Nucleic Acids Research, Vol. 20, No. 6 1201–1208

Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene

H.Steingrimsdottir, G.Rowley, G.Dorado[1], J.Cole and A.R.Lehmann*
MRC Cell Mutation Unit, University of Sussex, Falmer, Brighton BN1 9RR, UK and [1]Department Bioquímica y Biología Molecular, Facultad Veterinaria, Universidad Córdoba, 14071 Córdoba, Spain

Received February 6, 1992; Accepted February 20, 1992

ABSTRACT

A large proportion of mutations at the human *hprt* locus result in aberrant splicing of the *hprt* mRNA. We have been able to relate the mutation to the splicing abnormality in 30 of these mutants. Mutations at the splice acceptor sites of introns 4, 6 and 7 result in splicing out of the whole of the downstream exons, whereas in introns 1, 7 or 8 a cryptic site in the downstream exon can be used. Mutations in the donor site of introns 1 and 5 result in the utilisation of cryptic sites further downstream, whereas in the other introns, the upstream exons are spliced out. Our most unexpected findings were mutations in the middle of exons 3 and 8 which resulted in splicing out of these exons in part of the mRNA populations. Our results have enabled us to assess current models of mRNA splicing. They emphasize the importance of the polypyrimidine tract in splice acceptor sites, they support the role of the exon as the unit of assembly for splicing, and they are consistent with a model proposing a stem-loop structure for exon 8 in the *hprt* mRNA.

INTRODUCTION

The X-linked hypoxanthine-guanine phosphoribosyltransferase (*hprt*) gene is the most widely used system for studying mutations in mammalian cells. Mutants selected by their resistance to 6-thioguanine (6TG) are unable to incorporate this toxic analogue into nucleic acids because of a deficiency in HPRT activity. Although *hprt* is dispensible in cultured cells, germ-line mutations in this gene in males result in the X-linked Lesch-Nyhan syndrome, a severe neurological disorder. Recently several laboratories have analysed, at the DNA sequence level, the nature of the mutations in the *hprt* gene which cause the defective activity. Such analyses have been carried out on mutants induced in cultured cells by various mutagens (1–11), on mutant lymphocytes selected directly from the peripheral blood of human subjects (12, 13), and on cells from Lesch-Nyhan patients (14, 15).

In humans, the X-linked *hprt* gene is 44Kb in length and contains 9 exons. The mRNA is approximately 1.6Kb in length and the coding sequence of 651 bases codes for a protein of 217aas. A variety of types of mutations have been identified in the different studies. An unexpected finding is that a large proportion of mutations, whether induced in cultured cells (eg 2, 6, 8, 10) or occurring in individuals (12, 13 and our unpublished observations), resulted in aberrant splicing of the *hprt* mRNA.

Although much is now known about the mechanism of splicing, the reasons for selection of particular splice sites is still not understood. Alternative splicing is a widely used mechanism for varying gene expression during development and for regulating tissue specificity of gene expression. Alternative splicing involves both *cis* elements in the pre-mRNA as well as *trans*-acting protein factors (reviewed in 16). The mechanism of splicing involves the formation of a spliceosome complex which contains small ribonucleoprotein particles (U1,U2,U5 and U4/6 Sn RNPs) and other proteins. During splicing the 5′ splice site is cleaved in the consensus sequence $^{C}_{A}$AG↓GU PuAGU with the formation of a lariat structure, followed by cleavage at the 3′ consensus sequence (Py)nNPyAG↓G and ligation of the 5′ and 3′ exons. Studies on the mechanism of splice site selection have involved the analysis of naturally occurring mutations or the construction of minigenes with exons and introns in a variety of configurations (see Discussion).

In this paper we describe the effects of 34 mutations in the *hprt* gene which result in aberrant splicing. The mutations were generated either by UV-irradiation of SV40-transformed fibroblasts from a xeroderma pigmentosum (complementation group A) patient who was totally deficient in excision-repair of ultraviolet damage (6), or 'naturally' in the circulating lymphocytes of a variety of different donors (17). The full mutation spectra are presented elsewhere (6 and Steingrimsdottir *et al.*, manuscript in preparation).

MATERIALS AND METHODS

Cell culture

(a) Fibroblasts. SV40-transformed XP12ROSV40 cells were grown in Eagle's minimal essential medium + 10% fetal calf serum. The generation, selection and expansion of 6-TG-resistant mutants is described in Dorado *et al.* (6).

* To whom correspondence should be addressed

(b) Lymphocytes. Mononuclear cells were separated from the blood of a variety of donors and plated in microtitre dishes in 6-TG for the selection of mutants, as described by Cole *et al.* (17). Clones were picked and expanded in RPMI 1640 medium supplemented with 10% AB human serum and recombinant interleukin-2. The medium also contained irradiated RJK 853 human lymphoblastoid cells as feeders. These cells contain a total deletion of the *hprt* gene (18). Clones were expanded to at least 10^7 cells.

Analysis of cDNA

Cytoplasmic RNA was extracted from about 10^7 cells by lysis in 0.15M NaCl, 10mM tris-HCl pH7.4, 1mM $MgCl_2$, 0.5%NP40, spinning out the nuclei, and phenol extraction of the cytoplasm. After ethanol precipitation, the RNA was dissolved in H_2O at a concentration of 1mg/ml or greater. 2−5μg RNA was reverse transcribed with murine Moloney-leukaemia virus reverse transcriptase using the procedure of Yang *et al.* (19). After a 1h reaction at 37°, the 10μl reaction mixture was made up to 100μl. 10μl were then used in a 100μl polymerase chain reaction, using 30 cycles of 94° for 1′, 65° for 2′ and 70° for 3′. The reaction products were electrophoresed in 1% low melting-point gels, the bands excised and diluted with H_2O to give a concentration of 2 to 5 ng/μl of amplified product. 10μl of this 'agarose stock' were used in a second round of 25 cycles of PCR, in which one primer was biotinylated at the 5′ end. The final PCR product was purified using 'Geneclean' and the two strands were separated using streptavidin-coated magnetic 'Dynabeads' (20). Sequencing reactions were carried out using Sequenase 2 (United States Biochemicals) on the DNA strand attached to the Dynabeads.

Analysis of genomic DNA

Sequencing of the cDNA of many mutants revealed that either whole or part of one or more exons was missing from the cDNA. Genomic DNA from these mutants was extracted using standard procedures, and PCR carried out on 0.5μg DNA using primers which resulted in the amplification of the appropriate exon(s) together with its flanking splice sites. With all except exons 1 and 3 one round of 25 cycles of PCR was sufficient to provide enough product for sequencing. With exons 1 and 3 two rounds were required. Biotinylated primers were used as described above. The PCR products were cut out from 1% low melting point gels, Genecleaned and sequenced as described above.

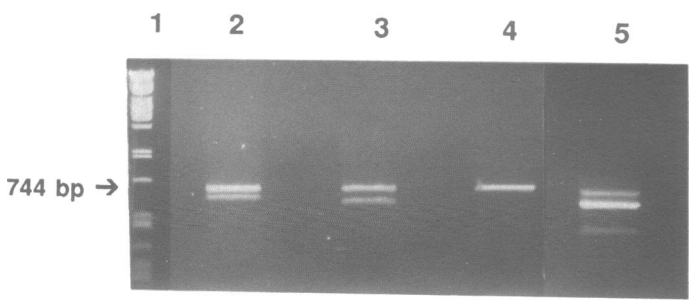

Figure 1. Agarose gel electrophoresis of PCR products. Lane 1, λ-Pst marker; lane 2, mutant 34M30 (see Table IV); lane 3, mutant 34M34 (Table IV); lane 4, mutant 34M7 (contains a point mutation); lane 5, mutant 34M14 (Table II).

Primers

The primers used for the PCR were as follows:

cDNA

1st round

H17:	CTG CGC CTC CGC CTC CTC CTC TG	(bases −80 to −58)
H16:	AGA ACT AGA ACA TTG ATA ATT TTA CTG GCG	(bases 735 to 706)

2nd round

H14:	GGC TTC CTC CTC CTG AGC AGT CAG C	(−47 to −23)
H13:	AGG ACT CCA GAT GTT TCC AAA CTC AAC TT	(697 to 669)

Genomic DNA (numbering system from reference 21)

Exon 1	ATGGGACGTCTGGTCCAAGGATTCA	(1224−1249)
	CCGAACCCGGGAAACTGGCCGCCC	(1851−1828)
Exon 2	TGCTGGGATTACACGTGTGAACC	(14575−14597)
	GACTCTGGCTAGAGTTCCTTCTTCC	(15149−15125)
Exon 3	TTTGCAGGCATGGGGTCTCACTATATT	(16480−16507)
	AATAAGTATATATCCTCCAAGGTGACTAG	(16982−16954)
Exon 4	CACTGTGATTGAAGATGGGTGGCTGT	(27550−27575)
	CCACAGAGGCAGACAGCAGTACTTG	(28191−28167)
Exon 5	GAGTCAGGAGACTATAAGAGACCAACT	(31259−31286)
	ATCTCACTCCTTTAGAACACAAGCCCA	(31769−31743)
Exon 6	TCCTGCACCTACAAAATCCAGTCCTG	(34727−34752)
	CCTCTGCCATGCTATTCAGGACAAAC	(35283−35258)
Exon 7	GTGCCTTGTCTGTAGTGTCAACTCA	(39601−39625)
and 8	TTATGTGACTAATGGGAACCATCAGT	(40235−40210)
Exon 9	ATTAAACTAATGTGATAGACTACTGCTTTG	(41256−41285)
	TTCAATGTTTCACTCAATAGTGCTGTGG	(41786−41759)

RESULTS

Mutants in the *hprt* gene were selected in 6-thioguanine, either following UV-irradiation of repair deficient XP fibroblasts (6) or after stimulation of circulating lymphocytes taken directly from the peripheral blood of a variety of donors (17). RNA was extracted from mutant cells and reverse transcribed. Two rounds of PCR were carried out on the resulting cDNA, using primers which flanked the coding region of the *hprt* gene (see Materials and Methods). The primers used for the first round amplified a DNA fragment of 815 bp with cDNA from wild-type cells e.g. Figure 1, lane 4. With many of the mutant cell lines the PCR product was either smaller or, less frequently, larger than the normal product. In other cases two or even three bands were obtained (eg Figure 1, lanes 2,3,5). Sequencing of the PCR products revealed that the alterations were the result of splicing out of exons or parts of exons, or of inclusion of parts of introns in the mRNA. After identification of the aberrantly spliced exons

Table I. Splice donor sites in the human *hprt* gene

Consensus sequence	$^{G}_{A}$AG↓	GU$^{G}_{A}$AGU	Scores[c]
Intron 1	GtG	GUGAG**c**	84.9
Intron 2	**c**AG	GUAAGU	100
Intron 3	**tgt**	GUGAGU	76.1
Intron 4	AAG	GUA**u**GU	88.7
Intron 5	GA**a**	GUAAGU	87.6
Intron 6	AAG	GUA**u**GU	88.0
Intron 7	**Act**	GUAAGU	79.0
Intron 8	AA**t**	GUAAGU	87.2
Cryptic 1[a]	**c**AG	GUG**gcg**	67.1
Cryptic 5[a]	AAG	GUAAG**c**	94.1
Mutated 3[b]	AAG	GUG**g**G**c**	80.1

Bases in lower case show deviations from the consensus sequence.
[a] Cryptic sites used when donor sites in introns 1 and 5 are mutated.
[b] Cryptic site generated by mutation in mutants 253M7 and 34M34 (see Table V).
[c] Scores using system of Shapiro and Senapthy (22).

by sequence analysis of the amplified cDNA, PCR was carried out with genomic DNA using primers which flanked these exons and the surrounding intron sequences. Sequencing of the amplified product from the genomic DNA revealed, in most cases, the nature of the mutation causing the aberrant splicing in the mRNA.

Table I shows the consensus sequence for splice donor sites in mammalian cells together with the splice donor sequences of the 8 introns of the human *hprt* gene (from reference 21). The donor sequences of introns 4,5,6 and 8 differ from the consensus sequence by 1 base, of introns 1 and 7 by 2 bases, and of intron 3 by 3 bases. Matching of splice donor sequences to the consensus can be quantified using a scoring system described by Shapiro and Senapathy (22) and the scores based on this analysis are shown in Table I. The donor site of intron 2 has the maximum score of 100 whereas in intron 3 it has a relatively low score of 76.1. The requirements of the splice acceptor sites are more complex, apart from the invariant AG, and they are discussed in more detail below.

Mutations in splice acceptor sites

The mutations that we have detected in splice acceptor sites are summarized in Table II. An obvious possible consequence of a mutation in the invariant AG of a splice acceptor site is that the acceptor site of the next intron downstream is utilised instead, resulting in splicing out of the intervening exon. Two mutations in the AG of intron 3 resulted in splicing out of exon 4 and a similar mutation in exon 6 resulted in loss of exon 7 in the mRNA (Table II, rows 2,3,6). With intron 5 the first base of the following exon is a G so that the G→A mutation of the terminal G of intron 5 in mutant 5 (Table II, row 4) resulted in the splice site being shifted downstream by 1 base, and a consequent alteration in the

Table II. Mutations in the splice acceptor sites of the *hprt* gene

	Exon	Mutant number	Position[a]	Mutation	Change in DNA
1	1	34M14	I1:−2	A→G	(a) Loss of exon 2 (b) Loss of exons 2,3 (c) Loss of 1st 5 bases of exon 2
2	3	8	I3:−1	G→A	Loss of exon 4
3	3	27	I3:−2	A→T	Loss of exon 4
4	5	5	I5:−1	G→A	Loss of 1st base of exon 6
5	5	30	I5:−1, E5:1	GG→AA	Loss of exon 6
6	6	253M89	I6:−2	A→T	Loss of exon 7
7	7	253M90	I7:−1	G→A	Loss of 1st 21 bases of exon 8
8	7	42M32	I7:−1	G→A	(a) Loss of 1st 21 bases of exon 8 (b) Loss of exon 8
9	7	32	I7:−1	G→A	Loss of exon 8
10	8	39	I8:−16	G→A	Insertion of 14 bases in front of exon 9

[a] Designations such as I6:−2 refer to the second base from the end of intron 6.

Table III. Mutations in the splice donor sites of the *hprt* gene

	Exon	Mutant number	Position[a]	Mutation	Change in cDNA
1	1	34	I1:5	G→A	Cryptic site I1:50
2	1	38	I1:5	G→T	Cryptic site I1:50
3	1	42M51	E1:−2 to I1:34	Deletion	Cryptic site I1:50
4	4	34M45	I4:1	G→A	Loss of exon 4
5	4	21	I4:1	G→A	Loss of exon 4
6	5	42M15	I5:2	T→C	Cryptic site I5:68
7	5	41	I5:3, 4	AA→GT	Cryptic site I5:68
8	5	11	I5:5	G→A	Cryptic site I5:68
9	6	253M2	I6:1	G→A	Loss of exon 6
10	6	42M9	E6:−1 (386)	G→A	Loss of exon 6
11	7	253M94	I7:1	G→A	Loss of exon 7
12	8	102M3	E8:−11 to I8:3	Deletion	Loss of exon 8

[a] Designations such as I6:1 refers to the 1st base of intron 6.

reading frame. This shift in the reading frame brings a TAA ochre codon in frame two aa downstream. In mutant 30 (Table II, row 5) both the terminal G of intron 5 and the adjacent G, the first base of exon 6, were tandemly mutated (GG→AA). This resulted in the splicing out of exon 6. In intron 7, mutation of the terminal G→A was found in three independent mutants. Two different PCR products were obtained as a result of this mutation, (1) in which the whole of exon 8 was spliced out from the mRNA or (2) in which the first 21 bases of exon 8 were lost. The latter occurs by utilising as cryptic acceptor site the AG residues at bases 557–558 in exon 8. Somewhat suprisingly the relative amounts of these two products differed between the three mutants, varying from all product (1) in mutants 32 (Table II, row 9), 50% of each in mutant 42M32 (Table II, row 8), and all of product (2) in mutant 253M90 (Table II, row 7). These differences may reflect genuine differences in splicing in two different cell types (mutant 32 is a fibroblast line, 42M32 and 253M90 are lymphocytes). Alternatively the PCR may preferentially amplify one of the forms of the mRNA molecules, thus distorting the actual percentages of spliced and unspliced molecules. Similarly, in intron 1, mutation of the A of the acceptor site to G in mutant 34M14 (Table II, row 1 and Figure 1, lane 5) resulted in a mixture of three PCR products resulting from: loss of exon 2 (60%), loss of exons 2 and 3 (20%) and loss of the first 5 bases of exon 2 (20%) by using the AG at positions 31 and 32 as cryptic acceptor sites.

The final mutation affecting the splice acceptor site involved a G→A transition in intron 8, 16 bases upstream from the normal splice acceptor site (Table II, row 10). This created a new acceptor site resulting in the inclusion of an extra 14 bases at the beginning of exon 9.

Mutations in splice donor sites (Table III)

Mutations affecting the splice donor sites resulted in either (a) a downstream cryptic splice donor site being utilised, with inclusion of the sequences between the mutated site and the cryptic site being spliced into the upstream exon, or (b) loss of the whole of the upstream exon. The former type was found with mutations in the donor sites of intron 1 or intron 5. Mutations altering the 5th base of intron 1 or a 36 base-pair deletion covering the splice donor site of intron 1 (Table III, rows 1–3) resulted in a cryptic splice site 49 base pairs downstream being used (see Table I). Likewise base substitutions at various positions in the donor site of intron 5 (Table III, rows 6–8) resulted in the use of a cryptic donor site 67 bases downstream (Table I). In contrast mutations in the donor sites of introns 4, 6, 7, or 8 resulted in the loss of the upstream exons (Table III, rows 4, 5, 9–12).

Mutations in exons (Table IV)

Mutations in splice donor and acceptor sites are anticipated to affect splicing of the mRNA molecules. Less expected however, was our finding of a series of point mutations in exon 3 and 8 which affected splicing of those exons.

(a) Exon 3. The sequence of the splice donor site of exon 3 has the lowest score using the system of Shapiro and Senapathy (22) (see Table I). It might therefore be predicted that this exon could be unstable. Mutants from two different donors (253M7 and 34M34) had the same mutation of the G at base 209 to T (Table IV, rows 3, 4). This G is the third in a run of 6 G's and the mutation creates a cryptic splice donor site (see Table I). In a proportion of the mRNA molecules this cryptic donor site was utilised with the consequent loss of 111 bases from exon 3 (Figure 1, lane 3). 20–40% of the PCR products resulted from this aberrant splicing event. In two other mutants there was a C→T transition at base 151, the 17th base of exon 3 and an A→T tranversion at base 163, the 29th base of exon 3 respectively (Table IV, rows 1, 2). These mutations resulted respectively in 50% and 10% of the PCR products being derived from mRNA molecules in which exons 2 and 3 were lost.

Table IV. Mutations in *hprt* exons resulting in aberrant splicing

	Exon	Mutant number	Position[a]	Mutation	Change in cDNA[b]
1	3	253M78	E3:17 (151)	C→T	(a) Loss of exon 2 and 3 (50) (b) Normal size (50)
2	3	253M92	E3:29 (163)	A→T	(a) Normal size (90) (b) Loss of exons 2 and 3 (10)
3	3	253M7	E3:75 (209)	G→T	(a) Normal size (70) (b) Loss of 111 bases from exon 3 (20) (c) Loss of exons 2 and 3 (10)
4	3	34M34	E3:75 (209)	G→T	(a) Normal size (60) (b) Loss of 111 bases from exon 3 (40)
5	8	15	E8:12 (544)	G→A	(a) Normal size (95) (b) Loss of exon 8 (5)
6	8	34M30	E8:19 (551)	C→T	(a) Normal size (70)
7	8	42M30			(b) Loss of exon 8 (30)
8	8	53	E8:−13 (597)	C→T	(a) Loss of exon 8 (90) (b) Normal size (10)

[a] Numbers in parenthesis denote position in cDNA.
[b] Numbers in parenthesis denote percentage of PCR product with indicated changes.

(b) Exon 8. We have detected mutations in the 12th, 19th and 65th bases of exon 8 (positions 544, 551, 597) (Table IV, rows 5–8). 5, 30 and 90% of the PCR products from the cDNA of these mutants had respectively lost exon 8. Of particular interest is mutant 53 with a C→T transition at base 597. This change of a TTC codon to TTT does not affect the amino acid sequence. The mutation resulted, however, in 90% of the PCR products being derived from mRNA molecules in which exon 8 was spliced out. This presumably caused the Hprt⁻ phenotype.

Undetected mutations resulting in aberrant splicing

In a few mutants aberrant splicing was detected by sequencing of amplified cDNA, but analysis of the genomic DNA in the donor and acceptor regions flanking the exons missing from the cDNA failed to reveal any mutations (see Table V). The causative mutations have not been identified in these mutants.

DISCUSSION

Analysis of mutations in the *hprt* gene, whether derived from Lesch-Nyhan patients (14, 15), from circulating human lymphocytes from a variety of donors (13; our unpublished observations) or from cultured human (4–6, 8, 9) or hamster cells (2, 10) has shown that a large proportion of the mutations result in aberrant splicing of the *hprt* mRNA. In only a few cases has the causative mutation been identified (10, 15).

Our analysis of 34 *hprt* mutants in which the *hprt* mRNA was aberrantly spliced has revealed causative mutations at splice acceptor sites (Table II), splice donor sites (Table III) or in exons some distance from the splice sites (Table IV). Two of the mutations were small deletions, the remainder being single (or in 1 case tandem) base-changes.

Mutations in splice acceptor sites can result either in use of the acceptor site of the next intron downstream, with consequent loss of the intervening exon, or in the utilisation of a cryptic splice acceptor site in the next exon downstream. The *in vitro* studies of Smith *et al.* (23) on splicing of the α-tropomyosin intron 1 suggested that the first AG downstream from the branch site was used as the splice acceptor site. AG sites are found within 30 bases downstream of the natural splice acceptor site in all *hprt* exons except exon 5 (see Table VI). We have found however that only the AG sites in exons 2, 8 and 9 appear to be capable of being used as cryptic acceptor sites if the upstream acceptor sites are mutated (see also reference 14). Exons 4, 6 and 7 are spliced out in their entirety if the upstream acceptor sites are mutated, even though in the case of exon 4 there is a potential cryptic site only 9 bases downstream. The reason for this can be seen if the splice acceptor sites are scored using the method suggested by Shapiro and Senapathy (22) (see Table VI). Using this system, the cryptic sites which are used score 72 or over, whereas those that are not used are in the 60–66 range. (We have not identified any acceptor site mutations for intron 2). The low scores for this latter group result principally from a lack of a polypyrimidine tract at bases −14 to −5 in the putative intron. Note that the scores for the naturally used splice sites are considerably higher than those for all the cryptic sites, the former being in the 81–96 range. Our results emphasize the importance of the polypyrimidine tract and also demonstrate the usefulness of the scoring system derived by Shapiro and Senapathy (22).

We have identified mutations in the splice donor sites of 6 out of the 8 introns of the *hprt* gene. In introns 1 and 5, downstream cryptic donor sites were used to replace the mutated sites, whereas in all other introns, the upstream exon was lost from the cDNA. Table I shows that the cryptic splice site in intron 5 has a very high score of 94.1—it closely matches the donor site consensus sequences. In intron 1 the cryptic splice donor site has a relatively low score of 67.1. As this is the first intron, however, splicing out of the upstream (first) exon is not feasible, so utilisation of the cryptic site, despite its low score, is presumably the only possible option. In all other introns, GT sequences within 300 bases of the natural site are in sequences with relatively low scores, between 45 and 68. These scores are presumably too low

Table V. Undetected mutations resulting in aberrant splicing

Mutant	Splice abnormality	Region in which mutation excluded[a]
42M39	Loss of exon 4	I3A; I4D
42M18	Loss of 1st 17 bases from exon 9	I8A
42M20	Loss of exon 8	I7A
325M42, others	Lossof exon 2 or exons 2 and 3	I1A, I2D, I2A, I3D, I3A, I4D

[a] A, acceptor site; D, donor site.

Table VI. Potential splice acceptor sites in *hprt* exons

Exon	Sequence[a]	distance from upstream splice site (bases)	Score for cryptic site	Score for natural site
2	ATTAGT	5	**76.3**	95.8
3	ACTGAACGTCTTGCTCGAGA	19	74.5	86.3
4	AATGACCAGT	9	60.0	86.5
5	–	87.8		
6	GATATAATTGACACTGGCAAAACAATGCAGA	30	64.0	87.8
7	CTTGCTGGTGAAAAGG	15	66.1	81.0
8	TTGTTGGATTTGAAATTCCAGA	21	**79**	82.5
9	CATGTTTGTGTCATTAGT	17	**72.1**	85.9

[a] The sequences shown are at the start of the indicated exons.

for these sequences to be selected as cryptic donor sites. Our results therefore support the model of the exon as the unit of assembly proposed by Robberson *et al.* (24) and Talerico and Berget (25). By analysing *in vitro* the splicing of exons 1 and 2 of the adenovious late transcription unit with various manipulations of exon size and splice site mutations, they concluded that the spliceosome "assembly machinery searches precursor RNA sequences hunting for 5′ splice sites lying a short distance downstream of 3′ intron sequences. In their presence, an exon is defined and stable assembly intermediates accumulate; in their absence no exon is defined and stable assembly intermediates are not created" (24). They suggested that this is the reason for the maximum size of 300 bases for internal exons in vertebrates, and showed that increasing the exon size beyond this limit reduced the efficiency of splicing of the upstream intron, and resulted instead in the exon being skipped. This also explains why potential cryptic splice donor sites are only utilised if they are close to the natural (mutated) site, never far downstream within the intron. In the absence of a suitable cryptic splice site close to the natural donor site, an exon of less than 300 nucleotides cannot be found, and the whole of the upstream exon is therefore spliced out. Our results (Table III) are consistent with this model. Talerico and Berget (25) list 18 splice donor site mutations reported in the literature, of which 14 resulted in skipping of the upstream exon (15, 25–36), 4 resulted in the utilisation of cryptic downstream donor sites (37–39) and in two cases both options were used (40, 41). More recently donor site mutations resulting in exon skipping have been reported for intron 12 of the retinoblastoma gene (42), and intron 7 of the human ornithine transcarbamylase gene (43).

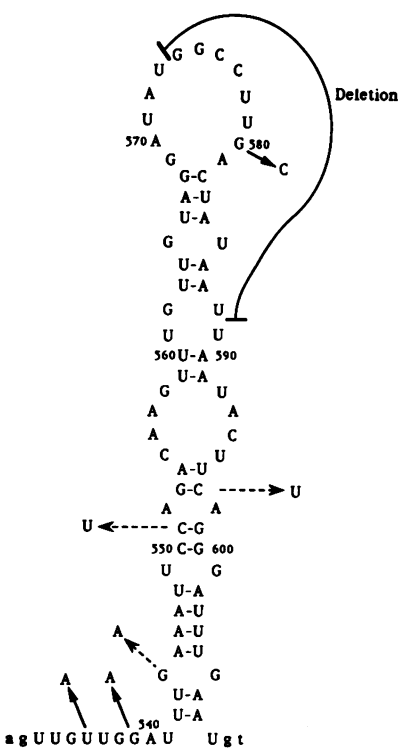

Figure 2. Exon 8 showing hypothetical stem-loop structure. Solid arrows indicate base changes which do not result in aberrant splicing, dashed arrows do result in aberrant splicing. Lower case letters denote the end and beginning of flanking introns.

Several mutants with base changes in exons 3 or 8 resulted in aberrant splicing of these exons in at least a proportion of the mRNA molecules (Table IV). The splice donor site of intron 3 diverges from the consensus sequence more than any of the other introns (score of 76.1), and in two mutants an exon 3 mutation resulted in the formation of a sequence which could be used as a cryptic donor site, with a relatively high score of 80.1. This new cryptic donor site was used in a fraction of the mRNA molecules in these mutants. None of the other changes had any obviously predictable effect on splicing.

An explanation for the effects of mutations in exon 8 may be found in terms of secondary structure. It is possible to draw a stem-loop structure for exon 8 which includes a region in which 12 out of 15 bases are paired (Figure 2), and which ends immediately upstream of the donor site of intron 8. We have identified a total of 7 mutations in exon 8. Four of these have no effect on splicing. These include 2 base-change mutations outside the stem-loop, and a base-change and 17bp deletion from bases 574–590 at the top part of the loop (see Figure 2). The three mutations which do affect splicing of exon 8 to varying degrees (see Table IV) all lie within the region in which base-pairing is strongest (Figure 2, indicated with dashed arrows). If we assume that the relative abundance of PCR products is an accurate reflection of the abundance of the corresponding mRNA species, alteration of the non base-paired G residue at 544 results in only 10% aberrant splicing, whereas mutations at paired C residues at 551 and 597 lead to much more aberrant splicing. Maintenance of this stem-loop may be necessary for exposure of the splice sites, and its disruption may provide the explanation for the splicing abnormalities resulting from mutations in exon 8. Our finding of the mutation at 597 which does not affect the coding properties of the gene, but in fact results in splicing out of exon 8 in most of the mRNA molecules, is as far as we are aware unprecedented in the literature.

Several studies have shown that splicing can be affected by manipulation of exons. Somasekhar and Mertz (44) showed that deletions or a 10bp insertion in the late region mRNA of SV40 could alter the pattern of splicing of this mRNA, and Reed and Maniatis (45) showed that large deletions or substitutions of exon sequences could inactivate adjacent splice sites in *in vitro* studies on artificial constructs of parts of the human β-globin gene. Using a minigene construct Mardon *et al.* (46) showed that an 81bp deletion in Exon EDIIIA of the human fibronectin gene resulted in its being spliced out of the mRNA in HeLa cells. Similarly Streuli and Saito (47) studied the mechanism of alternative splicing of exon 4 of the human leucocyte common antigen gene, using a minigene construct. This exon is spliced in in B cells, but out in thymocytes. A 95 base deletion in exon 4 resulted in exon 4 being spliced in in both tissues. Replacement of various sections of the exon with a 10bp Cla linker delineated regions of exon 4 which were essential to retain alternative splicing. Similar results were obtained for exon 6 of the same gene (48). In a similar type of study Libri *et al.* (49) studied alternative splicing of exon 6 in the chicken β-tropomyosin gene. Alterations of 15 base stretches in one of the alternatively spliced exons (6B) resulted in dramatic changes in the relative utilisation of the two alternative exons 6A and 6B. These changes were interpreted in terms of putative secondary structures. Likewise deletion of exon sequences in the early region 3 of adenovirus 2 resulted in the activation of a cryptic splice site in the downstream intron

(50). Evidence was presented to show that these exon sequences formed a large stem-loop structure immediately upstream from the normal splice donor site. Deletion of the exon sequences destroyed the stem-loop and thereby enhanced the use of a cryptic site. This is similar to our suggestions for the effects of mutations in the exon 8 of the *hprt* gene. Cooper and Ordahl (51) introduced a series of 4-base mutations into a minigene containing exon 5 of the cardiac troponin gene, and found that three different four base mutations resulted in exclusion of the exon from processed mRNA.

We are aware however of only three reports of a single base change in an exon affecting splicing. Ricketts et al. identified a mutation in the bovine thyroglobulin gene which resulted in cattle goitre (52). This mutation was a C→T transition at position 697 in exon 9, 28 bases away from the end of the exon, and resulted in loss of exon 9 from the mRNA. Ligtenberg *et al.* examined the episialin (MUCI) gene in which two variants of exon 2 were found (53). A variant with a G in a particular site in exon 2 resulted in the exon commencing 35 bases upstream, whereas in the variant with A at this site the exon began only 8 bases upstream. The authors proposed a stem loop structure for the exon to account for the choice of splice site. Very recently Zhang and Jenssen isolated *hprt* mutants induced in V79 hamster cells by N-methyl-N-nitrosourea (10). Four mutants contained point mutations in exon 4, about 10 bases from the beginning of the exon, which resulted in splicing out, either of exon 4, or of exons 2–4 from the mRNA.

Taken together, our own and these studies demonstrate that exon sequences contain *cis*-acting elements which affect mRNA splicing. In some cases dramatic changes can be brought about by single-base changes. Whether exon mutations disrupt splicing by altering secondary structures, as suggested in several studies including our own, or by directly affecting interactions with *trans*-acting factors has not yet been established. In some reports disruption of putative secondary structures has been proposed (49, 50, 53 and see above), whereas in other reports, no secondary structures were apparent (47, 51).

ACKNOWLEDGEMENTS

We are grateful to David Beare and Kay Aldridge for growing up the mutant clones, and to Dr H.Vrieling, State University of Leiden for invaluable help and advice. This work was supported in part by EC contracts EV4V.0037UK(H) and EV4V.0186UK(H). During the course of this work G Dorado was supported by an EMBO fellowship.

REFERENCES

1. Vrieling, H., Simons, J. W. I. M., and Van Zeeland, A. A. (1988b) *Mutation Res. 198,* 107–113.
2. Vrieling, H., Van Rooijen, M. L., Groen, N. A., Zdzienicka, M. Z., Simons, J. W. I. M., Lohman, P. H. M., and Van Zeeland, A. A. (1989) *Mol. Cell. Biol. 9,* 1277–1283.
3. Vrieling, H., Venema, J., Van Rooyen, M-L., Van Hoffen, A., Menichini, P., Zdzienicka, M. Z., Simmons, J. W. I. M., Mullenders, L. H. F., and Van Zeeland, A. A. (1991) *Nucleic Acids Res. 19,* 2411–2415.
4. Chen, R-H., Maher, V. M., and McCormick, J. J. (1990) *Proc. Natl. Acad. Sci. USA 87,* 8680–8684.
5. Chen, R-H., Maher, V. M., and McCormick, J. (1991) *Cancer Res. 51,* 2587–2592.
6. Dorado, G., Steingrimsdottir, H., Arlett, C. F., and Lehmann, A. R. (1991) *J. Mol. Biol. 217,* 217–222.
7. Bronstein, S. M., Cochrane, J. E., Craft, T. R., Swenberg, J. A., and Skopek, T. R. (1991) *Cancer Res. 51,* 5188–5197.
8. McGregor, W. G., Chen, R-H., Lukash, L., Maher, V. M., and McCormick, J. J. (1991) *Mol. Cell. Biol. 11,* 1927–1934.
9. Yang, J-L., Chen, R-H., Maher, V. M., and McCormick, J. J. (1991) *Carcinogenesis 12,* 71–75.
10. Zhang, L-H., and Jenssen, D. (1991) *Carcinogenesis 12,* 1903–1909.
11. Keohavong, P., Liu, V. F., and Thilly, W. G. (1991) *Mutation Res. 249,* 147–159.
12. Recio, L., Cochrane, J., Simpson, D., Skopek, T. R., O'Neill, J. P., Nicklas, J. A., and Albertini, R. J. (1990) *Mutagenesis 5,* 505–510.
13. Rossi, A. M., Thijssen, J. C. P., Tates, A. D., Vrieling, H., Natarajan, A. T., Lohman, P. H. M., and Van Zeeland, A. A. (1990) *Mutation Res. 244,* 353–357.
14. Gibbs, R. A., Nguyen, P. -N., McBride, L. J., Koepf, S. M., and Caskey, C. T. (1989) *Proc. Natl. Acad. Sci. USA 86,* 1919–1923.
15. Gibbs, R. A., Nguyen, P-N., Edwards, A., Civetello, A. B., and Caskey, C. T. (1990) *Genomics 7,* 235–244.
16. Smith, C. W. J., Patton, J. G., and Nadal-Ginard, B. (1989) *Annu. Rev. Genet. 23,* 527–577.
17. Cole, J., Green, M. H. L., James, S. E., Henderson, L., and Cole, H. (1988) *Mutation Res. 204,* 493–507.
18. Yang, T. P., Patel, P. I., Chinault, A. C., Stout, J. T., Jackson, L. G., Hildebrand, B. M., and Caskey, C. T. (1984) *Nature (London) 310,* 412–414.
19. Yang, J-L., Maher, V. M., and McCormick, J. J. (1989) *Gene. 83,* 347–354.
20. Hultman, T., Stahl, S., Hornes, E., and Uhlen, M. (1989) *Nucleic Acids Res. 17,* 4937–4946.
21. Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C. T., and Ansorge, W. (1990) *Genomics 6,* 593–608.
22. Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res 15,* 7155–7174.
23. Smith, C. W. J., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. (1989) *Nature 342,* 243–247.
24. Robberson, B. L., Cote, G. J., and Berget, S. M. (1990) *Mol. Cell. Biol. 10,* 84–94.
25. Talerico, M., and Berget, S. M. (1990) *Mol. Cell. Biol. 10,* 6299–6305.
26. Hidada, Y., Palella, T. D., O'Toole, T. E., Tarle, S. A., and Kelly, W. N. (1987) *J. Clin. Invest. 80,* 1409–1415.
27. Mitchell, P. J., Urlaub, G., and Chasin, L. (1986) *Mol. and Cell. Biol. 6,* 1926–1935.
28. Ohno, K., and Suzuki, K. (1988) *J. Biol. Chem. 263,* 18563–18567.
29. Kudo, S., and Fukudu, M. (1989) *Proc. Natl. Acad. Sci. USA 86,* 4619–4623.
30. Brandt, C. R., Morrison, S. L., Birshtein, B. K., and Milcarek, C. (1984) *Mol. Cell. Biol. 4,* 1270–1277.
31. Bakhshi, A., Guglielmi, P., Siebenlist, U., Ravetch, J. V., Jensen, J. P., and Korsmeyer, S. J. (1986) *Proc. Natl. Acad. Sci. USA 83,* 2689–2693.
32. Seidman, J. G., and Leder, P. (1980) *Nature 286,* 779–783.
33. Marvitt, J., DiLella, A. G., Brayton, K., Ledley, F. D., Robson, K. J. H., and Woo, S. L. C. (1987) *Nucleic Acids Res. 15,* 5613–5628.
34. Weil, D., Bernard, M., Combates, N., Wirtz, M. K., Hollister, D. W., Steinmann, B., and Ramirez, F. (1988) *J. Biol. Chem. 263,* 8561–8564.
35. Weil. D., D'Alessio, M., Ramirez, F., DeWet, W., Cole, W. G., Chan, D., and Bateman, J. F. (1989) *EMBO J. 8,* 1705–1710.
36. Weil, D., D'Alessio, M., Ramirez, F., Steinmann B., Wirtz, M. K., Glanville, R. W., and Hollister, D. H. (1989) *J. Biol. Chem. 264,* 16804–16809.
37. Wieringa, B., Meyer, F., Reiser, J., and Weissmann, C. (1983) *Nature 301,* 38–43.
38. Chen, E. Y., Liao, Y. C., Smith, D. H., Barrera-Saldana, H. A., Gelinas, R. E., and Seeburg, P. H. (1989) *Genomics 4,* 479–497.
39. Sikder, S. K., Kabat, E. A., and Morrison, S. L. (1985) *Proc. Natl. Acad. Sci. USA 82,* 4045–4049.
40. Treisman, R., Orkin, S. H., and Maniatis, T. (1983) *Nature 302,* 591–596.
41. Hodges, P. E., and Rosenberg, L. E. (1989) *Proc. Natl. Acad. Sci. USA 86,* 4142–4146.
42. Dunn, J. M., Phillips, R. A., Zhu, X., Becker, A., and Gallie, B. L. (1989) *Mol. Cell. Biol. 9,* 4596–4604.
43. Carstens, R. P., Fenton, W. A., and Rosenberg, L. R. (1991) *Am. J. Hum. Genet. 48,* 1105–1114.
44. Somasekhar, M. B., and Mertz, J. E. (1985) *Nuc. Acids Res. 13,* 5591–5609.
45. Reed, R., and Maniatis, T. (1986) *Cell 46,* 681–690.
46. Mardon, H. J., Sebastio, G., and Baralle, F. E. (1987) *Nucleic Acids Res. 15,* 7725–7733.
47. Streuli, M., and Saito, H. (1989) *EMBO J. 8,* 787–796.

48. Tsai, A. Y. M., Streuli, M., and Saito, H. (1989) *Mol. Cell. Biol. 9,* 4550–4555.
49. Libri, D., Goux-Pelletan, M., Brody, E., and Fiszman, M. Y. (1990) *Mol. Cell. Biol. 10,* 5036–5046.
50. Domenjoud, L., Gallinaro, H., Kister, L., Meyer, S., and Jacob, M. (1991) *Mol. Cell. Biol. 11,* 4581–4590.
51. Cooper, T. A., and Ordahl, C. P. (1989) *Nucleic Acids Res. 17,* 7905–7921.
52. Ricketts, M. H., Simons, M. J., Parma, J., Mercken, L., Dong, Q., and Vassart, G. (1987) *Proc. Natl. Acad. Sci. USA 84,* 3181–3184.
53. Ligtenberg, M. J. L., Gennisen, A. M. C., Vos, H. L., and Hilkens, J. (1991) *Nucleic Acids Res. 19,* 297–302.