



Universiteit
Leiden
The Netherlands

Positional cloning in Xp22 : towards the isolation of the gene involved in X-linked retinoschisis

Vosse, E. van de

Citation

Vosse, E. van de. (1998, January 7). *Positional cloning in Xp22 : towards the isolation of the gene involved in X-linked retinoschisis*. Retrieved from <https://hdl.handle.net/1887/28328>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/28328>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/28328> holds various files of this Leiden University dissertation.

Author: Vosse, Esther van de

Title: Positional Cloning in Xp22 : towards the isolation of the gene involved in X-linked retinoschisis

Issue Date: 1998-01-07

CHAPTER 3

Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product

Datson, N.A., Van de Vosse, E., Dauwerse, H.G., Bout, M., Van Ommen, G.J.B., and Den Dunnen, J.T.

Nucleic Acids Research 24:1105-1111, 1996

Reprinted with permission

Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product

Nicole A. Datson, Esther van de Vosse, Hans G. Dauwerse, Mattie Bout, Gert-Jan B. van Ommen and Johan T. den Dunnen*

MGC-Department of Human Genetics, Leiden University, Wassenaarseweg 72, 2333 Al Leiden, The Netherlands

Received November 17, 1995; Revised and Accepted February 7, 1996

ABSTRACT

To facilitate the scanning of large genomic regions for the presence of exonic gene segments we have constructed a cosmid-based exon trap vector. The vector serves a dual purpose since it is also suitable for contig construction and physical mapping. The exon trap cassette of vector sCOGH1 consists of the human growth hormone gene driven by the mouse metallothionein-1 promoter. Inserts are cloned in the multicloning site located in intron 2 of the hGH gene. The efficiency of the system is demonstrated with cosmids containing multiple exons of the Duchenne Muscular Dystrophy gene. All exons present in the inserts were successfully retrieved and no cryptic products were detected. Up to seven exons were isolated simultaneously in a single spliced product. The system has been extended by a transcription-translation-test protocol to determine the presence of large open reading frames in the trapped products, using a combination of tailed PCR primers directing protein synthesis in three different reading frames, followed by *in vitro* transcription-translation. Having larger stretches of coding sequence in a single exon trap product rather than small single exons greatly facilitates further analysis of potential genes and offers new possibilities for direct mutation analysis of exon trap material.

INTRODUCTION

Exon trapping has become a widely used method which is generally acknowledged as a versatile tissue-independent approach to detect genes in cloned DNA. In contrast to RNA-based methods, such as cDNA selection and direct screening of cDNA libraries, exon trapping is independent of tissue-specific gene expression. It uses cloned DNA directly to select sequences surrounded by functional splice sites (1-3). Original exon trapping protocols have been improved with respect to speed and efficiency and improvements have been made to reduce the background consisting of cryptically spliced

products and products arising from vector-vector splicing (4). However, some of the limitations of the original systems have remained unaddressed.

A major limitation of current systems is the need for subcloning of the region of interest in a vector with a capacity for inserts typically measuring 1-2 kb. This has several consequences: (i) due to the small insert size after subcloning, multiple exons will only rarely be present in one insert, resulting in exon trap clones containing only a single exon. Consequently, many of the exon trap probes derived are small (~80-150 bp) and frequently give poor signals or a high signal-to-noise ratio in subsequent experiments, e.g. the screening of cDNA-libraries or probing of Northern blots. Furthermore, since the individually trapped exons require the use of cDNA libraries in the next step to further define the gene, the initial advantage of working with an expression independent system is to a large extent lost in the subsequent step. (ii) Due to subcloning into plasmid-based exon trap vectors the gene(s) present are scattered into many separate, disconnected pieces. Any exons thus obtained have to be aligned to reconstruct their original order. Reconstruction of the gene from individually trapped exons requires a significant amount of time and effort and implies a major loss of information originally contained within the input material prior to subcloning. (iii) Subcloning disrupts the genomic context around the exons. Cloning of regions which are never transcribed or of intronic sequences without their naturally flanking exons often results in activation of cryptic splice sites, leading to recognition of false exons and a background of false positives. On the other hand, genuine exons will be missed due to poor recognition of the host system or due to unfavourable factors resulting from the cloning (e.g. spacing of restriction sites). (iv) Current exon trapping systems can only be used in combination with specific cell lines (e.g. COS cells), since they require a system of replication in the host cell, commonly based on the SV40 origin of replication (2). It is imaginable that some exons of genes with a highly tissue-specific expression pattern will not be included in the mature transcripts generated in a completely different cell type (5).

Although the 3' exon trapping recently described (6,7) has some advantages in that it allows larger exons to be trapped, specifically identifies the end of a gene, and selects exons based

on two independent criteria i.e. splicing and polyadenylation; it does not, however, address the other limitations of small-insert exon trapping.

We have designed a large-insert exon trapping vector capable of scanning 25–40 kb genomic regions for exons. The vector has a dual use: as cosmid vector for contig construction and physical mapping, and as exon trap vector for isolation of coding sequences. In the vector, inserts are cloned into intron 2 of the human growth hormone gene (hGH) and transcription is driven by a mouse metallothionein-1 promoter (mMT-1). This is a strong, ubiquitously expressed promoter which allows many different cell types to be used, thus obviating the restriction to COS cells applying to the SV40-based systems used so far. During exon trapping the genomic context is maintained over the entire 25–40 kb region, reducing the false positive rate while yielding processed transcripts with multiple exons spliced together in the correct order. The efficiency of the system is demonstrated using cosmids containing up to seven exons of the duchenne muscular dystrophy gene (DMD). We believe that the system should greatly increase the speed and reliability of gene isolation by exon trapping by offering a solution for most major limitations of current exon trapping systems.

MATERIALS AND METHODS

Vector construction

sCOGH1, schematically drawn in Figure 1, was constructed as follows: cosmid vector sCos1(8) was digested with *EcoRI* and the 7.9 kb vector fragment was separated from the *EcoRI* linker by agarose gel electrophoresis and elution. Similarly, plasmid pXGH5(9) was digested with *EcoRI* and the 4 kb fragment containing the mouse metallothionein-1 promoter (mMT-1) and the human growth hormone gene (hGH) was isolated by gel-purification. Both fragments were combined by ligation, resulting in the isolation of sCOGH0a and sCOGH0b, differing in the orientation of the mMT1/hGH-insert in sCos1. Subsequently a linker composed of two complementary oligonucleotides (5'-AGCGGCCGCGAATTCCGATCCGGCGGCCGC-3' and 5'-CTGCGGCCGCGGATCCGAATTCCGCGGCCG-3') was synthesized containing *NotI*, *BamHI* and *EcoRI* sites as well as *AccI* sticky ends, and introduced into intron 2 of the hGH gene by digestion of sCOGH0b with *AccI* and ligation. The resulting vector was designated sCOGH1.

Subsequently specific variants of sCOGH1 were constructed (Fig. 1). To facilitate screening of human positive cosmids sCOGH2, a sCOGH1 variant without the *Alu* sequence in the 3'-UTR of hGH was constructed. Three sCOGH1 fragments were ligated: fragment 1 was produced by digesting sCOGH1 with *EcoRI*. The sticky ends were filled in using T4 DNA polymerase. After *Clal* digestion the 7.9 kb vector fragment was isolated from a gel. Fragment 2 was the gel-purified 2.4 kb *Clal*-*BamHI* fragment of sCOGH1 containing the 5' part of hGH. To obtain the third fragment the 1.6 kb *EcoRI* fragment of sCOGH1, containing the 3' part of hGH, was first subcloned in pGEM7zf (Promega). The resulting clone, pGHE1.6, was digested with *BamHI* and *SspI* and the 0.9 kb 3' hGH fragment was gel purified. Combination of the three fragments resulted in sCOGH2.

sCOGH3 is a promoterless variant of sCOGH1, constructed for trapping promoter/first exon regions of tissue-specific genes. sCOGH2 was digested with *EcoRI* and ligated, resulting in

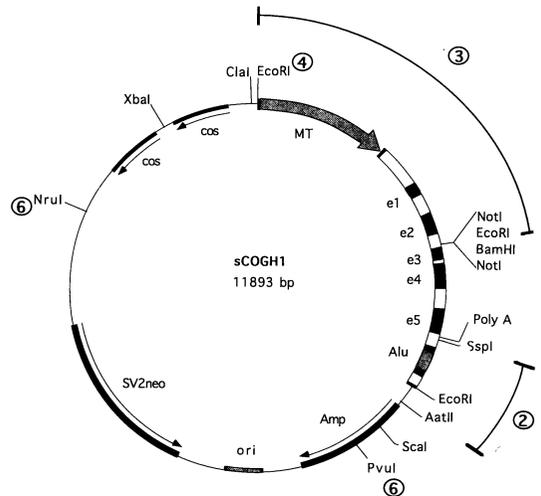


Figure 1. Vector sCOGH1 (11893 bp) contains the entire hGH exon trap cassette driven by a mMT-1 promoter. The MCS located in intron 2 of hGH contains sites for *EcoRI* and *BamHI* flanked by two *NotI* sites, which can be used to release the cosmid insert. Several variants of sCOGH1 were constructed, marked in the outer ring of sCOGH1 by a corresponding number. In sCOGH2, sCOGH3, sCOGH4 and sCOGH6 the *Alu* repeat in the 3'-UTR of hGH has been removed (2). sCOGH3 lacks the mMT-1 promoter and exons 1 and 2 of hGH (3). sCOGH4 and sCOGH6 lack the *EcoRI* site at the 5' end of hGH (4). In sCOGH6 the SV2neo selectable marker has been removed between the *PvuII* and *NruI* sites and has been replaced by a fragment from pUC19 (6).

sCOGH3 lacking the 2.4 kb 5' hGH fragment. A variant lacking the *EcoRI* site at the 5' end of the mMT-1 promoter, sCOGH4, was constructed to allow use of the *EcoRI* site in the polylinker for cloning of cosmid inserts. sCOGH2 was partially digested with *EcoRI* followed by filling in of the ends with T4 DNA polymerase. The linear band of 11.3 kb was isolated from a gel and ligated, resulting in sCOGH4. In another variant, sCOGH6, the SV2neo marker was removed to allow cloning of larger cosmid inserts. For this purpose sCOGH4 was digested with *NruI* and *PvuII*, the 6.6 kb vector fragment was isolated from gel and ligated to the 1.3 kb *PvuII*-*PvuII* fragment containing the ampicillin resistance gene and *E. coli* origin of replication of pUC19 which was gel purified.

Construction of cosmid libraries in sCOGH1

All sCOGH-derivatives were propagated in *E. coli* strain HB10B (kindly provided by Pieter de Jong). For cosmid cloning, vector DNA was linearised with *XbaI*, dephosphorylated and subsequently digested with *BamHI*. Agarose plugs containing genomic yeast DNA and YAC DNA of yDMD(0-25)C, containing the human DMD-gene from 100 kb upstream of the brain exon 1 to 100 kb downstream of exon 79 (10), were partially digested with *MboI*, size fractionated and ligated into the *BamHI*-site of sCOGH1. The ligated material was packaged using Gigapack II Plus Packaging Extract (Stratagene) and used to infect *E. coli* 1046. Cosmids containing specific regions of the DMD-gene were isolated and analysed using standard protocols

(11) by hybridization with specific DMD cDNA sequences (10). The exon content was established by PCR with exon primers and by hybridisation of the *Hind*III-digested cosmids with the DMD cDNA. The inserts of screened cosmids were reversed by *Not*I-excision of the insert, religation and transformation to *E. coli*. The orientation of the insert was determined by restriction digestion.

Cell culture and electroporation

Initially COS-1 cells were used for transfection experiments. We found, however, that exon trapping results strongly improved using hamster V79 cells. Higher yields were obtained of full length PCR fragments. Therefore later experiments were performed with this cell line. We explain the improvement by a lower degree of homology between the hGH-primers and the endogenous hamster growth hormone gene compared to the corresponding sequences in COS-1 cells. The cells were cultured in DMEM with 10% inactivated fetal calf serum (Gibco-BRL).

Cosmid DNA was introduced by electroporation: actively growing cells were collected by centrifugation, washed in cold PBS (without bivalent cations) and resuspended in cold PBS at a density of 2×10^7 cells/ml. Cell suspension (0.5 ml) was added to 20 μ l of PBS containing 10 μ g cesium-chloride purified cosmid DNA and placed in a pre-chilled electroporation cuvette (0.4 cm chamber, BioRad). After 5 min on ice, the cells were electroperated in a BioRad Gene Pulser [300 V (750 V/cm); 960 μ F], and placed on ice again. After 5 min the cells were transferred gently to a 100 mm tissue culture dish containing 10 ml of pre-warmed, equilibrated DMEM + 10% FCS. Transfection efficiency was monitored by assaying the hGH concentration in 100 μ l of the culture medium of cells transfected in parallel with pXGH5 using the Allégro hGH Transient Gene Assay kit (Nichols Institute, San Juan Capistrano, USA) (9).

RNA isolation, RNA-PCR and product analysis

48–72 h after transfection, the cells were harvested and total RNA was isolated using RNazoIB (CINNA/BIOTECH). First-strand cDNA synthesis was performed by adding 50 pmol of primer hGHf to 2 μ g total RNA in a volume of 16 μ l. The mixture was incubated at 65°C for 10 min and chilled on ice. 14 μ l of a mix containing 3 μ l 0.1 M DTT, 3 μ l 10 mM dNTPs, 0.5 μ l RNasin (40 U/ μ l; Promega), 6 μ l 5 \times RT buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM MgCl₂; Gibco-BRL) and 150 U SuperScript Reverse Transcriptase (Gibco-BRL) were added to a final volume of 30 μ l, and incubated at 42°C for 1 h. Subsequently, the solution was heated to 95°C for 5 min and chilled on ice. RNase H (2.25 U; Promega) was added and the solution was incubated at 37°C for 20 min. An aliquot of the solution (10 μ l) was used in a PCR reaction containing 12.5 pmol of primer hGHe, 50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris-HCl pH 8.0, 0.2 mM dNTPs, 0.2 mg/ml BSA and 0.25 U SuperTaq (HT Biotechnology Ltd) in a reaction volume of 50 μ l, followed by an initial denaturation step of 5 min at 94°C, 30 cycles of amplification (1 min at 94°C, 1 min at 60°C and 2 min at 72°C) and a final extension of 10 min at 72°C. No additional hGHf primer was added in the PCR reaction. Nested PCR, using either internal hGH primers or combinations of a hGH primer and a DMD primer, was performed on 1 μ l of the primary PCR material with 12.5 pmol of each primer and PCR conditions identical to the

first PCR. The internal hGH primers used were hGHa and hGHb. When RNA-PCR products were used for *in vitro* transcription-translation, primer hGHa was replaced by hGHORF1, hGHORF2 or hGHORF3. Direct sequencing of PCR products was performed using the SequenaseTM PCR Product Sequencing kit (USB).

Oligonucleotides and hybridisation probes

hGHa: 5'-CGGGATCCTAATACGACTCACTATAGGCGTCTG-CACCAGCTGGCCTTTGAC-3'
 hGHb: 5'-CGGGATCCCGTCTAGAGGGTTCTGCAGGAATG-AAATACTT-3'
 hGHe: 5'-ACGCTATGCTCCGCGCCATCGT-3'
 hGHf: 5'-ACAGAGGGAGGTCTGGGGGTTCT-3'
 D69F1: 5'-GCCATAAAAATGCACTATCCA-3'
 D72F1: 5'-CCTCAGCTTTCACACGATGA-3'
 D72R1: 5'-TCATCGTGTGAAAGCTGAGG-3'
 D73R1: 5'-ATCCATTGCTGTTTTCCATTTC-3'
 D74R1: 5'-GCAGGACTACGAGGCTGG-3'
 polyT-REP: 5'-GGATCCGTCGACATCGATGAATTC(T)₂₅-3'
 hGHORF1: 5'-CGGGATCCTAATACGACTCACTATAGGAC-GACCACCATGTCAGCTGGCCTTTGACACCTACCAGG-AG-3'
 hGHORF2: 5'-CGGGATCCTAATACGACTCACTATAGGAC-AGACCACCATGTCAGCTGGCCTTTGACACCTACCAGGAG-3'
 hGHORF3: 5'-CGGGATCCTAATACGACTCACTATAGGAC-GACCACCATGGGCAGCTGGCCTTTGACACCTACCA-GGAG-3'
 hGHUTR1: 5'-CAGGAGAGGCACTGGGGA-3'

hGH primers were designed from the sequence M13438 and DMD primers from the sequence M18533 (EMBL sequence database). cDNA probe 63-1/3 is a subclone of the DMD cDNA and was used to screen for cosmids containing specific regions of the DMD gene. Probe 63-1/3 contains exons 65–74. Probe P20 contains exon 45 and part of intron 44 of the DMD gene (12).

In vitro transcription-translation

Modified primers, containing a T7 promoter and an eukaryotic translation initiation sequence, were used to generate PCR products suitable for *in vitro* transcription-translation. T7-PCR product (200–400 ng) was added to the TnT/T7 coupled reticulocyte lysate system (Promega). The synthesized protein products were separated on a 15% SDS-polyacrylamide minigel system. Fluorography was obtained by washing the gels in DMSO/PPO. Dried gels were exposed 16–40 h for autoradiography.

RESULTS

Outline of cosmid-based exon trapping procedure

Vector sCOGH1 contains all the essential elements of a cosmid vector, i.e. origin of replication, antibiotic resistance marker (ampicillin and neomycin) and two *cos* sites (Fig. 1). In addition it contains an exon trap cassette consisting of a mMT-1 promoter driving expression of the hGH gene, containing a multicloning site (MCS) located between exons 2 and 3 (see Materials and Methods for details of vector construction). The ubiquitous mMT-1 promoter allows the use of many cell types. The vector is

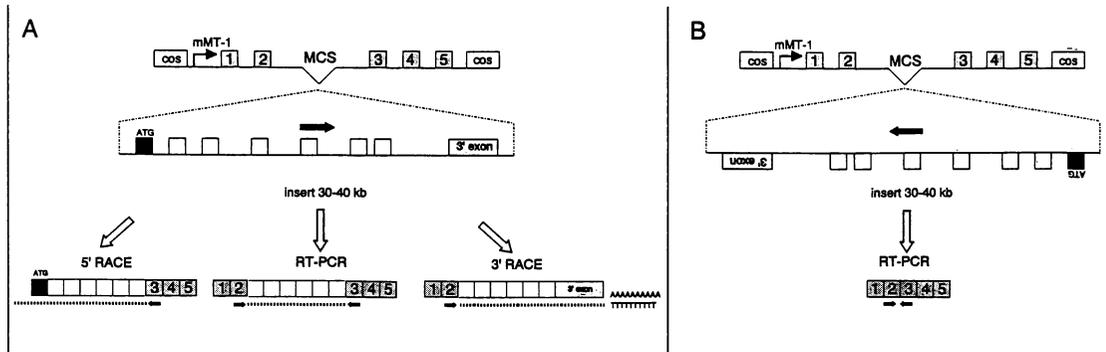


Figure 2. Entire cosmids constructed in vector sCOGH1 are transfected into V79 cells in which the introduced DNA is transcribed and processed. Spliced transcripts are amplified from isolated RNA. (A) Internal exons are amplified using 2 vector-derived primers. 5' first exons can be amplified in a 5' RACE using a single vector-derived primer. Similarly, 3' terminal exons can be isolated in a 3' RACE. (B) Trapping of cosmids containing inserts in the antisense orientation with respect to hGH will not result in formation of a chimeric hGH/unknown gene product. Instead, an 'empty' hGH transcript will be amplified by RT-PCR.

constructed such that the inserts can easily be excised and religated to obtain the opposite transcriptional orientations.

Cosmids are introduced into the cell type of choice by electroporation. We have tested and compared various cell lines and found V79 Chinese hamster lung cells to be a very efficient general host cell type. Upon expression, the hGH-initiated transcript will incorporate putative exons from the insert, cloned between exons 2 and 3 of the hGH gene, thus giving a chimeric product. After processing of the primary transcript, the putative RNA containing the exons to be trapped is amplified by RT-PCR using flanking vector-derived primers (Fig. 2A). In the specific event of 5' and 3' ends of genes being present in the insert, these will be skipped by the processing system or lead to alternatively initiated or terminated transcripts. They can be detected in the same mixture by 5' or 3' RACE, using opposite vector-derived primers separately (13). Gene inserts cloned in an antisense orientation will not be trapped, resulting in amplification of hGH sequences only (Fig. 2B).

RT-PCR analysis of V79 cells transfected with the sCOGH1 and sCOGH2 vectors yielded the expected 'empty' products (data not shown). A PCR with primers hGHa and hGHb gave the spliced 132 bp product containing hGH exons 2 and 3. Similarly, a 3' RACE using primers hGHa and polyT-REP yielded a spliced 720 bp product starting in hGH exon 2 and ending in the poly(A) tail of the transcript. These experiments show that the insertion of the MCS into hGH intron 2 does not effect splice site selection.

Trapping of multiple DMD-exons in a single spliced product

To demonstrate the ability of the present method to isolate exonic gene segments from eukaryotic mammalian DNA, we subcloned YAC yDMD(0-25)C known to contain the human DMD gene (10). Two cosmids were isolated and used for exon trapping: cDMD2 and cDMD3. cDMD2 contains exons 72-76 and cDMD3 exons 68-74 (Fig. 3). RNA was isolated from V79 cells transfected with the cosmids and vector-derived transcripts were amplified by RT-PCR using either two vector-derived primers (hGHa and hGHb; Fig. 3, lane A,) or a combination of a

vector-derived and DMD exon primers (Fig. 3; lanes B and C). In all cases, RNA-PCR analysis yielded products containing the expected exonic DMD-segments.

In the case of cDMD2, RNA-PCR analysis using two vector-derived primers produced a product of 0.85 kb containing DMD exons 72-76 (Fig. 3). Direct sequencing confirmed that the PCR products contained the expected DMD exons, spliced together between growth hormone exons. All exon-exon transitions between DMD exons were perfect and DMD exon 72 was spliced correctly to hGH exon 2. DMD exon 76 was not spliced to hGH exon 3, but instead to a unknown sequence of 18 bp preceding the cloning site, resulting in insertion of an extra 61 bp.

Analysis of cDMD3 yielded a product of 0.88 kb containing DMD exons 68-74. DMD exon 68 was spliced correctly to hGH and all DMD exon-exon transitions were correct. Sequence analysis showed that not the hGH exon 3 splice acceptor site was used but a new site directly at the *Bam*HI cloning site. This results in a 43 bp insert of MCS and hGH intron 2 between DMD exon 74 and hGH exon 3.

Inserts in antisense orientation

Exon trapping of cDMD2r, containing the exonic DMD segments in the antisense orientation, gave no insert-derived products. The only product amplified was the 132 bp empty hGH exon 2/exon 3 product (Fig. 4, lane 1). This shows that, using this system, the false-positive rate of an entire 30 kb insert is effectively zero. Exon trapping of cDMD3r resulted in a PCR product of ~0.25 kb (Fig. 4, lane 4), either corresponding to a cryptic product or to an unknown exon derived from the antisense strand. Hybridisation of this product to a *Hind*III-digest of cDMD3 showed that it mapped to the cosmid and was spliced. The 339 bp product visible in lanes 2 (cDMD2r) and 3 (cDMD3) represents unspliced hamster growth hormone and results from traces of contamination of V79 genomic DNA in the RNA preparation.

In vitro transcription-translation

The 0.73 kb RT-PCR product of cDMD3 (Fig. 3B, lane B) was reamplified, replacing the hGH exon 2 forward primer with three

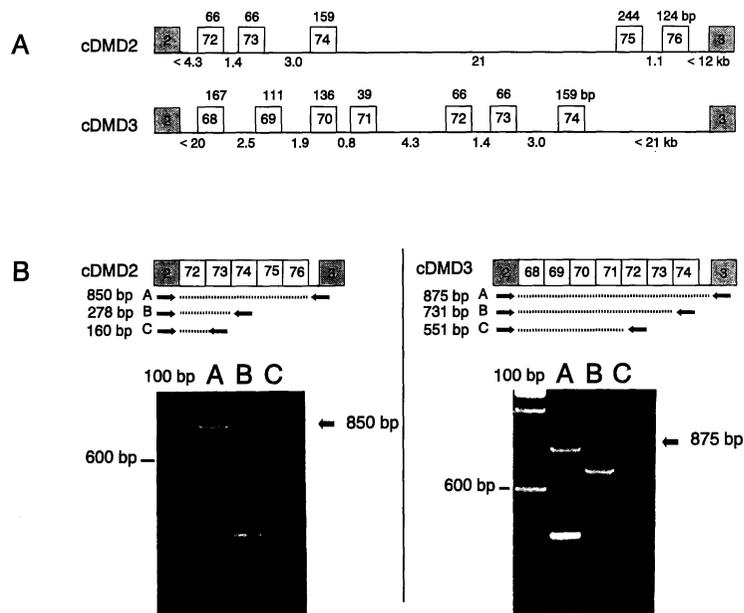


Figure 3. Exon trapping using DMD cosmids. (A) The genomic content of cosmids cDMD2 and cDMD3. The boxes represent exons and the lines introns. The shaded boxes represent hGH exons 2 and 3 and the open boxes DMD exons. The exon size in bp is indicated above the exons, the intron size in kb below the line. (B) RT-PCR of trapped exons. Lane A shows the RT-PCR product amplified using vector primers hGHa and hGHb. Lanes B and C of cDMD2 show the product amplified using primers hGHa in combination with D74R1 and hGHa with D73R1 respectively. Similarly lanes B and C of cDMD3 represent the product obtained with primers hGHa with D74R1 and hGHa with D72R1 respectively. The length of the products obtained is indicated. The 339 bp product visible in lane A of cDMD3 results from DNA contamination in the RNA sample and represents the endogenous growth hormone gene. Marker, 100 bp ladder (Gibco-BRL).

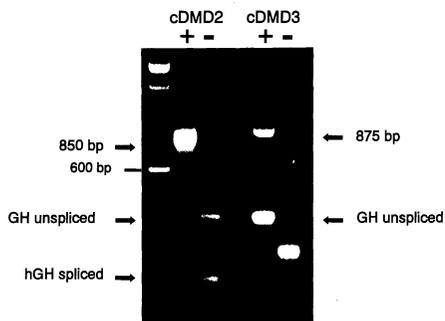


Figure 4. The effect of orientation of the cosmid insert. +, cosmid insert in sense orientation; -, cosmid insert in antisense orientation. Marker, 100 bp ladder (Gibco-BRL).

different primers, hGHORF1-3, containing a T7 promoter, a translation initiation sequence and either no, one or two additional nucleotides inserted between the ATG translation initiation codon and the hGH-sequence. The resulting RT-PCR products, each introducing a different reading frame, were used in an *in vitro* transcription-translation assay to scan for the presence of an open

reading frame (ORF). As a control, a 0.6 kb PCR product of the hGH gene was synthesized in the three reading frames, using the same forward primers in combination with primer hGHUTR1, located in the 3'-UTR of the hGH gene. The control hGH product synthesized using primer hGHORF1 (Fig. 5), was predicted to contain an ORF of 172 amino acids and yielded the expected peptide of ~20 kDa, while no product was obtained in the two other reading frames. Similarly, *in vitro* transcription-translation of the cDMD3-derived hGHORF1 RT-PCR product yielded a peptide slightly over 30 kDa, as expected for the 230 amino acid ORF (Fig. 5). This system is based on our earlier published 'protein truncation test' (PTT) system for the detection of open reading frames by *in vitro* transcription-translation (14).

DISCUSSION

The cosmid-based exon trapping method described in this paper copes with several limitations of currently available exon trapping methods. Using large genomic inserts of 30 kb and larger, we isolated all exons present as a complete set, eliminating the need of subcloning and reordering of individually isolated exons and verification of their continuity from isolated cDNAs. If the cosmid inserts were in the antisense orientation, either nothing or a small product (i.e. cDMD3r) was trapped. The relevance of the latter product is still unclear; it either contains several cryptic exons or is part of a newly identified transcription unit. We did not trap any

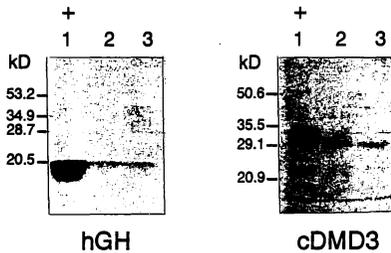


Figure 5. *In vitro* transcription-translation of RT-PCR products. The left side of the figure shows the translated peptide obtained with hGH and the right side the translated peptide obtained with the cDMD3 hGH/DMD chimeric transcript. Lanes 1-3, products amplified with primer hGHORF1, 2 and 3 respectively. Marker, pre-stained SDS-PAGE Low Range Marker (BioRad). The + indicates which of primers hGHORF1-3 should result in an in frame product.

false exons and the false positive rate obtained was in fact zero. Splicing was perfect between all DMD exon-exon transitions and hGH exon 2 and DMD exons. In both cosmids analysed the last DMD exon was not spliced to the splice acceptor of hGH exon 3, but to a site directly upstream of or in the multiple cloning site. This indicates the existence of *cis*-active 'higher order' effects in splicing, further underscoring the advantage of concerted trapping of a series of unknown exons, selected during evolution to cooperate in their parent gene. When separate exons are inserted in an 'alien' context this fine-tuning will be lost, which is probably the explanation for the differences in trapping efficiency of different exons using current systems. Alternatively, but not mutually exclusive, the selection of cryptic splice sites could be related to the maintenance of an open reading frame which has recently been shown to be an important factor influencing splice site selection (15).

We have constructed several variants of sCOGH1. In sCOGH2 the *Alu* repeat in the 3'-UTR of hGH has been removed, facilitating the screening of human positive cosmids with radiolabelled human DNA after subcloning of, for example, YACs from a mixture of YAC and total yeast DNA. In sCOGH6 a 4.7 kb fragment containing the SV2neo selectable marker has been removed, facilitating cloning of larger inserts. sCOGH3 differs from sCOGH1 by a deletion of the mMT1/hGH-exons 1 to 2 region (i.e. the promoter and 5' end of the gene). Due to the removal of the promoter and 5' end of the hGH-gene no RNA will be produced unless an insert contains a 5'-first exonic gene segment and a promoter which is active in the chosen cell line. These 5'-exonic sequences can be isolated efficiently from the RNA using a 5' RACE protocol (13).

RNA production can be boosted by super-inducing the mMT-1 promoter with heavy metal ions such as Zn^{2+} and Cd^{2+} (16). Neomycin selection can be used to select transfected clones specifically, but the system as described works so efficiently that we have never applied this selection, and in fact removed the neo gene in the sCOGH6 vector to generate 4.7 kb more space. The vectors used do not require a specific system for replication in the host cell and can be used in combination with any *in vivo* or *in vitro* system able to produce correctly processed RNA. In particular, due to the use of a strong ubiquitously expressed promoter, the necessity to use COS-1 cells for the initiation of transcription from the SV40 promoter is eliminated. The

sCOGH-system allows one to use other cell types (e.g. hamster V79), opening up several possibilities including targeting of tissue specific genes, e.g. in combination with sCOGH3, and functional complementation in specific cell types.

The results described in this paper deal with single cosmids containing part of a large gene. In gene rich regions more than one gene might be present in the cosmid insert and it is unclear what would happen in such an event. Most likely, transcripts initiated at the strong mMT-1 promoter will overrun cloned promoters, a situation similar to known genes with multiple promoters (17). We expect that the presence of cloned 3' exons will usually cause transcription termination. Still, examples are known where genes have multiple 3' exons, often expressed in a tissue specific manner, indicating that transcription can proceed and trap downstream sequences. Since with this system two RT-PCR reactions are standard, one with hGH exon 2 and 3 primers and one 3' RACE reaction, in most cases where multiple genes are cloned one should at least trap sequences from the most upstream gene. The identification of all genes from gene rich regions will depend on the use of a highly redundant cosmid contig covering the region. To scan large regions with the sCOGH-system, one has two possibilities: perform one experiment with a mixture of cosmids or use every cosmid in a single experiment. The feasibility of using complex mixtures remains to be tested. However, the situation will not differ significantly from that using small-insert vectors, where the high complexity of the input material introduces several technical problems. First, a large proportion of the clones will be empty and produce a PCR-favoured small product. Secondly, a wide range of products will be trapped with large size differences making it difficult to recognise the individual products. Consequently, PCR conditions should be chosen carefully to allow amplification of a wide size-range of RT-PCR products, especially for the cosmid-based system e.g. by using long-range PCR protocols. Since each cosmid contains a 25-40 kb insert, covering extensive regions with a manageable number of clones should be possible. Therefore, we would opt to use multiple cosmids simultaneously but in a miniaturised exon trap experiment where the cosmids are not mixed.

As demonstrated using RT-PCR and *in vitro* transcription-translation of products synthesized in all three possible reading frames derived from the hGH control and cDMD3, the exon trapping system can be coupled with a direct transcription-translation test (TTT) to detect the presence of large ORFs in the isolated sequences. This TTT approach provides an efficient tool to discriminate *bona fide* coding sequences from false positives. At the same time, this assay facilitates the identification of mutations by comparison of translated products derived from different sources of input genomic DNA, e.g. normal versus patient samples. Recently, we have shown that such a test can be performed even when only limited parts of a newly identified coding sequence have been elucidated (18). Since the proper connection of adjacent exons provides for correct translation, any disturbance in patient samples will become immediately apparent and highlight the area to be sequenced. In this way we could identify the CBP gene as the gene involved in Rubinstein-Taybi by the detection of translation terminating mutations in some patient-derived products.

ACKNOWLEDGEMENTS

We are grateful to Paola van der Bent-Klootwijk for technical assistance. This work was supported by grants from the Netherlands

Organisation for Scientific Research (NWO), Council for Medical and Health Research, project nos 900-716-818 and 900-716-830.

REFERENCES

- 1 Duyk,G.M., Kim,S., Myers,R.M. and Cox,D.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8995-8999.
- 2 Buckler,A.J., Chang,D.D., Graw,S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman,D.E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005-4009.
- 3 Auch,D. and Reth,M. (1990) *Nucleic Acids Res.* **18**, 6743-6744.
- 4 Church,D.M., Stotler,C.J., Rutter,J.L., Murrell,J.R., Trofatter,J.A. and Buckler,A.J. (1994) *Nature Genet.* **6**, 98-105.
- 5 Andreadis,A., Nisson,P.E., Kosik,K.S. and Watkins,P.C. (1993) *Nucleic Acids Res.* **21**, 2217-2221.
- 6 Datson,N.A., Duyk,G.M., Van Ommen,G.J.B. and Den Dunnen,J.T. (1994) *Nucleic Acids Res.* **22**, 4148-4153.
- 7 Krizman,D.B. and Berget,S.M. (1993) *Nucleic Acids Res.* **21**, 5198-5202.
- 8 Evans,G.A., Lewis,K. and Rothenberg,B.E. (1989) *Gene* **79**, 9-20.
- 9 Selden,R.F., Howie,K.B., Rowe,M.E., Goodman,H.M. and Moore,D.D. (1986) *Mol. Cell. Biol.* **6**, 3173-3179.
- 10 Den Dunnen,J.T., Grootsholten,P.M., Dauwerse,J.D., Monaco,A.P., Walker,A.P., Butler,R., Anand,R., Coffey,A.J., Bentley,D.R., Steensma, H.Y. *et al.* (1992) *Hum. Mol. Genet.* **1**, 19-28.
- 11 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- 12 Wapenaar,M.C., Kievits,T., Blonden,L.A.J., Bakker,E., Den Dunnen,J.T., Van Ommen,G.J.B. and Pearson,P.L. (1987) *Cytogenet. Cell Genet.* **46**, 711.
- 13 Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.
- 14 Roest,P.A.M., Roberts,R.G., Sugino,S., Van Ommen,G.J.B. and Den Dunnen,J.T. (1993) *Hum. Mol. Genet.* **2**, 1719-1721.
- 15 Dietz,H.C. and Kendzior,R.J., Jr. (1994) *Nature Genet.* **8**, 183-188.
- 16 Searle,P.F., Stuart,G.W. and Palmiter,R.D. (1985) *Mol. Cell. Biol.* **5**, 1480-1488.
- 17 Ahn,A.H. and Kunkel,L.M. (1993) *Nature Genet.* **3**, 283-291.
- 18 Petrij,F., Giles,R.H., Dauwerse,J.G., Saris,J.J., Hennekam,R.C.M., Masuno,M. Tommerup,N., Van Ommen,G.J.B., Goodman,R.H., Peters,D.J.M. *et al.* (1995) *Nature* **376**, 348-351.

