elements are more highly conserved. Like P elements, transposable elements of the foldback (FB) family are highly variable in size. However, different members of the FB family can carry completely unrelated internal sequences, and their terminal repeats, while inverse, are of variable length, being made up of tandemly repeated copies of simple sequence DNA (Truett et al., 1981; Levis et al., 1982b).

Are the 2.9 kb P Elements P Factors?

We believe that the 2.9 kb P elements correspond to P factors, although the equivalence of the biochemically defined 2.9 kb P elements with the genetically defined P factors is difficult to establish unequivocally. The 2.9 kb P elements are multicopy, dispersed, and conserved in structure, as is expected for P factors. We also expect that P factors encode proteins for the control and catalysis of transposition. The DNA sequence of the 2.9 kb P element suggests that it could encode a number of proteins. That the 2.9 kb P elements have at least some of the functions of P factors has been demonstrated in a series of experiments in which 2.9 kb P elements were injected into developing M strain embryos (Spradling and Rubin, 1982; Rubin and Spradling, 1982). The injected elements were capable of causing the hybrid-dysgenesis-induced mutation of the *singed* locus, *singed weak* (*sn*^{w}), to mutate further and revert; of catalyzing their own transposition from the injected plasmid onto the chromosomes of the injected embryo; and of catalyzing the transposition in *trans* of a coinjected P element constructed in vitro to carry the *rosy* gene. These results show that the 2.9 kb P elements are proficient for the production of transposition function(s). Further transpositions of the introduced 2.9 kb elements to new, secondary sites have been observed (A. C. Spradling, G. M. Rubin, and W. R. Engels, unpublished observations), so it is not clear if they encode functions for the control of transposition. A critical number of 2.9 kb P elements may be required for the establishment of repression of transposition (see below).

Previous studies have shown a correlation between the position of strain-specific hotspots for dysgenesis-induced chromosomal rearrangement and the sites of P-element-homologous sequences (Bingham et al., 1982). However, the frequency of rearrangement varies widely for different hotspots (Berg et al., 1980; Engels and Preston, 1981). In these studies it was not possible to determine if the frequency with which a particular P element was associated with rearrangement was correlated with its structure. The 2.9 kb P element in pπ25.1 has been localized by in situ hybridization to 17C (Spradling and Rubin, 1982), and that in pπ12.20 to 5E (W. R. Engels, personal communication). These are two of the most frequent sites for rearrangement on the π_2 X chromosome during hybrid dysgenesis. In contrast, rearrangement associated with the presumably defective element at *sn*^{w} (Engels, 1979b) occurs at a much lower frequency (Engels and Preston, 1981).

Relationship between the Small P Elements and the 2.9 kb P Element

The sequences of the small P elements can, for the most part, be derived by a single deletion of internal sequences of the 2.9 kb element. The most likely mechanism for the generation of the observed deletions is that they occur during transposition as a consequence of an error-prone replication process. Transposition almost certainly involves DNA replication, and it may be that this replication is more prone to error than normal chromosomal replication. The structure of the deletion endpoints is consistent with this hypothesis. First, a sequence found only once in the 2.9 kb element at the site of the deletion is tandemly repeated three times in the P element of $\lambda\pi$51 (Figure 5), suggesting a mechanism involving strand slippage and recopying of the same template sequence by a DNA polymerase. Second, although there is very little sequence variation between P elements other than single large deletions, base changes (in $\lambda\pi$49) and insertions (in *w*^{hd81b9} and $\lambda\pi$51) are found at the deletion endpoints. It therefore seems likely that these changes are a consequence of the mechanism of deletion.

Most of the upper strand sequences of the 2.9 kb P element are represented in spliced poly(A)-containing RNAs from dysgenic embryos (R. E. Karess and G. M. Rubin, unpublished results). The 5′ ends of these transcripts map at approximately position 85 (R. E. Karess and G. M. Rubin, unpublished results), and, as for other eucaryotic transcription start sites (reviewed in Breathnach and Chambon, 1981), a TATA sequence can be found some 30 bp upstream. Further evidence that sequences close to the left end are important comes from the observation that when the first 38 bp of the P element in pπ25.1 are deleted, the residual sequences do not provide *trans*-acting functions for transposition (G. M. Rubin and A. C. Spradling, unpublished results).

Although the coding potential of the small P elements is considerably reduced with respect to that of the 2.9 kb element, the dysgenesis-induced *white* locus mutations do revert during hybrid dysgenesis (Rubin et al., 1982). This suggests that the inserted P elements retain all sequences required to respond to the mobilization of P elements during hybrid dysgenesis (i.e., to respond to transposase provided in *trans* by 2.9 kb P elements). At most, only the first 138 and the last 216 bases of the 2.9 kb P element appear to be required for transposition or excision. As the element in $w^{\#6}$ has inserted into the *white* locus and can excise during hybrid dysgenesis, the single base change it carries at position 33 does not appear to affect the transposition process. However, it seems likely that more than just the 31 bp inverse repeats is required in *cis* for transposition. The introduction of DNA into the leftmost Hind III site (position 39) of the P element from $w^{\#6}$ appears to block transposition in the presence of *trans*-acting functions provided by the 2.9 kb P element (G. M. Rubin and A. C. Spradling, unpublished results). This observation suggests that some *cis*-acting sequence necessary for transposition is located in this region of the P element. A further interesting feature of the DNA sequence of the 2.9 kb P element is that two copies of a 17 bp sequence are found as perfect inverse repeats (positions 718–734 and 764–783). However, as this is a region that is often deleted in the smaller elements (Figure 4), this sequence arrangement cannot be required in *cis* for mobilization.

Sequence Specificity of P Element Insertion

There is considerable variation in the frequency at which mutations arise in different genes during hybrid dysgenesis (Golubovsky et al., 1977; Green, 1977; Simmons and Lim, 1980). We have shown that three insertions into the *white* locus are at the same nucleotide, implying a high degree of specificity of insertion at the local DNA sequence level. The two sites where P elements have inserted in the *white* locus both fall within large open reading frames (unpublished observations), and it seems likely that the white phenotype is due to disruption of a protein product of the *white* locus. While one could imagine that insertions at some positions (e.g., in the 3′ noncoding region) might not produce a white phenotype, it is probable that an insertion at any position in the coding sequences would. Thus the observed specificity of insertion does not appear to be a consequence of phenotypic selection.

We have shown that the 8 bp sequences adjacent to many P elements are similar and have derived a consensus sequence (GGCCAGAC) identical with that at the *white* locus hotspot for insertion. It is likely that the DNA sequence of the target influences both the position of insertion and the rate at which insertions occur.

Excision of P Elements

The results described here show that P elements can precisely excise during hybrid dysgenesis to restore the original DNA sequence. This does not mean that they always excise precisely, as any imprecise excisions of the *white* locus P elements would probably not have changed the eye color from white, so imprecise excisions would not have been detected. Apparent imprecise excision of P elements from the *Rpo11* locus has been reported (Searles et al., 1982). Phenotypically wild-type revertants of the *white* locus mutants were obtained at high rates only during hybrid dysgenesis (Rubin et al., 1982), suggesting that P factor function(s) are required for the excision of P elements. Precise excision of an FB element, a different D. melanogaster transposable element family, has also been observed (Collins and Rubin, 1983). This contrasts with procaryotic transposable elements, in which precise excision occurs at a much lower frequency and is apparently independent of transposon-encoded functions (Kleckner, 1981).

Precise excision may simply occur by the reversal of the first steps of the insertion mechanism. The direct repeats found bounding all transposable elements are thought to be produced by a staggered cut made at the site of integration which is subsequently repaired during integration (for references see Kleckner, 1981). If the enzyme which made the 8 bp staggered cut for insertion of a P element made the same staggered cut at both copies of the same 8 bp sequence now found bounding the P element, then the P element would be excised. Both the P element and the site of insertion would have complementary 8 bp "sticky ends" that could be ligated to restore the original sequence of the insertion site and form a circular P element with one copy of the 8 bp bounded by the 31 bp inverse repeats from the ends of the element. Such circular structures could be intermediates in transposition, or have some role in the determination of cytotype as suggested by Engels (1981).

Cytotype and the Control of Transposition

It is clear that transposition of P elements is genetically controlled. This control of transposition appears to be ultimately determined by the chromosomal constitution of the fly, but its inheritance is complex. Female offspring from a P♂ × M♀ cross have the same chromosomal composition as the female offspring of an M♂ × P♀ cross. However, only the daughters from the former cross give dysgenic progeny when mated to a P male, and they are said to have inherited M cytotype from their female parent. The daughters from the latter cross have inherited P cytotype and, like their mothers, are resistant to the action of P factors. The M cytotype can be thought of as either the presence of factors that allow P element transposition or the lack of factors that suppress transposition. However, this inheritance of cytotype through the female line is only partially independent of chromosomal constitution, and so is not strictly cytoplasmically determined. Switches from P to M and vice versa can occur in a single generation (Kidwell, 1981) and are correlated with the number of P-derived chromosomes (for an M to P switch) or the number

of M-derived chromosomes (for a P to M switch). For a detailed discussion of the determination of cytotype see Engels (1979a) and Kidwell (1981).

Transposition of P elements is repressed in P strains but occurs when they are introduced into an M cytotype egg. This is similar to the zygotic induction of phage λ when a lysogenic Hfr male bacterium conjugates with a nonlysogenic female (discussed in Hayes, 1964). The prophage is stably integrated in the lysogen because lytic functions are controlled by a phage-encoded repressor; but when the DNA is transferred in conjugation, then lysis normally occurs, much as in a viral infection. Various models have been proposed whereby some cytoplasmic or nucleoplasmic factor(s), analogous to phage λ repressor, are required to repress transposition in P strains. Such simple models cannot explain the limited heritability of cytotype through the maternal line, so it has been proposed that the cytotype-determining factors can self-replicate to some extent (Engels, 1981). Replication is hypothesized since a single female can produce several thousand eggs, so that any nonreplicating factor would be subject to enormous dilution. It has been suggested that the spatial organization of the nucleus may be propagated in this way (Sved, 1976), or more recently that cytotype is determined by an episome with a limited potential for self-replication (Engels, 1981). As we have pointed out above, precise excision of chromosomal P elements could be a mechanism for the generation of such episomes.

The continued presence of P factors does ultimately lead to the establishment of P cytotype, but this may take several fly generations (Engels, 1979a; Kidwell, 1981). We propose that P factors code for at least two functions, a transposase and a regulator that can lead to the suppression of transposition. The nonreciprocity of the two P × M crosses requires us to hypothesize the second function. If the regulator positively feeds back to stimulate its own activity, then this could explain the self-reproduction of cytotype and thus be an alternative to the episomal model. In P cytotype, there is a high activity of the regulator so that transposition is switched off (see Figure 9). In a dysgenic cross, P factor DNA is introduced into a background where there is neither transposase nor regulator, and expression of both functions leads (at least initially) to transposition. With time, positive feedback by the regulator on its own level of activity produces enough regulator to shut off transposition and the P cytotype is established. Depending on how effectively the regulator positively feeds back to increase its activity, and on how much regulator is required to shut off transposition efficiently, this could explain why cytotype has limited inheritance through the female line. The regulator produced during the growth of the female progeny of a P♂ × M♀ cross could be insufficient to change the cytotype—i.e., insufficient to suppress transposition when they are mated to a P strain male. The postulated positive feedback of the regulator could also account for the observation that intermediate cytotypes are rarely found, as there would be a sharp transition

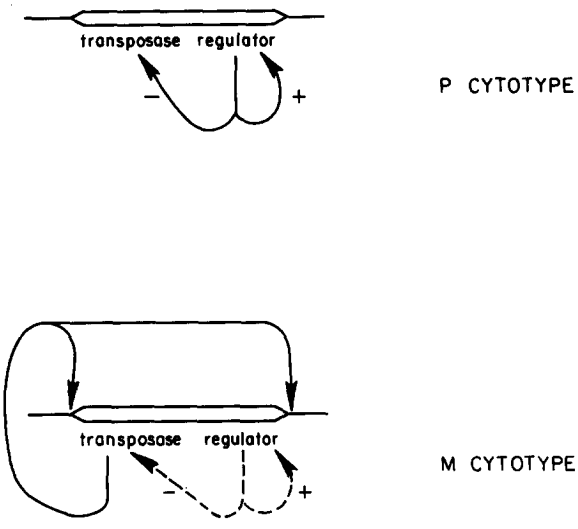

Figure 9. Hypothesis for the Germ-Line Control of Transposition by Transposon-Encoded Functions

The proposed activities of the two element-encoded functions are indicated in P and M cytotype. In P cytotype, the amount of regulator is such that, as a result of its positive feedback on itself and negative feedback on transposase, transposition does not occur. When P elements are first introduced into the M cytotype, the regulator is insufficient for efficient negative feedback on transposase, so transpositions of both 2.9 kb and small P elements occur. However, as the regulator accumulates (in a process that may take several generations), it will feed back positively on itself and negatively on transposase, and a level will eventually be reached at which the cytotype will switch to P.

between low (M cytotype) and high (P cytotype) levels of regulator. A further consequence of this model is that there would tend to be a similar number of P factors in all P strains, as a low number of templates for the production of regulator would lead to transposition and thus to an increased number of P factors, until enough regulator had accumulated to suppress transposition. The molecular details of the mechanism whereby transposition might be repressed by such a regulator remain to be elucidated.

Experimental Procedures

Drosophila Strains and λ Cloning

The construction of λ libraries and isolation of phages containing inserts from the *white* loci of the strains π_2, $w^{\#6}$, $w^{\#12}$, $w^{hd60k17}$, w^{hd81b9}, and $w^{\#6/}$ have been described (Rubin et al., 1982). The π_2 library was screened as described by Levis et al. (1982) using a mixture of fragments from the *white* loci of strains $w^{\#6}$ and $w^{\#12}$, then counterscreened with the equivalent fragments from a wild-type *white* locus to isolate phages containing sequences homologous to P elements (see Levis et al., 1982a, and Rubin et al., 1982, for maps of these *white* locus regions).

The phages λS25 and λS12 were isolated from a λ library of the M strain Canton S (Maniatis et al., 1978), using as probes the plasmids pπ25.1 and pπ12.20. Screening, purification, growth, and DNA extraction of the phages were as described (Levis et al., 1982a).

Analysis of Cloned DNAs

DNA blotting and the construction of restriction enzyme maps were as previously described (Levis et al., 1982a). Bam HI digests of λ clone DNAs were ligated with a pBR322 vector (prepared by digestion with Bam HI and treatment with bacterial alkaline phosphatase), and subclones containing the Bam HI fragments of interest were isolated and grown by standard procedures (Maniatis, Fritsch, and Sambrook, 1982).

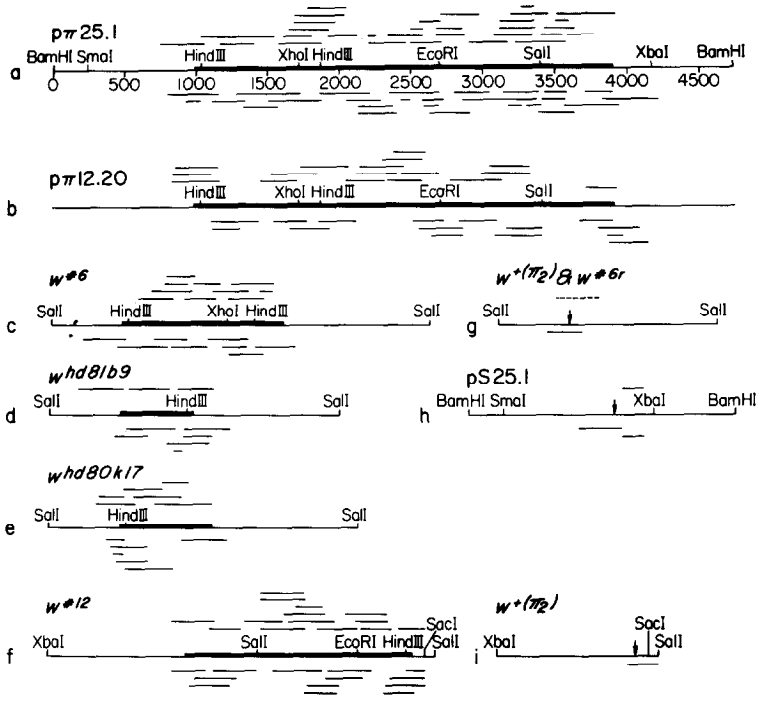

Figure 10. Determination of the DNA Sequences of P Elements and Their Sites of Insertion

Restriction site maps of portions of plasmids containing P elements or their sites of insertion and excision are shown as follows: (a) pπ25.1; (b) pπ12.20; (c) $w^{\#6}$; (d) w^{hd81b9}; (e) $w^{hd80k17}$; (f) $w^{\#12}$; (g) w^{+}, and $w^{\#6r}$; (h) pS25.1; (i) w^{+}. Lines represent individual gel readings. Above the maps are readings of the upper (5′–3′) strand; below are readings of the lower (3′–5′) strand. Heavy lines show the P elements; arrows show the sites of insertion. The dotted line in (g) is the gel reading from the wild-type *white* locus.

M13 Cloning and DNA Sequencing

To determine the complete sequence of DNA fragments, M13 subclones were constructed containing either restriction enzyme fragments or fragments whose ends were generated by DNAase I (Anderson, 1981) or sonication (P. Deininger, personal communication). The fragment was gel purified, self-ligated, and then treated with DNAase I or (more frequently) sonicated to introduce breaks at random and so produce fragments whose mean size was about 500 bp (range of 100–1500 bp). The ends of the fragments were then repaired using T4 DNA polymerase (Challberg and Englund, 1980) before size fractionation by agarose gel electrophoresis. Several pools covering the size range 250–1000 bp were recovered by trough elution (Yang et al., 1979). These preparations were ligated with M13mp8 that had been linearized with Sma I and treated with bacterial alkaline phosphatase. Restriction enzyme fragments were subcloned using M13mp7, M13mp8, or M13mp9 cut with the appropriate enzyme(s) (Sanger et al., 1980; Gardner et al., 1981; Messing and Vieira, 1982).

Growth of recombinant phages and preparation of DNA templates were as described by Sanger et al. (1980). DNA sequencing using dideoxynucleotides was as described by Sanger et al. (1980) except that the sequence reactions were not in glass capillaries, but in Eppendorf tubes as described by Heidecker et al. (1980).

Individual gel readings (see Figure 10) were assembled into complete sequences using the BATIN, DBCOMP, and DBUTIL programs of Staden (1980). Although positions in an individual reading could not always be unambiguously assigned, the ambiguities were resolved by sequencing the other strand or in other gel readings of the same strand. The sequences of the *white* locus P elements, the P element in pπ25.1, and the partial sequence of the P element in pπ12.20 were independently determined, and the few differences detected are noted in the text. The region of the P element in pπ25.1 that was sequenced on only one strand (positions 551–585) was sequenced on the second strand in pπ12.20 and on both strands in the element from $w^{\#6}$, with no differences detected. There is an Eco K restriction site at position 483 of the 2.9 kb P element, and as the initial subcloning was from unmodified plasmid DNA (grown in HB101) into the restricting host JM101, no clones covering this position were recovered from either pπ25.1 or pπ12.20. To complete the sequence of the P element in pπ25.1, restriction enzyme fragments were cloned from modified λ DNA (grown in K802). To circumvent this problem, all later subcloning was from modified DNA.

Acknowledgments

We thank Alan Bankier, Prescott Deininger, Bart Barrell, and other members of the MRC Molecular Biology Laboratory, Cambridge, for their generosity in teaching DNA sequencing methods to K. O'H. This work would not have been possible without the excellent technical assistance of Christine Murphy. We thank our colleagues in this laboratory and Bill Engels, Michael Simmons, and Margaret Kidwell for stimulating discussions and critical reading of the manuscript and Mark Krasnow for suggesting the mechanism of precise excision. This work was supported by a grant from the National Institutes of Health to G. M. R.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 29, 1983; revised May 19, 1983

References

Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase-I generated fragments. Nucl. Acids Res. *9*, 3015–3027.

Berg, R., Engels, W. R., and Kreber, R. A. (1980). Site-specific X chromosome rearrangements from hybrid dysgenesis in Drosophila melanogaster. Science *210*, 427–429.

Bingham, P. M., Kidwell, M. G., and Rubin, G. M. (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell *29*, 995–1004.

Bregliano, J.-C., and Kidwell, M. G. (1983).Hybrid dysgenesis determinants. In Mobile Genetic Elements, J. A. Shapiro, ed. (New York: Academic Press), pp 363–410.

Breathnach, R., and Chambon, P. (1981). Organisation and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. *50*, 349–383.

Challberg, M. D., and Englund, P. T. (1980). Specific labelling of 3′ termini with T4 DNA polymerase. Meth. Enzymol. *65*, 39–43.

Collins, M., and Rubin, G. M. (1983). High frequency precise excision of the Drosophila foldback transposable element. Nature *303*, 259–260.

Engels, W. R. (1979a). Hybrid dysgenesis in Drosophila melanogaster: rules of inheritance of female sterility. Genet. Res. Camb. *33*, 219–236.

Engels, W. R. (1979b). Extrachromosomal control of mutability in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA *76*, 4011–4015.

Engels, W. R. (1981). Hybrid dysgenesis in Drosophila and the stochastic loss hypothesis. Cold Spring Harbor Symp. Quant. Biol. *45*, 561–565.

Engels, W. R., and Preston, C. R. (1980). Components of hybrid dysgenesis in a wild population of Drosophila melanogaster. Genetics *95*, 111–128.

Engels, W. R., and Preston, C. R. (1981). Identifying P factors in Drosophila by means of chromosome breakage hotspots. Cell *26*, 421–428.

Gardner, R. C., Howarth, A. J., Hahn, P., Brown-Luedi, M., Shepherd, R. J., and Messing, J. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. Nucl. Acids Res. *9*, 2871–2888.

Golubovsky, M. D., Ivanov, Yu. N., and Green, M. M. (1977). Genetic instability in Drosophila melanogaster: putative multiple insertion mutants at the singed bristle locus. Proc. Nat. Acad. Sci. USA *74*, 2973–2975.

Green, M. M. (1977). Genetic instability in Drosophila melanogaster: de novo induction of putative insertion mutants. Proc. Nat. Acad. Sci. USA *74*, 3490–3493.

Hayes, W. (1964). The Genetics of Bacteria and Their Viruses. (New York: Wiley and Sons).

Heidecker, G., Messing, J., and Gronebom, B. (1980). A versatile primer for DNA sequencing in the M13mp2 cloning system. Gene *10*, 69–73.

Kidwell, M. G. (1979). Hybrid dysgenesis in Drosophila melanogaster: the relationship between the P-M and the I-R interaction systems. Genet. Res. Camb. *33*, 205–217.

Kidwell, M. G. (1981). Hybrid dysgenesis in Drosophila melanogaster: the genetics of cytotype determination in a neutral strain. Genetics *98*, 275–290.

Kidwell, M. G., Kidwell, J. F., and Sved, J. A. (1977). Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics *86*, 813–833.

Kleckner, N. (1981). Transposable elements in prokaryotes. Ann. Rev. Genet. *15*, 341–404.

Levis, R., Bingham, P. M., and Rubin, G. M. (1982a). Physical map of the *white* locus of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA *79*, 564–568.

Levis, R., Collins, M., and Rubin, G. M. (1982b). FB elements are the common basis for the instability of the w^{DZL} and w^{c} Drosophila mutations. Cell *30*, 551–565.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eucaryotic DNA. Cell *15*, 687–701.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Messing, J., and Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene *19*, 269–276.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science *218*, 348–353.

Rubin, G. M., Kidwell, M. G., and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell *29*, 987–994.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161–178.

Searles, L. L., Jokerst, R. S., Bingham, P. M., Voelker, R. A., and Greenleaf, A. L. (1982). Molecular cloning of sequences from a Drosophila RNA polymerase II locus by P element trnsposon tagging. Cell *31*, 585–592.

Shapiro, J. A. (1983). Mobile Genetic Elements (New York: Academic Press).

Simmons, M. J., and Lim, J. K. (1980). Site specificity of mutations arising in dysgenic hybrids of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA *77*, 6042–6046.

Spradling, A. C., and Rubin, G. M. (1981). Drosophila genome organisation: conserved and dynamic aspects. Ann. Rev. Genet. *15*, 219–264.

Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science *218*, 341–347.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. Nucl. Acids Res. *8*, 3673–3694.

Sved, J. A. (1976). Hybrid dysgenesis in Drosophila melanogaster: a possible explanation in terms of spatial organisation of chromosomes. Austral. J. Biol. Sci. *29*, 375–388.

Truett, M. A., Jones, R. S., and Potter, S. S. (1981). Unusual structure of the FB family of transposable elements in Drosophila. Cell *24*, 753–763.

Yang, R. C.-A., Lis, J., and Wu, R. (1979). Elution of DNA from agaorse gels after electrophoresis. Meth. Enzymol. *68*, 176–183.