The EMBO Journal vol.6 no.8 pp.2401-2408, 1987

A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A+T-rich DNA

Olivier Hyrien, Michelle Debatisse, Gérard Buttin and Bruno Robert de Saint Vincent

Unité de Génétique Somatique (UA CNRS 361), Institut Pasteur, 25, rue du Dr. Roux, 75724 Paris Cedex, 15, France

Communicated by G.Buttin

We have identified, in the amplified domain of adenylate deaminase (AMPD) overproducing Chinese hamster fibroblasts, a 2.6 kb recombinogenic DNA region which is frequently involved in amplification-associated rearrangements. The nucleotide sequence reveals a mosaic organization of four *Alu*-equivalent repeats of the B1 and B2 families and eight long A+T-rich DNA segments. Part of this region is enriched with long imperfect palindromes. The center of one palindrome contains a putative topoisomerase I cleavage site and this site defines the position of a novel junction which was formed by illegitimate recombination with another A+T-rich DNA sequence located far apart on the amplified DNA. These findings and their significance are discussed in the context of related data from other systems and in the light of current models for eukaryotic DNA recombination, replication and organization.

Key words: *Alu*-like repeats/A+T-rich DNA/gene amplification/illegitimate recombination/inverted and directed repeats/adenylate deaminase

Introduction

DNA sequence amplification has been described as both a normal and abnormal process in a wide range of living organisms (for reviews see Schimke, 1982, 1984; Cowell, 1982; Stark and Wahl, 1984; Hamlin *et al.*, 1984; Kafatos *et al.*, 1985; Stark, 1986).

Formation of amplified arrays involves replication, recombination and rearrangement of vast amounts of DNA since estimations of the amplified unit size vary from the minimal estimate of 135 kb in a multistep highly amplified cell line (Milbrandt *et al.*, 1981) to possibly 10 000 kb in first step mutants (Giulotto *et al.*, 1986). The characteristics of these recombination processes are not known in detail. Studies of the CAD and DHFR gene amplification systems did not reveal 'hotspots' for rearrangements (Caizzi and Bostock, 1982; Ardershir *et al.*, 1983; Federspiel *et al.*, 1984).

We have previously reported AMPD overproduction and amplification of several genes (designated X, W, Y_1, Y_2) in Chinese hamster fibroblasts exposed to stepwise increasing doses of coformycin in the presence of adenine and azaserine (Debatisse *et al.*, 1984, 1986). Independent lineages of mutants isolated at each selection step were analysed. Amplified DNA was found to be chromosomally integrated at every step (B.R.S.V., unpublished *in situ* hybridization data). A genetic map of the region based on the segregation properties of these genes was deduced (Figure 1). We observed and studied the appearance and degree of amplification of novel joints most of which presumably result from recombination events that resolve the newly amplified copies into tandem chromosomal arrays. In contrast to the CAD and DHFR gene amplification systems, these novel joints were not randomly distributed along the amplified DNA but mapped mainly within the W gene (Debatisse *et al.*, 1986).

Here, we report the cloning of the wild-type W gene and the nucleotide sequence of a short region where most of the W-associated novel joints cluster. We also report the cloning and sequencing of one of these novel joints. These nucleotide sequences reveal unusual features both at and around the breakpoint, which are most probably related to the high frequency of recombination in the W region, and possibly to the mechanism of gene amplification.

Results

Cloning of the wild-type W region

All our amplified cell lines present rearrangements in the W gene. Although some of them also possess apparently wild-type copies of the W region, we chose to clone this gene from the GMA32 (wild-type) line in order to avoid ambiguity in the construction of its restriction map.

A genomic library of GMA32 DNA was constructed in λ EMBL4, amplified, and screened with a full length W cDNA probe. Twenty out of 9×10^5 plated recombinant phages were strongly positive on duplicate filters at the first round of screening. However, clones were difficult to isolate and after 3-5 rounds of screening only five were purified. One of them had rearranged during cloning and is not shown here. The four others overlap and cover 30 kb of wild-type W DNA (Figure 1). The accuracy of this restriction map was verified by using DNA from

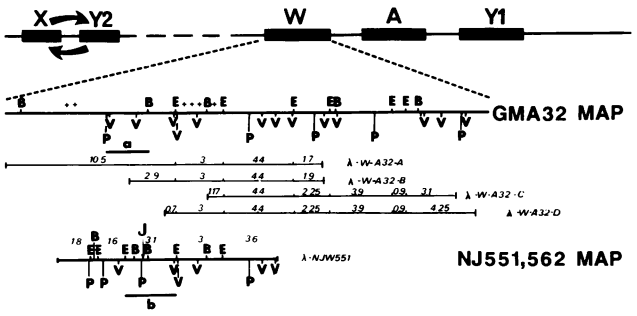

Fig. 1. Molecular clones used in this study and their location in the amplified region. **Upper line**: genetic map of the amplified region. A: AMPD gene. X, W, Y_1, Y_2: coamplified genes. The W-A-Y_1 linkage has been confirmed by a chromosome walk (M.Debatisse, in preparation). **Middle line**: restriction map of the wild-type (GMA32) W region. E: *Eco*RI; B: *Bam*HI; P: *Pst*I; V: *Pvu*II; the signs (+) indicate the approximate location and extent of the exons. The positions of the overlapping phages and the sizes of their *Eco*RI fragments are shown under the middle line. **Bottom line**: restriction map of λNJW551. the sizes of the *Eco*RI fragments are indicated in italics. The position of the junction is indicated (J). Fragments a and b represent respectively the 2.6 kb *Pst*I-*Bam*HI fragment of λ-W-A32-A and the 3.1 kb *Eco*RI fragment of λNJW551 which were subcloned for sequencing.

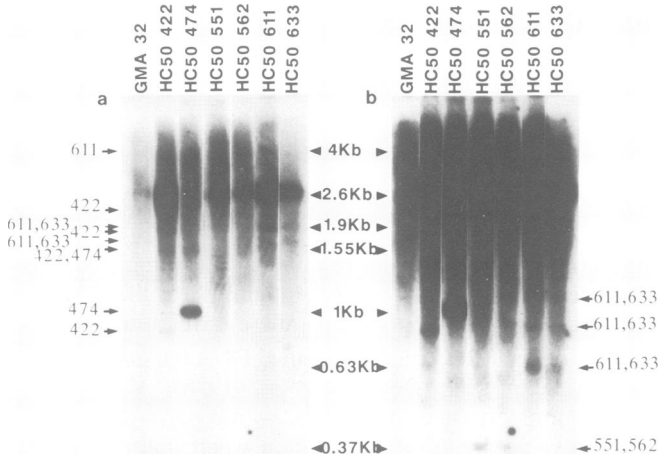

Fig. 2. Novel joints in the 2.6 kb *Bam*HI−*Pst*I fragment of the wild-type W region. Genomic DNAs (10 μg) from the cell lines indicated were digested with *Bam*HI + *Eco*RI + *Pst*I, separated on a 1.5% agarose gel, transferred to Hybond-N, and probed with the 2.6 kb W region *Bam*HI−*Pst*I fragment (probe a on Figure 1). **a** after 36 h, and **b** after 6 days exposure. The position of each novel joint is indicated by an arrow with the number(s) of the cell line(s) in which it is seen. The sizes of some fragments are indicated in the middle. A *Bst*eII digest of lambda DNA was used as a size marker.

the recombinant phages to probe a blot of GMA32 DNA digested with the four restriction endonucleases used to establish the phage restriction maps: the fragments thus revealed were of the sizes predicted from the restriction map; the approximate location of the exons was determined by probing various digests of the recombinant phage DNAs with the W cDNA (not shown).

Mapping of W-associated novel joints inside the W gene

At least nine novel joints in six different cell lines were previously detected in the W region with the full-length W cDNA probe, which allows ~20 kb of DNA to be scanned for the presence of rearrangements (Debatisse *et al.*, 1986 and unpublished data). In order to localize these novel junctions within the W gene, *Bam*HI + *Pst*I + *Eco*RI triple digests of the DNAs of the highly amplified lines were performed and a Southern transfer of these restricted DNAs was probed with various subfragments of the W region genomic clones. Figure 2 shows hybridization of this blot with the 2.6 kb *Pst*I−*Bam*HI fragment of λ-W-A32-A (probe a on Figure 1). If novel joints are distributed randomly within the W gene, the novel fragments detected by this probe on the triple digests should be a small subfraction of those detected with the W cDNA on single digests since the DNA length scanned here is reduced to 2.6 kb instead of 20 kb. Actually, the 2.6 kb *Pst*I−*Bam*HI fragment detected twelve different novel joints on the triple digests. Not all are visible on a single exposure of the blot because of the elevated background due to the high enrichment of the probe in repetitive sequences (see below). In contrast, no novel joint was detected when the same blot was probed with either the 4.5 kb *Bam*HI−*Pst*I fragment (to the left of the 2.6 kb *Pst*I−*Bam*HI fragment), with the 1.6 kb *Bam*HI−*Eco*RI fragment (to the right) or with the 2 kb *Eco*RI−*Bam*HI fragment (further to the right). Therefore the W-associated novel joints cluster in the 2.6 kb *Pst*I−*Bam*HI fragment. Some of these twelve novel joints escaped detection by the W cDNA probe − which revealed nine of them − probably because they were obscured by highly amplified fragments. The selection step at which each of these twelve novel joints appeared is reported in Table I. Two were formed at the third step. Three were formed in second step mutants and reamplified in their third step derivatives. The seven others were formed in first step mutants and reamplified in their second and third step derivatives. This explains why the same novel joint (e.g. the 0.37 kb NJ 551,562, initially formed in HC5) was sometimes observed in two different cell lines (e.g. $HC_{50}551$ and $HC_{50}562$). Each novel joint in this case was scored above only once.

Table I. Distribution in the mutant lines of the novel joints formed in the 2.6 kb *Pst*I−*Bam*HI fragment

Step I	**HC4** 1.55 kb		**HC5** 0.37 kb		**HC6** 0.63 kb, 0.85 kb, 1.1 kb 1.7 kb, 1.9 kb	
Step II	**$HC_{10}42$** same as in HC4 + 1.8 kb and 0.8 kb	**$HC_{10}47$** same as in HC4 + 1 kb	**$HC_{10}55$** same as in HC5	**$HC_{10}56$** same as in HC5	**$HC_{10}61$** same as in HC6	**$HC_{10}63$** same as in HC6
Step III	**$HC_{50}422$** same as in $HC_{10}42$ + 2.3 kb	**$HC_{50}474$** same as in $HC_{10}47$	**$HC_{50}551$** same as in HC5	**$HC_{50}562$** same as in HC5	**$HC_{50}611$** same as in HC6 + 4 kb	**$HC_{50}633$** same as in HC6

HC4, HC5 and HC6 are three different first step mutants isolated from the GMA32 line. From each, two second step mutants were isolated. A third selection step yielded one step III mutant from each step II line. Each novel joint is identified by its size (in kb).

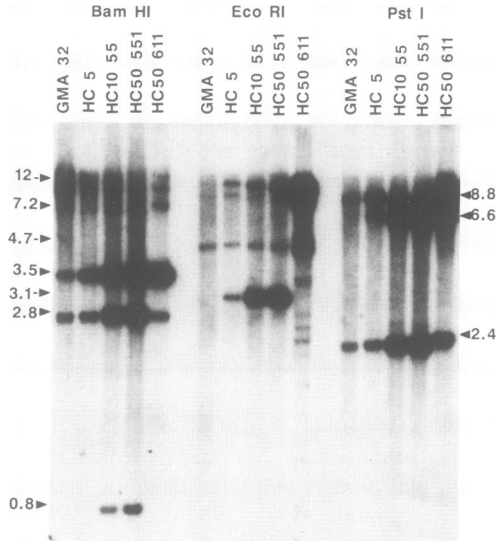

Fig. 3. Analysis of the novel joint NJ 551,562. Genomic DNAs from the indicated cell lines were digested with the restriction endonuclease indicated, separated on a 0.7% agarose gel, transferred to Hybond-N and probed with the 3.1 kb fragment of λNJW551 (probe b on Figure 1). A *Bst*eII digest of lambda DNA was used as a size marker. Note that lane $HC_{50}611$ of the *Bam*HI digest contains about one third of the amount of DNA loaded in other lanes.

Cloning of the novel junction NJ 551,562

NJ 551,562 is a novel junction which was previously detected in the W region as a 6.6 kb novel *Pst*I fragment. It appeared during the first selection step (line HC5) as a single copy and was selectively reamplified upon selection for higher resistance

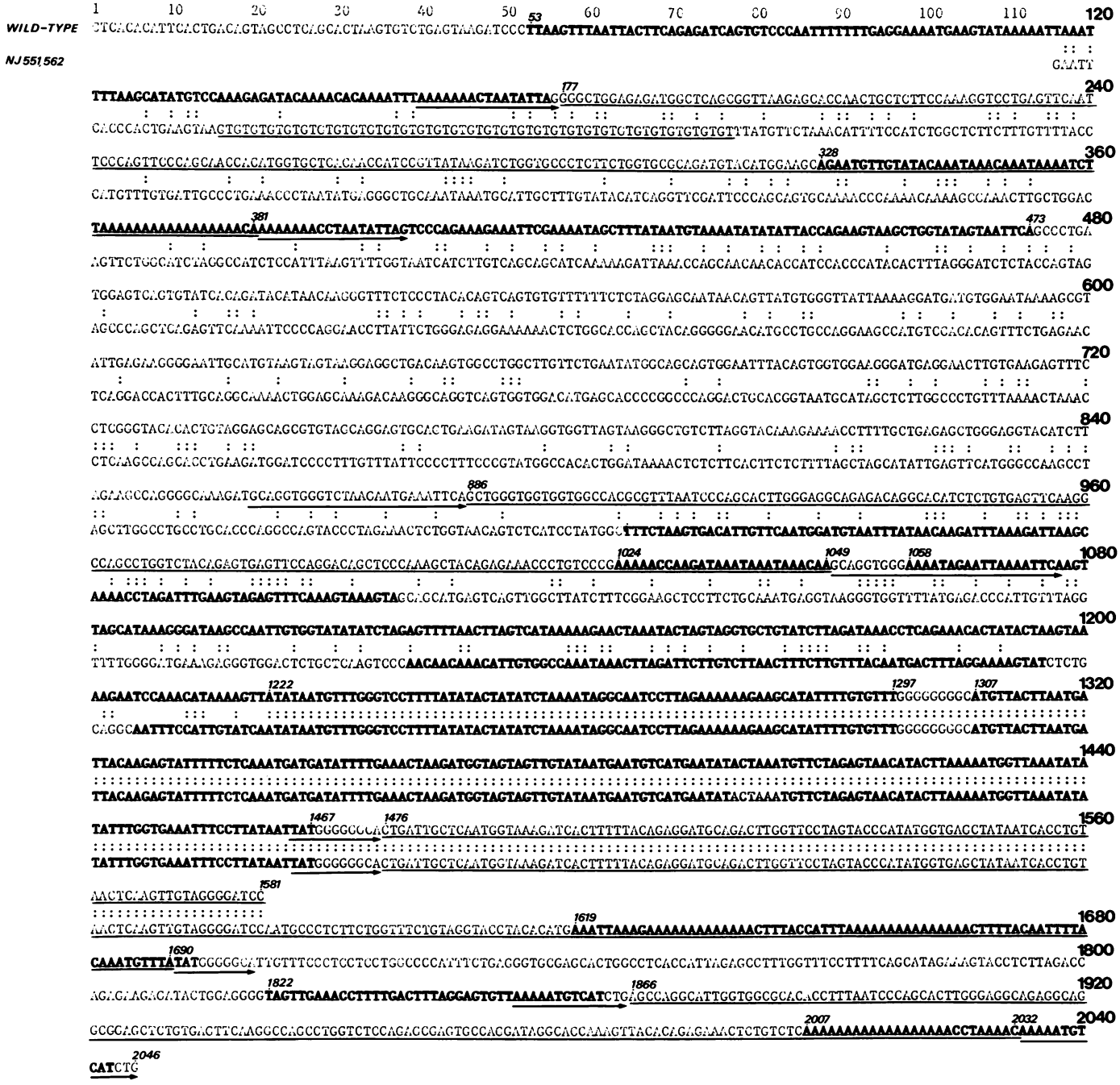

Fig. 4. Alignment of the wild-type W sequence and the NJ 551,562 sequence. Nucleotide numbering is that of the wild-type W sequence starting 1581 bp upstream of the *Bam*HI site of fragment a (Figure 1). The sign (:) indicates homologous bases. A+T-rich segments are printed in bold types. *Alu*-equivalent repeats are underlined and their framing direct repeats are marked by an arrow. A $(GT)_{31}$ simple repeat in NJ 551,562 is also underlined.

(second step lines $HC_{10}55$ and $HC_{10}56$ and third step derivatives $HC_{50}551$ and $HC_{50}562$). This novel junction was cloned after two rounds of screening, with the W cDNA probe, of a genomic $HC_{50}551$ library constructed in λ EMBL4 (1.5×10^4 recombinants were plated). The restriction map of λ NJW 551, the recombinant phage carrying the junction, is shown aligned under the wild-type map in Figure 1.

The right hand part of this map is identical to the GMA32 map. The leftmost restriction site common to both GMA32 and NJ 551,562 maps is a *Bam*HI site. A novel *Pst*I site is present in λ NJW 551 approximately 0.5 kb to the left of this *Bam*HI site, hence implying that the rearrangement took place in the 2.6 kb *Pst*I−*Bam*HI fragment of GMA32 DNA, no more than 0.5 kb to the left of the *Bam*HI site. Since this position corresponds to an intron in the W map, the previously deduced intronic location of the rearrangement (Debatisse *et al.*, 1986) is confirmed.

The rearrangement has deleted the sequences extending to the left of this point and replaced them by a new DNA fragment. The origin of this new DNA fragment was investigated by a blot analysis (Figure 3) using as a probe the novel 3.1 kb *Eco*RI fragment of λ NJW 551 (probe b on Figure 1). This fragment contains the junction between the W sequences and the newly recombined sequences. DNAs of the wild-type and several amplified cell lines were analyzed after digestion with *Bam*HI,

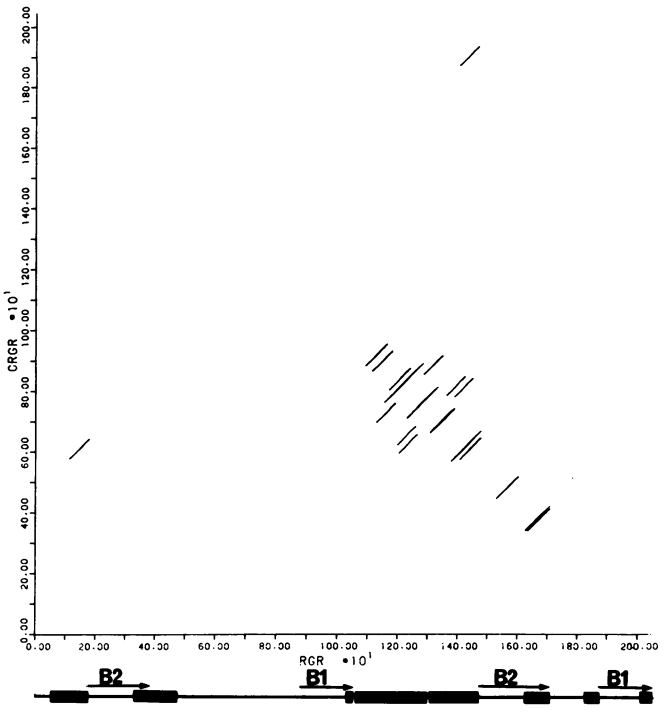

Fig. 5. Features of the wild-type W sequence. Segments of A+T-rich DNA are indicated by black boxes and *Alu*-equivalent (B1 and B2) repeats by arrows with the arrowhead on the 3′ side. The upper diagram is a dot-matrix analysis which compares the W sequence with its complementary strand. A dash indicates at least 53% homology over more than 60 bp. Nucleotide numbering as in Figure 4.

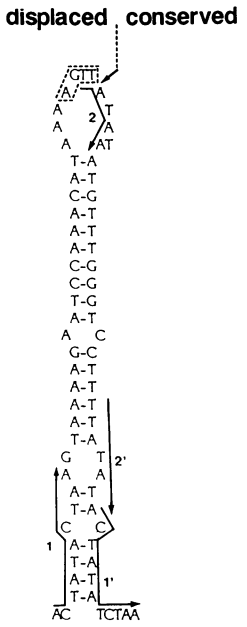

Fig. 6. Possible stem-loop structure and other sequence features around the NJ 551,562 breakpoint in the W DNA. The stem-loop structure is that predicted for fragment 1187−1257 by the program of Zuker and Stiegler (1981). 1,1′ and 2,2′ indicate direct repeats. An arrow is placed between sequences displaced or conserved in the novel junction sequence. This coincides with a topo I cleavage consensus sequence (boxed in dotted lines).

*Eco*RI or *Pst*I. The sizes of the genomic fragments revealed by this probe in the $HC_{50}551$ line are identical to those expected from the restriction maps, ruling out the possibility of a cloning artifact in this region. In addition to the wild-type fragments from the W gene (12 kb *Eco*RI, 3.5 kb and 7.2 kb *Bam*HI, 8.8 kb *Pst*I), this probe detected in the wild-type DNA a 4.7 kb *Eco*RI fragment, a 2.8 kb *Bam*HI fragment and a 2.4 kb *Pst*I fragment. These non-W fragments represent the second wild-type counterpart of the junction. In HC5 and its derivatives, and in this cell family only, the probe detects the novel fragments (3.1 kb *Eco*RI, 6.6 kb *Pst*I and 0.8 kb *Bam*HI) bearing the junction between the W and non-W DNAs: these fragments are not present in wild-type DNA; as already noted in previous analyses, they are present at single copy level in HC5 and selectively amplified in its derivatives $HC_{10}55$ and $HC_{50}551$. The non-W fragments present in the wild-type DNA are not amplified in HC5 and its derivatives (Figure 3), nor in many other amplified lines (unshown), but are nevertheless slightly (3−4 fold) amplified in $HC_{50}611$ DNA. These results indicate that the corresponding region can infrequently be incorporated in amplified units created in different cell lineages: it may lie on one edge of an amplifiable domain common to all cell lines.

We did not detect in HC5 or its derivatives any fragments which could from their size correspond to the reciprocal joint. The 10 kb *Eco*RI fragment seen in HC5 may correspond to a W or non-W novel joint created during the first amplification step and lost upon further amplification. Based on their sizes, the novel fragments appearing in the $HC_{50}611$ lanes correspond to previously identified W novel joints.

Nucleotide sequence of the W recombinogenic region and of NJ 551,562

The 2.6 kb *Pst*I−*Bam*HI wild-type fragment (fragment a, Figure 1), which contains the breakpoint corresponding to NJ 551,562 and the 3.1 kb novel *Eco*RI fragment carrying the NJ 551,562 junction (fragment b) were subcloned in M13 vectors. The 1581 bp sequence ending at the *Bam*HI site of the 2.6 kb *Pst*I−*Bam*HI fragment and the 1908 bp sequence beginning at the left *Eco*RI site of the 3.1 kb *Eco*RI fragment were determined on both strands and are shown in Figure 4. Homology between the NJ 551,562 sequence and the wild-type sequence starts at position 1222. The strict identity of both sequences in the segment 1222−1581 and the identity of the restriction maps of GMA 32 and NJ 551,562 DNA to the right of the junction suggest that the NJ 551,562 sequence down from nucleotide 1222 can be taken as a faithful copy of the corresponding wild-type sequence. Therefore 2046 bp of the wild-type W sequence are available for analysis.

The main features of this wild-type sequence are summarized in Figures 4 and 5. A first peculiarity is the presence of eight long A+T-rich segments: segment 53−175 (78% A+T), segment 328−473 (79.5% A+T), segment 1024−1049 (84.6% A+T), segment 1058−1297 (73.3% A+T), segment 1307−1467 (78.1% A+T), segment 1619−1693 (89.3% A+T), segment 1822−1862 (70.7% A+T) and segment 2007−2043 (86.5% A+T). These are frequently framed by G+C-rich boxes: GGGGCTGG in position 176−183, GCCCTGATGG in position 474−482, CCCTGTCCCG in position 1014−1023, GCAGGTGGG in position 1050−1058, GGGGGGGGC in position 1298−1306, GGGGGGC in position 1468−1474, GGGGGC in position 1694−1699, GGAGGGG in position 1815−1821, and GCCAGGC in position 1867−1873.

Second, a homology search in the Alamos GenBank database

revealed the presence of four Chinese hamster *Alu*-equivalent repeats (Haynes *et al.*, 1981a, 1981b): (i) two members of the B2 family (177−381 and 1476−1690) framed by respectively 17−18 bp and 10−11 bp imperfect direct repeats and (ii) two members of the B1 family (886−1049 and 1866−2032) framed by respectively 26−28 bp and 14 bp direct repeats. These four repetitive elements are oriented in the same direction, with the A-rich tail in the 3′ direction and are flanked on either or both sides by A+T-rich DNA. The A-rich tails fully account for only three of the eight A+T-rich DNA segments.

Finally, a dot-matrix analysis revealed an enrichment in long imperfect palindromes in region 1100−1500 (Figure 5). This is not solely due to the bias in nucleotide composition, since both other A+T-rich regions from the W sequence or obtained by random simulation (unshown) failed to produce a similar pattern. The number of palindromes thus revealed varies with the stringency of the search: 12 are detected when at least 53% of homology is asked over more than 60 nucleotides. Only three of these are retained at minimum match of 56%, whilst the best palindrome which is centered on position 1221 (see below) is still observed when 70% homology is required.

The wild-type sequence and the NJ 551,562 sequence show no significant homology to the left of the junction (nucleotide 1221). The wild-type W sequence around the recombination site shows several interesting features: first, it consists of a long A+T-rich region framed by *Alu*-equivalent repeats and enriched in long imperfect palindromes (Figure 5). Second, the crossover site is located at the center of the most perfect of these palindromes (Figure 6). This palindrome is itself framed by 11 bp long imperfect (10/11) direct repeats located at positions 1187−1196 and 1247−1257. Other short direct repeats were also identified within the palindrome. Notably, the heptamer TTATATA at the crossover position is repeated 21 bp downstream. Third, the tetranucleotide 5′ AGTT is found immediately 5′ to position 1222. The tetranucleotide consensus found immediately 5′ to most topoisomerase I (topo 1) cleavage sites is 5′ A/T C/G A/T T (Been *et al.*, 1984). The end of the displaced W DNA therefore coincides with a possible topo I cleavage site (Figure 6).

The non-W sequence of NJ 551,562 failed to reveal any convincing homology with the W sequence by various computer searches. The only feature it shares with the W DNA is the presence of A+T-rich stretches at and around the junction (Figure 4: segment 905−997 (71% A+T), segment 1119−1195 (70.1% A+T) and segment 1206−1221 (75% A+T)). Another remarkable feature is the simple $(GT)_{31}$ repeat sequence found 1 kb away from the junction (Figure 4).

Predicted secondary structure around the breakpoint

The secondary structure prediction program of Zucker and Stiegler (1981) was applied to fragment 1187−1257 of the GMA32 W sequence. The resulting stucture is shown in Figure 6. It can be seen that the topo 1 consensus site at the crossover site is placed at the top of a stem-loop structure of significant stability ($\Delta G° = -18.4$ kcal.mol^{-1}). This energy value is exceptionally low since the mean energy ± standard deviation of predicted secondary structures for 10 random sequences of the same length and base composition is -7.3 ± 2 kcal.mol^{-1}. The same structure was predicted using varying windows including up to 300 bp of the surrounding DNA, although overlapping palindromes are present in the neighborhood (Figure 5).

Discussion

A short recombinogenic region is present in the amplified domain of AMPD overproducing fibroblasts

We previously described a non-random distribution of novel joints along the amplified unit of AMPD overproducing Chinese hamster fibroblasts. Using cDNA probes and/or cosmid probes covering 120 kb of coamplified DNA out of the W gene (Debatisse *et al.*, 1986, and unpublished data) we detected only four different novel joints in the DNA of six highly amplified cell lines. On the contrary, at least nine novel joints were detected in the same cell lines with the W cDNA probe (which scans 20 kb of DNA). The W gene is included in the amplified unit but is not the gene under selection. We considered the possibility that overproduction of the W protein might be toxic and hence cause selection for rearrangements which turn off transcription of the gene. However selection for deleterious rearrangements of the W gene cannot explain the non-random distribution of novel joints inside the W gene itself: twelve distinct novel joints were detected within 2.6 kb of a W intron in the same six cell lines, and none in 8.1 kb of the immediate surroundings (this work). The frequency of novel joints in this short region is therefore much higher than in the rest of the W gene and is about 150 fold higher than in the rest of coamplified DNA (12 joints in 2.6 kb versus 4 joints in 120 kb): all the observations are simply accounted for by the presence of a recombinational hotspot within the 2.6 kb of the W gene intron where these novel joints map. This is also supported by the fact that these joints were formed in different cell lineages and at various stages of the amplification process and may explain the difficulties experienced in cloning this region if it is also recombinogenic in *Escherichia coli*.

The clustering of novel joints discovered in the AMPD system contrasts with their distribution observed in CAD and DHFR amplification systems, where hotspots were not detected. Each system has been studied in cells of different species (Chinese hamster, Syrian hamster, mouse), which could affect the results observed as much as the difference in locus involved. Alternatively, the possibility remains that hotspots exist in these systems but escaped so far detection (Giulotto *et al.*, 1986).

Why is this region a hotspot?

The W hotspot is enriched in *Alu*-like repetitive sequences and palindromes. Genomic instability has repeatedly been found to be associated with clusters of human *Alu* repeats (Calabretta *et al.*, 1982; Barsh *et al.*, 1983; Lehrman *et al.*, 1985, 1986; Hobbs *et al.*, 1985); several deletion endpoints in the β-globin gene cluster also fall in *Alu* repeat sequences (Henthorn *et al.*, 1986; see also Lehrman *et al.*, 1987a, 1987b, for a review of human mutations involving breakage of DNA in *Alu* repeats); and transfected non-viral DNA integrates preferentially in repetitive sequences (Kato *et al.*, 1986). Palindromes have also been found at sites of DNA breakage associated with different types of chromosomal recombination (Henthorn *et al.*, 1986; Krawinkel *et al.*, 1986). The clustering of amplification novel joints in the intron of the W region indicates that *Alu*-like repeat sequences and palindromes may also be involved in recombination associated with gene amplification.

The A+T-rich segments of the W hotspot may also define preferred recombination substrates. In fact, retroposons (Daniels and Deininger, 1985; Furano *et al.*, 1986) and possibly RNA- and DNA-tumour proviruses (Scherrer and Moreau, 1985; Shi *et al.*, 1984) have a preference for A+T-rich integration sites.

The lower melting energy of A+T-rich DNA might facilitate the opening of the duplex and the entry of recombination enzymes. It has also been noted that long polydA.polydT segments cannot wrap around nucleosomes (Kunkel and Martinson, 1981; Prunell, 1982). Moreover DNA is curved at places where A_n tracts recur with a helical periodicity (Koo *et al.*, 1986). Both nucleosome free gaps and sites of curvature could provide signals or entry sites for recombination enzymes. Sullivan and Lilley (1986) have demonstrated that the presence of A+T-rich DNA fragments can profoundly alter over significant lengths of DNA the kinetic parameters of structural transitions such as extrusion of a cruciform from an inverted repeat. This may be relevant to our case since one breakpoint is at the center of a long palindrome located within a long A+T-rich segment. These authors suggest that A+T-rich DNA is likely to be structurally polymorphic, flexible, and deformable. Also, A+T-rich DNA may belong to 'cryptically simple' sequences which have probably originated by slippage-like mechanisms (Tautz *et al.*, 1986) and may contain recombination signals (Jeffreys *et al.*, 1985).

Sequence peculiarities at the NJ 551,562 breakpoint

To determine if the breakpoints clustered in the W region could be associated with local peculiarities, we sequenced the novel junction NJ 551,562. We found the surroundings of the NJ 551,562 breakpoint enriched in imperfect palindromes. The best palindrome could form a stable stem-loop structure with the breakpoint at the top. This situation is strikingly similar to that recently observed in the human β-globin gene cluster: the 3′ breakpoint of a deletion falls 3 bp from the center of a 160 bp perfect palindrome (Henthorn *et al.*, 1986).

Association of palindromic secondary structures with different types of recombination has been repeatedly reported (Bullock *et al.*, 1984; Glickman and Ripley, 1984 and references therein; Krawinkel *et al.*, 1986; Nalbantoglu *et al.*, 1986) but the breakpoints were not always located at the top of a potential stem-loop structure. While it is possible that extrusion of the cruciform does not occur *in vivo* (Sinden *et al.*, 1983; Courey and Wang, 1983) and that the location of the breakpoint simply reflects the preference of dimeric DNA binding proteins for sites with a dyad symmetry, enzymes which cut at the base or inside the single strand loop of a cruciform are known (Panayotatos and Wells, 1981; Lilley, 1980, 1981; West and Körner, 1985; Mizuuchi *et al.*, 1982; Lilley and Kemper, 1984; de Massy *et al.*, 1984) and related enzymes could act during illegitimate recombination. It is therefore of interest that we found a possible topo I cleavage site associated with the breakpoint. Bullock *et al.* (1984, 1985) pointed out that many illegitimate crossover points in viruses and chromosomes are associated with potential topo I recognition sites and established that crossover points for SV40 excision are associated wtih eukaryotic topo I cleavage sites *in vitro*. It is also possible that the short direct repeats present around the breakpoint played a role in the recombination process as suggested by Hasson *et al.* (1984) for SV40 illegitimate recombination.

In conclusion, the NJ 551,562 breakpoint is found at a remarkable location within the W hotspot. We must await sequencing of additional novel joints to determine if breakpoints fall frequently within this palindrome. The center of this palindrome could define a preferred but not obligatory site in the recombinational hotspot.

The DNA at the left of the NJ 551,562 junction is not from the W region. Blot hybridization indicated that this non-W DNA can sometimes be incorporated in some of the amplified units of different cell lines and therefore probably lies on the edge of the amplified domain. However, it is not an inverted duplicated copy of the sequences present at the other side of the junction. Therefore, this junction between adjacent segments of amplified DNA is not of the 'head-to-head' type described by others (Ford and Fried, 1986; Saito and Stark, 1986; Nalbantoglu and Meuth, 1986). Probably both 'head-to-head' and 'head-to-tail' joints coexist in the amplified DNA because we also found inversion joints in the W region (unpublished).

A $(GT)_{31}$ simple repeat is present in the non-W sequence 1 kb upstream of the junction. Although circumstantial evidence linking poly(GT) with increased recombination has been reported (Flanagan *et al.*, 1984; Slightom *et al.*, 1980; Stringer, 1982, 1985; Swanstrom *et al.*, 1983; Mager *et al.*, 1985), such elements are dispersed in the eukaryotic genome and present on average at a frequency of one for every 30 kb (Hamada *et al.*, 1982): given the relatively long distance between the $(GT)_{31}$ element and the actual crossover point, its presence may therefore be coincidental.

We failed to find any significant homology between the wild-type W sequence and the non-W part of the recombinant junction, but this conclusion cannot be made definitive until we have cloned and sequenced the wild-type non-W DNA around both sides of the breakpoint. However, as discussed above, a common feature of the non-W sequence with the W sequence is the presence of A+T-rich (>70%) DNA at and around the breakpoint. It has also been found that some novel joints in the CAD system were formed by recombination between A+T-rich segments (O.Brison, personal communication).

Recombination hotspots, chromosome structure, and gene amplification

DNA in the nucleus is organized in loops 50−100 kb in size which are affixed to the 'nuclear matrix' or 'chromosomal scaffold' and are believed to correspond to separate replicons (Benyajati and Worcel, 1976; Paulson and Laemmli, 1977; Vogelstein *et al.*, 1980; Cook and Brazell, 1976). Specific matrix attachment sites have recently been identified in several instances (Mirkovitch *et al.*, 1984; Cockerill and Garrard, 1986; Gasser and Laemmli, 1986) and were found to be A+T-rich DNA segments. Moreau *et al.* (1981, 1982) have described clusters of A+T-rich segments which appear to punctuate eukaryotic genomes (including rat and mouse) at regular intervals (50−100 kb in vertebrates) and were proposed to define structural and functional domains in the eukaryotic DNA. Although recent hypotheses concerning the origin of DNA regions with biased nucleotide composition do not predict a regular organization (Tautz *et al.*, 1986), it is possible, as discussed by Scherrer and Moreau (1985), that these A+T-rich segments represent attachment sites for the nuclear scaffold of chromatin loops. Interestingly, it was also found that the map of repetitive sequences coincides with that of the A+T-rich segments in the globin gene domains of *Xenopus* and the chicken (Kretsovali *et al.*, 1986). It is therefore possible that novel joints cluster in the W hotspot because that DNA sequence defines some physical boundary in the chromatin. This boundary could correspond to the end of a chromosomal loop, as previously considered by Hamlin *et al.* (1984). Although amplified domains appear to be frequently longer than a single chromosomal loop and may cover several loops, the basal portion of each loop may be a preferred site for recombination.

We also would like to point out that the W hotspot may contain a replication origin since there is evidence suggesting that mammalian replication origins are located in DNA regions enrich-

ed in *Alu* repeat sequences and snap back sequences (Zannis-Hadjopoulos *et al.*, 1984, 1985; Anachkova *et al.*, 1985), and A+T-rich DNA (Montiel *et al.*, 1984; Roth *et al.*, 1983) and are permanently attached to the nuclear matrix (Dijkwell *et al.*, 1986; Razin *et al.*, 1986). Illegitimate recombination hotspots in prokaryotes also coincide with a nick present at a replication origin, in a region which is enriched in palindromes (Michel and Ehrlich, 1986a, 1986b).

Materials and methods

Cell lines

The GMA32 line of Chinese hamster fibroblasts and its coformycin-resistant variants have been described previously (Debatisse *et al.*, 1982, 1984, 1986).

Phage cloning and mapping

Genomic libraries of GMA32 and HC_{50}551 were prepared in λ EMBL4 according to Frischauf *et al.* (1983). These libraries were screened using the method of Benton and Davis (1977) with a full length W cDNA probe (Debatisse *et al.*, 1986) ^{32}P-labelled by nick-translation (2×10^8 c.p.m./μg). Recombinant phage DNAs were prepared either by a large scale procedure (Maniatis *et al.*, 1982) or by the quick-prep procedure of Davis *et al.* (1980). Phage restriction maps were obtained by cos-mapping (Rackwitz *et al.*, 1984) and refined when necessary by analysis of single digests of phage DNAs on ethidium bromide stained agarose gels.

DNA sequencing and analysis

Relevant restriction fragments were purified from an agarose gel by electrophoresis onto a DEAE NA45 membrane (Schleicher and Schull) followed by elution in 1 M NaCl, 0.05 M L-arginine at 70°C for two hours and ethanol precipitation. These fragments were subcloned in M13 vectors tg130, tg131 (Kieny *et al.*, 1983) or mp18 (Norrander *et al.*, 1983) for sequencing of both strands. The method of Henikoff (1984) was used to create a set of overlapping unidirectional deletion subclones for each strand. Nucleotide sequences were determined on both strands by the method of Sanger *et al.* (1977). Sequences were assembled and analyzed with the computer facilities of the Institut Pasteur (SASIP: Claverie, 1984).

Nucleic acids hybridization

Southern transfers were performed employing Hybond-N membranes (Amersham) as indicated by the supplier. Repeated DNA sequences were removed from the probes when necessary by reannealing with wild-type GMA32 DNA as described (Giulotto *et al.*, 1986) until a Cot value of 10 was reached.

Acknowledgements

We thank M.-A.Akimenko, B. Mariamé and J.-J.Panthier for helpful suggestions, the members of the Unité d'Informatique Scientifique for their help with the computer, K.Scherrer and F.Schaeffer for discussions, M.Berry for expert technical assistance, P.Avner and A.Dautry-Varsat for critical reading of the manuscript, and M.Mahérou for typing the manuscript. This work was supported in part by the Université P. et M.Curie, the INSERM (grant no. 861022), the Ligue Nationale Française contre le Cancer and the Fondation pour la Recherche Médicale Française.

References

Anachkova,B., Russev,G. and Altmann,H. (1985) *Biochem. Biophys. Res. Commun.*, **128**, 101–106.
Ardeshir,F., Giulotto,E., Zieg,J., Brison,O., Liao,W.S.L. and Stark,G.R. (1983) *Mol. Cell. Biol.*, **3**, 2076–2088.
Barsh,G.S., Seeburg,P.H. and Gelinas,R.E. (1983) *Nucleic Acids Res.*, **11**, 3939–3958.
Been,M.D, Burgess,R.R. and Champoux,J. (1984) *Nucleic Acids Res.*, **12**, 3097–3113.
Benton,W.E. and Davis,R.W. (1977) *Science*, **196**, 180–182.
Benyajati,C. and Worcel,A. (1976) *Cell*, **9**, 393–407.
Bullock,P., Forrester,W. and Botchan,M. (1984) *J. Mol. Biol.*, **174**, 55–84.
Bullock,P., Champoux,J.J., Botchan,M. (1985) *Science*, **230**, 954–958.
Caizzi,R. and Bostock,C.J. (1982) *Nucleic Acids Res.*, **10**, 6597–6618.
Calabretta,B., Robberson,D.L., Barrera-Saldana,H.A., Lambrou,T.P. and Saunders,G.F. (1982) *Nature*, **296**, 219–225.
Claverie,J.M. (1984) *Nucleic Acids Res.*, **2**, 397–407.
Cockerill,P.N. and Garrard,W.T. (1986) *Cell*, **44**, 273–282.
Cook,P. and Brazell,I. (1976) *J. Cell Sci.*, **22**, 287–302.
Courey,A.J. and Wang,J.C. (1983) *Cell*, **33**, 817–829.
Cowell,J.K. (1982) *Annu. Rev. Genet.*, **16**, 21–59.
Daniels,G.R. and Deininger,P.L. (1985) *Nucleic Acids Res.*, **13**, 8939–8954.
Davis,R.W., Botstein,D. and Roth,J.R. (1980) *A Manual For Genetic Engineering: Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory Press, New York.
Debatisse,M., Berry,M. and Buttin,G. (1982) *Mol. Cell. Biol.*, **2**, 1346–1353.
Debatisse,M., Robert de Saint Vincent,B. and Buttin,G. (1984) *EMBO J.*, **3**, 3123–3127.
Debatisse,M., Hyrien,O., Petit-Koskas,E., Robert de Saint Vincent,B. and Buttin,G. (1986) *Mol. Cell. Biol.*, **6**, 1776–1781.
Dijkwel,P.A., Wenink,P.W. and Poddighe,J. (1986) *Nucleic Acids Res.*, **14**, 3241–3249.
Federspiel,N.A., Beverley,S.M., Schilling,J.W. and Schminke,R.T. (1984) *J. Biol. Chem.*, **259**, 9127–9140.
Flanagan,J.C., Lefranc,M.P. and Rabbitts,T.H. (1984) *Cell*, **36**, 681–688.
Ford,M. and Fried,M. (1986) *Cell*, **45**, 425–430.
Frischauf,A.-M., Lehrach,H., Poutska,A. and Murray,N. (1983) *J. Mol. Biol.*, **70**, 827–842.
Furano,A.V., Somerville,C.S., Tsichlis,P.N. and d'Ambrosio,E. (1986) *Nucleic Acids Res.*, **14**, 3717–3727.
Gasser,S.M. and Laemmli,U.K. (1986) *Cell*, **46**, 521–530.
Gilson,E., Clément,J.-M., Brutlag,D. and Hofnung,M. (1984) *EMBO J.*, **3**, 1417–1421.
Giulotto,E., Saito,I. and Stark,G.R. (1986) *EMBO J.*, **5**, 2115–2121.
Glickman,B.W. and Ripley,L.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 512–516.
Hamada,H., Petrino,M.G. and Kakunaga,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6465–6469.
Hamlin,J.L., Milbrandt,J.D., Heintz,N.H. and Azizkhan,J.C. (1984) *Int. Rev. Cytol.*, **90**, 31–82.
Hasson,J.F., Mougneau,E., Cuzin,F. and Yaniv,M. (1984) *J. Mol. Biol.*, **177**, 53–68.
Haynes,S.R. and Jelinek,W.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6130–6134.
Haynes,S.R., Toomey,T.P., Leinwand,L. and Jelinek,W.R. (1981) *Mol. Cell. Biol.*, **1**, 573–583.
Henikoff,S. (1984) *Gene*, **28**, 351–359.
Henthorn,P.S., Mager,D.L., Huisman,T.H.J. and Smithies,O. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5194–5198.
Hobbs,H.H., Lehrman,M.A., Yamamoto,T. and Russell,D.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7651–7655.
Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) *Nature*, **314**, 67–72.
Kafatos,F.C., Orr,W. and Delidakis,C. (1985) *Trends Genet.*, **1**, 301–306.
Kato,S., Anderson,R.A. and Camerini-Otero,R.D. (1986) *Mol. Cell. Biol.*, **6**, 1787–1795.
Kieny,M.P., Lathe,R. and Lecocq,J.P. (1983) *Gene*, **26**, 91–99.
Koo,H.-S., Wu,H.-M. and Crothers,D.M. (1986) *Nature*, **320**, 501–506.
Krawinkel,U., Zoebelein,G. and Bothwell,A.L.M. (1986) *Nucleic Acids Res.*, **9**, 3871–3882.
Kretsovali,A., Marcaud,L., Moreau,J. and Scherrer,K. (1986) *Mol. Gen. Genet.*, **203**, 193–201.
Kunkel,G.R. and Martinson,H.G. (1981) *Nucleic Acids Res.*, **9**, 6869–6888.
Lehrman,M.A., Schneider,W.J., Südhof,T.C., Brown,M.S., Goldstein,J.L. and Russel,D.W. (1985) *Science*, **227**, 140–146.
Lehrman,M.A., Russel,D.W., Goldstein,J.L. and Brown,M.S. (1986) *Proc. Natl. Acad. Sci. USA*, **831**, 3679–3683.
Lehrman,M.A., Russel,D.W., Goldstein,J.L. and Brown,M.S. (1987a) *J. Biol. Chem.*, **262**, 3354–3361.
Lehrman,M.A., Goldstein,J.L., Russel,D.W. and Brown,MS. (1987b) *Cell*, **48**, 827–835.
Lilley,D.M.J. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6468–6472.
Lilley,D.M.J. (1981) *Nature*, **292**, 380–382.
Lilley,D.M.J. and Kemper,B. (1984) *Cell*, **36**, 413–422.
Mager,D.L., Henthorn,P.S. and Smithies,O. (1985) *Nucleic Acids Res.*, **13**, 6559–6575.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.
Marsden,M.P.F. and Laemmli,U.K. (1979) *Cell*, **17**, 849–858.
de Massy,B., Studier,F.W., Dorgai,L., Appelbaum,E. and Weisberg,R.A. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 715–726.
Michel,B. and Ehrlich,S.D. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 3386–3390.
Michel,B. and Ehrlich,S.D. (1986b) *EMBO J.*, **5**, 3691–3696.
Milbrandt,J.D., Heintz,N.H., White,W.C., Rothman,S.M .and Hamlin,J.L. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6043–6047.
Mirkovitch,J., Mirault,M.-E. and Laemmli,U.K. (1984) *Cell*, **39**, 223–232.
Mizuuchi,K., Kemper,B., Hays,J. and Weisberg,R.A. (1982) *Cell*, **29**, 357–365.
Montiel,J.F., Norbury,C.J., Tuite,M.F., Dobson,M.J., Mills,S.J., Kingsman,A.J. and Kingsman,S.M. (1984) *Nucleic Acids Res.*, **12**, 1049–1068.

Moreau,J., Matyash-Smirniaguina,L. and Scherrer,K. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1341–1345.
Moreau,J., Marcaud,L., Maschat,F., Kejzlarova-Lepesant,J., Lepesant,J.-A. and Scherrer,K. (1982) *Nature*, **295**, 260–262.
Nalbantoglu,J., Hartley,D., Phear,G., Tear,G. and Meuth,M. (1986) *EMBO J.*, **5**, 1199–1204.
Nalbantoglu,J. and Meuth,M. (1986) *Nucleic Acids Res.*, **14**, 8361–8371.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101–106.
Panayotatos,N. and Wells,R.D. (1981) *Nature*, **289**, 466–470.
Paulson,J.R. and Laemmli,U.K. (1977) *Cell*, **12**, 817–828.
Prunell,A. (1982) *EMBO J.*, **1**, 173–179.
Rackwitz,H.-R., Zehetner,G., Frischauf,A.-M. and Lehrach,H. (1984) *Gene*, **30**, 195–200.
Razin,S.V., Kekelidze,M.G., Lukanidin,E.M., Scherrer,K. and Georgiev,G.P. (1986) *Nucleic Acids Res.*, **14**, 8189–8207.
Roth,G.E., Blanton,H.M., Hager,L.J. and Zakian,V.A. (1983) *Mol. Cell. Biol.*, **3**, 1898–1908.
Saito,I. and Stark,G.R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8664–8668.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scherrer,K. and Moreau,J. (1985) In *Proceedings of the 16th. FEBS Congress, Part B.* VNU Science Press, pp. 105–122.
Schimke,R.T. (ed.) (1982) *Gene Amplification*. Cold Spring Harbor Laboratory Press, New York.
Schimke,R.T. (1984) *Cell*, **37**, 705–713.
Shih,C.K., Linial,M., Goodenow,M.M. and Hayward,W.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4697–4701.
Sinden,R.R., Broyles,S.S. and Petitjohn,D.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1797–1801.
Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell*, **21**, 627–638.
Stark,G.R. and Wahl,G.M. (1984) *Annu. Rev. Biochem.*, **53**, 447–491.
Stark,G.R. (1986) *Cancer Surveys*, **5**, 1–23.
Stringer,J.R. (1982) *Nature*, **296**, 363–366.
Stringer,J.R. (1985) *Mol. Cell. Biol.*, **5**, 1247–1259.
Sullivan,K.M .and Lilley,D.M.J. (1986) *Cell*, **47**, 817–827.
Swanstrom,R., Parker,R.C., Varmus,H.E. and Bishop,J.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2519–2523.
Tautz,D., Trick,M. and Dover,G. (1986) *Nature*, **322**, 652–656.
Vogelstein,B., Pardoll,D.M. and Coffey,D.S. (1980) *Cell*, **22**, 79–85.
West,S.C. and Körner,A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6445–6449.
Zannis-Hadjopoulos,M., Kaufmann,G. and Martin,R.G. (1984) *J. Mol. Biol.*, **179**, 577–586.
Zannis-Hadjopoulos,M., Kaufmann,G., Wang,S.-S., Lechner,R.L., Karawya,E., Messe,J. and Martin,R.G. (1985) *Mol. Cell. Biol.*, **5**, 1621–1629.
Zuker,M. and Stiegler,P. (1981) *Nucleic Acids Res.*, **9**, 133–148.

Received on March 25, 1987; revised on May 11, 1987