Nucleic Acids Research, 1996, Vol. 24, No. 20 **4023–4028**

Inappropriate splicing of a chimeric gene containing a large internal exon results in exon skipping in transgenic mice

Robin L. Davisson, Nina Nuutinen[+], Sean T. Coleman and Curt D. Sigmund*

Departments of Internal Medicine and Physiology & Biophysics, University of Iowa College of Medicine, Iowa City, IA 52242, USA

Received June 11, 1996; Accepted September 3, 1996

ABSTRACT

We generated transgenic mice containing a chimeric construct consisting of the α-cardiac myosin heavy chain (αcMHC) promoter and the human renin (hRen) gene in order to target hRen synthesis specifically to the heart. The construct consisted of three segments: (i) an αcMHC DNA segment including 4.5 kb of 5′ flanking DNA and an additional 1.1 kb of genomic DNA encompassing exons I–III (non-coding) and the first two introns; (ii) a partial hRen cDNA consisting of exons I–VI; and (iii) a hRen genomic segment containing exons VII through IX, their intervening introns, and 400 bp of 3′ flanking DNA. This results in the formation of a 909 bp internal fusion exon consisting of αcMHC, polylinker, and hRen sequences. Despite the presence of splice acceptor and donor sites bracketing this exon, transcription of this transgene resulted in a major alternatively spliced mRNA lacking the exon and therefore a majority of the hRen coding sequence. Cloning and sequencing of RT–PCR products from several heart samples from two independent transgenic lines confirmed accurate and faithful splicing of αcMHC exon II to hRen exon VII thus bypassing the internal fusion exon. All other exons (αcMHC exons I and II and hRen exons VII, VIII and IX) were appropriately spliced. These results are consistent with the hypothesis on exon definition which states that internal exons have a size limitation. Moreover, the results demonstrate that transgenes present in the genome at independent insertion sites and in either a single copy or multiple copies can be subject to exon skipping. The implications for transgene design will be discussed.

INTRODUCTION

Transgenic mice have been and continue to be important tools for the analysis of gene regulation, developmental expression, and as models for studying human disease. Although the most essential step in the generation of a transgenic model is the design of the transgene construct, by and large there are no concrete rules governing this process (reviewed in 1). The minimal requirements for all transgene constructs include: (i) a transcriptional initiation sequence composed of a promoter and associated regulatory sequences; (ii) the protein coding region of the gene of interest; and (iii) a polyadenylation sequence for appropriate 3′ end processing. In addition, studies have demonstrated that constructs which contain introns are expressed more reproducibly and at a higher level than constructs lacking introns (2–4). This either reflects a requirement for splicing for the normal maturation, transport and stability of a mRNA or the presence of important transcriptional regulatory sequences within introns. These hypotheses are supported by data demonstrating that (i) a heterologous generic intron can increase the transcriptional efficiency of a human histone H4 promoter-chloramphenicol acetyltransferase fusion transgene (2); (ii) expression of the rat growth hormone gene under the control of the mouse metallothionein promoter requires endogenous growth hormone introns in their normal position or a heterologous intron between the promoter and growth hormone gene (3); (iii) expression of the p53 gene under the control of the SV40 promoter/enhancer requires endogenous p53 introns (5); and (iv) the presence of enhancers or silencers within introns of numerous genes are necessary for appropriate transcriptional regulation of many promoters in transgenic mice (4,6,7).

Transgene constructs derived from genomic clones that include 5′ and 3′ flanking DNA and all associated exons and introns of a gene are often reliably expressed in multiple independent transgenic lines if sufficient 5′ flanking DNA is employed in the construct. Nevertheless, transgene expression among lines, which differ in integration site and copy number, can often be widely variable and disproportional to transgene copy number. Exceptions to this have been noted in transgenic mice containing large genomic constructs derived from cosmids, P1 clones, or yeast artificial chromosomes (8–10). However, in contrast to genomic transgenes, the level and reproducibility of expression among lines becomes less predictable when fusion transgenes consisting of the promoter from one gene and coding region of a heterologous gene are used. These types of transgenes are often generated when one wishes to confer a highly restricted tissue-specific, cell-specific and/or developmental-stage specific expression profile on a protein of interest. Often the heterologous gene is encoded in the form of a cDNA lacking all intron sequences. As such, the simple fusion of the promoter and cDNA does not always reproducibly result in appropriate tissue-specific or high level expression in transgenic

* To whom correspondence should be addressed

[+]Present address: Department of Pharmacology and Toxicology, University of Oulu, Kajaanintie 52 D, 90220 Oulu, Finland

mice even if the promoter was previously shown to direct appropriate expression of a genomic construct in mice (4,11). This may be due to important transcriptional or other maturation functions provided by intron sequences. The addition of a generic cassette containing the donor and acceptor sites of a heterologous intron followed by a polyadenylation signal has been used effectively to enhance the expression of some transgenes (2) but not others (3).

Taken together, the accurate targeting of a reporter gene or protein to a specific cell or tissue type may require the construction of a complex chimeric transgene consisting of (i) a promoter, including 5′ flanking DNA, the transcriptional initiation site, and a portion of the exon/intron region downstream of the promoter which contains regulatory elements necessary for promoter function, (ii) one or more exons (and introns) of the heterologous gene encoding the desired protein, and (iii) a polyadenylation signal either provided by the 3′ flanking DNA of the heterologous gene or as an added cassette. Unfortunately, the increased complexity of such chimeric constructs may lead to less predictability concerning the expression of transgenes of varying design.

We generated such a chimeric construct containing the α-cardiac myosin heavy chain (α-cMHC) promoter and the human renin (hRen) gene coding sequence in order to specifically target the expression of hRen to cardiac myocytes. In designing this construct we generated a large artificial internal exon within the construct which fuses α-cMHC sequences to hRen coding sequences. This exon is surrounded by an upstream intron from the α-cMHC gene and a downstream intron from the hRen gene. Herein, we demonstrate high level tissue-specific expression but inappropriate splicing resulting in exon skipping of the chimeric exon. Because of its size (909 bp) this chimeric exon potentially violates the ‘exon definition’ hypothesis which states that exon sequences may be the recognition unit of splice site selection and that the presence of internal exons > 300–400 nucleotides (nt) in length could lead to either exon skipping during splicing or the recognition and use of cryptic splice sites (reviewed in 12). Our results suggest that the size of internal exons should be considered as an additional factor when designing chimeric constructs for use in transgenic animals.

MATERIALS AND METHODS

Generation of the αcMHC-hRen construct

The αcMHC promoter segment (13) was cloned (generous gift of Jeffrey Robbins, University of Cincinnati) as an *Xho*I to *Hin*dIII fragment into pBluescript II SK– (Stratagene, La Jolla, CA) to form the plasmid pMHC-1. This construct contains 4.5 kb of αcMHC 5′ flanking DNA and a 1071 bp segment containing exons I, II and 9 bp of exon III along with their intervening intron sequences. Initiation of αcMHC translation normally starts further downstream within exon III. The *Hin*dIII site is the terminal restriction site in an 80 bp polylinker segment (*Not*I, *Kpn*I, *Apa*I, *Sal*I, *Xba*I, *Sac*I, *Pst*I, *Eco*RV, *Nco*I, *Hin*dIII) which was carried along with the αcMHC DNA.

The hRen cDNA/genomic chimera was generated by first cloning a *Hin*dIII to *Eco*RI fragment from the plasmid pRhR1100 (generous gift of Tim Reudelhuber, Clinical Research Institute of Montreal, Montreal, Canada) into pBluescript II SK– to form the plasmid phRen[I–V]. This fragment contains a *Hin*dIII linker placed at the –9 position relative to the translational start site for hRen. The presence of an *Eco*RI site within exon VI (first six bases) facilitated the construction. The hRen genomic segment was cloned from a previously described genomic clone (14) as an *Eco*RI to *Bgl*II fragment and cloned into the *Eco*RI and *Bam*HI sites of phRen[I–V] to form the plasmid phRen[I–V–IX]. This chimeric hRen gene was then cloned as a *Hin*dIII to *Spe*I fragment into pMHC-1. All cloning junctions were confirmed by sequencing. The presence of the splice acceptor site at the 3′ end of αcMHC intron 2 and the splice donor site at the 5′ end of hRen intron 6 were confirmed by sequencing.

Generation of transgenic mice

The αcMHC-hRen transgene was purified for microinjection by digestion with *Xho*I and *Spe*I followed by gel electrophoresis and purification of the transgene DNA away from the prokaryotic vector DNA. The transgene was purified using a SpinBind column (FMC BioProducts, Rockland, Maine) and the DNA concentration was estimated by agarose gel electrophoresis using standards of similar molecular weight. The purified transgene DNA was diluted to 2 µg/ml in 10 mmol/l Tris–HCl pH 7.5, 0.1 mmol/l EDTA made with embryo culture certified water (Sigma) and the concentration was confirmed using a DNA Dipstick (Invitrogen, San Diego, CA). All mice were fed standard mouse chow and water ad libitum. Care of the mice used in the experiments met or exceeded the standards set forth by the National Institutes of Health in their guidelines for the care and use of experimental animals. All procedures were approved by the University Animal Care and Use Committee at the University of Iowa.

Transgene DNA was microinjected into the male pronucleus of 1-cell fertilized mouse embryos derived from [C57BL/6J× SJL/J]F_2 (B6SJL) as previously described (15,16). Transgenic founders were detected by PCR amplification using primer set 4 (Table 1) of DNA samples isolated from tail biopsy samples as previously described (15). An 893 bp band was diagnostic of the presence of the transgene. Transgenic offspring were obtained by breeding founder transgenic mice to nontransgenic B6SJL/J or by backcross breeding to C57BL/6J mice. All mice were obtained from the Jackson Laboratory (Bar Harbor, Maine). Transgenic offspring were differentiated from their non-transgenic littermates by PCR using the same primer set.

Dot blots were performed by first denaturing 2.0 µg of genomic DNA isolated from tail biopsies with 0.4 N NaOH, neutralizing the solution and then applying the samples to nylon-supported nitrocellulose using a dot blot manifold. Blots were crosslinked with ultraviolet light and hybridized under standard conditions (17) with an [α-^{32}P]dCTP random primer labeled complete hRen cDNA probe. Blots were washed stringently (0.1× SSC, 0.1% SDS at 65°C) and exposed to X-ray film overnight.

Gene expression analysis

Total tissue RNAs were isolated by homogenization in guanidine isothiocyanate followed by phenol emulsion extraction at pH 4.0 using a modification of the method previously described (18,19). Homogenizations were scaled up to 2.5 ml to increase RNA yield and quality. Twenty micrograms of total tissue RNA was separated on 1.5% agarose formaldehyde gels and transferred to nylon-supported nitrocellulose as previously described (17). Blots were hybridized to either a 5′ hRen cDNA or 3′ hRen cDNA probe (Fig. 1) which were labeled by generating a single-stranded antisense RNA. The 3′ hRen probe has been described previously (14). The 5′ cDNA probe was cloned as a *Hin*dIII to *Eco*RI fragment into pGEM-3 (Promega, Madison, WI) and was derived

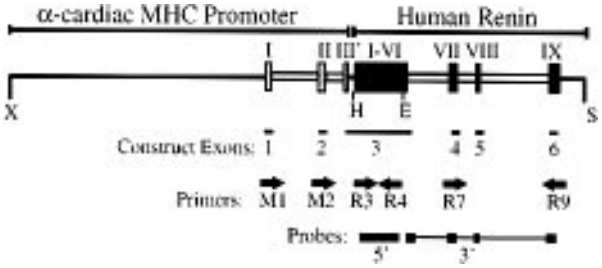

Figure 1. Schematic representation of the transgene. A map of the chimeric construct used to generate transgenic mice is shown. The αcMHC and hRen exons are denoted by hatched bars and solid bars, respectively, and are numbered in roman numerals to reflect endogenous exons in the parent genes. The re-defined exons of the constructs is shown below the map along with the approximate location of the PCR primers and their priming direction. The position of the 5′ and 3′ probes used for the northern blot analysis is also shown. The 3′ probe is broken into segments to reflect its origin from a cDNA clone. X, *Xho*I; H, *Hin*dIII; E, *Eco*RI; S, *Spe*I.

from exons I–V of the hRen cDNA. To insure the specific detection of hRen transcripts, northern blots were treated with 1.0 µg/ml RNAse A in 2× standard saline citrate for 15 min at room temperature. We have previously demonstrated that this procedure removes non-specific hybridization of single stranded RNA probes (14).

For reverse transcriptase polymerase chain reaction (RT–PCR), 10 µg of total heart RNA was treated with 2.0 U RQ1 DNase (Promega) in 40 mM Tris–HCl pH 7.9, 10 mM NaCl, 6 mM $MgCl_2$, 10 mM $CaCl_2$ and 200 U RNasin RNase inhibitor (Promega) for 30 min at 37°C. The RNA was extracted twice with phenol–chloroform–isoamyl alcohol (25:24:1, v/v/v) and once in chloroform–isoamyl alcohol (24:1 v/v) followed by ethanol precipitation to inactivate the DNase prior to RT–PCR. RT–PCR was performed by mixing 1.0 µg of total heart RNA in reverse transcriptase buffer containing, 1.0 mM nucleotide triphosphates, 10 mM dithiothreotol, 200 U RNAsin, 200 ng random hexanucleotides and 200 U MMLV reverse transcriptase in a 20 µl volume. Reverse transcriptase was left out of control reactions. The reactions were incubated at 25°C for 10 min, 42°C for 1 h, 95°C for 5 min, and then were immediately placed on ice. One to five microliters of the RT reaction was used for PCR amplification. For PCR, the reactions were scaled to 100 µl containing 1× *Taq* buffer, 2.5 mM $MgCl_2$, 200 µM nucleotide triphosphates, 20.0 pmol of forward and reverse primers (see Fig. 1) and 5.0 U *Taq* polymerase (Boehringer Mannheim or Perkin-Elmer). Amplification was carried out under oil in a Perkin-Elmer 480 or MJ Research PTC-100 thermal cycler using the following thermal profile: 94°C for 1 min, 55°C for 1 min and 72°C for 2 min for 30 cycles. The 94°C incubation was increased to 5 min for the first cycle and the extension was increased to 5 min for the last cycle. PCR products were visualized by ethidium bromide staining after agarose gel electrophoresis and were cloned into the pCRII vectors using the TA cloning kit and methodology recommended by the vendor (Invitrogen). DNA sequencing was either performed using the Sequenase Kit (United States Biochemical) or by automated sequencing using the University of Iowa DNA Core Facility.

The primers for the PCR analysis were: M1-CAGAGATTTC-TCCAACCC, M2-CACTGTGGTGCCTCGTTCCAG, R3-TGA-CACTGGTTCGTCCAATG, R4-ATAGCGGAGGGTAGTTCTG, R7-GGGTCATCCACCTTGCTC, R9-CTTTCGGATGAAGGT-GGC.

RESULTS

We generated a construct consisting of the αcMHC promoter fused to the hRen gene in an effort to specifically target hRen production to the heart in order to examine the pathophysiological consequences of elevated heart-specific activity of the renin–angiotensin system. When designing this construct we were conscious of the need to include a highly tissue-specific and developmentally-regulated promoter to target expression to the desired tissue and cell types. We used a 5.6 kb segment of the αcMHC gene containing 4.5 kb of its 5′ flanking DNA and 1071 bp of contiguous genomic DNA terminating at position 9 of exon III (Fig. 1). This segment has previously been shown to target myocardial-specific expression of a reporter gene in adult animals and to mimic the developmental expression of the endogenous αcMHC gene (13). The hRen gene segment consisted of a cDNA encoding exons I–V fused to a genomic sequence encoding exons VI–IX. This allowed us to reduce the size of the hRen gene which spans 14 kb on human chromosome 1 to ~3 kb while retaining several spliceable introns and a viable 3′ end for poly A addition (Fig. 1). We felt this strategy to be important since the presence of introns has been previously reported to be important for high level expression of transgenes in mice (2–4), and fusion transgenes containing endogenous introns in their normal position of the protein coding portion of the transgene are reproducibly expressed while similar constructs containing cDNAs are not (3,5). Moreover, we previously demonstrated that a hRen genomic construct including all exons, introns and 400 bp of 3′ flanking DNA was expressed in tissue-specific, cell-specific and developmentally regulated fashion in transgenic mice (14,20). This unique gene fusion resulted in the generation of a 909 bp internal exon (exon 3 of the construct, Fig. 1).

Table 1. Size predictions for PCR products

	Primers		Product size		
Primer set	Upstream	Downstream	Full length mRNA	Truncated mRNA	Genomic DNA
1	M1	R9	1348	439	N/A
2	M2	R9	1321	412	N/A
3	R7	R9	339	339	N/A
4	M2	R4	602	–	893
5	M1	R4	629	–	1601
6	R3	R9	859	–	N/A

The primer sets used in the study and the predicted sizes of the amplification products are shown. The primer names shown in the table refer to those shown in Figure 1.

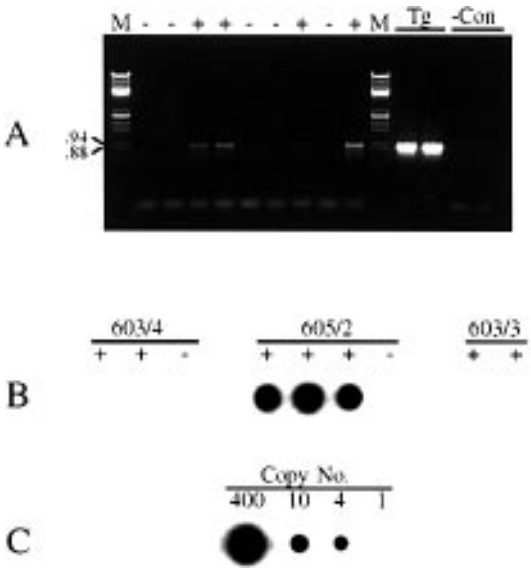

Figure 2. Identification of transgenic founders by PCR and dot blot analysis. (**A**) Genomic DNA isolated from tail biopsy samples was amplified using primer set 4 (Table 1). PCR products were resolved on 1.5% agarose gels and stained with ethidium bromide. M, marker lane composed of a mixture of *Hin*dIII, and *Hin*dIII + *Eco*RI digested phage λ DNA; –, non-transgenic; +, transgenic founder; Tg, PCR product derived from purified transgenic DNA (10 pg), –CON, negative control DNA. The order of transgenic founders on the gel is 603/3, 603/4, 604/3 and 605/2. (**B**) Dot blot analysis of tail DNAs from transgenic offspring. Each dot is from a separate animal from the indicated transgenic line. +, transgenic offspring; –, non-transgenic littermate. (**C**) Dilution series of purified transgene DNA was used to estimate the copy number of the transgene in each line. The approximate genome equivalent is list above each dot.

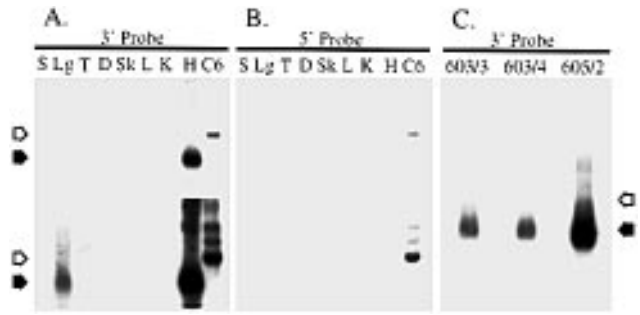

Figure 3. Northern blot analysis of transgene expression. Total tissue RNA from a 605/2 transgenic mouse was purified and blotted as described in the Materials and Methods and probed with either the 3′ probe (**A**, **C**) or the 5′ probe (**B**). The upper and lower blots in (A) and (B) are 18 h and 6 day exposure, respectively. The open and closed arrowheads point to the location of the full length hRen RNA and truncated transgene mRNAs, respectively. All three lanes in (C) were derived from the same northern blot and the transgenic lines are indicated. S, spleen; Lg, lung; T, testes; D, diaphragm; Sk, skeletal muscle; L, liver; K, kidney; H, heart; C6, Calu-6 cells.

Fifty-nine offspring were screened for the presence of the transgene by PCR analysis of partially purified DNA samples isolated from tail biopsies using primer set 4 (Table 1) which spanned the αcMHC promoter and hRen cloning junction (Fig. 2A). Four transgenic founders were identified and each was bred to establish independent lines. These founders were designated 603/3, 603/4, 604/3 and 605/2. The presence of an 893 bp amplified band in the founders (Fig. 2A) and their offspring (data not shown) when primer set 4 was used, and a 1.6 kb band in their offspring when primer set 5 (Table 1) was used (data not shown) eliminated the possibility that a major rearrangement occurred during transgenesis. Dot blot analysis of DNA from several transgenic offspring from three independent lines demonstrated that the number of transgene copies integrated into the genome was highly variable among the lines (Fig. 2B). A comparison with known amounts of a standard DNA sample (Fig. 2C) revealed the copy number to be ~1 for line 603/4, 1–2 for line 603/3 and >100 for line 605/2. This variation in transgene copy number is a natural consequence of the microinjection process (reviewed in 1).

In order to evaluate the tissue-specific expression of the transgenes, total tissue RNA from offspring from each transgenic line was examined by northern blot hybridization. A representative northern blot from transgenic line 605/2 is shown in Figure 3. Probes for this analysis were derived from hRen cDNAs encoding exons VI–IX (3′ probe) or I–V (5′ probe). RNA from a pulmonary carcinoma cell line (Calu-6) which expresses full length, appropriately processed endogenous hRen mRNA (21) was used as a control (Fig. 3, lane C6). High level expression of a faster migrating band was evident in heart when the 3′ probe was used. Low level expression was also evident in lung. Fully processed transgene mRNA should be ~70 bp longer than hRen mRNA based on the design of the construct. No expression was evident in brain, spleen, testes, liver, kidney, or in other muscle types including diaphragm and skeletal muscle consistent with the tissue-specific expression of the αcMHC promoter (13). An identically fast migrating hRen hybridization band was observed in the heart in transgenic lines 603/3 and 603/4 (Fig. 3C), while no expression was detected in the 604/3 transgenic line (data not shown). In order to evaluate whether the truncated hRen mRNA observed in the heart was due to aberrant splicing of the chimeric exon (exon 3) we repeated the northern hybridization using the hRen 5′ probe. This resulted in a nearly complete absence of hRen transcripts in the heart and lung strongly suggesting that exon 3 was being alternatively spliced in these mice (Fig. 3B). The 5′ probe still recognized the full length hRen mRNA present in Calu-6 cells.

RT–PCR analysis was performed to precisely examine the utilization of splice sites in the construct. RT–PCR of heart RNA using primer set 3 (Table 1) correctly amplified a band of 339 bp suggesting appropriate splicing of hRen exon VII to VIII and VIII to IX (Fig. 4A). This was confirmed by sequencing the 3′ end of cloned PCR products generated using primer sets 1 and 2. Sequencing the 5′ end of cloned PCR products generated using primer set 1 similarly demonstrated appropriate splicing of MHC exon I to II. Although the sequence at the ends of the PCR products generated with primer sets 1 and 2 were appropriate, the size of the PCR products themselves (439 versus 1348 bp for primer set 1 and 412 versus 1321 bp for primer set 2) were consistent with an internal deletion of ~900 bp which could be accounted for by a deletion of mRNA encoded within exon 3 of the construct (Fig. 4A). This was confirmed by sequence analysis which clearly demonstrated the accurate and faithful splicing of MHC exon II to hRen exon VII (Fig. 5). Identical sequence results were obtained from several independently cloned PCR products from heart samples from different animals and from two independent transgenic lines (Fig. 4B).

In order to assess whether any correctly spliced transgene mRNA was present in the heart, RT–PCR was performed using

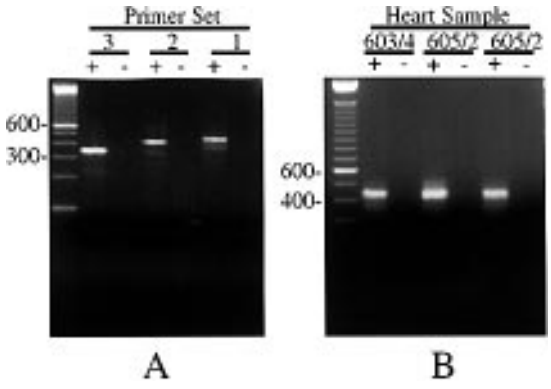

Figure 4. RT–PCR analysis of splice variants. RT–PCR was performed on total RNA isolated from the heart of transgenic mice as described in the Materials and Methods. +, reactions performed with reverse transcriptase; –, control reactions performed in the absence of reverse transcriptase. (**A**) RT–PCR was performed using the indicated primer set (Table 1) on heart RNA isolated from a 603/4 line transgenic mouse. 100 bp DNA ladder is shown in the leftmost lane. (**B**) RT–PCR was performed using primer set 1 (Table 1) on heart RNA from a transgenic mouse of the indicated line. 100 bp DNA ladder is shown in the leftmost lane.

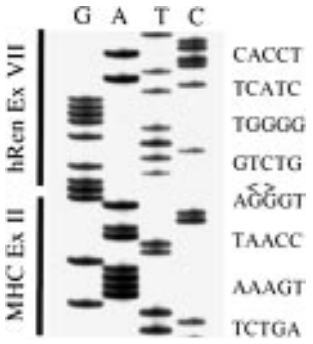

Figure 5. Sequence of the alternatively spliced transgene mRNA. A DNA sequence ladder showing the accurate splicing of αcMHC exon II to hRen exon VII is shown. <,> indicates the splice junction. This sequence shown was derived from a cloned PCR product derived from primer set 1 (line 603/4) using an SP6 sequencing primer. This sequence was confirmed on two independent clones from each sample and on two independent heart samples from the 603/4 and 605/2 transgenic lines.

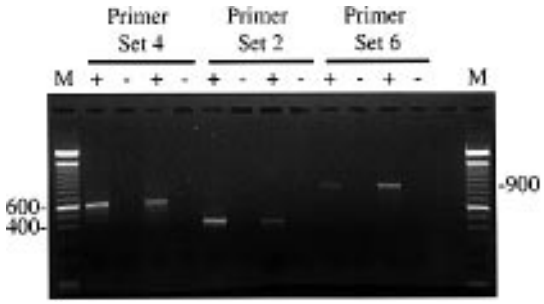

Figure 6. RT–PCR analysis of fully processed transgene mRNA. RT–PCR was performed on total heart RNA of two independent 605/2 line transgenic mice using the indicated primer sets. +, reactions performed with reverse transcriptase; –, control reactions performed in the absence of reverse transcriptase. M, 100 bp DNA ladder.

primer sets which would amplify the exon 2 to 3 junction (primer set 4) and the exon 3 to 4 junction (primer set 6). Both primer sets utilize a PCR primer present within exon 3 of the construct and would therefore only amplify a band of a specific size (602 bp for set 4 and 859 bp for set 6) if transgene mRNA containing exon 3 was present. Specific RT–PCR products of the expected size were evident in two independent heart RNA samples when both primer sets were used (Fig. 6). Sequence analysis of these products confirmed appropriate splicing of αcMHC exon II to III and hRen exon VI to VII. The results strongly suggest that a small proportion of the transgene precursor mRNA was appropriately spliced into full length message. Nevertheless, we estimate that >95% of the transgene mRNA lacks exon 3 sequences based on the northern blot and RT–PCR data.

DISCUSSION

The first and arguably most critical step in the development of a novel transgenic model is the design of the transgene construct, as it is the construct which ultimately dictates the overall pattern of tissue-specific, cell-specific, developmentally-regulated and hormonally responsive expression observed in transgenic animals. In theory, transgene design simply requires a choice of which protein should be expressed and the pattern of expression desired. In practice however, this is limited by the availability of promoters to target appropriate temporal and spatial expression of the transgene, and the need to not only generate a primary transgene transcript but also to process and transport a stable mature transgene mRNA from the nucleus to the cytoplasm. Although examples of appropriate expression of intronless transgenes in mice have been reported (22,23), it is well documented that transgenes which contain either native (3–5) or heterologous (2,3) introns are expressed more reliably than transgenes consisting merely of a cDNA. Therefore, one could reasonably postulate that mRNA splicing may be an important part of the maturation, transport and stability of a mRNA. Interestingly, Brinster *et al.* (4) demonstrated that the rate of transcription was significantly higher in intron-containing constructs than in those lacking introns, suggesting that the intron plays an additional role in regulating transcriptional activity. Indeed, numerous examples of enhancers and silencers that influence expression of transgenes in mice have been reported in introns (6,7). In addition, introns have been reported to align nucleosomes on a transcribed gene (24), direct nuclear matrix attachment (25), and cause enhanced transcript stability in response to spliceosome assembly (26).

Whatever the mechanism, it is clear that introns should be considered a critical component when designing transgene constructs. The data reported to date suggests that elements which regulate expression of a gene, such as flanking DNA and promoters, function in concert with sequences present within the coding region and in introns at the level of chromatin, suggesting that the context in which introns are presented can be a crucial feature. Indeed, this is supported by experiments in which heterologous introns are able to rescue expression of an intronless transgene when placed between the promoter and cDNA but not when placed downstream of the cDNA (3). This would suggest that in the generation of chimeric constructs the most desirable design would be to retain as much of the normal genomic structure as possible.

In the experiments described in this report, we attempted to generate a construct which contains as many of the features of the

native α-cMHC and hRen genes as possible while eliminating the cumbersome nature of a 14 kb section of genomic DNA containing several large introns and numerous six base restriction sites which can complicate cloning. In doing so, we generated a chimeric construct which retains the essential 5′ flanking DNA and exon/intron regions necessary for appropriate function of the α-cMHC promoter, and reduced the 14 kb genomic clone to approximately 3 kb by fusing a cDNA encoding exons I–V with a genomic segment encoding the remainder of the protein and 3′ flanking DNA containing the polyadenylation site. Although high level tissue-specific expression of the construct was evident in multiple independent transgenic lines, the mRNA was inappropriately spliced and therefore lacked a majority of the protein coding region.

Vertebrate internal exons (i.e. exons which are not attached to CAP or polyadenylation sites) average 137 bp and are rarely $>$300–400 bp in length (reviewed in 12). Therefore, the chimeric exon described in these studies is above the average size of internal mammalian exons and may be outside the size range of exons which are efficiently recognized by the splicing machinery. It is likely that the skipping of this exon is due primarily to its size because: (i) the acceptor and donor sites on the upstream and downstream introns, respectively, are unaltered; and (ii) upstream and downstream exons, which are well within the size range proposed by the model, are efficiently and appropriately spliced together. It is unlikely that our observations result from a defect in the transgene because no gross rearrangements in transgene structure were detected by PCR in genomic DNA from transgenic founders and offspring, and because a small proportion of the total transgene mRNA was appropriately processed.

The mechanism of splice site recognition in eukaryotic genes remains poorly understood. This is largely due to the fact that sequences surrounding vertebrate splice sites are poorly conserved and that introns can vary in size from $<$100 bases to hundreds of kilobases. Although simple consensus sequences appear at the donor and acceptor terminals of all introns, the selection of splice sites must be far more complex in order to explain the faithful generation of fully processed mRNAs from genes which can span great distances of the chromosome. Berget *et al*. (12) has proposed the exon definition model to explain splice site selection in vertebrate genes. The model proposes that the exon is the unit of recognition and that the splicing machinery recognizes a pair of splice sites (an acceptor site at the 3′ end of the upstream intron and a donor site at the 5′ end of the downstream intron) surrounding an exon. One prediction of this model is that the exon would be defined as a sequence of a limited size between donor and acceptor sites. Mechanistically, the model proposes that intron splice sites are identified after the splicing machinery, including U1 and U2 snRNPs and 5′ and 3′ splice site recognition factors, recognizes and binds the intron/exon junction of internal exons. Intron definition then ensues through the assembly of a splicing complex with the subsequent joining of upstream and downstream exons. It remains possible that the limitation on exon size along with the normal small size of most vertebrate exons reflects a maximum distance in which splicing factors can interact across an exon, and may have arisen to prevent the recognition of cryptic splice sites within internal exons which would disrupt protein coding.

In conclusion, the results described in the report are consistent with the skipping of large internal exons as proposed by the exon definition model. Our results suggest that minimizing the size of internal chimeric exons should be considered in the design of transgenes. Unfortunately, the outcome of any particular experimental transgene design is still difficult to predict *a priori*.

ACKNOWLEDGEMENTS

We would like to thank Drs Andrew Russo and Scott Moye-Rowley for their critical review of the manuscript, and to Julie Lang, Norma Sinclair and Lucy Robbins for their superb technical assistance. This work was funded by grants from the NIH (HL48058-03) and American Heart Association. Curt D. Sigmund is an Established Investigator of the American Heart Association. Robin L. Davisson is funded by an institutional post-doctoral fellowship from the NIH (HL07121-20). Transgenic mice were generated and maintained at the University of Iowa Transgenic Animal Facility which is supported in part by the College of Medicine and the Diabetes and Endocrinology Research Center. DNA sequencing was performed by the University of Iowa DNA Core Facility.

REFERENCES

1 Sigmund,C.D. (1993) *Hypertension* **22**, 599–607.
2 Choi,T., Huang,M., Gorman,C. and Jaenisch,R. (1991) *Mol. Cell Biol.* **11**, 3070–3074.
3 Palmiter,R.D., Sandgren,E.P., Avarbock,M.R., Allen,D.D. and Brinster,R.L. (1991) *Proc. Natl Acad. Sci. USA* **88**, 478–482.
4 Brinster,R.L., Allen,J.M., Behringer,R.R., Gelinas,R.E. and Palmiter,R.D. (1988) *Proc. Natl Acad. Sci. USA* **85**, 836–840.
5 Lozano,G. and Levine,A.J. (1991) *Mol. Carcinogenesis* **4**, 3–9.
6 Brooks,A.R., Nagy,B.P., Taylor,S., Simonet,W.S., Taylor,J.M. and Levy-Wilson,B. (1994) *Mol. Cell. Biol.* **14**, 2243–2256.
7 Barash,I., Nathan,M., Kari,R., Ilan,N., Shani,M. and Hurwitz,D.R. (1996) *Nucleic Acids Res.* **24**, 602–610.
8 Grosveld,F., van Assendelft,G.B., Greaves,D.R. and Kollias,G. (1987) *Cell* **51**, 975–985.
9 Gaensler,K.M.L., Kitamura,M. and Wai Kan,Y. (1993) *Proc. Natl Acad. Sci. USA* **90**, 11381–11385.
10 Peterson,K.R., Clegg,C.H., Huxley,C., Josephson,B.M., Haugen,H.S., Furukawa,T. and Stamatoyannopoulos,G. (1993) *Proc. Natl Acad. Sci. USA* **90**, 7593–7597.
11 Sola,C., Tronik,D., Dreyfus,M., Babinet,C. and Rougeon,F. (1989) *Oncogene Res.* **5**, 149–153.
12 Berget,S.M. (1995) *J. Biol. Chem.* **270**, 2411–2414.
13 Subramaniam,A., Jones,W.K., Gulick,J., Wert,S., Neumann,J. and Robbins,J. (1991) *J. Biol. Chem.* **266**, 24613–24620.
14 Sigmund,C.D., Jones,C.A., Kane,C.M., Wu,C., Lang,J.A. and Gross,K.W. (1992) *Circ. Res.* **70**, 1070–1079.
15 Hogan,E., Costantini,F. and Lacy,E. (1986) *Manipulating the mouse embryo.* Cold Spring Harbor Laboratory, Cold Spring Harbor.
16 Yang,G., Merrill,D.C., Thompson,M.W., Robillard,J.E. and Sigmund,C.D. (1994) *J. Biol. Chem.* **269**, 32497–32502.
17 Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) *Current protocols in molecular biology.* Wiley, New York.
18 Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.* **161**, 156–159.
19 Burson,J.M., Aguilera,G., Gross,K.W. and Sigmund,C.D. (1994) *Am. J. Physiol.* **267**, E260–E267.
20 Sigmund,C.D. (1993) *Pediatr. Nephrol.* **7**, 639–645.
21 Lang,J.A., Yang,G., Kern,J.A. and Sigmund,C.D. (1995) *Hypertension*, **25**, 704–710.
22 Jaenisch,R. (1988) *Science*, **240**, 1468–1474.
23 Hanahan,D. (1989) *Science*, **246**, 1265–1275.
24 Liu,K., Sandgren,E.P., Palmiter,R.D. and Stein,A. (1995) *Proc. Natl Acad. Sci. USA* **92**, 7724–7728.
25 Boulikas,T. (1992) *J. Cell. Biochem.* **50**, 111–123.
26 Dreyfuss,G., Swanson,M.S. and Serafin,P.R. (1988) *Trends Biochem. Sci.*, **13**, 86–91.