Oncogene (1991), **6**, 73–80

© Macmillan Press Ltd, 1991

Transcriptional activation of the melanoma inducing *Xmrk* oncogene in *Xiphophorus*

Dieter Adam, Winfried Mäueler & Manfred Schartl

Genzentrum, Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-8033 Martinsried, Federal Republic of Germany

The melanoma inducing locus of *Xiphophorus* encodes a tumorigenic version of a novel putative receptor tyrosine kinase (*Xmrk*). To elucidate the mechanism of oncogenic activation of *Xmrk*, we compared the structure and expression of two oncogenic loci with the corresponding proto-oncogene. Only minor structural alterations were found to be specific for the oncogenic *Xmrk* genes. Marked overexpression of the oncogene transcripts in melanoma, which are approximately 1 kb shorter than the proto-oncogene transcript, correlates with the malignancy of the tumors. The tumor transcripts are derived from an alternative transcription start site that is used only in the oncogenic loci. Thus, oncogenic activation of the melanoma inducing *Xmrk* gene appears primarily to be due to novel transcriptional control and overexpression.

Introduction

Proto-oncogenes constitute essential members of the vertebrate genome. Upon quantitative and/or qualitative alteration they become involved in tumor development as oncogenes (reviewed in Sinkovics, 1988; Tronick & Aaronson, 1988; Boettiger, 1989). Most activated oncogenes have been detected and isolated from solid tumors or tumor-derived cell lines. Such tumor cells have undergone a dramatic evolution which makes them different from the first neoplastically transformed cell, being the tumor founder cell. It has been shown in many incidences that a variety of cooperative events has to occur to enable such a founder cell (or founder cell population) to give rise to what is recognized as a tumor (Weinberg, 1989). In studying the primary events of neoplastic transformation, the problem arises as to whether the activated oncogene isolated from the tumor is actually involved in the primary steps of transformation or in later steps of tumor progression. The situation is much better in experimental systems where the gene causally responsible for tumor induction is clearly defined by classical Mendelian genetics. The Central American teleost genus *Xiphophorus* represents such a vertebrate model uniquely suited for studies on the primary events of melanoma formation (for review see Anders *et al.*, 1984; Schwab, 1987; Vielkind *et al.*, 1989).

The pigmentation of *Xiphophorus* is composed of two melanophore types: micromelanophores responsible for the uniform grey body coloration and macromelanophores which form characteristic spot patterns e.g. present in *X. maculatus* (Gordon, 1927) however absent in several populations of *X. helleri*. In hybrid crosses of *X. maculatus* with unspotted *X. helleri*, the F_1-offspring show enhanced macromelanophore pigmentation diagnosed as benign melanoma. Backcrossing of the F_1-hybrids with *X. helleri* leads to an uncontrolled macromelanophore growth associated with malignant melanoma (Gordon, 1927; Häussler, 1928; Kosswig, 1928) which was explained by abnormal 'regulation' of a single sex chromosomal genetic locus of *X. maculatus* (Gordon, 1958; Atz, 1962; Kosswig, 1965; Zander, 1969), designated *Tu* (Anders, 1978). Based on extensive genetic studies, it has been proposed that *Tu* is normally under the control of one or several regulatory genes (*R*) acting as tumor suppressors (Ahuja & Anders, 1976). In *X. helleri* these suppressor genes are absent. Thus, the *Tu* locus might become activated in hybrid fish in which the suppressor genes have been diluted out by two or more backcrosses with *X. helleri*. Reintroduction of *R* by repeated backcrossing to *X. maculatus* leads to the benign phenotype of non-proliferative macromelanophore spots indicating that the *Tu* locus remains structurally unchanged during the process of hybridization conditioned oncogenic activation (Schwab, 1987).

We have recently cloned the melanoma inducing gene from the *Tu* locus of *X. maculatus*. It encodes a novel receptor tyrosine kinase which is closely related to the EGF-receptor (Wittbrodt *et al.*, 1989). This gene, which we designated *Xmrk*, is present in one copy each on the X- and Y-chromosome (designated X and Y), and both copies are shown to cause melanoma in the absence of the other gene. In addition, a third copy of *Xmrk* is present in all individuals including the hybrids of *X. maculatus* and *X. helleri* (designated INV). This copy shows no association with melanoma formation and may represent the proto-oncogene.

To obtain information on the possible mode of melanoma oncogene activation in the hybrid genome versus the situation in the purebred parental fish, we have analysed structure and expression of the proto-oncogene locus INV and the sex-chromosomal oncogene loci X and Y. Here, we present evidence that oncogenic activation is linked to overexpression of the oncogene loci due to an altered transcriptional control.

Results

Structural features of the three Xmrk *loci of* X. maculatus

The restriction fragment length polymorphism of the three *Xmrk* genes (Schartl, 1988; Adam *et al.* 1988, Zechel *et al.*, 1988) originates from polymorphic EcoRI sites within their 3′ regions (Wittbrodt *et al.*, 1989). We have cloned genomic DNA which represents the INV-,

Correspondence: M. Schartl

Received 23 July 1990; accepted in revised form 15 August 1990

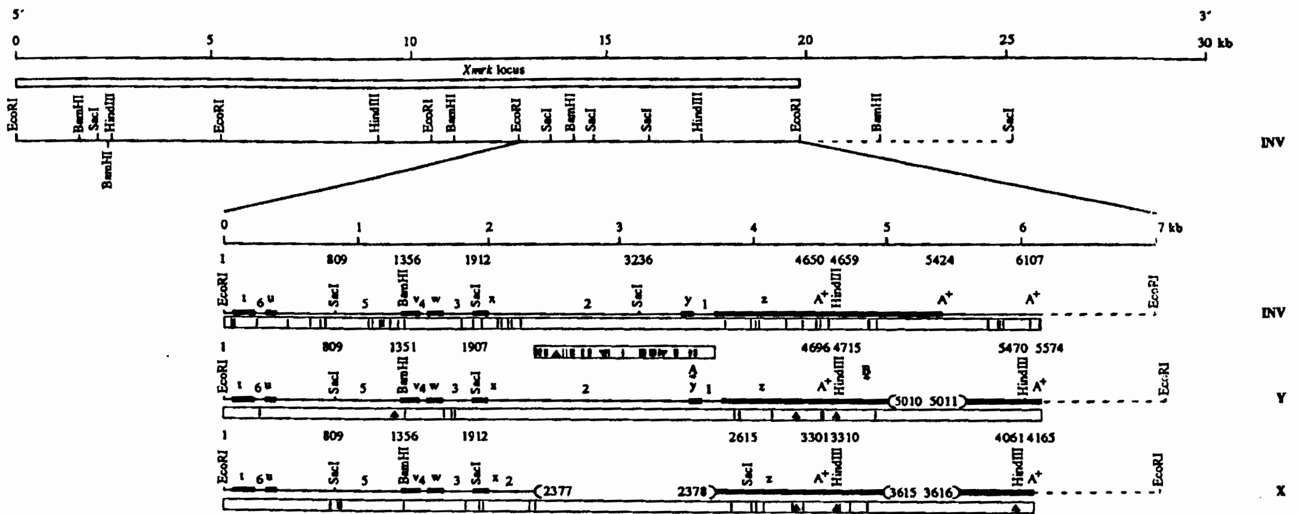

Figure 1 Top: The genomic region of the *Xmrk* gene (shown for the INV-locus). The open bar represents the extent of the *Xmrk* gene. The area represented by a solid line is cloned. Center: The polymorphic 3′ ends of the INV-, X- and Y-locus. The sequenced part is shown by solid lines. Exons t to z are indicated by black boxes, the intervening introns are numbered 1–6. The distribution of point mutations is indicated by bars in the boxes below each locus. Minor deletions in the X- or Y-locus are given by triangles, additional sequence present in the Y-locus by inverted triangles. Large deletions in the oncogenic loci are bounded by semi-circles. Sequence differences between the INV- and Y-locus located in the fragment deleted in the X-locus are shown in between both loci. Positions of polyadenylation consensus sites are indicated by A⁺. The location of oligonucleotides A and B is represented by arrows

X- and Y loci. Figure 1 shows the physical maps of the genomic DNA of the three *Xmrk* loci. No structural differences between the three loci have been detected at the 5′ end of the SacI site (nt 1912) so far. Polymorphic sites were restricted to the 3′ part of the *Xmrk* loci, suggesting only minor alterations, if any, in the region upstream of the SacI site. As the physical mapping already indicated structural differences between the INV, Y- and X-loci, we determined the sequence of the polymorphic regions (Figure 2).

The exon/intron arrangement of the *Xmrk* gene is identical to the human EGF-receptor and *neu* gene (Semba *et al.*, 1985) as well as to *ERBB3* (Kraus *et al.*, 1989). Splice donor and acceptor sites correspond to the published consensus (Mount, 1982). The most striking difference between the three *Xmrk* loci is a deletion of 1344 bp in the X-locus which comprises most of intron 2, the entire exon y and intron 1. Two imperfect direct repeats of 13/12 nt at the borders of the missing sequence are indicative of a recombinative deletion (Jeffreys & Harris, 1982) in the X-chromosomal copy.

The large deleted fragment in the X-locus includes the regular splice acceptor site of exon z (Figure 2). Comparison of the X-locus cDNA (B. Malitschek, personal communication) to the genomic sequence revealed that for splicing of the X-locus transcript a cryptic acceptor site 14 nt downstream of the regular acceptor site in exon z is used, thus generating an in-frame deletion. A second large deletion in the trailer removes 581 bp from the two oncogenic loci (see below).

Including several minor insertions and deletions, 122 sequence differences were found in all three loci, 23 of which affect the coding sequence. Ten mutations result in amino acid changes: three conservative, four semiconservative and three nonconservative changes. Two of the nonconservative changes consist of the X- and Y-loci having proline converted to leucine (aa 960, exon x) and leucine converted to glutamine (aa 1132, exon z). The Y-locus has a tyrosine converted to an asparagine (aa 1014, exon y) which is totally missing in the X-locus (positions refer to the previously characterized Y-chromosomal cDNA sequence [Wittbrodt *et al.*, 1989]). The position of this tyrosine does not correspond to one of the five published autophosphorylation sites of the EGF-receptor (Margolis *et al.*, 1989; Walton *et al.*, 1990) but could represent a *Xmrk* homolog of other potential, as yet uncharacterized autophosphorylation sites that have also been proposed for the EGF-receptor (Honegger *et al.*, 1989).

To obtain information on the phylogenetic relationship between the three *Xmrk* loci, the nucleotide sequences were compared. Based on the distance matrix method of Fitch and Margoliash (Fitch & Margoliash, 1967), the INV-locus shows a larger genetic distance from the X- (69 differences) and Y-locus (58 differences) than these loci display to each other (35 differences).

Different expression of the Xmrk *loci*

In all non-tumorous fish carrying only the INV-locus, low expression of *Xmrk* was found in skin, fins and gills represented by a 5.8 kb transcript (Figure 3a, lanes 1–3). All other tissues tested showed no detectable expression. Northern analyses of melanoma bearing fish revealed two major transcripts, one of 5.8 kb in the corresponding non-tumorous tissues and an additional one of 4.7 kb within the tumorous tissues (Figure 3a, lanes 4–8). To find out if the melanoma-specific 4.7 kb transcript was actually derived from the melanoma inducing X- or Y-locus, we used oligonucleotide A specific for an exon (y) which is present in the INV- and Y loci but not in the X-locus (see Figure 1). Even after overexposure of the filter, oligonucleotide A did not detect the 4.7 kb transcript in melanoma caused by the presence of the X-chromosomal gene locus, while in Y-chromosomal melanoma this transcript was readily detected (Figure 3b, lanes 1 and 2). In a control experiment with oligonucleotide B which is specific for all three *Xmrk* loci (see Figure 1), the 4.7 kb transcripts were detected in both tumor RNAs (Figure 3b, lanes 3 and 4). This shows that

```
INV gaattcaaat atttatttat ctatttagcg gttagttctg gtttgttaaa tgacacactc tgtcctgctg gttcagGGAA TGAACTACCT GGAAGAGCGC CACCTGGTGC ACCGCGACCT  120
Y                                                                   g                    G
X                                                                   g                    G
INV GGCAGCCAGG AACGTCCTGC TGAAAAACCC GAACCACGTC AAGATCACAG ACTTCGGTCT GTCCAAGCTG CTGACGGCTG ACGAGAAGGA ATACCAAGCC GACGGAGGAA AGgtgccatg  240
INV gcaacgcctg actggtttct gtttgctgtt cggactgaaa acatgtcaga gatgaatcac tgctgcatct ctgtgagcag GTTCCCATTA AGTGGATGGC TTTGGAGTCG ATCCTCCAGT  360
Y       t                         c
X       t                         g
INV GGACCTACAC CCATCAGAGC GACGTGTGGA GCTACGgtga ggaatcgtcc ccacagcgcc acctacctgc cttcaccctc tgcttcctgt tagccggttt aaacatctga aaccacacag  480
Y                                                                                                                                 g
X                                                                                                                                 g
INV ccagaggatt tagactgaat aaaatacttc tacactatta gtgaatattg gttgcacatt atcatctctt gttcaaaatg ttaaatgctt gagactcaaa ccagagtgag atttttgtac  600
INV actgaacaga tgctgaacat cagtgttagc atgtaacoct ggttatattg atttattcct gattatattt aatctgatca tggaaatctt accagttttc ctgtttgaaa ttagactgag  720
Y                                                a                                                                              a
X                                                a                                                                              a
INV tgatttgtat tactaaaaat taaaacttac ggcccttagt ttagctaaaa tcccatctta cgttgaaagg acgtggcgtt catttatgga gctccagtag cagtcagtca agccacaaat  840
Y                                  g    t                                      g      g
X                                  g    t                                      c      a
INV ctgaagaagc ctctgactgg agactgttgc agcttaacag ctccacatcc aggcaccaga aacattagta acatagcaac acatcctgga tgactccatc cgttagcata ctctgctaat  960
Y           g                         g t
X           c                         c c
INV ccgcctcatt tgcttttgct gctcgtcttt aaacctctgg ttggctgtga ggtcagaaag atgttggagt aggggatatg atcgatggaa cttgaacttc ttccttctct ctctgctctg 1080
Y                                                                                                                                      g
X                                                                                                                                      c
INV catccatgaa gcctcctttc tacctcctct gggatttggg gtcaaagttc aggtgatatt ttatctacct gaggtccaca aatcgggtag acgaatccgt tgacaggaca gctatgagtc 1200
Y                               a                                                              a          g                     c
X                               a                                                              a          g                     c
INV attctaacta taaatggcta agtgtgttcc tgcagaaaat gaggaaattg ttctttttcc ttttgtgcgg tgttattgct aggctaggct aggctaacca agctgaagtc actcctgtct 1320
Y                                                  a                                   *** **                       a
X                                                  a                                                                 a
INV ctgcagGTGT GACGGTTTGG GAGTTGATGA CGTTCGGATC CAAACCGTAC GACGGCATCC CGGCCAAGGA GATCGCCTCG GTGCTGGAGA ACGGGGAGCG GCTGCCGCAG CCTCCCATCT 1440
Y                                     A  C            A
X                                     A  T            G
INV GCACCATCGA AGTCTACATG ATCATCCTGA AGTgtaagag ctgaccctgc aggtggcgct gcaggctggc tgggctgttt cctcatggta gtgtgttgtc cccgtctccc tgcagGCTGG 1560
INV ATGATCGACC CGTCCAGCAG ACCCAGGTTC AGAGAGCTGG TGGGCGAGTT CTCCCAGATG GCCCGGGACC CGTCCAGGTA CCTGGTCATA CAGgtacgcc accgcccccа cctgacccta 1680
Y                                                                                                                  t
X                                                                                                                  c
INV ctataaaaca ctaacagaac aattctgggt tttcacttta cgtttcagaa aatatatcca aagagtccaa atcaaaatct aaatgttttt cattaagtaa ctgaacatta agaagcgaat 1800
Y                                               t                           t                                                 a
X                                               c                           c                                                 a
INV ctaatcaaat atggaataat aaagtatatt tgactgtttt ggttgcctct gacttgtttc ccagGGCAAC CTGCCGAGTC CGTCTGATCG GAGGCTTTTC TCCCGCCTGC TGAGCTCTGA 1920
Y                              a                                                            T
X                              G                                                            T
INV TGACGACGTG GTCGACGCCG ATGAATACCT GCTGCCGTAC AAACGGATAA ACCGCCAGGG CAGCGAGCCC TGCATCCCGC CGgtaaaaac tctgatgcta accgaacagc aacctgaaac 2040
Y                      C                   A       C
X                      G                   A       T
INV aagtcctgga gtttctggag ttttcattct gacccggaaa gtccgaggtt cctcaccggc tgggtctcac aatccaacat ggctgccgct catacagaca gctgtcaaga ttttcatctg 2160
Y                                         t                                     g                                                    t
X                                         t                                     g                                                    t
INV aatttgtgtc tcaataatct gattcttata aaaccataag accagctaat aaaactgttg ccacaaaatt aaattacata aatcaggatt tctgagcagt ctgggaaaat ctggacttta 2280
Y                            c                                                                              t
X                            c                                                                              t
INV acttcagtaa ctttccaggt ctggatcagc aaaaacaact aaagtgtgga ggatgttgta tttgtaactg cttttccatt tggccatgtt tcatgttcgt ccttctttcc tgttccttct 2400
Y                                                 a                                                      c     t
X                                                 c                                                      t     *** ********** **********
INV tcctatactt ccttgtgtcc ttcaatcctt cattgtcttc ttgtctcatt gaataagaga aacgagaaac ttttatttat ttattccttt gtccgtcatt ctttcatttg tacttttgtt 2520
Y   ag                                          a                g
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV ttttccttcc ttccttcctt cttctcctgt ccttcctacc ttccttaatt actttccttg ttccttcttt caattctttt cgttgtctct tgttttgca  gcttagatat ttaaacoatt 2640
Y    ****                                                       t                           a                                     a
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** 2760
INV cctttattt  acatatgtac cataaattta tcaacatttg gtgttttata ttttaactgg ttttcaaatt tacaatatag ttgttgttgg tattttagcc gtagacagta aagaagctct
Y    a      c        t                                                                        g        g
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV ctcacctgac gagacgtaga gcagtgaatc cgtctctcct taaagacatt tcccagttct tcctcacatc gtgtctttat gaaaccttca aaataaaaac agaa****** ********** 2880
Y           t    *t                                                                                                   gatttg gtgtcttact
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV ********** ***aacatat aaataatatt ctttgccacc tgtgtaatat ttgctttgga ttgttataac acataatatt agaataatat atttataatt ataaataaaa ttgttttatt 3000
Y   tctgtgaaag aatg                            t
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV tctaagatag cgctcttgtg tcaaataatt taaatcaacg ttagggaata attcaaactt tttgaaaatt tacaaaatct cagttatcct tcatagcgtg gactgaacac cggtgttaca 3120
Y                                    c
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV tcaagcacta aagctaatat tagctggcgt ttcatggtgg acaaaaatac aatataataa cattctaatt ataaaatata aacaataata tgttcatctt acttgtgaaa cattatcccc 3240
Y                                a                t        **     a t                                                        g
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV *gagatctca tgtaaatagt ccgtcgagct caacaaaaat a********t gccgagtcta cttct***** ********** ********** ggaacgacgg ctccaagatg gccgccgtat 3360
Y   t           g               t                tccggcag   t     t        gctgt tttggtttag ccaagtagga
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV ttcctgtgca gccaatgcgg ggtttactct ttactatgtc tatgatgtta gcacacatgt cctgttactg ccgaccacca ggggcagaa  caatacgtaa agtgaaataa tgggcatgtt 3480
Y                                                          a             c       a
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV ttatccaaca tggacactga tccatctgaa ctgtgtgtgt ttttttagAA TGGGCATCCA GTGAGAGAGA ACAGCATGGC TCTCCGGTAC ATCTCCGACC CGACCCAGAA CGCGCTGGAG 3600
Y                                               *           A                          CA        A
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV AAAGACCTGG ATGGTCACCg taggttctga ccagcaatca ctttcttcaa gaaatcagga agtgatcaca ataatggagt caaacagaca ggatacgagt cagatgcaaa caaggagttt 3720
Y
X   ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** ********** **********
INV caccatagca accgagtcca ggctctaaat aaatcttcct gatgtttgtg ttccagAGTA CGTTAACCAG CCTGGGAGTG AAACCAGCAG CAGGCTGTCG GATATCTACA ATCCCAACTA 3840
Y                                                                                                                                A
X   ********** ********** ********** ********** ********** ********** ****                                                       G
INV CGAGGACCTG ACGGACGGCT GGGGCCCGGT GTCGCTGTCC TCCCAGGAGG CGGAGACCAA CTTCTCCAGA CCGGAGTACC TGAACACCAA CCAAAACTCG CTCCCCCTGG TGTCCAGTGG 3960
Y           A                  C                                        G        A
X           A                  G                                        T        G
INV CAGCATGGAC GACCCCGACT ACCAGGCGGG CTACCAGGCT GCCTTCCTAC CGCAGACTGG AGCGCTCACT GGGAACGGCA TGTTTCTCCC TGCAGCAGAG AACCTGCAGT ACCTGGGACT 4080
Y                              C                                      C                       C                               G  A
X                              T                                      A                       C                               A  A
INV GGAGGCGCCA CTGTACACTC CTGTCCGGTA GGGGGCAGCA GCGCTCCGAC TGAAGCTGTG GCTTTGCTCA GATAAACAGT GGTCTGTTTT AAATTGGACA GATAACTGGA TTTAACTCAG 4200
Y                               C                                                     G
X                               C                                                     A
INV ATGGTTCACT GTCGGTTCTC CAGAGGAACT GAAACAGAAC AAGAACTTCC TGTTCCTTCA GGCAAAGGTC CAATCTTTGG CTCTAAGCTG ACGGGATGAA CGCATCTGGA AAATGGAGCC 4320
Y                                                                                                         G                   *****
X                                                                                                         T                   *****
INV AAACATTACC CAGATAAACT GTAAAAACAT AAGCTAATTC AGGTTACATT AAAAAAGACT CGCTCAACTC TCCTGCTCTC ACATGTGGGT TTGGATGCTG CAGGACGATG CAGATGTTTT 4440
Y   *                                                                     A
X   *                                                                     G
INV AATCCTGGGC GCTTCTGTTT CCAACATTTA CATTCATGAA CTGAAAATAT TTAAGTTTTC TGAAGATATG ATTTGACTGT CAGTCAACAT GATGTTTGTA ACCTTAAATC TTTCAGAATA 4560
Y                                                    C                                                    A C
X                                                    C                                                    A C
INV AGAGAAAACA GGTTGAAACT GGATTCTCAG GCCACACTGC TCTACTCCAG ATATTTATAA CTGCCATTTT G***TAAAAA AAAGAATCTG CATTCACACA ATCTGTTGGT TTTCTATCAG 4680
Y               *                                                                 TAAT                              T
X               *                                                                 T*AA                              C
INV AACAATTTTA TACAAGCCTT GCAAGAGCAG GTGGAAATAA AATTAAGCTT ACTAAAATTG AAATTTTGTA GAAGACAACG TTGGGTGAAC GAAATAACAC AAATGGTTCC GTTTTATAGG 4800
Y                                                                                                                      C
X                                                                                                                      T
```

Figure 2 Nucleotide sequence of the INV-, X- and Y-locus: Bases identical in all three loci are shown only for the INV locus. Exons appear in uppercase, introns in lowercase letters. Trailer regions in exon z are represented by italics. Asterisks indicate nucleotides not present in the corresponding locus. The imperfect direct repeats at the beginning and at the end of the large deletion in the X-locus are marked by arrows. Polyadenylation sites appearing to represent the 3′ termini are boxed. Numbering of nucleotides refers to the sequence of the INV locus

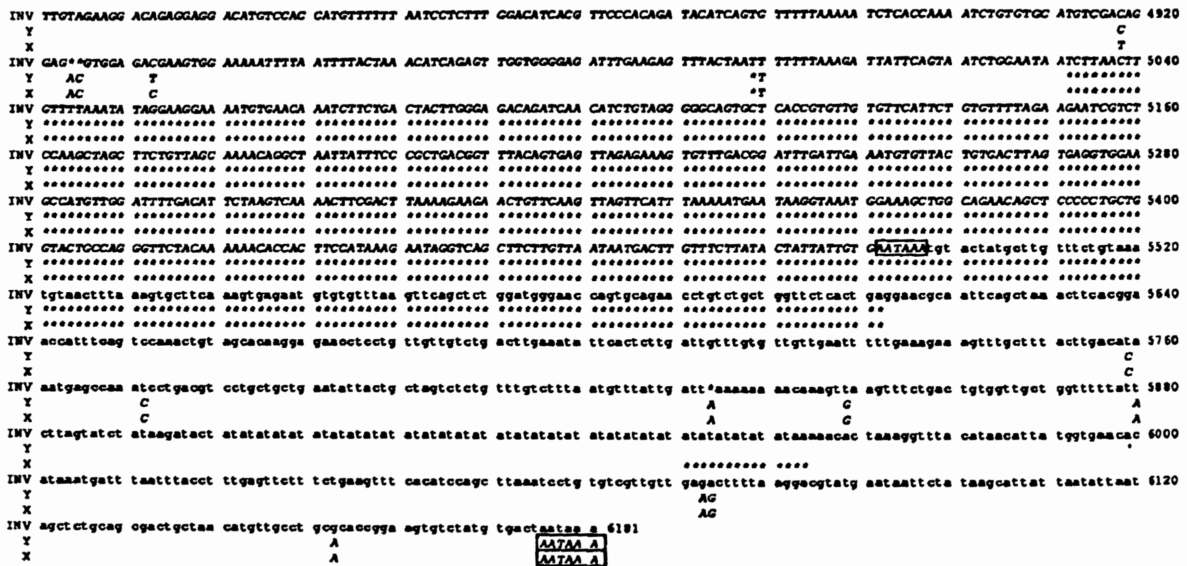

Figure 2 *Continued*

in the X-chromosomal melanoma the 4.7 kb mRNA is derived from the X-chromosomal *Xmrk* locus and lacks the 91 bases of exon y. By analogy we conclude that the 4.7 kb transcript observed in Y-chromosomal melanomas — as it does carry exon y — is derived from the Y-chromosomal *Xmrk* locus. Therefore, the 5.8 kb transcript originates from the proto-oncogene locus (INV), while the X- and Y-loci produce shorter mRNAs of similar size (4.7 kb) with internal structural differences (lack or presence of the 91 bases of exon y).

In tumor bearing fish, the oncogenic X- or Y-transcripts were found at high level in malignant melanoma (Figure 3a, lanes 5, 8, 9), while benign melanoma showed lower amounts of the oncogenic transcripts (Figure 3a, lane 4). The INV-locus was transcribed at low level in all melanomas irrespective of the degree of malignancy. Together with earlier findings (Wittbrodt *et al.*, 1989), these data show that the abundance of the oncogenic X- and Y-locus transcripts correlates with the malignancy of melanoma. To see if elevated expression of *Xmrk* can be detected in tissue other than malignant melanoma, we analysed gills of several *Xiphophorus* genotypes either without melanoma or with melanoma of differing malignancies. Only transcripts of the INV-locus were detected (Figure 3a, lanes 3, 6, 7) at uniform levels regardless of the presence or absence of melanoma or of tumor malignancy.

To determine if the observed size difference between proto-oncogene and oncogene transcripts was due to different 3′ ends we analysed the sequence of the corresponding regions of all three loci. Due to the 581 bp deletion the corresponding polyadenylation consensus present in the INV locus is removed from the oncogenic loci. However, the next following consensus sequence 683 bp downstream is used (see Figure 1) giving rise to a trailer being even 88/102 bp longer than in the INV locus transcript. Northern hybridizations using oligonucleotides confirmed these results (data not shown). This excludes the possibility that the observed size difference is due to alternate 3′ end formation in the oncogene transcripts.

We performed primer extension analyses to determine the extent of the 5′ noncoding leader sequence of the different transcripts. We found that the 5′ untranslated regions of both X- and Y-locus transcripts comprise 68 bp while the corresponding region of the proto-oncogene mRNA is approximately 1 kb long (Figure 4, lanes 1, 2, 4, 5). The position of the transcription start site of the oncogenic transcripts was confirmed using a different primer (Figure 4, lane 3). Sequencing of RNA from the tumor cell line PSM revealed exactly the same 5′ end for the oncogenic transcripts as was detected by the primer extension analyses (Figure 5). In addition, the RNA sequence gave an explanation for the 5′ truncation of the *Xmrk* cDNA sequence described elsewhere (Wittbrodt *et al.*, 1989) by displaying strong unspecific terminations at a position corresponding to the incomplete 5′ end of the cDNA sequence.

Thus, the size difference between the two transcripts is explained by additional sequence in the INV transcript and points to alternative transcription start sites in the tumorigenic mRNAs.

Discussion

In the *Xiphophorus* genome three different copies of the *Xmrk* gene coexist, one of which — as a proto-oncogene — appears to serve a so far unknown physiological function, and two additional ones which are oncogenic. In order to contribute to an understanding of the molecular basis of the oncogene activation process, we studied the structural organization and transcription of these *Xmrk* loci.

Based on the phylogenetic analysis of the sequence differences, we propose a duplication of the ancestral INV-locus creating a new sex-chromosomal *Xmrk* copy which subsequently diverged into the recent X- and Y-copies. This is in agreement with INV being the precursor locus of X and Y as a proto-oncogene. Further evidence for this issue comes from the observation that INV is expressed in very early embryogenesis at high levels and at low levels in several normal tissues without leading to transformation. In the absence of the tumor suppressor locus *R* expression of INV is not altered in any tissue of melanomatous fish. In addition, in an

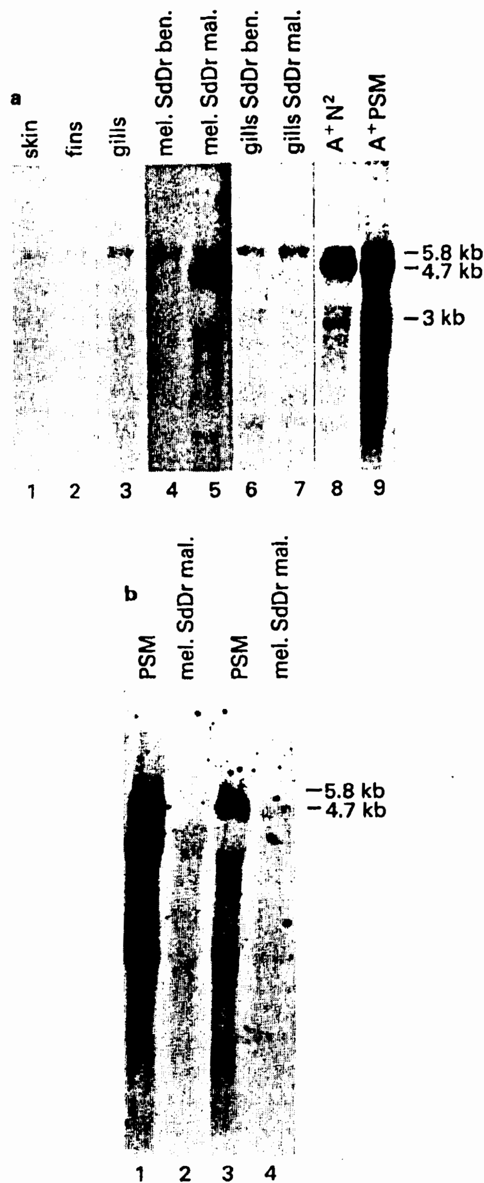

Figure 3 Northern analysis of the three *Xmrk* loci. **a**) Lanes 1–3 (exposure 6 days): Expression of *Xmrk* in skin, fins and gills of *X. helleri* (Rio Lancetilla stock) carrying only the INV-locus. Lanes 4, 5 (exposure 14h): Expression of *Xmrk* in benign or malignant melanoma from hybrid fish carrying the X-locus and in gills from these hybrids (lanes 6, 7; exposure 2.5 days). Lanes 8, 9 (exposure 5h): Expression of the Y-locus from *X. maculatus* (N^2) and from PSM cells. Transcripts below 3 kb represent polyadenylated transcripts of different length from the same gene generated by differential splicing or alternate transcription starts and polyadenylation signals, respectively. This phenomenon is observed frequently with highly expressed genes in tumor cell lines. For hybridization, the Y-locus cDNA served as a probe. (**b**) Lanes 1, 2: Hybridization with oligonucleotide A. Lanes 3, 4: Control hybridization with oligonucleotide B using the same filter. Overexposure for 5 days. The 5.8 kb INV transcript which is also recognized by both oligonucleotides is not visible due to its low abundance as compared to the 4.7 kb transcripts. In each case 20 μg of total RNA were analysed, except in (**a**), lanes 8 and 9 (polyA$^+$-RNA)

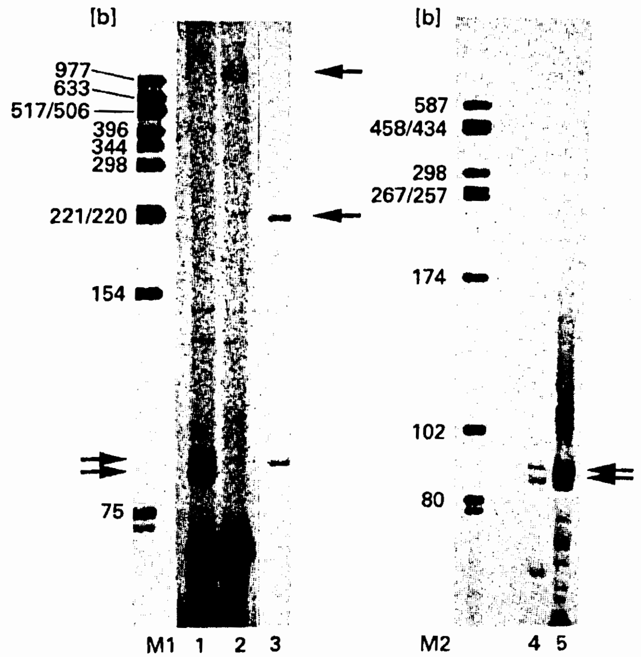

Figure 4 Primer extension analyses using RNA from melanoma containing the X locus transcript (lanes 1, 4), from oocytes containing the INV locus transcript (lane 2) or from PSM cells containing the Y locus transcript (lanes 3, 5) and primers C (lanes 1, 2, 4, 5) or D (lane 3). Major start sites are marked by arrows, multiple bands below 80 bp in lanes 4 and 5 as well as the band at 90 bp in lane 3 are located already in the coding region of the gene and therefore likely to be unspecific. Included size markers were labeled pBR322 cut with HinfI (lane M1) or pUC18 cut with HaeIII (lane M2). Gels were exposed for 3 weeks (lanes 1, 2), 5 h (lane 3) or overnight (lanes 4, 5). The observed start site at 220 bp (lane 3) using primer D located farther 3′ corresponds to the start sites at 90 bp (lanes 1, 4, 5) obtained with primer C which is located more 5′

extensive survey on over 500 wildtype fish from feral populations the INV locus was always found without exception in all individuals, while X and Y alleles were found only in those few animals which are predisposed to develop melanoma following the appropriate crossing (Schartl, 1990).

The additional potentially tumorigenic *Xmrk* copies did not evolve as pseudogenes, as they show a ratio of effective to silent changes of 1:1 each instead of 3:1 which is the expected rate for pseudogenes (Nei, 1987). The clustered distribution of the mutations displays selection against mutations in certain areas of the genes, e.g. the kinase domain, where the X- and Y-locus show only two silent mutations. Therefore, the tumorigenic copies of *Xmrk* may serve a physiological role in wildtype fish (e.g. directing macromelanophore pigmentation or pattern formation) as long as they are under control of the *R* locus.

The question if the X- and Y-copies have acquired a so called 'activating' mutation during the process of gene duplication and divergence is not of primary relevance to the understanding of the process of tumor formation in the *R*-lacking hybrid. Also the wildtype fish which never develop tumors would carry such an 'activated' allele as is clear from the formal genetics, but this will not lead to a phenotypic effect. We concentrated on the activation of the potentially tumorigenic *Xmrk* X and Y alleles in the hybrids after loss of the *R* locus, not on a possible 'activation' of such alleles by acquiring mutations during the process of their generation by duplication of the INV locus. However, such information will be important to understand how neoplastic transformation of pigment cells is mediated by *Xmrk*. Therefore the structure of all three loci concerning the region encoding the intracellular portion of the receptor

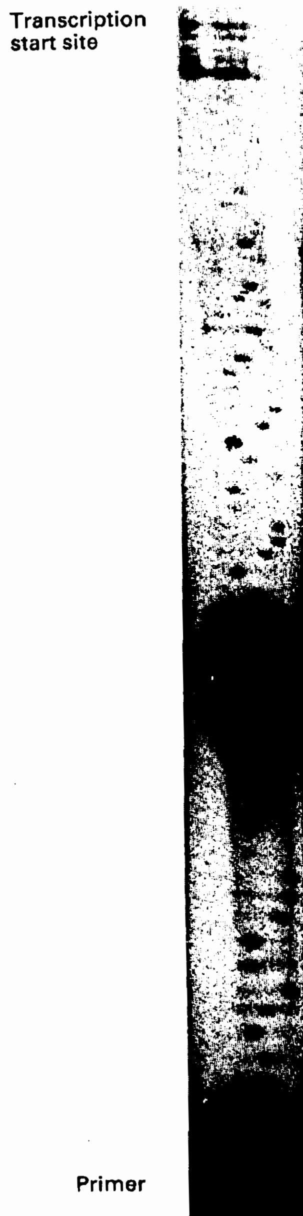

Figure 5 RNA sequence of the Y-locus transcript. 20 μg of total RNA from PSM cells were analysed using primer C. The position of primer C and the transcription start site is indicated. The region of unspecific termination (see 'Results' section) is clearly visible

and the untranslated 3′ trailer was analysed in detail because of the obvious divergence of this region. Physical mapping revealed considerable structural differences only in the 3′ regions of the *Xmrk* loci. Sequencing of the polymorphic regions led—besides one other large deletion affecting the trailer regions of both oncogenic transcripts—to the detection of a large deletion in the X-locus which includes one carboxyterminal exon of the encoded receptor tyrosine kinase. However, this structural alteration cannot be the cause for the transforming potential, as the Y-chromosomal *Xmrk* copy, which is tumorigenic as well, does contain this exon. On the other hand, loss of this exon, which contains one (aa 1032) of the five putative tyrosine phosphorylation sites of EGFR-like receptor tyrosine kinases (Margolis *et al.*, 1989; Walton *et al.*, 1990), may explain the observed higher malignancy of X-chromosome encoded melanomas as compared to Y-chromosomal copies. For the EGF-receptor it has been shown in vitro that C-terminal deletions affecting autophosphorylated tyrosines do not alter the transforming potential in general, but confer higher sensitivity to ligand induced mitogenesis when compared to the wild type situation (Honegger *et al.*, 1988; Khazaie *et al.*, 1988). On the other hand, minor structural alterations, e.g. point mutations, have been found to be responsible for the transforming potential of a gene (Bargmann *et al.*, 1986; Seeburg *et al.*, 1984; Hunter, 198[illegible]). Within the sequenced regions we detected a total of 10 amino acid changes, only three of which are non-conservative and therefore likely to alter the biochemical properties of the *Xmrk* protein. At present, we cannot decide if these or other, so far unknown point mutations in the 5′ region of the gene contribute to the oncogenic potential of the X- and Y-locus, as these sorts of questions can only be addressed by functional assays, e.g. by production of transgenic fish. It has been shown for the closely related EGF receptor that only the cytoplasmic domains carry the structural determinants that define its transforming potential (Riedel *et al.*, 1989). Therefore, point mutations in the 5′ region of *Xmrk* are unlikely to contribute to the transforming mechanism. Concerning the loss of tyrosine 1014 in the oncogenic loci, differences in carboxyterminal tyrosine phosphorylation of receptor tyrosine kinases are not regarded as instrumental in primary oncogenic activation (Honegger *et al.*, 1988; Khazaie *et al.*, 1988) but may modulate the tumorigenicity of receptor tyrosine kinases. At present, it is not clear if any of the observed point mutations do contribute at all to the oncogenic potential of the X- and Y-locus. The difference in tumorigenicity between X and Y on one, and INV on the other side after loss of the *R* locus may as well be due solely to the degree of expression of these loci.

The X- and Y-locus transcripts are 1.1 kb smaller than the proto-oncogene transcript. This is not exclusively due to differences in 3′ end formation, because all three *Xmrk* loci terminate within 102 bp. Primer extension analyses revealed that the size differences are due to a different start site of the INV-transcript, giving rise to a transcript of approximately the size determined from Northern analyses. At present it is not possible to define the exact position of the INV start site, as the corresponding genomic regions are not cloned. Therefore, it is not possible to perform primer extension analyses with primers located more 5′. However, it is clear that the oncogenic *Xmrk* copies use a start site definitely different from the INV start site, as they were extended only by 68 nt, which is in perfect accordance to the observed size of 4.7 kb. The observed differences in the 5′ end of the transcripts affect only untranslated sequence and do not lead to a structurally altered protein.

The INV-locus is expressed at low level only in skin, fins and gills and its expression is not influenced by the presence of melanoma. The low expression of the INV-locus in the melanotic tissue itself is—unlike the expression of the X- and Y-locus—not correlated to the state of malignancy. Therefore, the elements governing expression of the INV-locus must be different from those of the X- and Y-locus. The finding of two different start sites of transcription for the oncogenic and proto-oncogenic loci points to two different potential promoters. Southern analyses showed that the overall

structure of the 5′ region of all three loci is alike. This weakens, but does not rule out the possibility that the X- and Y-loci have acquired new regulatory sequences from the gene duplication/translocation event that led to their generation which would be reminiscent of the situation found with c-*myc* in Burkitt's Lymphoma (reviewed in Bornkamm *et al.*, 1988). As outlined above, the genomic region containing the promoter of the INV-locus is not cloned. The possibility of an additional intron in the 5′ region of the INV-locus cannot be ruled out at present. Therefore it is possible that one of the breakpoints of the proposed duplication/translocation event affects genomic regions farther upstream which are not yet accessible to Southern analyses. On the other hand, we cannot exclude that minor structural changes (e.g. point mutations) have occurred, which created the new potential promoter in the 5′ region of the oncogenic *Xmrk* loci.

Accumulation of tumor specific transcripts in melanoma can be explained by enhanced transcription from these tumor specific promoters. The observed overexpression of the oncogenic loci and its correlation with the malignancy of the tumors is intriguing in the light of findings that for RTKs of subclass I overexpression in tissue-culture has been shown to lead potentially to oncogenesis, and increased levels are frequently associated with certain human malignancies (reviewed in Yarden & Ullrich, 1988). Such overexpression obviously would result from the activity of the potential X- and Y-locus tumor specific promoters, thus giving an explanation for the mode of *Xmrk* activation.

Again, it is important to note that the structure of the tumorigenic *Xmrk* loci including the 5′ regions in the tumorous hybrids is not different from that of the non-tumorous *X. maculatus* wildtype. Therefore, minor structural changes in the *Xmrk* gene product are not sufficient to explain the mode of oncogenic activation because a structurally altered *Xmrk* protein would exert its action in the parental fish as well as in the hybrid. Differences in gene regulation leading to overexpression are, on the other hand, in accordance with the genetic data. The most important difference between the hybrids and the wildtype fish with regard to melanoma formation is the absence of the tumor-suppressor locus. It is therefore tempting to speculate that this locus is somehow involved in a negative control of the potential promoters of the oncogenic *Xmrk* loci.

Materials and methods

Cloning of the Xmrk *loci*

The polymorphic EcoRI fragments of the INV-, X- and Y-locus were cloned as previously described (Wittbrodt *et al.*, 1989). The INV-, X- and Y-loci were derived from male *X. maculatus* (stock Rio Jamapa X: *Tu-Sd*, Y: *Tu-Sr*). Subfragments of the INV, Y- and X-locus were subcloned into pBluescript (Stratagene) for sequencing.

Sequencing

Nested deletions were generated using either T4 DNA polymerase (with oligonucleotides adapted for pBluescript) (Dale *et al.*, 1985), or Exonuclease III (Henikoff, 1984). Sequences were determined by dideoxy-sequencing using T7 polymerase (Pharmacia) or Sequenase (United States Biochemical Corporation). Sequence analysis was performed using the GCG sequence analysis software package (Devereux *et al.*, 1984). The BESTFIT alignment of the GCG package was used to determine if an amino acid change was conservative, semi- and nonconservative.

RNA sequencing

RNA was sequenced using dideoxynucleotides and AMV reverse transcriptase (New England Biolabs) in a modified primer extension reaction (Carpenter & Simon, 1990).

Northern blot analyses

RNA preparation, Northern blots and quantitation of transferred RNA were performed essentially as described (Mäueler *et al.*, 1988). For size calibration a RNA-ladder (BRL, Bethesda) was included. Probes were random priming labeled (Feinberg & Vogelstein, 1983). For hybridization with the Y-cDNA probe, filters were washed at 65°C, 0.1 × SSC, 1% SDS. Kinased oligonucleotides were hybridized at 42°C in 6 × SSC, 1% SDS with subsequent washing done at 45°C, 6 × SSC, 1% SDS. Oligonucleotides were kinased according to Chaconas and van de Sande (Chaconas & van de Sande, 1980).

Oligonucleotides

Oligonucleotides A and B were designed as antisense probes complementary to nt 3530 to 3569 and nt 4832 to 4860 of the Y-locus, respectively (see Figure 1). Oligonucleotides C, spanning nt 35 to 56 (in antisense orientation) and D, spanning nt 160 to 179 (in antisense orientation) of the Y-chromosomal cDNA sequence (Wittbrodt *et al.* 1989), were used for primer extension reactions.

Primer extension analyses

20 μg total RNA from *X. helleri* oocytes and from PSM cells (Wakamatsu, 1981) together with 20 μg poly A^+ RNA from *X. maculatus* melanoma (*Tu-Sd*), containing abundant transcripts of either INV-, Y- or X-locus, were added to 1.5×10^5 cpm of kinased primer C or D (770 fmoles), precipitated in 300 μl 0.3 M NaOAc (pH 5.2), rinsed sequentially with 80% and with 96% EtOH and then dried. The precipitates were resuspended in 10 μl of 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 250 mM KCl, incubated for 60 min at 60°C and brought to 42°C to anneal the primers. After addition of 23 μl of a freshly prepared buffer containing 10 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, 5 mM DTT, 330 μM dNTPs, 100 μg ml^{-1} actinomycin D, 20 units RNAsin (Amersham) and 10 units MMLV reverse transcriptase (New England Biolabs), the primer extension reactions were incubated for 60 min at 42°C and precipitated with 300 μl EtOH. Pellets were rinsed with 300 μl of 70% EtOH, dried and loaded onto 16 cm polyacrylamide gels (8%). Dried gels were exposed with intensifying screens at −70°C.

Acknowledgements

We thank Herbert Jäckle for discussions and critical comments, Christof Steiner for valuable support concerning primer extensions, Catherine Schindewolf for critical reading of the manuscript and Andrea Oswald for quick and expert synthesis of oligonucleotides.

This work was supported by grants to M.S. from the Bundesministerium für Forschung und Technologie through 'Schwerpunkt: Grundlagen und Anwendungen der Gentechnologie', the Deutsche Forschungsgemeinschaft through 'Schwerpunkt: Klassische und molekulare Tumorzytogenetik' and the Max-Planck-Gesellschaft. D.A. is a fellow of the Fonds der Chemischen Industrie.

References

Adam, D., Wittbrodt, J., Telling, A. & Schartl, M. (1988). *Nucleic Acids Res.*, **16**, 7212.

Ahuja, M.R. & Anders, F. (1976). *Prog. Exp. Tumor Res.*, **20**, 380–397.

Anders, A. & Anders, F. (1978). *Biochim. Biophys. Acta*, **516**, 61–95.

Anders, F., Schartl, M., Barnekow, A. & Anders, A. (1984). *Adv. Cancer Res.*, **42**, 191–275.

Atz, J.W. (1962). *Zoologica*, **47**, 153–181.

Bargmann, C.I., Hung, M. & Weinberg, R.A. (1986). *Cell*, **45**, 649–657.

Boettiger, D. (1989). *Curr. Top. Microbiol. Immunol.*, **147**, 31–78.

Bornkamm, G.W., Polack, A. & Eick, D. (1988). *Cellular Oncogene Activation*. G. Klein (ed.) Marcel Dekker: NY, pp. 223–273.

Carpenter, C.D. & Simon, A.E. (1990). *BioTechniques*, **8**, 26–27.

Chaconas, G. & van de Sande, J. (1980). *Meth. Enzymol.*, **65**, 75–85.

Dale, R.M.K., McClure, B.A. & Houchins, J.P. (1985). *Plasmid*, **13**, 31–40.

Devereux, J., Haeberli, P. & Smithies, O. (1984). *Nucleic Acids Res.*, **12**, 387–395.

Feinberg, A.P. & Vogelstein, B. (1983). *Anal. Biochem.*, **132**, 6–13.

Fitch, W.M. & Margoliash, E. (1967). *Science*, **155**, 279–284.

Gordon, M. (1927). *Genetics*, **12**, 253–283.

Gordon, M. (1958). *Ann. NY Acad. Sci.*, **71**, 1213–1222.

Häussler, G. (1928). *Klin. Wochenschr.*, **33**, 1561–1562.

Henikoff, S. (1984). *Gene*, **28**, 351–359.

Honegger, A., Dull, T.J., Bellot, F., Van Obberghen, E., Szapary, D., Schmidt, A., Ullrich, A. & Schlessinger, J. (1988). *EMBO J.*, **7**, 3045–3052.

Honegger, A., Kris, R.M., Ullrich, A. & Schlessinger, J. (1989). *Proc. Natl. Acad. Sci. USA*, **86**, 925–929.

Hunter, T. (1987). *Cell*, **49**, 1–4.

Jeffreys, A.J. & Harris, S. (1982). *Nature*, **296**, 9–10.

Khazaie, K., Dull, T.J., Graf, T., Schlessinger, J., Ullrich, A., Beug, H. & Vennström, B. (1988). *EMBO J.*, **7**, 3061–3071.

Kosswig, C. (1965). *Zool. Anz.*, **175**, 21–50.

Kosswig, C. (1928). *Z. Indukt. Abstammungs-V[illegible]rbungsl.*, **47**, 150–158.

Kraus, M.H., Issing, W., Miki, T. & Popesc[illegible] N.C. (1989). *Proc. Natl. Acad. Sci. USA*, **86**, 9193–9197.

Margolis, B.L., Lax, I., Kris, R.M., Dombalagi[illegible] M., Honegger, A., Howk, R., Givol, D., Ullrich, A. & [illegible]lessinger, J. (1989). *J. Biol. Chem.*, **264**, 10667–10671.

Mäueler, W., Raulf, F. & Schartl, M. (1988). *On[illegible]ene*, **2**, 421–430.

Mount, S.M. (1982). *Nucleic Acids Res.*, **10**, 459[illegible]2.

Nei, M. (1987). *Molecular Evolutionary Genet[illegible]* Columbia University Press: NY.

Riedel, H., Dull, T.J., Honegger, A., Schlessinge[illegible] & Ullrich, A. (1989). *EMBO J.*, **8**, 2943–2954.

Schartl, M. (1988). *Genetics*, **119**, 679–685.

Schartl, M. (1990). *Genetics*, (in press).

Schwab, M. (1987). *Trends Genet.*, **3**, 38–42.

Seeburg, P.H., Colby, W.W., Capon, D.J., Goe[illegible]el, D.W. & Levinson, A.D. (1984). *Nature*, **312**, 71–74.

Semba, K., Kamata, N., Toyoshima, K. & Y[illegible]amoto, T. (1985). *Proc. Natl. Acad. Sci. USA*, **82**, 6497–65[illegible].

Sinkovics, J.G. (1988). *Crit. Rev. Immunol.*, **8**, 217–[illegible]98.

Tronick, S.R. & Aaronson, S.A. (1988). *Adv. Sec. [illegible]ess. Phosphoprot. Res.*, **21**, 201–214.

Vielkind, J.R., Kallman, K.D. & Morizot, D.[illegible] (1989). *J. Aquat. Anim. Health*, **1**, 69–77.

Wakamatsu, Y. (1981). *Cancer Res.*, **41**, 679–680.

Walton, G.M., Chen, W.S., Rosenfeld, M.G. & Gill, G.N. (1990). *J. Biol. Chem.*, **256**, 1750–1754.

Weinberg, R.A. (1989). *Cancer Res.*, **49**, 3713–3721.

Wittbrodt, J., Adam, D., Malitschek, B., Mäueler W., Raulf, F., Telling, A., Robertson, S.M. & Schartl, M. (1989). *Nature*, **341**, 415–421.

Yarden, Y. & Ullrich, A. (1988). *Ann. Rev. Biochem.*, **57**, 443–478.

Zander, C.D. (1969). *Mitt. Hamburg. Zool. Mus. Inst.*, **66**, 241–271.

Zechel, C., Schleenbecker, U., Anders, A. & Anders, F. (1988). *Oncogene*, **3**, 605–617.

The sequences have been deposited in the EMBL/GenBank database and are available under the following accession numbers: INV: X56317; X: X56318; Y: X56319.