ELSEVIER

Mutation Research 411 (1998) 179–214

Reviews in Mutation Research

Mutations that alter RNA splicing of the human *HPRT* gene: a review of the spectrum

J. Patrick O'Neill [a,*], Peter K. Rogan [b], Neal Cariello [c], Janice A. Nicklas [a]

[a] *University of Vermont Genetics Laboratory, 32 North Prospect Street, Burlington, VT 05401, USA*
[b] *Department of Human Genetics, Allegheny University of the Health Sciences, 320 E. North Avenue, Pittsburgh, PA, 15212, USA*
[c] *Glaxo, 5 Moore Drive, Research Triangle Park, NC, 27709, USA*

Received 25 October 1997; revised 18 June 1998; accepted 19 June 1998

Abstract

The human *HPRT* gene contains spans approximately 42,000 base pairs in genomic DNA, has a mRNA of approximately 900 bases and a protein coding sequence of 657 bases (initiation codon AUG to termination codon UAA). This coding sequence is distributed into 9 exons ranging from 18 (exon 5) to 184 (exon 3) base pairs. Intron sizes range from 170 (intron 7) to 13,075 (intron 1) base pairs. In a database of human *HPRT* mutations, 277 of 2224 (12.5%) mutations result in alterations in splicing of the mRNA as analyzed by both reverse transcriptase mediated production of a cDNA followed by PCR amplification and cDNA sequencing and by genomic DNA PCR amplification and sequencing. Mutations have been found in all eight 5′ (donor) and 3′ (acceptor) splice sequences. Mutations in the 5′ splice sequences of introns 1 and 5 result in intron inclusion in the cDNA due to the use of cryptic donor splice sequences within the introns; mutations in the other six 5′ sites result in simple exon exclusion. Mutations in the 3′ splice sequences of introns 1, 3, 7 and 8 result in partial exon exclusion due to the use of cryptic acceptor splice sequences within the exons; mutations in the other four 3′ sites result in simple exon exclusion. A base substitution in exon 3 (209G → T) creates a new 5′ (donor) splice site which results in the exclusion of 110 bases of exon 3 from the cDNA. Two base substitutions in intron 8 (IVS8-16G → A and IVS8-3T → G) result in the inclusion of intron 8 sequences in the cDNA due to the creation of new 3′ (acceptor) splice sites. Base substitution within exons 1, 3, 4, 6 and 8 also result in splice alterations in cDNA. Those in exons 1 and 6 are at the 3′ end of the exon and may directly affect splicing. Those within exons 3 and 4 may be the result of the creation of nonsense codons, while those in exon 8 cannot be explained by this mechanism. Lastly, many mutations that affect splicing of the *HPRT* mRNA have pleiotropic effects in that multiple cDNA products are found.

Keywords: *HPRT*; *HPRT* mutation; Human mutation; Splice site mutation; Exon exlusion or exon skipping

* Corresponding author. Tel.: +1-802-656-8332; Fax: +1-802-656-8333; E-mail: patrick.oneill@uvm.edu

1383-5742/98/$ - see front matter
PII: S1383-5742(98)00013-1

1. Introduction

1.1. Splicing in general

Most genes are divided into coding regions (exons) and intervening non-coding regions (introns). Transcription of genomic DNA creates large pre-RNAs from which the intervening non-coding regions must be precisely removed (spliced) to create functional mRNAs. Splicing is performed by what has been called the spliceosome of which 5 small nuclear RNA (snRNAs) [U1, U2, U4, U5 and U6] are the major components (reviewed in Refs. [1–5]). A number of proteins are also important in this process including the serine-rich proteins (SR proteins) (reviewed in Ref. [4]) and proteins that bind heterogenous nuclear RNA (hnRNP proteins) (reviewed in Ref. [6]). SR proteins contain an amino-terminal RNA recognition motif which is widely present in RNA binding proteins and an arginine-serine dipeptide repeat region (RS domain) at their C terminus which is considered diagnostic of splicing proteins. Two examples are SF2/ASF and SC35. While these proteins were originally thought to be non-specific as they could be switched between organisms, more recent work has shown specific binding and regulation of different genes by different SR proteins [7]. The SR proteins bind to exon sequences, especially those 25 bases upstream of the 3′ end of the exon [5]. The hnRNP are found only in organisms with large introns and contain an RNP domain and protein–protein recognition domains [6]. They appear to recognize intronic sequences and could be involved in exon juxtaposition [8].

The snRNAs and proteins recognize consensus sequences on the DNA at or near the splice junctions. These consensus sequences will be discussed in detail below but the most important elements are: (1) a gt as the first 2 bases of the intron (the 5′ splice or splice donor site), (2) an ag as the last two bases of the intron (the 3′ splice or splice acceptor site), (3) a pyrimidine-rich tract usually just upstream of the acceptor site and (4) an 'a' upstream of the pyrimidine-rich tract (branch site).

Briefly, splicing proceeds by ordered binding of 4 complexes (E, A, B, C) at the splice sites. Initial binding of U1 occurs at 5′ splice sites (E complex) while U2 binds to the branch point region (A complex) by complementary base pairing with the help of U2 auxiliary factor (U2AF) bound to the pyrimidine rich sequence. U1 is replaced by U4–U5–U6 with complementary base-pairing of U6 with the 5′ splice site. Complementary base-pairing of U2 and U6 bridges across the intron (B complex) with release of U4 [9]. Splicing commences by U2 facilitating a bulging out of the branch A and the 2′ hydroxyl of A's 'attack' of the phosphodiester bond between the last base of the proceeding exon and the first base of the intron (C complex). This causes a lariat to be formed with the branch A attached to the first base of the intron. The 'free' 3′ OH of the 5′ exon, held by U5 [10], then 'attacks' the last base of the intron, attaching the 2 exons together and releasing the intron and associated snRNPs and proteins as a lariat (I complex).

At least one way that the SR proteins appear to facilitate these steps is by strengthening binding of the complexes to the RNA, especially by recruiting U1. They are especially crucial when the 5′ and 3′ site sequences are less than optimal. In fact, U1 is apparently dispensable if sufficient SR proteins are present [11,12]. Heterogenous ribonucleoprotein particles (hnRNPs) appear to compete with SR binding and are important in regulating alternative splicing (see below). For example, adjusting relative levels of SF2/ASF and hnRNP A1 promotes exon inclusion or exclusion, respectively [13,14]. SR proteins can be phosphorylated at the serines in the RS domains and this may play an important part in their function, regulation or specificity [4]. Polypyrimidine tract binding protein (PTB) binds in vitro to sequences that are very similar to the polypyrimidine tract in the alternatively spliced exons of rat α and β tropomyosin and can apparently suppress U2AF binding to pyrimidine sequences [15].

When introns were first discovered, discussion focused on the intron as a recognized unit that is spliced out; however, there are a number of difficulties with this model. Introns can be hundreds to thousands of basepairs long, and to scan down the DNA from one end to find the next is not a feasible plan. Also, this model would predict that mutations of consensus splice sites would cause inclusion of introns in the final mutant mRNA because they

cannot be spliced out. This is rarely observed. Instead, the most common result of a splice site mutation is exclusion of the adjacent exon from the mRNA (i.e., the upstream exon for 5′ or donor site mutations and the downstream exon for 3′ or acceptor site mutations). In mammalian genes, splice mutations occur at these frequencies: exon skipping—51%, splicing at a nearby spot—32%, creation of a pseudo exon in an intron—11% and intron retention—only 6% [16]. Also, in in vitro studies, mutations in the donor site affect splicing of both the previous intron and the intron in which they occur [17]. Lastly, surveys of exons have demonstrated that exons have a clear maximum size of about 300 bp; very few exons are larger than this (95% are less than 300 bp) [8,18]. In vitro studies of constructs with increasing exon lengths show that splicing either is diminished or cryptic sites within the exons are utilized when the exon reaches 300 bp [8,19]. These studies led investigators to develop the exon definition model of splicing whereby binding of splicing factors at either end of the exon defines the exons; these exons are then spliced together to form the mRNA (exon juxtaposition) [8,17,19]. If either one of the splice sites at the ends of the exon do not function, then the exon is not recognized—it is just considered part of the new large intron made by joining what were the two smaller introns on either side of that exon.

The exon definition model is supported by evidence showing that SR proteins bind to exon sequences, apparently bridging the acceptor and donor binding snRNPs across the exon [20]. In addition, in vitro experiments show that deletion of certain exon sequences abrogates splicing in of the exon [21]. These sites are called exon recognition sequences (ERS) or exonic splicing enhancers or exon splicing elements (ESE) [22–24,26]. Splicing enhancers are genomic sequences in exons or nearby in introns that facilitate splicing (Table 1). These sequences are generally purine rich.

Poison sequences (the opposite of enhancers) also exist [28,29]. Del Gatto et al. [28] found that the exonic sequence TAGG in the alternative K-SAM exon of fibroblast-growth factor receptor-2 inhibits splicing of that exon; this sequence inhibits splicing in other genes as well, possibly because it is the consensus 3′ splice sequence (as discussed below).

Table 1
Splicing enhancers

Gene	Exon	Site	Sequence	Comments	Reference
Mouse IgM	M2	5′ region of exon	GGAAGGACAGCA	works in other genes	[22]
Cardiac troponin	exon 5(alt exon)	in exon	AAAGAGGA	exon 5 is only 30 bp, works in other genes	[23]
Fibronectin	ED1(alt exon)	center, 81 bp	GAAGAAGAC	works in other genes	[24]
Bovine growth hormone	last exon	in exon	GGAAGGA	binds SF2/ASF	[14,25]
Rat fibronectin	EIIIB(alt exon)	downstream intron	TGCATG—highly repeated	works in other genes, found in other genes	[26]
Chicken cardiac troponin	exon 17	134 bases in downstream exon	purine rich, GGGGCTG	exon is only 6 bases, five of which are purines	[27]

Furdon and Kole [29] found that a 5′ region of the rabbit globin intron 2 inhibited splicing when downstream of the 3′ splice site.

Some genes undergo different (alternative) splicing patterns in different tissues or during development. The classic example is of *tra* and *dsx* in Drosophila, which are spliced differently in the two sexes, thereby controlling sexual development (reviewed in Ref. [30]). For *tra*, choice of an upstream splice site leads to inclusion of an exon with a stop codon in the mRNA causing a non-functional protein and male development (default splice). Use of a more 3′ splice site excludes this exon and the resulting protein controls female development. The gene Sex-lethal (*Sxl*), which is active only in females, codes for a snRNP that binds to the U-rich pyrimidine sequence at the upstream splice sequence and suppresses splicing. *Sxl* is also alternatively spliced in males and females through autoregulation. *Tra* and *tra-2* regulate the doublesex gene (*dsx*) by binding to 13 base repeats in the fourth exon and causing the female alternative splicing pathway. A purine rich sequence can substitute for this 13 base repeat. These examples show both positive and negative regulation of splicing.

Consensus sequences for the 5′, 3′ (includes pyrimidine tract) and branch sequences have been determined and a number of groups have developed methods to score and/or search for sites within genes or DNA sequences [31–39]. In one model, the 5′ consensus DNA sequence is: AGgt(a/g)agt and the 3′ consensus DNA sequence is: 10(t/c)n(t/c) agG. Different groups suggest different lengths for the acceptor associated pyrimidine tract [31,33,35]. The most important feature appears to be the number of pyrimidines in the stretch; consecutive T's strengthen the acceptor while purines weaken it [40,41]. Chu et al. [42] describe a polymorphism in the pyrimidine tract of exon 9 of the cystic fibrosis gene which affected the level of exon 9-transcript. In order to rank potential splice sites and determine possible cryptic sites, Senapathy et al. [31] used tables of base frequencies at each position to develop ad hoc measures of the respective 'strength' of 3′ and 5′ sites (a Senapathy score).

Stephens and Schneider [35] have discussed the features of splice sites in human genes and reviewed the models offered for this process. Based on the sequence of 1799 donor sites and 1744 acceptor sites, they have developed an information theory-based model of splicing in which information curves define the relative constraints on the specific base in that sequence [34,37–39]. This study argues that previous models for splice sites do not comprehensively define functional splice sites and does not include many functional sites with different sequences. The information in bits for a splice site (R_i) is defined as the dot product of a weight matrix derived from the nucleotide frequencies at each position of the splice site from the database and the vector of a particular splice junction sequence. According to information theory and the second law of thermodynamics, sites with $R_i < 0$ cannot be recognized or bound by the splicing proteins. In practice, sites with $R_i < 2.36$ bits are not utilized. The sequence logo graphically illustrates the average information content (in bits) flanking each site of the exon.

Using this model to study 1799 donor and 1744 acceptor sites, Stephens and Schneider [35] found donor sites had an average value of 8.4 bits and showed a significant sequence specificity in the last 3 bases in the exon and the first 6 bases of the intron and the acceptor splice sequence had an average value of 9.3 bits and included the last 25 bases of the intron and the first 2 bases of the exon. Since the information content of the acceptor excedes that of the donor; this supports the exon definition model of the spliceosome, in which the acceptor site is recognized first prior to finding an appropriate donor site downstream. Once 2 exons have been located, the intervening intron is removed. (This model does not explain how the first exon's donor site is recognized or how the last exon's acceptor site is spliced.) The donor and acceptor sequence logos contain the common sequence CAG/GT which may have been the sequence of a proto-splice junction that evolved early. The idea of a single splice junction recognition sequence is attractive because it would maximize the information required for recognition on the intronic component of both splice sites allowing largely unrestricted codon choices throughout exons.

Further analysis of 130 mutations in 42 genes showed significant losses of information (-4.0 bits for acceptors and -6.7 bits for donors) with more severe mutations showing more significant decreases

[38]. By contrast, polymorphic variants do not significantly change the R_i value of a site. The consensus donor and acceptor sequences have the maximum possible R_i values of 13.1 and 21.4 bits, respectively while actual donor and acceptor sequences have average R_i values of 7.1 and 9.4 bits, respectively. Cryptic sites were also detected and generally have R_i values comparable to the natural splice sequences.

Whereas a large number of donor (5′) and acceptor (3′) splice sites have been sequenced and collated, the branch site sequence containing the branch whose 'attack' forms the lariat has only rarely been experimentally determined. The branch consensus sites given in the literature are: ctray or tnctrac (where a is the branch A described above) [32,43]. Nelson and Green [43] list 31 branch sites; however, a number of these are from in vitro constructs, lower organisms or viruses and 8 of the remaining 14 are from one family of genes (globin). Table 2 lists 25 vertebrate branch sites found in the literature. At the 20 constitutive sites, the branch A is located 18 to 38 bases from the acceptor.

A special feature of several alternatively spliced exons is that the branch site is located far away from the acceptor (up to 200 bases) (Table 2). These branch sites have a very long pyrimidine tract located just downstream [43–49]. While studies of acceptor sites have generally assumed that the pyrimidine tract is associated with the 3′ acceptor, these cases suggest instead that the pyrimidine tract is in reality associated with the branch point and that a very long pyrimidine tract can allow usage of a distant branch site. Reed [41] also performed in vitro experiments showing that a branch point 116 bases away but with a long pyrimidine stretch was used over a closer branch point that had only a purine stretch.

The importance of the different sequences (5′, 3′, branch and pyrimidine tract) is debated. Clearly, weaknesses in one can be made up for by strengths in the other [40,50]. One hypothesis is that the 3′ site is simply the first ag downstream of the branch point although if a stronger ag is close (<30 bases further down) then that will be chosen (sites are rated

Table 2
Branch sites

Gene	Intron	Sequence	Distance to AG	Reference
Rat insulin	1	CCTCAAC	18	[21]
Human α globin	2	CACTGAC	18	[44]
Human α globin	1	CCCTCAC	19	[44]
H. chorionic somatomammotropin		CCTCCAT	22	[43]
Human calcitonin/CGRP 1	3	TACTGTC *	23	[43]
rat β tropomyosin	7	CACTGAC	24	[45]
rat β tropomyosin	5	GATATAT or CTCTGAT	25 or 29	[45]
Human growth hormone (pituitary)	2	CGTAGAC or AAGGAAC	26 or 22	[46]
Human growth hormone (placental)	2	CATAGAC	26	[46]
Human growth hormone	1	GGCTCCC *	27	[43]
Human γ globin	1	TTCTGAC	30	[44]
Human ϵ globin	1	CTCTAAT	30	[44]
Rabbit β globin	1	TGCTGAC	34	[44]
Mouse β globin	1	CACTAAC	35	[44]
Human calcitonin/CGRP 1	4	CACTCAC	36	[43]
Human β globin	1	CACTGAC	37	[44]
Human growth hormone	4	ACCCAAG	38	[43]
cardiac troponin	4	GGCTAAC	38	[47]
Chicken β tropomyosin * *	5A	CCTTGAC	48	[48]
Chicken β tropomyosin * *	AB	TCTCAAC	105	[48]
Rat β tropomyosin * *	6	CTCTCAT or TCATCAC	147 or 144	[45]
Rat α tropomyosin * *	2	GGCTAAC	177	[49]

* Underlined base is a non-consensus base.
* * Alternative splice site.

cag > tag > aag > gag) [51]. This is supported by the fact that ag's are not commonly found 10–20 bases upstream of acceptor sites [36] and when present they are associated with weaker sites. However, other work where the branch points are mutated shows that a new branch point (always an A) is chosen 22 to 37 bases upstream of the acceptor implying that the 3′ site is the more important site [52]. Senapathy et al. [31] state that the 5′ (donor) site must be recognized first, however, based on information content, as stated previously, Stephens and Schneider [35] suggest just the opposite.

As discussed above, internal exons have a maximum size of about 300–400 bp with a mean of 245 bases and a peak of about 125 bases [8,18]; however, they also have a minimum size of approximately 50 bases [53]. In vitro experiments with constructs have shown that when exons drop below this size, they are skipped [27,29]. Small exons can be spliced however, and exons as small as 3 bases are known [8,47]. These small exons apparently all require specific enhancer sequences in order not to be skipped. For example in the cardiac troponin gene the 30 base exon 5 has a purine rich enhancer [47]. In vitro, small exons that are skipped can be spliced in if purines are removed from the pyrimidine tract [53].

The above size constraints concern only internal exons. The average size of the 5′ (first) exon is 200 bases with the highest percentage less than 100 bases and ranging up to 1 kb. Deletions in the first exon do not appear to affect splicing of exon 2 [29]. The 3′ (last) exon averages 649 bases with a peak at 300 bases and a large percentage over 900 bases [18]. These features imply that the splicing signals for the first and last exon must be different than for internal exons. For example, Liu and Mertz [54] found that an acceptor signal that worked well at an internal exon did not work for the last exon unless the pyrimidine tract was strengthened or sequences were deleted in the last exon. First exons apparently require the presence of the 5′ 7-methyl-guanosine cap (and its binding proteins) which is present on all polymerase II transcripts [8,55]. Last (3′) exons end at the poly A tail and studies indicate that splicing and polyadenylation factors interact (i.e., mutation of the last exon inhibits polyadenylation and mutation of the adenylation site inhibits last exon splicing) [8,56].

A minimum intron size has also been determined [33,49,52,57]. Smith and Nadal-Ginard [49] found a minimum 5′ splice site to branch distance of 51–59 bases in the α-tropomyosin gene while Wieringa et al. [57] found a minimum intron size of 80 bases (did not function at 69 bases or less) in the rabbit β-globin gene and their survey of the literature found no introns of less than 50 bases. An unusual mutation in the donor site of exon 8 of the COL1A1 gene [58] causes not only skipping of exon 8 but inclusion of 96 bases on the 5′ end of intron 7 into the mRNA. Looking at the genomic DNA, the new donor site (Ggtaaga, Senapathy score—86.5) is better than the old site (Tgtgagt, Senapathy score—75.8); the reason that the better site must not normally be used is because this would reduce the size of intron 7 to only 63 bases (too small for an intron); however, in the mutant where exon 8 is not recognized then 'intron 7' extends until the beginning of exon 9 (280 bases).

There is also information that suggests that some exons are recognized as a cluster or cooperative unit [59–61]. Sterner and Berget [59] found that a mini-exon requires the preceding exon to be spliced in; mutation of the 5′ splice site of the upstream exon leads to intron inclusion. Ge et al. [60] found cooperation of two 3′ sites in order to splice a 48 bp exon. Both exon 8 (44 bp) and exon 9 are skipped with a mutation in the exon 9 donor site in the *p*67-*PJOX* gene [61].

Splicing can also be influenced by pre-mRNA secondary structure [46,48,62]. In the human growth hormone gene, stabilization of a stem loop causes shifting to the alternative splice site [46] while a single nucleotide polymorphism in exon 2 of the episialin gene [62] causes use of different splice acceptor sites; this polymorphism is hypothesized to cause the loss of a hairpin loop. In the tropomyosin gene, a large hairpin encompassing the alternative exon 6B causes it to be spliced out. Mutations that destabilize the loop cause the exon to be spliced in Ref. [48].

Mutations leading to splicing errors are a significant proportion of mutations leading to human disease (i.e., 101 of a total of 659 point mutations causing human diseases are in splice junctions [63]). Krawczak et al. [63] and Nakai and Sakamoto [16] review 101 human and 209 mammalian splicing

mutations, respectively. Krawczak et al. [63] report 62 donor (5′) splice site mutations, 26 acceptor (3′) splice site mutations and 13 creation of novel sites. For the donor (5′) splice site mutations, 60% involve the invariant gt with increased mutations at IVS + 1 and IVS + 2. For the acceptor (3′) splice site, 87% involve the invariant ag with an excess of mutations at IVS-2. Of interest, 12 of 13 of the new created novel sites are upstream of the normal site and only 4 of 13 have lower Senapathy scores than the original site. For the 5′ splice site mutations, 16 lead to skipping, 7 to cryptic utilization and 5 to both. For the 3′ splice mutants, 4 result in exon skipping and 6 in cryptic site utilization. For both sites, purines are most likely to be introduced in the mutation. Nakai and Sakamoto [16] found that 92% of mutations occur at consensus sites, 97 at 5′ sites and 55 at 3′ sites with 15% causing novel sites. For the 5′ site mutations, 48 lead to skipping and 33 to cryptic and for the 3′ site 29 lead to skipping and 21 to cryptic. They also found that novel sites are all upstream of the normal site. As indicated above, intron retention is rare. Among the intron retention mutants, 4 are very short introns and 3 are terminal introns; however, 3 are large internal introns.

Mutations at splice sites often lead to the observation of multiple cDNA species after RT/PCR analysis of intracellular mRNA. Sometimes utilization of a cryptic site is incomplete with some normal message observed or both skipped and normal or skipped and cryptic site mRNAs seen. These indicate competition between splice sites. In addition, sometimes mRNAs with multiple skipped exons are observed. Nonsense mutations in exons have also been found to lead to both low levels of mRNA and/or mRNAs (sometimes multiple) usually excluding the exon containing the nonsense mutation ([64–72]; reviewed in Ref. [73]). It has been suggested that the premature termination of translation caused by the nonsense mutation causes mRNA instability and, thus, low levels of mRNA. This is supported by the observation that nonsense mutants do not affect mRNA levels if they occur at the very 5′ end of the gene (re-initiation of transcription possible) or at the very 3′ end (transcription goes long enough that the mRNA is not unstable). Often, if the exon containing the nonsense mutation is skipped and an inframe deletion results, premature termination is avoided and that mRNA will be stable. In general, it is not suggested that the nonsense mutation causes the exon skipping, but merely that it stabilizes what is normally a low level product that is seen in the mutant only because the major full length (nonsense containing) product is degraded. The exon skipped product seen in nonsense mutants has been demonstrated at a low level in normal cells for some genes [68,69]. However, some groups suggest that the nonsense mutation does directly affect splicing because changes in mRNA are observed in the nucleus and translation effects must be cytoplasmic [64,70]. It is difficult to model how a nonsense mutation could be recognized by the spliceosome as this requires recognition of mRNA frame in the nucleus and many 'nonsense codons' obviously exist out of frame in normal mRNAs. Some of these nonsense/exon skipping mutations could occur because the nonsense mutation just happens to be within an enhancer which is disrupted [71]. Models have been proposed whereby a link exists between splicing and translation; the ribosome pulls the hnRNA through the nuclear pore driving splicing [65]. If the ribosome releases the RNA, splicing ceases. *Caenorhabditis elegans* with mutations in their *smg* genes do not have nonsense mutation reductions in mRNA in other genes [72] implying that specific genes exist to degrade prematurely terminated mRNAs.

1.2. The HPRT gene

The X-linked hypoxanthine-guanine phosphoribosyltransferase (*HPRT*) gene is widely used in mutation studies because of the ease of selection of mutant cells. HPRTase phosphoribosylates hypoxanthine and guanine in the purine salvage pathway. Although constitutively produced, the HPRT enzyme product is non-essential as purines can also be created de novo. HPRTase deficient mutants can be selected in the presence of purine analogues such as 6-thioguanine (TG) which kill normal, wild type cells; only cells containing a mutation in their *HPRT* gene are able to survive and proliferate. Although HPRT is dispensable at the cellular level, germinal HPRT deficiency does lead to human disease, either Lesch–Nyhan syndrome (complete deficiency) or X-linked gout (partial deficiency), apparently caused by the high levels of purine metabolites such as uric

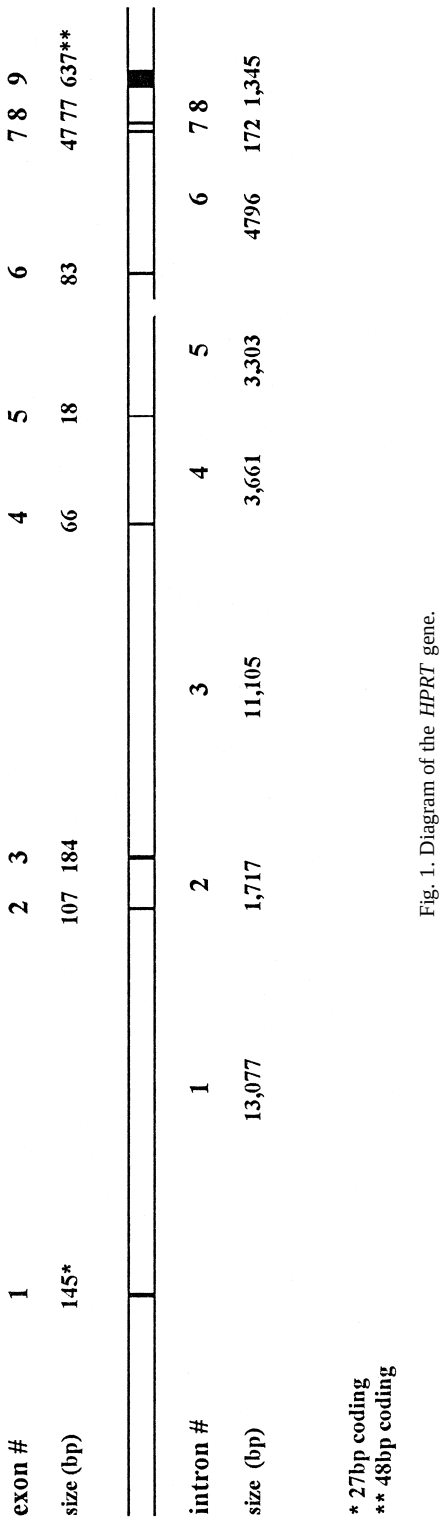

Fig. 1. Diagram of the *HPRT* gene.

acid in the blood. The symptoms of Lesch–Nyhan disease include an auto-mutilation syndrome in hemizygous males. A reversion assay for selection against *HPRT* mutations also exists, which is the ability to grow in hypoxanthine-aminopterin-thymine (HAT) medium. Studies using the *HPRT* gene include: in vitro analyses of radiation and chemical effects in human lymphoblastoid cell lines or peripheral T-cells, in vivo studies of mutations induced in humans after radiation or chemical exposures, in vivo studies of individuals with DNA repair defects and analysis of mutations in Lesch–Nyhan individuals. Numerous in vivo and in vitro studies have also been performed using the *hprt* gene in animals and animal cells including mice, rats and CHO cells.

The human *HPRT* gene is located at Xq26 and its 9 exons cover 43 Kb (Fig. 1). The mRNA is 900 bp with a coding region of 657 bp. Of note is that *HPRT* has two large introns (introns 1 and 3) and one relatively small one (intron 7). Also, exon 5 is very small (only 18 bases); this is well below the 'minimum' exon size of 50 bases and suggests that exon 5 should possess splice sites with high scores and probably also an enhancer. Exon 7 is also small (47 bases). A 57 kb region containing the entire gene has been sequenced and linked DNA markers have been mapped, allowing for detailed molecular studies [74,75]. A world-wide database and international repository has been developed describing human *HPRT* mutation molecular spectra [76]. Currently this database (release 6) includes 2224 mutations, 638 of which (28.7%) are listed as 'splice' mutations. This is the largest collection of splicing mutations from a single gene. This paper reviews mRNA splicing at the *HPRT* gene including the sequences and structures of the splice sites, predicted and actual cryptic sites, the locations and frequencies of splice mutations and the production of multiple mRNA products. We will especially focus on mutations with unusual or unexpected consequences which reveal information on splicing mechanisms.

2. Materials and methods

Mutations in the *HPRT* gene are selected through the phenotype of resistance to the purine analogue 6-thioguanine (TG). In general, TG^r mutants are

selected by growth in the presence of 10 μM TG. The mutations described here are from a variety of human sources. Most are in vitro derived from either T-lymphocytes or lymphoblastoid cell lines, often treated with mutagens. However, many are in vivo derived somatic mutations in T-lymphocytes from normal or mutagen-exposed individuals and some are from individuals with Lesch–Nyhan syndrome or X-linked gout (germinal mutations). The *HPRT* gene is an ideal source to probe hypotheses of splicing recognition. Analyses of mutations that are known to affect splicing provide unequivocable information on the nature of splice recognition in this human gene.

The splice mutations discussed in this review are a compilation of the mutations described in the references listed in the Appendix A to this paper. They are part of an *HPRT* mutation database (Ref. [76], release 6) containing 2224 mutations, of which 638 (28.7%) are classified as splice mutations. Most of the *HPRT* splice mutations were discovered because they gave abnormal products after RT-PCR analysis; however, some were originally discovered based on genomic sequencing. The RT products are often smaller than the expected full length cDNA, with sometimes several products due to multiple splicing events. DNA sequencing of the RT product reveals the absence of one or more exons (exon exclusion) or the presence of inserted intron sequences (intron inclusion). When coupled with genomic PCR studies, it is clear that the excluded exons are present in the genomic DNA and the alterations in mRNA splicing must be due to point mutations in splice sites. Genomic DNA sequencing is necessary to define the exact mutation. This review focuses on the 277 of the 638 splice mutations for which the actual genomic basis (i.e., the exact basepair(s) changes) for the splicing alteration has been determined. (For 341 mutants the genomic mutation is not known, only that the mRNA is altered; for example, the cDNA lacks a single exon. This could be due to a small deletion, a point mutation in the 5′ splice site, a point mutation in the 3′ splice site, etc.)

Donor and acceptor splice site Senapathy scores were calculated at each nucleotide of the entire *HPRT* gene (bases 1701–41,501) using the equations and tables in [31] and the program Mathematica (Wolfram). The information content in bits for donor and acceptor sites (R_i) was also calculated at each nucleotide of the HPRT gene using the equations in [35]. To assess the effects of mutations, R_i were computed for the normal and mutated site using the program Scan and displayed with MakeWalker, DNAplot and Lister as in Ref. [33].

3. Results

This paper attempts to summarize the types of splice site mutations observed in the human *HPRT* gene and to discern any 'rules' which apply to this particular gene. In this review, we will use the following splice recognition sequence terminology. The splice sequences in the intron are designated 5′ and 3′ splice sites. Relative to an exon, the intron 3′ splice sequence is termed 'donor' and the intron 5′ splice sequence is termed 'acceptor.' Table 3 gives the *HPRT* 3′ (donor) and 5′ (acceptor) site sequences and their Senapathy scores [31] and R_i values [38]. By the Senapathy method, a perfect score requires a donor sequence of AGgt(a/g)agt and an acceptor sequence of 10(t/c)n(t/c)agG. Using the information theory based model, a minimal R_i of 0 is required for splicing but in general a cutoff of 2.36 bits is utilized; however, values can be as high as 21 although most sites have R_i values in the 7–12 bit range [38]. Also given is the rank of that site among all possible sites of the *HPRT* gene (donor and acceptor splice site scores and values were calculated for each base from the beginning of exon 1 to the end of exon 9 with the genomic sequence of 43 kb). Although some sites are very good (the exon 2 donor splice site has a Senapathy score of 100 and an R_i value of 12.7, both ranking it #1), others are quite poor (the exon 8 acceptor has a Senapathy score of only 77.1 and is ranked #571 and an R_i value of 2.8 and is ranked #733). Nevertheless, these rankings indicate that the exon 3 donor, exon 7 donor, and exon 8 acceptors might be particularly sensitive to mutation. Exon 2 has a very strong acceptor and donor indicating it might be more resistant to mutation. As for the small exon 5, both acceptor and donor are strong.

The determination of Senapathy scores and R_i values for the entire *HPRT* gene allows speculations

Table 3
hprt splice sites

Intron	5′ splice (donor) [AGgt(a/g)agt]	Senapathy score (rank)	R_i value in bits (rank)	3′ splice (acceptor) [10 (t/c)N(t/c)agG]	Senapathy score (rank)	R_i value in bits (rank)
1	TG|gtgagc	83.0 (48)	7.6 (49)	tatttctttttcag|AT	95.8 (6)	9.7 (50)
2	AG|gtaagt	100.0 (1)	12.8 (1)	ttttatttctgtag|GA	89.8 (55)	11.3 (18)
3	GT|gtgagt	76.1 (145)	5.1 (161)	ttttttttaactag|AA	86.6 (120)	10.0 (40)
4	AG|gtatgt	88.7 (10)	9.5 (11)	tctttttcttctag|AA	87.7 (94)	14.7 (5)
5	AA|gtaagt	87.6 (11)	8.0 (39)	ttctttttgaaag|GA	86.4 (125)	7.1 (159)
6	AG|gtatgt	88.7 (9)	9.5 (10)	tttgtaattaacag|CT	81.1 (323)	4.6 (423)
7	CT|gtaagt	79.0 (86)	7.6 (50)	atgattctttttag|TT	77.1 (571)	2.8 (733)
8	AT|gtaagt	87.2 (13)	9.2 (12)	attttttttatag|CA	85.3 (150)	8.9 (74)

as to the locations of likely cryptic sites. Cryptic sites are usually close to the original site (as otherwise the exon becomes too large or small) and would be expected to have a high score (although usually lower than the normal site). Table 4 lists sites within 50 bases of the normal donor (5′) and acceptor (3′) sites that have Senapathy scores of at least 70 or R_i values of at least 2.36 bits. The acceptor sites at +50 of exon 6, at +21 of exon 8, and at −42 of exon 9 all have higher values than the normal site, while none of the putative donor sites do. The exon 2 donor at +5 has fairly high values; however, since this overlaps with the normal site, mutations in the normal site will also affect +5 site.

Table 4
Putative cryptic splice sites (Within 50 bp of normal site and Senapathy score > 70.0 or R_i value > 2.36

5′ Splice site				3′ Splice site			
Intron	Distance	Senapathy score	R_i value in bits	Intron	Distance	Senapathy score	R_i value in bits
1	+49	67.2	5.0	1	−17	76.8	
					+5	76.3	4.4
					−48	75.9	
2	+5	81.9	6.0	2	−47	75.8	
	−32		2.9		−3	70.9	
					+20	72.6	
3	−2		2.5	3	+16	75.9	
	+39		4.4		−4		4.4
4	−39	73.2	6.0	4	−25	71.6	
	+16	76.1	2.8		+2		2.9
					+3	70.7	3.9
5	none			5	−22	71.0	
					−4	70.2	
					−1	70.4	
					+50	92.8	7.5
6	−38	77.6	2.8	6	−39		5.8
	+4		2.9		−28	70.3	
7	−40		5.1	7	−24	70.2	
	+4		3.6		+6	73.2	
					+21	79.3	3.4
8	+4	73.2	4.6	8	−47	75.7	
					−42	91.2	12.2
					+3		2.8
					+17	72.0	2.7

Use of the calculated Senapathy scores across the gene also allows prediction of exons within the region (i.e., would the observed exons be predicted, and/or do other 'exon-like' sequences exist that would be predicted to be exons but are not?). Using acceptors with Senapathy scores greater than 85.2 and donors with Senapathy scores greater than 83.0 (7 of 9 *HPRT* exons have these scores), and the criterion that an exon must be between 50 and 300 bp, leads to the prediction of 26–27 exons (there are two that overlap) of which only 3 are actual *HPRT* exons (exons 2, 4 and 6). Two exons would be predicted around the 18 bp exon 5 (the overlapping set). These new exons are defined by either the exon 5 donor or acceptor splice sequence being used with another 'splice sequence' to yield a larger exon of 190 or 85 bases, respectively. This reaffirms that something must be enhancing exon 5 splicing since other better choices exist. For exons 3 and 8, the acceptor and donor splice sites, respectively, are present, however, they have no cognate 5′ or 3′ splice site, respectively, within the 50–300 base range. For exon 7, neither splice site makes the list.

An advantage of *HPRT* mutations for analyzing splice recognition site mutations is that polymorphisms are not a complication since a mutant phenotype has been selected. In fact, no polymorphisms in the coding region of the human *HPRT* gene have been described. The wild type sequence can be directly compared to that in the mutant cell. There are, however, some complications. *HPRT* splice mutations are often defined by RT/PCR methods; determining the exact mutation often requires a genomic DNA PCR assay and sequencing. Unfortunately, for many of the 'splice mutations' in the *HPRT* database only the RT analysis or RT analysis and genomic (multiplex) PCR (but not sequencing of the genomic PCR products) has been performed. Thus, many of these 'splice' mutants show the phenotype of exon exclusion in cDNA with the exon's presence only confirmed in genomic DNA. Therefore, this database is a work in progress, as more mutations will certainly be determined; at present, the precise mutation has been determined in 277 splicing mutants.

As discussed in the introduction, the human *HPRT* gene contains 8 introns and 9 exons (Fig. 1). Exclusion of a single exon due to aberrant splicing can yield an in-frame reading frame (exon 4 or 5 only) or out-of-frame reading frame (rest of exons) in the mRNA. Since the usual method of analysis is RT/PCR with primers 5′ and 3′ to the *HPRT* coding sequence, a splicing alteration which excludes exon 1 or 9 will not be detected because there will be no cDNA produced by the RT/PCR procedure.

The bases in the *HPRT* gene are designated by two numbering systems in this review. A nucleotide's position in the *HPRT* gene is indicated by subscripts after the base. The base numbers are given both according to the genomic sequence of 56,736 bases of Edwards et al. [74] (Note: the GenBank file M26434 has an inserted T between 22,227 and 22,228 and thus numbering differs by one between the Edwards system and M26434 after base 22,227) and according to the cDNA sequence (base 1 is the A in the AUG initiation codon). To easily discriminate between the two numbering systems, if the number is less than 700 it is cDNA numbering while those numbers greater than 1700 are genomic numbering. Intronic bases may also labeled by an intervening sequence nomenclature; the intron (IVS) bases are numbered away from the nearest exon, bases 5′ to the exon given as negative (−) and bases 3′ to the exon given as positive (+). For example, the first g of the intron 1 splice 5′ (donor) site is designated base number 1704 in genomic DNA or intron 1, base + 1 (IVS1 + 1). The first base 5′ to exon 2 (exon 2 acceptor site) is designated base number 14,779 or intron 1, base −1 (IVS1-1). Exonic bases are designated by uppercase letters and intronic bases by lowercase letters.

3.1. Mutations in intron splice recognition sequences

Table 5 lists the number of mutations seen for each of the 5′ (donor) and 3′ (acceptor) splice sites and the data on splice mutations are presented individually below for each of the 9 *HPRT* exons.

3.1.1. Exon 1 (Table 6: cDNA bases 1–27), 9 donor mutations

The first exon of the *HPRT* gene contains the first 27 coding bases (1–27) plus approximately 60 bases 5′ to the AUG initiation triplet. The splice

Table 5
Frequencies of splice site mutations

Intron	No. of 5′ Splice Site (donor) Mutations (%)*	No. of 3′ Splice Site (acceptor) Mutations (%)*
1	9 (8.0)	9 (10.1)
2	5 (4.5)	5 (5.6)
3	3 (2.7)	11 (12.4)
4	15 (13.4)	5 (5.6)
5	21 (18.8)	7 (7.9)
6	13 (11.6)	10 (11.2)
7	18 (16.1)	21 (23.6)
8	28 (25.0)	21 (23.6)
Total	112	89

* An equal distribution among the 9 exons would be 14 5′ (donor) splice site mutations (112/8 = 14, 12.5%) and 11 3′ (acceptor) splice site mutations (89/8 = 11, 12.5%).

donor sequence is GTG_{27}gtgagc. Mutations at IVS1 +1, +2 or +5 have been observed to result in the inclusion of the first 49 bases of intron 1 in the cDNA. This is the result of the use of a cryptic splice donor sequence g_{1753}tggcg (Senapathy score 67.2, R_i value 5.0 bits). Although the values for the cryptic site are low, it is the best in the area and it is essential that very large intron 1 (13,075 bp) be spliced out to make a mRNA. From these results, it appears that an exon 1 splice donor sequence mutation always results in the use of this cryptic donor sequence and creates a 49 base intron inclusion in the mRNA. This inclusion is out of frame and yields a chain terminating codon TAG at new codon number 27 (initiation codon AUG still designated codon 1).

In summary, exon 1 splice donor site mutations have been found in base 1, 2 or 5 of the 6 splice sequence bases and in the last base of the exon (27G) and all lead to an intron inclusion. This included sequence contains the mutated base.

3.1.2. Exon 2 (Table 7: cDNA bases 28–134), 9 acceptor and 5 donor mutations

The splicing of exon 1 to exon 2 removes the largest *HPRT* intron (IVS1—13,075 bases). The intron 1 5′ splice (exon 2 acceptor) sequence is cagA_{28}T and mutations at IVS1-1 or -2 have been reported. Mutations at both sites have pleiotropic effects, yielding multiple species of mRNA. The largest cDNA is missing exon 2 bases 1–5 (*HPRT* bases 28–32:ATTAG) due to the use of a cryptic splice acceptor sequence (TAG_{32}/T) (a potential cryptic splice sequence, CAG_{46}/G, does not appear to be used). In addition, cDNA species missing exon 2 or exon 2 and 3 have been reported.

Mutations in the exon 2 acceptor splice site can cause different mRNA spectra depending on the particular type of nucleotide replacement. Changing of the invariant g at IVS1-1 (IVS1-1G) leads to use of a cryptic site 5 bases into exon 2 [normal—tatttct-ttttcagATTAGTG; mutant—tatttctttttcanATTAGTG which is spliced as tatttctttttcanattagTG]. The new alternate site has Senapathy scores of 82.6 (R_i value of 7.3 bits), 81.1 (R_i value of 7.1 bits) and 75.8 (R_i value of 4.6 bits) for g → t, g → c, and g → a, respectively [vs. 95.8 (R_i value of 9.7 bits) for the usual unmutated site and 75.8 (R_i value of 4.4 bits) for the unmutated site]. The g → t and g → c mutations lead to exclusive use of the cryptic site while the g → a mutation leads to mixed exon 2 loss and cryptic site mRNAs. This is understandable based on the lower values for the g → a mutation site. Changing of the invariant a at IVS1-2 to t (IVS1-2 A → T) leads to use of the cryptic site (Senapathy score 76.6,

Table 6
Splice site mutations involving exon 1 (27 Bases; 1–27)[a] $A_1TG \ldots GTG_{27}$gtgagc

5′ (Donor) Splice Sequence	Mutation	Effect in cDNA
GTGg_{1704}tgagc	g_{1704} → a or t or t_{1705} → a or g_{1708} → a or t	Inclusion of intron 1, b 1–49 due to use of a cryptic splice sequence (cag$_{1752}$|gtggcg)
	Δ (TG_{27}gtgagc … c_{1737})	Deletion of E1, b 26–27 and insertion of intron 1, b 35–49 due to use of a cryptic splice sequence (cag$_{1752}$|gtggcg)

[a] Nine donor site mutations have been reported. A mutation at 27G (G → A) has been reported to affect splicing (see Table 15).

Table 7
Splice site mutations involving exon 2 (107 bases; 28–134)[a] atatttcttttt cag $A_{28}TT \ldots AG_{134}$gtaagt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
$cag_{14779}ATT$	$g_{14779} \rightarrow$ c or a or t or $a_{14778} \rightarrow$ g or t	Loss of E2, b 28–32 due to use of a cryptic splice site ($ATTAG_{32}$|T) and/or loss of E2 and/or loss of E2 + 3
5′ (Donor) Splice Sequence	Mutation	Effect in cDNA
AG g_{14887}taagt	$g_{14887} \rightarrow$ a or $t_{14888} \rightarrow$ g	Loss of E2

[a] Nine acceptor and five donor site mutations have been reported. No mutations in 28A have been reported; mutations at 134G (G → A or T) have been reported with no effect on splicing.

R_i value of 6.5 bits) while changing to a g leads to exon 2 and/or exon 2–3 loss (old mutated site Senapathy score 71.1, R_i value of 4.3 bits). The latter still has an acceptable R_i value, it is not clear why it is not used.

The intron 2 5′ splice site sequence is CAG_{134}gtaagt and mutations at IVS2 + 1 and IVS2 + 2 have been reported. All reported mutations result in the exclusion of exon 2 from the cDNA. Mutations at 134G do not affect splicing. Mutations in the exon 2 splice donor sequence do not lead to the use of a cryptic splice site in intron 2, although the sequences gtaaga and gtggga exist 4 and 191 bases downstream, respectively.

In summary, mutations in either base of the ag dinucleotide in the exon 2 splice acceptor sequence or mutations in base 1 or 2 of the 6 bases in the exon 2 donor sequence lead to exclusion of part or all of exon 2 in the cDNA. The exclusion of exon 2 bases 1–5 results in an immediate chain terminating TGA at new codon 10. The exclusion of exon 2 results in a chain terminating codon TGA at new codon 11. The exclusion of exon 2 + 3 results in an in-frame mRNA lacking 291 bases. The observed exclusion of both exon 2 and 3 only with acceptor site mutations suggests that exons 2 and 3 are spliced together prior to splicing to exons 1 and 4 (this is discussed below).

3.1.3. Exon 3 (Table 8: cDNA bases 135–318), 5 acceptor and 3 donor mutations

The splicing of exon 2 to exon 3 removes the relatively small intron 2 of 1715 bases (IVS2). Mutations in the exon 3 tagG_{135}A acceptor splice site have been found only at IVS2-1G. Mutations in the exon 3 donor splice TGT_{318}gtgagt site have been found only at IVS3 + 1G. Mutations in both splice sequences result in the exclusion of exon 3 and/or exon 2 and 3 from the cDNA. Mutations of g → a at IVS3 + 1 cause exon 2–3 loss while g → t mutations cause only exon 3 loss. There appears to be no explanation for this. The differences could be purely coincidence related to variations in reporting of multiple cDNAs by different researchers. The use of a potential cryptic donor site, gtaggt, 39 bases downstream has not been reported. The exclusion of exon 3 results in a chain terminating T_{341}AA at new codon 53. As pointed out above, the exclusion of exon 2 and 3 results in the in-frame loss of 291 bases from the mRNA.

The exclusion of exon 2 + 3 is quite common with mutations in the 3′ splice sequences of both

Table 8
Splice site mutations involving exon 3 (184 bases; 135–318)[a] attttatttctg tagG_{135}ACT…TGT_{318}gtgagt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
$tag_{16602}GAC$	$g_{16602} \rightarrow$ c or a or t	Loss of E3 and/or E2 + 3
5′ (Donor) Splice Sequence	Mutation	Effect in cDNA
TGTg_{16787}tgagt	$g_{16787} \rightarrow$ a or t	Loss of E3 and/or E2 + 3

[a] Five acceptor and three donor site mutations have been reported. Mutations at 135G (G → T or C) and at 318T (T → A) have been reported with no effect on splicing.

intron 1 and 2 and the 5′ splice sequence in intron 3. This exclusion is not reported with intron 2 5′ splice sequence mutations, which suggests that the resultant splicing of exon 1 to exon 4 is the frequent result of interference with the splicing between exon 2 and 3. Alternatively, the mRNA missing both exon 2 and 3 is probably more stable than the mRNA missing exon 2 or 3 alone because of the chain terminating codons created by the loss of either single exon.

In summary, mutations affecting the splicing of exon 3 are relatively infrequent and all reported so far cause the exclusion of exon 3 and/or exons 2 and 3. It should be noted, however, that the exclusion of exons 2 and 3 in cDNA has been reported often in the database and many of these could also be splice sequence mutations. However, since the genomic alteration has not been defined, these could be V(D)J recombinase mediated deletions—which are common in T-lymphocytes [77,78]—or other genomic deletions of exons 2 and 3. In conclusion, exon 3 splice mutations make up only 4% of the total-well below the expected 12.5%.

3.1.4. Exon 4 (Table 9: cDNA bases 319–384), 11 acceptor and 15 donor mutations

The splicing of exon 3 to exon 4 removes the 11,103 base intron 3 (IVS3). Mutations of the acceptor sequence tagA_{319}A at IVS3-1 or -2 result in the loss of exon 4 or exon 4, bases 319–327 due to the use of a cryptic splice sequence (CAG_{327}|T). Mutations in the donor splice sequence AAG_{384}gtatgt have been reported only at IVS4 + 1 or + 2. Both result in exclusion of exon 4 from the cDNA which results in an in-frame deletion of 22 amino acids, while the exclusion of bases 319–327 results in an in-frame deletion of 3 amino acids. There is no reported use of cryptic 5′ splice sequences in IVS4 although potential sites do exist; i.e., gtatct and gtgtgc, 4 and 57 bases downstream, respectively.

In summary, exon 4 splicing errors occur at an average frequency (12.5% for each site) with acceptor splice site mutations showing partial or complete exon exclusion. Donor splice site mutations show simple exon 4 exclusion.

3.1.5. Exon 5 (Table 10: cDNA bases 385–402), 5 acceptor and 21 donor site mutations

The splicing of exon 4 to exon 5 removes the 3659 bases of intron 4. Mutations in the acceptor splice sequence tagA_{385}A have been reported at IVS4-1 or -2 and result in the exclusion of exon 5. Mutations in the donor splice sequence, GAA_{402}gtaagt have been reported at IVS5 + 1, + 2, or + 5. These mutations cause the inclusion of 67 bases of IVS5 in the cDNA, due to the use of a cryptic splice site at aag$_{31701}$/gtaagc. Mutations at IVS5 + 5G → T cause exon 4–5 exclusion and use of the cryptic site, while IVS5 + 5G → A mutations cause use of only the cryptic site. The differences could be purely coincidence related to variations in reporting of multiple cDNAs by different researchers. The 67 base intron inclusion results in a chain termination codon, taa, at new codon 136. In

Table 9
Splice site mutations involving exon 4 (66 bases; 319–384)[a] ttttttttaactag A_{319}AT...AAG_{384}gtatgt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
tag$_{27890}$AAT	g_{27890} → c or a or t or a_{27889} → g Δ(A_{319}ATG)	Loss of E4 or loss of E4, b 319–327 due to use of a cryptic splice site (CAG_{327}|T) Loss of E4 or loss of E4, b 319–322 (tag|A_{322}) or loss of E4, b 319–327 (CAG_{327}|T)
Donor Splice Sequence	**Mutation**	**Effect in cDNA**
AAGg_{27957}tatgt	g_{27957} → a or t or t_{27958} → g or a	Loss of E4

[a] Eleven acceptor and 15 donor site mutations have been reported. No mutations at 319A or at 384G have been reported.

Table 10
Splice site mutations involving exon 5 (18 bases; 385–402)[a] ctctttttcttctagA$_{385}$AT…GAA$_{402}$gtaagt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
tag$_{31616}$AAT	g$_{31616}$ → c or t or a$_{31615}$ → g or c	Loss of E5
5′ (Donor) Splice Sequence	**Mutation**	**Effect in cDNA**
GAAg$_{31635}$taagt	g$_{31635}$ → a or t or t$_{31636}$ → g or c or a or g$_{31639}$ → a or t	Inclusion of intron 5, b 1–67 due to use of a cryptic splice site (aag$_{31701}$|gtaagc) and/or loss of E4 + 5
	Δa$_{31637}$ or Δa$_{31638}$	Inclusion of 66 bases of intron 5
	aa$_{31638}$ → gt	Inclusion of 67 bases of intron 5

[a] Five acceptor and 21 donor site mutations have been reported. No mutations have been reported at 385A or 402A.

addition, multiple cDNA products have been reported which lack exons 4 and 5 and/or exons 2–6.

In summary, exon 5 donor site mutations result in the use of a cryptic splice donor site in intron 5 (similar to exon 1 donor site mutations). In addition, the frequency of exon 5 donor site mutations is 18.8% of the total, higher than the 12.5% that would be expected if all 8 donor sites were equally mutated (100% ÷ 8 = 12.5%) (Table 5).

3.1.6. Exon 6 (Table 11: cDNA bases 403–485), 7 acceptor and 13 donor site mutations

The splicing of exon 5 to exon 6 removes the 3301 bases of intron 5. Mutations in the acceptor sequence, aagG$_{403}$A, have been reported at IVS5-1. When the IVS5-1 g is mutated to a t or c there is exon 6 skipping; however, when IVS5-1 (g) is mutated to a, an aG dinucleotide is created using the first base of exon 6 (403G) and this aG is used as a splice acceptor. The net result is loss of 403G in the mRNA [i.e., normal sequence—tctttttgaaagGA (Senapathy score of 86.4; R_i value of 4.9 bits), mutated sequence—tctttttgaaaaGA which splices as cttttttgaaaagA (Senapathy score of 81.6; R_i value of 4.9 bits)]. The mutated site have an a at IVS-3 which is non-consensus; however, of note, the normal exon 6 acceptor also an a at IVS-3 instead of the expected c or t at IVS-3. Thus, for whatever reason for this exon an a at IVS5:-3 is acceptable. The exclusion of 403G results in a termination codon T$_{407}$AA at new codon 136. The exclusion of exon 6 results in an out-of-frame mRNA with a termination codon, T$_{522}$AA, at new codon 140. Mutations in the donor splice sequence, AAG$_{485}$gtatgt have been reported at IVS6 + 1 or + 2, and result in the exclusion of exon

Table 11
Splice site mutations involving exon 6 (83 bases; 403–485)[a] cttctttttgaaagG$_{403}$AT…AG$_{485}$gtatgt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
aag$_{34936}$GAT	g$_{34936}$ → a	Loss of E6 and/or loss of E6, b 403 due to use of *new* splice site (aaa$_{34936}$G|A)
	g$_{34936}$ → c or t	Loss of E6
5′ (Donor) Splice Sequence	**Mutation**	**Effect in cDNA**
AGg$_{35020}$tatgt	g$_{35020}$ → a or t or t$_{35021}$ → g or c or a	Loss of E6

[a] Seven acceptor and 13 donor site mutations have been reported. Mutations at 403G (G → C or A or T) have been reported with no effect on splicing. Mutations of 485G → A or T results in the loss of E6 (see Table 15).

6. There is no reported use of a cryptic donor site, such as $g_5$tatga (the next closest donor sequence is g_{310}tgagg).

In summary, mutations affecting exon 6 splicing make up 10.0% of the total splicing mutations (12.5% expected by equal distribution) and demonstrate no use of cryptic sites.

3.1.7. Exon 7 (Table 12: cDNA bases 486–532), 10 acceptor and 18 donor site mutations

The splicing of exon 6 to exon 7 removes the 4794 bases of IVS6. Mutations in the exon 7 acceptor sequence cagC_{486}T, have been reported at IVS6-1, -2, -3, and -12. All result in the loss of exon 7 from the cDNA. The IVS6-12 T → A mutation is unusual in that it creates a purine dinucleotide (ag) which could be recognized as a new acceptor splice site but is not. However, this change apparently does cause interference with recognition of the normal acceptor sequence nine bases downstream. This sensitivity may be related to the normal existence of 4 purines in the 11 base stretch from IVS6-5 to -15, which by consensus should be all pyrimidines.

The exon 7 donor splice sequence, ACT_{532}gtaagt has been found mutated at 5 of the 6 intron bases. Only IVS7 + 4 has not been found mutated. All donor site mutations cause the exclusion of exon 7 and there is no acceptable cryptic sequence in the 170 bases of intron 7. The exclusion of exon 7 yields an out-of-frame mRNA with a chain termination codon TGA at new codon 165.

In summary, the splice sequences flanking exon 7 contain at least 9 bases which can be mutated to cause splicing errors. Despite the larger number of 'targets', the incidence of acceptor (11.2%) and donor (16.1%) mutations are close to the expected 12.5%.

3.1.8. Exon 8 (Table 13: cDNA bases 533–609), 21 acceptor and 28 donor site mutations

The splicing of exon 7 to exon 8 removes the 171 bases of IVS7, the smallest intron in the *HPRT* gene. The acceptor splice sequence tagT_{533}T has been mutated at IVS7-1 or -2 and results in the loss of exon 8 and/or in the loss of the first 21 bases of exon 8 due to the use of a cryptic acceptor site AAG_{553}/A. In addition, mutations at IVS7-3, -5, -10, -13 or -96 have been shown to cause the loss of exon 8. These last 3 mutations change an at dinucleotide to an ag dinucleotide. The loss of exon 8 results in an out-of-frame mRNA with a chain terminating TAG at new codon 187. The use of the cryptic splice acceptor site yields an in frame mRNA which predicts a protein lacking 7 amino acids.

The donor splice site for exon 8 is AAT_{609}gtaagt and all 6 bases have been reported to be mutated and result in the loss of exon 8. There is no use of a cryptic site in IVS8 reported. The finding of mutations in all 6 bases of the IVS8 donor sequence may reflect the high frequency of IVS8 donor site mutations in general; i.e., 28 of 112 or 25% of the total compared to the expected 12.5%. In addition, 8 mutations in exon bases have been reported to affect splicing (Table 16—discussed below).

In summary, the splice sequences flanking exon 8 show a high frequency of mutations which interfere with normal splicing. Of the total of 201 reported

Table 12
Splice site mutations involving exon 7 (47 bases; 486–532)[a] ttttgtaattaacagC_{486}TTG...GACT_{532}gtaagt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
cag$_{39814}$C	g_{39814} → c or a or t or a_{39813} → t or ac_{39812} → ag or $t_{39803}g$ → ag	Loss of E7
5′ (Donor) Splice Sequence	**Mutation**	**Effect in cDNA**
Tg$_{39862}$taagt	g_{39862} → a or t or t_{39863} → c or a or a_{39864} → t or g_{39866} → c or a or t or t_{39867} → c	Loss of E7

[a] Ten acceptor and 18 donor site mutations have been reported. Mutations at 486C (C → G or A) and 532T (T → C) have been reported with no effect on splicing.

Table 13
Splice site mutations involving exon 8 (77 bases; 533–609)[a] tatgattcttttagT$_{533}$T...AAT$_{609}$gtaagt

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
tag$_{40032}$TT	g$_{40032}$ → a or t or a$_{40031}$ → g or t or t$_{40030}$ → a or t$_{40028}$ → g or a or a t$_{40023}$ → ag or t$_{40020}$g → ag or a t$_{39937}$ → ag	Loss of E8, b 533–553 due to use of a cryptic splice site (AG$_{553}$|A) and/or loss of E8; Loss of E8
5′ (Donor) Splice Sequence	Mutation	Effect in cDNA
AATg$_{40110}$taagt	g$_{40110}$ → c or a or t or t$_{40111}$ → g or c or a or a$_{40112}$ → t or a$_{40113}$ → g or g$_{40114}$ → c or a or t$_{40115}$ → g or c or a or Δa$_{40112}$	Loss of E8

[a]Twenty-one acceptor and 28 donor site mutations have been reported. Mutations at 533T (T → G or C) have been reported with no effect on splicing. No mutations have been reported at 609T.

splice site mutations, 49 (24.4.%) reside in these sequences which is twice the expected value.

3.1.9. Exon 9 (Table 14: cDNA bases 610–657), 21 acceptor site mutations

The splicing of exon 8 to exon 9 removes the 1343 bases of IVS8. The splice acceptor site, tagC$_{610}$, has been found to be mutated at IVS8-1, -2 or -3 and these mutations result in the loss of exon 9, bases 1–17 due to the use of a cryptic splice site, TAG$_{626}$/T. An IVS8-3T → G mutation change causes the acceptor to move upstream by 2 bp [normal sequence—ggatttttttttatagCA, mutated sequence—ggatttttttttagagCA which splices as ggatttttttttagAGCA] and adds 2 bases to 5′ exon 9 (IVS8-1G and -2A). The new splice site has a Senapathy score of 81.5 (R_i values of 7.4 bits) as compared to 81.9 (R_i value of 4.5 bits) if the old (but mutated site) was used. While the Senapathy scores are equivalent for the natural and cryptic sites, the new site has a higher information content than the natural site. The Senapathy score of the original unmutated site is 85.3 (R_i value of 8.9 bits). IVS8-3 T → C mutations are not known at this site and would not be expected as they would actually increase the values.

The loss of exon 9, bases 1–17, results in an immediate chain terminating TGA at new codon 204. The insertion of IVS8-1 and -2 results in a larger protein product with the first chain terminating codon T$_{749}$AG appearing at new codon 251. The insertion of IVS8-1 to -14, results in a chain terminating codon (TAG) at the same new codon but a larger protein product of 255 amino acids.

Table 14
Splice site mutations involving exon 9 (48 bases; 610–657)[a] gatttttttttatagC$_{610}$AT...TAA$_{657}$

3′ (Acceptor) Splice Sequence	Mutation	Effect in cDNA
tag$_{41453}$C	g$_{41453}$ → a or t	Loss of E9, b 610–626 due to use of a cryptic splice site (TAG$_{626}$|T)
	or a$_{41452}$ → g or t or t$_{41451}$ → a	
	tat$_{41451}$ → tag	Inclusion of intron 8, b-1 and -2 due to use of a new splice sequence (tag$_{41451}$|ag)
	tg$_{41438}$g → tag	Inclusion of intron 8, b-1 to b-14 due to use of a *new* splice site (tag$_{41439}$|a)

[a]Twenty-one acceptor site mutations have been reported. Mutations have been reported at 610C (C → G or T) with no effect on splicing.

In summary, the exon 9 acceptor site shows an elevated frequency of mutations (21 of 89 = 23.6%) in comparison with the 3′ sites in IVS1 to IVS6. Only the IVS7 3′ site shows a similar elevated frequency and 47.2% of the total acceptor site mutations reside in these 2 introns. Mutations in the IVS8 3′ sequence which block splicing of exon 8 to exon 9 cannot be determined because such a mRNA will not be detected by the conventional RT/PCR assay.

3.2. Mutations in internal (non-splice site) exon sequences that appear to affect splicing

Mutations within exon sequences have also been reported to affect splicing. This effect is usually manifested as exon(s) exclusion from the cDNA. Exons 3, 4, 6 and 8 have been reported to show these types of mutations (Tables 15 and 16). Exon 1 and 6 show mutations at the last base of the exon which effect splicing (lines 1 and 10).

A mutation in exon 3 has been found that creates a new splice 5′ (donor) site (Table 15, line 2). This G209 → T transversion in the run of 6 Gs (G_{207}GG-GGG$_{212}$) yields the sequence AG$_{207}$/GT$_{209}$GGGC which is recognized as a donor splice site (Senapathy score of 80.1; R_i value of 7.7 bits). [The normal exon 3 donor has a Senapathy score of 76.1 (R_i value of 5.13 bits)]. The use of this splice site results in the in-frame loss of bases 208–318 (37 codons). A second, larger cDNA product is also found that contains the entire exon 3 with the codon change GG$_{209}$G to GT$_{209}$G (gly to val).

Three other mutations in exon 3 have been reported to result in cDNA products with exons 2 and 3 excluded (Table 15, lines 3–5). All 3 create chain terminating stop codons which will probably result in an unstable mRNA. The mRNA missing exons 2 and 3 is in frame and would be expected to be more stable. It seems reasonable to conclude that these mutations do not directly affect splicing. A similar explanation might explain the exclusion of exon 4 from cDNA with either of 2 mutations that also create stop codons (Table 15, lines 6 and 7) because the exclusion of exon 4 also results in an in frame mRNA. The third exon 4 mutation, 377delC, results in a stop codon, T$_{392}$GA, after 6 codons (Table 15,

Table 15
Single base pair changes in exon sequence that affect splicing (Exons 1, 3, 4 and 6)

Exon	Position	Change	Splice Effect
1	GTG$_{27}$gtgagc	GTG → GTA (val → val)	Inclusion of bases 1–49 of intron 1 into cDNA due to use of the cryptic sequence cag$_{1752}$gtggcg
3	GG$_{209}$G GGC TAT	GGG → GTC (gly → val)	± Loss of bases 208–318 in exon 3 due to creation of a new splice donor site AG$_{207}$|GT$_{209}$GGGC
3	GCT C$_{151}$GA GAT	GGA → TGA (arg → STOP)	Loss of exon 2 + 3 (probably due to chain terminating codon resulting in unstable mRNA. Loss of exon 2 + 3 results in *in frame* mRNA which would be stable)
3	ATG A$_{163}$AG GAG	AAG → TAG (lys → STOP)	Loss of exon 2 + 3 (as above)
3	CTC TGT$_{198}$GTG	TGT → TGA (cys → STOP)	Loss of exon 2 + 3 (as above)
4	GGT G$_{355}$GA GAT	GGA → TGA (gly → STOP)	± Loss of exon 4 (probably due to chain terminating codon resulting in unstable mRNA. Loss of exon 4 results in an *in frame* mRNA which would be stable)
4	ACT TT$_{374}$A ACT	TTA → TGA (leu → STOP)	± Loss of exon 4 (as above)
4	TTA A(C$_{377}$)T GGA	Δ C$_{377}$	Loss of exon 4–8
6	GTC GC$_{482}$A AGg	GCA → GAA (ala → glu)	Loss of exon 6
6	AG$_{485}$gtatgt	G → A (AGC → AAC) (ser → asn) or G → T (AGC → ATC) (ser → ile)	Loss of exon 6

Table 16
Single base pair changes in exon sequence that affect splicing (Exon 8)

Exon	Position	Change	Splice Effect
8	GTT G_{538}GA TTT	GGA → TGA (gly → STOP)	± Loss of exon 8
		orGGA → AGA (gly → arg)	± Loss of exon 8
8	GTT GG_{539}A TTT	GGA → GTA (gly → val)	Loss of exon 8
8	GTT G_{544}AA ATT	GAA → AAA (glu → lys)	Loss of exon 8
		orGAA → TAA (glu → STOP)	± Loss of exon 8
8	ATT CC_{551}A GAC	CCA → CTA (pro → leu)	± Loss of exon 8
8	CTT G_{580}AC TAT	GAC → TAC (asp → tyr)	Loss of exon 8
8	AAT G_{589}AA TAC	GAA → AAA (glu → lys)	± Loss of exon 8
		orGAA → TAA (glu → STOP)	± Loss of exon 8
8	AAT GA_{590}A TAC	GAA → GTA (glu → val)	± Loss of exon 8
8	TAC TTC_{597} AGG	TTC → TTT (asp → val)	± Loss of exon 8
8	AGG GA_{602}T TTG	GAT → GTT (asp → val)	± Loss of exon 8

line 8). However, the reported cDNA is missing exons 4–8 (also in frame) and not the simple exon 4 exclusion.

Exon 6 contains 2 bases which, when mutated, result in exon 6 exclusion. Both bases are near the exon 6 splice donor site (Table 15, lines 9 and 10). The mutation at 482C does not reside within the conventional splice site, but the effect on splicing implies that this 'weak' 5′ site relies on this base as well. Changing 482C to A does not appear to create a new donor or a hairpin which might be hypothesized to effect splicing. The splice donor sequence is not very close to the consensus sequence with a pyrimidine t at IVS6 + 4 in place of the more common purine (a or g). G is the consensus base for the last exonic base in a donor site and 485G is the last base of exon 6. G_{485} → A or T splice mutations have been reported 3 times each and so this is not a rare event. Both mutations change the R_i value from 9.5 to 6.4 bits which is a significant decrease and the Senapathy score drops from 88.7 to 75.4.

Exon 8 is unique in that 22 exon mutations at 9 different internal (nonsplice site) bases have been reported to affect splicing, resulting in mixed RNA

Table 17
Use of cryptic splice sequences

(A) Splice Donor Site Mutations (Intron Inclusion)		
Intron 1	(Exon 1) GTGgtgagc	Include IVS1 + 1 to + 49 (cag_{1752}|gtggcg)
Intron 5	(Exon 5) GAAgtaagt	Include IVS5 + 1 to + 67 (aag_{31701}|gtaagc)
(B) Splice Acceptor Site Mutations (Partial Exon Exclusion)		
Intron 1	cagA (Exon 2)	Exclude Exon 2, b 28–32 ($ATTAG_{32}$|T)
Intron 3	tagA (exon 4)	Exclude Exon 4, b 319–327 (CAG_{327}|T)
Intron 5	aagG (Exon 6)	Exclude Exon 6, b 403 (aag_{403}|A)
Intron 7	tagT (Exon 8)	Exclude Exon 8, b 533–553 (CAG_{553}|A)
Intron 8	tagC (Exon 9)	Exclude Exon 9, b 610–657 (TAG_{626}|T)
(C) Exon Mutations Creating New Donor Sites		
Exon 3	GG_{209}G GGC → GT_{209}G GGC	Exclude Exon 3, b 208–318 (AAG_{207}|GTGGGC)
(D) Intron Mutations Creating New Acceptor Sites		
Intron 8	tg_{-16}g → ta_{-16}g	Inclusion of Intron 8, IVS8-1 to -14 (tag_{-41439}|a)
Intron 8	tat_{-3} → tag_{-3}	Inclusion of Intron 8, IVS8-1 and -2 (tag_{-41451}|a)

species (± exon 8) of variable percentages of full length and exon 8 exclusion (Table 16). Five mutations each at 538G, 544G and 551C have been reported. Steingrimsdottir et al. [79] have presented a model of exon 8 secondary structure to explain this unusual splicing sensitivity. It is possible that other exon 8 mutations also affect splicing but have not been reported because a full length cDNA product was also obtained and only it was noted.

In summary, mutations in exon sequences have been reported to affect splicing in exons 1, 3, 4, 6 and 8. In exon 3, a new splice donor site can be created by a single base change and, in exons 1 and 6, mutation of the exon base G adjacent to the splice donor sequence blocks splicing.

3.3. Use of cryptic or created splice sequences

Table 17 summarizes the mutation sites which result in the use of cryptic splice sequences or in the creation of new splice sites. The use of existing, cryptic donor sequences has been found in introns 1 and 5, when the normal sequence is mutated. In both instances, mutations at donor site bases +1, +2 or +5 have been reported (Table 17A). Mutations in five different acceptor splice sequences result in partial exon exclusion (Table 17B). A mutation in exon 3 creates a new donor splice site (Table 17C) and 2 different mutations in intron 8 create *new* exon 9 acceptor splice sites (Table 17D).

3.4. Pleiotropic effects of splice sequence mutations

It is becoming clear that multiple cDNA products are often found in RT/PCR studies of *HPRT* mutations. In most investigations, these products have not been characterized and it is not clear that these are not simply 'artifacts' of the methodology. Evidence is mounting that splice sequence mutations often yield characteristic cDNA multiple product patterns and these patterns might point to the most appropriate exon for genomic DNA studies. Table 18 lists the multiple products reported as a result of mutations in splice recognition sequences for several introns. There are probably other examples that have not been reported since most investigators have not sequenced or reported multiple products.

3.5. Other splice mutation events in the HPRT gene

The above mutations all occurred in the broadly defined consensus sites (or in the exons) and led either to exon skipping or use of a nearby cryptic site; however, a number of *HPRT* mutations affecting splicing occur outside of the consensus splice sites. These mutations are of several types: (1) creation of a nearby cryptic 5′ or 3′ splice site that is better than the normal site, (2) deletion of the normal and most often used cryptic site so a secondary cryptic site is utilized, (3) production of a new exon by creation of a new 5′ or 3′ splice site which utilizes a preexisting consensus 3′ or 5′ splice site, respectively, and (4) creation of an ag which suppresses splicing. These unique mutations will be summarized below.

A 14 bp deletion (Δ) at the exon 6 5′ border (ΔIVS5-7 to exon 6, 409A) deletes most of the exon 6 acceptor site (mutant-HS2 [80]); (note: the description of the mutant here differs slightly from the original paper, A. Grosovsky, personal communication). A new cryptic acceptor is utilized 30 bp into exon 6 at bp 432 [normal sequence—ttttttgaaaag-GATATAATTGACATGGCAAAACAATGCAGA-C, mutant sequence—ttttTTGACATGGCAAAACA-

Table 18
Pleiotropic effect of splice sequence mutations

Site of Mutation	Effect in cDNA
Intron 1 3′	(1) Exclusion of Exon 2, 28–32 (2) Exclusion of Exon 2 (3) Exclusion of Exons 2 + 3
Intron 2 3′	(1) Exclusion of Exon 3 (2) Exclusion of Exons 2 + 3
Intron 3 5′	(1) Exclusion of Exon 3 (2) Exclusion of Exons 2 + 3
Intron 5 5′	(1) Inclusion of Intron 5, IVS5 + 1 to +67 (2) Exclusion of Exons 4 + 5 (3) Exclusion of Exons 2 to 6
Intron 7 3′	(1) Exclusion of Exon 8, 533–553(2) Exclusion of Exon 8

ATGCAGAC which is spliced as ttttttgacatggcaaaacaatgcag|AC]. The Senapathy score for this new cryptic site is only 63.9 and the R_i value is negative. There is the 10 basepair stretch of pyrimidines (shown) just upstream of the scored region which must strengthen the site and/or allow use of the mechanism utilized to splice alternative splice sites. This deletion is believed to have occurred because a repeated TTGA flanks the site.

Several unusual deletions occur 5′ to exon 9 with related consequences. These are: ΔIVS8-11 to -2, ΔIVS8-16 to E9:670, ΔIVS8-21 to E9:613 and ΔIVS8-5 to E9:616 (FTB33 [81]; Case 5 [82]; HX305 [83,84]). Mutations in the exon 9 acceptor usually result in use of the cryptic site 17 bases into exon 9 at bp 627 (first bp of 'new exon 9'). This cryptic site has an Senapathy score of 72.0 and an R_i value of 2.7 bits (Table 19). The ΔIVS8-16 to E9:670 mutation deletes this cryptic site at 627. Splicing moves to an alternative cryptic site at IVS8-42. This site is listed in Table 4 (putative cryptic sites) and has a Senapathy score of 91.2 and a R_i value of 12.2 bits. Deletions of IVS8-21 to E9:613 and IVS8-11 to -2 also result in the use of this alternate cryptic site at IVS8-42, although for both the original cryptic site at bp 627 is still present; however, in both cases the IVS8-42 site is now closer and with its much higher score must now be favored. Lastly, the deletion of IVS8-5 to E9:616 uses the 627 acceptor site and not the IVS8-42 site; however, the deletion has removed part of the 627 site and replaced it with a 'better' stretch of four T's raising the Senapathy score to 76.8 and the R_i value to 2.8 bits. This score is apparently sufficient to cause it to be favored over the IVS8-42 site. Of note, also, is that one IVS8-2 A → G (of 8 in the database) is noted to have 2 mRNAs, the usual −17 bp of exon 9 (60%) and the +42 bp of IVS8 (40%) (PU29 [78]). It is likely that the other IVS88-2A mutations have low levels of the +42 bp product as well.

Three mutations [(lsmg38 [85]; I19-2A [86]; DDR4.7 [87]) result in a g → a transition mutation at IVS8-16 (IVS8-16 G → A) and lead to the use of a new created cryptic acceptor site at IVS8-14 [normal sequence—ctttttaatgtgaatttctggatttttttttatagT (mutated base is underlined), new sequence—ctttttaatgtgaatttctagatttttttttatagT which is spliced as ctttttaatgtgaattctagATTTTTTTTTTATAGT]. This new site has a Senapathy score of only 69.8 (R_i value of 3.5 bits) as compared to the normal site (which is still present and unmutated) score of 85.3 (R_i value of 8.9 bits) Why this new site is favored is unclear, except it may simply be closer to the branch site or is recognized first because it is 5′ of the natural site. It also has a stretch of 6 pyrimidines 5′ which may improve the site above what would otherwise be indicated.

A T → G mutation at IVS7-96 causes exon 8 loss. The mechanism for exon 8 loss is unclear, but this change does create a new ag dinucleotide and, thus, a possible new splice acceptor site. This 'cryptic' site has a Senapathy score of of only 60.8 and a negative R_i value. Also, use of this sequence as an acceptor site would create a 75 bp intron which is of borderline size.

A deletion of exon 4, bases 319–322 (319–322delAATC) (the first 4 bases of exon 4) causes the production of 3 mRNAs, exon 4 loss, −4 bp of 5′ exon 4 and −9 bp of 5′ exon 4 (Case 4 [82]). The first is exon loss, the second splicing at the usual

Table 19
Actual *HPRT* cryptic site sequences

Site	Distance	New site	Senapathy score	Normal site Senapathy score	R_i value in bits	Normal Site R_i value in bits
Exon 1 donor	+49 bp	AG|gtggcg	67.2	83.0	5.0	7.6
Exon 5 donor	+67 bp	AG|gtaagc	94.2**	87.6	11.0**	8.0
Exon 2 acceptor	+5 bp	tcttttcaaattag|TG*	75.4	95.8	4.6	9.7
Exon 4 acceptor	+9 bp	actataatgaccag|TC*	66.4	86.6	1.0	10.0
Exon 8 acceptor	+21 bp	atttgaaattccag|AC	79.3**	77.1	3.4**	2.8
Exon 9 acceptor	+17 bp	gtttgtgtcattag|TG	72.0	85.3	2.7	8.9

*Shows sequence and scores with the normal acceptor mutated.
**Cryptic site has a better Senapathy score/higher R_i value than normal site!

point except for the genomically deleted 4 bp and the last, selection of a new cryptic acceptor site created by the deletion [normal site—gttttttttttttttaactagAATGACCAGTCAACA, mutated site—gttttttttttttttaactag ACCAGTCAACA which is spliced as gttttttttttttttaactagaccagTCAACA]. The original exon 4 acceptor had a Senapathy score of 86.6 (R_i value of 10.0 bits), but with the 4 bp deletion this site still has a score of 86.6 (R_i value of 9.5 bits) while the new cryptic has a score of only 72.4 (R_i value of 3.6 bits). It is unclear why the mutated but original site is not used exclusively since its strength is essentially unchanged. While single base substitution mutations in the exon 4 acceptor usually result in exon loss, one IVS3-1G → T mutation is reported to give a 9 bp loss of 5′ exon 4 as in the above Δ319–322delAATC mutation [82]. None of the other g → t mutations or g → c or g → a mutations are reported to use this cryptic site. An interesting difference is the strengths of the cryptic site for three types of mutations at IVS3-1, i.e., g → t (72.4; R_i value of 1.0 bit), g → c (64.5; R_i value of 0.4 bit) and g → a (60.5; the R_i value is negative). This may explain why for the g → t mutation the cryptic site is used, but in the other two cases it is not.

Two mutations in intron 5 lead to creation of new exons spliced between exons 5 and 6. The first is a deletion of 4 base pairs, IVS5 + 1172–1175 or 32806–32,809 (M119 [88]). This creates an acceptor site at 32,814 [tttccttttattgccaaattgtatggacag-GTGTTC...] with a Senapathy score of 74.9 (R_i value of 1.5 bits). A donor site at IVS5 + 1300 (32,935) [TCgtaagt, Senapathy score 78.6 (R_i value of 5.9 bits)] is utilized to complete the 120 bp 'exon'. While the acceptor site has a low score there is a run of 10 pyrimidines 5′ to the acceptor site. The other mutation is an a → g at IVS5 + 1229 (32,863) (TL [89]). This results in a mixture of two mRNAs with different 'new exons'. The mutation creates a new donor site at 32,862 (TCgtaagt, Senapathy score of 78.7, R_i value of 5.9 bits) and utilizes two different acceptors. The 5′ one is at 32,747 (ttaatgttt-tcaagG, Senapathy score of 80.4, Ri value of 5.8 bits) and the second is at 32,815 at the same place as the first mutant but without the 4 bp deletion (tgtatg-gatagacagG, Senapathy score of 77.3, R_i value is negative) creating mRNAs with 116 bp and 48 bp 'new exons', respectively.

Recently a technique called 5′ RACE (**R**apid **A**mplification of **c**DNA **E**nds) has been used to look for the production of chimeric (fusion) mRNAs between *HPRT* and 3′ 'genes' in mutants that have deletions of the 3′ end of the *HPRT* gene [90]. These types of fusions are found in some oncogenes in cancers [91] and it was hypothesized that in reality these might be common events which could be found at *HPRT*. Seven of nineteen 3′ deletion mutants (36%) were found to make specific chimeric mRNAs, some into new 'genes' and some into *Alu* or LINE sequences. Several of the mutants revealed interesting spliced transcripts. Six of the seven mutants used the normal donor site 3′ of the last undeleted exon to splice to an 'acceptor' of an 'exon' from the other side of the deletion breakpoint. For the remaining mutant, the breakpoint actually occurred within exon 3 and a downstream sequence then became fused to exon 3. There were two exon 7–9 deletions; in one mutant exon 6 spliced to a LINE element downstream of the breakpoint while for the other mutant, apparently no acceptable acceptor existed and a cryptic polyadenylation site in intron 6 was used to complete the mRNA. This has interesting ramifications for the sequence of splicing

Table 20
Putative splicing enhancers in *HPRT*

Location	Exon	Sequence
Exonic		
14,840	exon 2	GAGGA
14,848	exon 2	GGAAAGGG
16,630	exon 3	GAAGGAGA
16,673	exon 3	AAGGGGGG
39,824	exon 7	GAAAAGGA
Intronic		
1874	171 bp 3′ exon 1	AGGGAGGA
1877	174 bp 3′ exon 1	GAGGA
15,034	148 bp 3′ exon 2	AAAAGAGGAGGG
15,035	149 bp 3′ exon 2	AAAGAGGA
15,038	152 bp 3′ exon 2	GAGGA
27,969	12 bp 3′ exon 4	AAAGGGAAGA-AAAAAAAG
31,478	139 bp 5′ exon 5	AAAGGAGAG
31,709	75 bp 3′ exon 5	AGGGAGAGAAA
34,753	184 bp 5′ exon 6	GGGGCTG
34,841	96 bp 5′ exon 6	GAAAAGAA
35,059	40 bp 3′ exon 6	GAAGGGGGA
40,163	54 bp 3′ exon 8	AGGAAAGAGAA

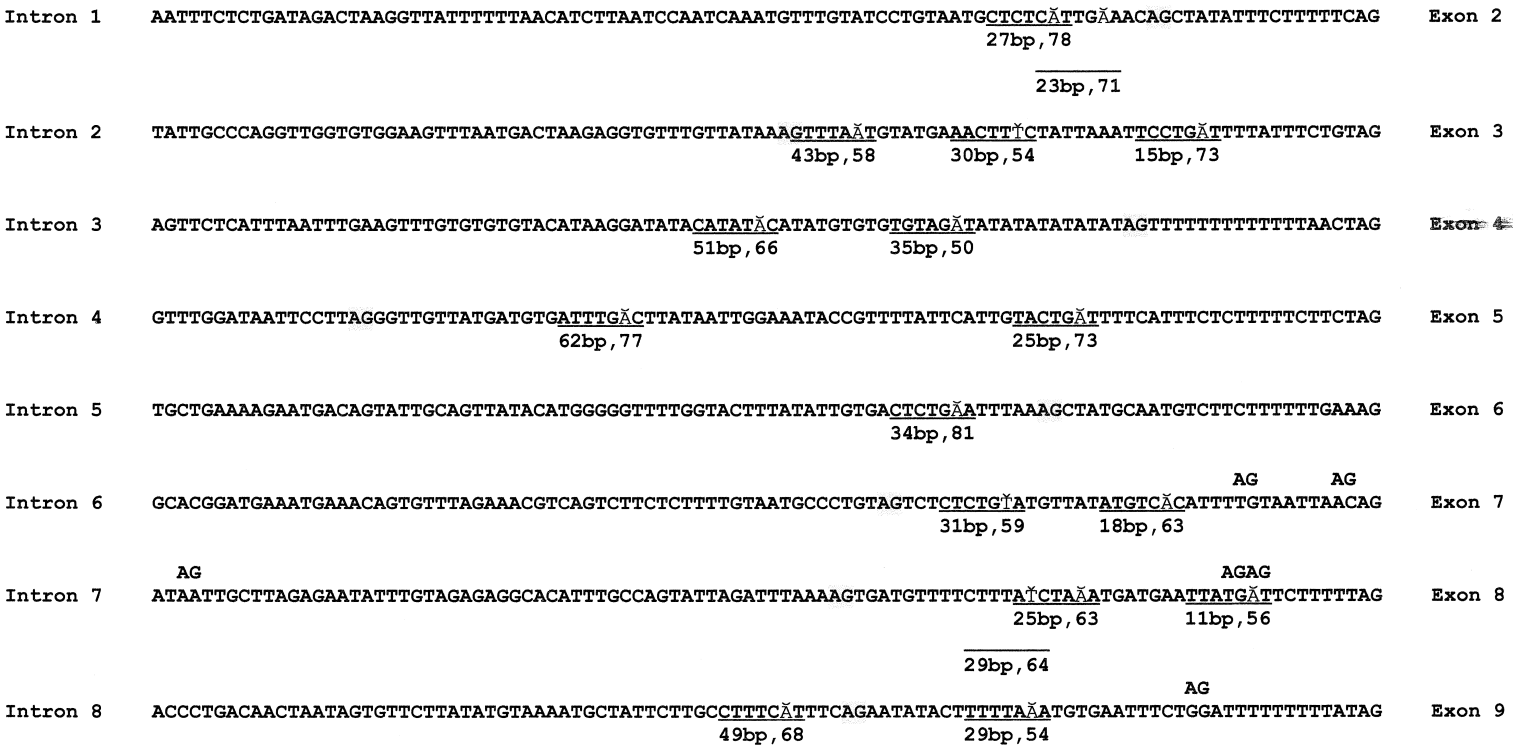

Fig. 2. Putative branch points are underlined with branch indicated as Ă or Ť. The numbers below each site indicate distance of the branch from the acceptor AG and Senapathy type score. The acceptor site is the AG at the right end of the sequence. The nearest AG upstream of that acceptor AG is shaded (the branch 'A' would be expected to be downstream of that shaded AG). For introns 1, 3, 5 and 8, the upstream AG is also shaded. Mutations which create a new AG are shown above the normal sequence.

and polyadenylation and recognition of the last exon. Splicing must occur first or otherwise *HPRT* mRNAs would always end at exon 6. However, if that intron 6 polyadenylation site is quickly spliced out, polyadenylation cannot occur. Only when a downstream acceptor is not found and no splicing occurs, does polyadenylation occur in intron 6. For one mutant, it could be shown that exon 8 was spliced to a 117 bp sequence located 10kb 3′ to exon 9 in the wildtype *HPRT* gene. There is an excellent acceptor on the 5′ end of the sequence (ttttttttttttttt-tgtagA, Senapathy score 88.2, R_i value of 16.4 bits) and a good donor on the 3′ end (TGgtaact, Senapathy score 77.9, R_i value of 6.0 bits) which spliced in turn to a downstream microsatellite repeat.

3.6. Other aspects of HPRT splice recognition

The *HPRT* gene was scanned for the enhancers shown in Table 1 as well as a generic eight purine stretch (most enhancers are purine-rich). Table 20 shows a partial listing of those results. Since most enhancers are exonic, all the exonic enhancer-like sequences are listed first. Sequences were found in exon 2, exon 3 and exon 7. In the bottom half of the table are shown enhancers-like sequences within 200 bp of the exons. In their original descriptions, the TGAATG [26] and GGGGCTG [27] enhancers were repeated; although these sequences were found in *HPRT* there were only single copies near any exon. There are no potential enhancer elements near exon 8 which appears to be the exon most susceptible to mutation resulting in splicing errors. Enhancers would also be expected near the 'too small', 18 basepair exon 5 and there are several nearby, although none is in exon 5 itself. Exon 5 itself is in fact fairly T rich (AATGTCTTGATTGTGGAA) and has no run of purines greater than 4 bases.

The branch sites for *HPRT* have not been experimentally determined; however, based on the consensus sequence (tnctray or ctray), educated conjectures as to the positions of branch points can be made. Fig. 2 gives the 100 bases upstream of each exon (the consensus acceptor ag dinucleotide is the last 2 bases shown). Putative branch points are underlined with their distance from the exon shown as well as a Senapathy type score [31] based on the frequencies of nucleotides at the different positions for the constitutive branch sites in Table 2. Good putative branch sites are found for exon 2, exon 5, exon 6 and possibly exon 3 (this site is only 15 bases away, however). Only poor sites are found for exons 7 and 9 (scores < 70). The exon 4 branch site is especially unusual. There is a pyrimidine stretch of 13 t's immediately preceded by an ag dinucleotide (ags are thought to be rare within 10–20 bases of an acceptor). Upstream of that are two stretches of 8–9 at dinucleotides with 4 gt dinucleotides in the middle; this could form a hairpin. The presence of the ag dinucleotide suggests that the branch should be very close upstream but this would place it in the hairpin in a low score region of ats. For introns 1, 3, 5 and 8, the ag upstream of the putative branch points is the second ag upstream of the acceptor ag. In intron 6, 7 and 8, mutations that create new ag dinucleotides either cause exon exclusion (intron 6 and 7) or create a new splice acceptor site resulting in intron inclusion (intron 8).

4. Discussion

Several conclusions can be drawn from study of *HPRT* splicing mutations. The first is that the *HPRT* splice sites often show large deviations from the consensus sequences especially in the first and last bases of the exons. This probably reflects the lower information content contributed by these bases. According to the consensus sequence, the first and last bases of the exon should be G. However, only exons 3 and 6 have a G as the first exonic base (3′ splice site) and for both of these, mutations at the G are missense mutations, with no reported effect on splicing. In the donor splice sites, only 4 of the 8 sites (exons 1, 2, 4, 6) have a consensus G as the last base of the exon. Exon 1 has one reported mutation at the last base G which because it is 3rd position wobble causes no amino acid change (Val → Val) but does affect splicing. Exon 2 has missense mutations at that site that do not affect splicing; this is not surprising as the Senapathy score remains a high 86.7 (R_i value of 9.7 or 9.4) even with a mutation at the G. No mutations, either splice mutations or missense, have been reported at 384G (exon 4). (The exon 4 G is a partial wobble, a 384G → C or T would change lysine to asparagine but 384G → A

remains lysine). Mutation at the last base of exon 6 cause splicing changes; of note, there is also a report that mutation three bases into the exon also causes exon 6 loss. Thus, exon 6 is unique in that several bases at the 3′ end of the exon appears critical for splicing. Table 21 summarizes the Senapathy scores and R_i values for observed mutations (splice and non-splice) at the exonic consensus G's. While all mutations reduce scores, there are some 'interesting' observations. The exon 1 donor G → A mutation reduces the Senapathy score to 70.6 and a R_i value of 4.6 bits, while the cryptic has a Senapathy score of 67.2 and a R_i value of 5.0 bits. The Senapathy score would predict that splicing would not be affected while information theory would predict that the cryptic site would be utilized which is what is observed. Thus, information theory more accurately predicts the result here. Of special note, the scores and mutations are identical for the exon 4 and exon 6 donors, but only for the exon 6 donor has a splice mutation been observed at the exonic G. The reason for this is unclear but may be due to exon 4's slightly better acceptor or to enhancer sequences or simply to chance.

Fig. 3 summarizes all the simple base changes in the consensus donor and acceptor sites. The relative frequencies at the various bases should be taken lightly as these mutations include both spontaneous and mutagen induced; i.e., some mutagens may have 'hotspots' that will skew the frequencies. However, from this figure, it can be seen at which bases mutation affects splicing and which probably do not. Obviously, saturation has not been reached as one invariant acceptor a (exon 3; IVS2-2A), one invariant acceptor a (exon 6; IVS5-2A) and one invariant donor t (exon 3; IVS3 + 2G) have not been found mutated. Considering only the 'spontaneous' mutations, the frequencies of base changes at the four invariant sites (2 G:C basepairs and 2 A:T basepairs, 45 mutations) are: 7% G:C → C:G, 44% G:C → A:T, 11% G:C → T:A, 15% A:T → T:A, 4% A:T → C:G, and 18% A:T → G:C.

We found that in our 201 simple splice site mutations that 56% were in the donor sites and 44% in the acceptor sites. Other compilations of splice sites have found much higher frequencies of donor mutations, e.g., 79% [39], 68% [16], and 70% [63]. The high frequency of acceptor site mutations at *HPRT* is predominantly due to the high number of exon 8 and exon 9 acceptor mutations (22 and 21 mutations, respectively).

The information theory model predicts that acceptors are recognized first and that cryptic acceptors either preexisting (which are only used if the normal site is mutated) or those created by mutation can be either upstream or downstream of the the normal

Table 21
Mutations in exonic consensus G's (last and first base of the exon)

Site	Normal	Senapathy	R_i value in bits	Mutation *	Senapathy	R_i value in bits	Results
Ex 1 donor	TG gtgagc	83.0	7.6	TA gtgagc	70.6	4.6	no effect on splicing [no use of cryptic score 67.2, 5.0 bits]
				TT gtgagc	70.6	4.3	no effect on splicing [no use of cryptic score 67.2, 5.0 bits]
Ex 2 donor	AG gtaagt	100.0	12.8	AA gtaagt	86.7	9.7	no effect on splicing
				AT gtaagt	86.7	9.4	no effect on splicing
Ex 4 donor	AG gtatgt	88.7	9.5	AA gtatgt	75.4	6.4	no effect on splicing
				AT gtatgt	75.4	6.4	no effect on splicing
Ex 6 donor	AG gtatgt	88.7	9.5	AT gtatgt**	75.4	6.4	ex 6 skip
				AA gtatgt**	75.4	6.4	ex 6 skip
Ex 3 acceptor	ttttatttctgtag G	89.8	11.3	ttttatttctgtag A	86.4	10.3	no effect on splicing
Ex 6 acceptor	ttctttttgaaag G	86.4	7.1	ttctttttgaaag A	83.0	6.2	no effect on splicing
				ttctttttgaaag T	80.1	4.9	no effect on splicing
				ttctttttgaaag C	80.6	5.2	no effect on splicing

*Mutated base is underlined and change is shown.
**The only mutations which do affect splicing.

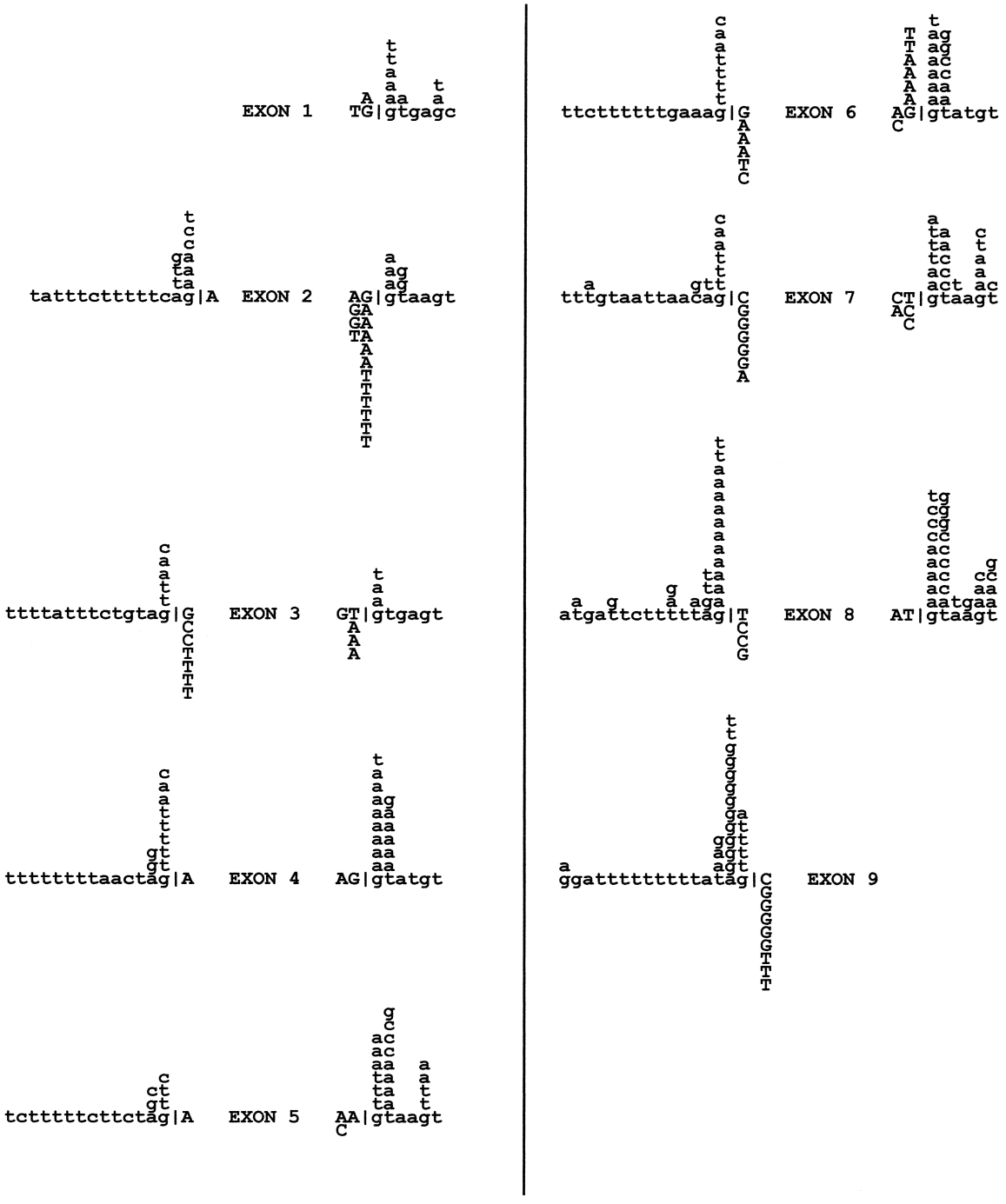

Fig. 3. The wild type splice site junctions are given. Those letters above the main line indicate mutations to that designated base from the base listed on the main line at that position which cause splicing errors. Letters below the main line indicate missense/nonsense mutations (i.e., these do not cause splicing alterations).

site. Scanning then proceeds downstream through the exon to find a donor site. Preexisting cryptic donors used after mutation of the normal site must thus be downstream of the normal site while newly created donors must be upstream of the normal donor site. Human *HPRT* mutations support this model. The preexisting cryptic donor sites (exons 1 and 5) are both downstream of the normal donor while the created donor site (exon 3) is upstream. *HPRT* cryptic acceptors are both upstream and downstream although all of the preexisting cryptics are downstream and the newly created cryptics are upstream. This may be coincidence as both preexisting and created cryptic acceptors occur both upstream and downstream in other genes [39].

All 8 intron 3′ splice sequences have shown mutations in the invariant base IVS-1G and 6 of 8 (exceptions are introns 2 and 5) in the invariant base IVS-2A. Three introns (6, 7 and 8) show mutations in base IVS-3 also. Two introns have the more prevalent c at IVS-3 (introns 1 and 6). Mutation has been observed only at the intron 6 site. The intron 7 and 8 IVS-3 mutations change the lower consensus t to an a or g. No mutations were found in base IVS-4 as expected from the consensus sequence hypothesis of no specific base preference at this site. In exons 3 and 6, which have a G as the first base, g to a mutations have been reported in IVS-1 (exon 3, tagG → taaG and exon 6, aagG → aaaG). However, this potential new splice acceptor site is used only in exon 6 and not in exon 3.

The polypyrimidine tract residing in bases −5 to −23 (shown in Fig. 2) is fairly variable among the 8 introns. Introns range from 94.7% polypyrimidine (intron 4) to 57.9% polypyrimidine (introns 6 and 7) in this 19 base stretch. As discussed in the introduction, ag's are rare in this region as presumably they become putative acceptors. Only introns 1 and 3 normally contain an ag dinucleotide in this stretch (at IVS2-18 and IVS3-20, respectively); possibly these are too close to the branch 'a' to be utilized. The creation of an ag dinucleotide within the normal splice site causes splicing errors in introns 6, 7 and 8. In introns 6 and 7, the exon is excluded, while only in intron 8 is this new ag sequence used as a splice acceptor site. (See Table 22 for a summary of Senapathy splice site scores and R_i values.) Both IVS8-3T → G and IVS8-16G → A cause utilization of the new ag as a new cryptic acceptor site. The IVS8-3 mutation yields reasonable new R_i values although the original site is still better. Use of the IVS-14 site is surprising based on its poor values; however, if exon 9 were skipped, no mRNA would have been detected; we may be detecting only a minor product which utilizes the cryptic site. That the other mutant ag's in introns 6 and 7 are not used as acceptors is not surprising based on their values. Somehow these new ag dinucleotides in the other introns must interfere with splicing so the whole exon is skipped. The other 5 introns contain at least 16 possible sites to create a new ag dinucleotide, while introns 6, 7 and 8 contain at least 11 other sites. The sensitivity of these sequences to mutation may be determined as additional splice mutations are defined.

For exon 8, mutations also occur at IVS7-5T → G or A and IVS7-3T → A (Table 13). The scores for the exon 8 acceptor site are by far the lowest of all the *HPRT* acceptor sites. Loss of either of these t's lowers the values considerably which appears to be sufficient to cause bypass of the exon. The two other mutations mentioned above at IVS7-10 and IVS7-13 also cause loss of t's and would reduce values similarly (see Table 22), thus, they both create ag dinucleotides and reduce values.

The intron 5′ splice site sequences all fit a gt(a/g)agt consensus sequence, except for intron 1 (gtgagc), intron 4 (gtatgt) and intron 6 (gtatgt). Mutations in all 8 intron sequences have been found in the invariant base +1 (g; IVS + 1G) and in 7 of the 8 introns in the invariant base +2 (t; IVS + 2T). Mutations in intron 1, 5, 7 and 8 at base +5 (g; IVS + 5G) have also been found. Three splice sites have mutations of their a at IVS + 3 and 2 have mutations of their t at IVS + 6. Only exon 8 has a mutation at IVS + 4 at its a although 5 other exons have a at this position. Tables 23 and 24 show the mutations that have and that have not been observed, respectively, in the IVS + 3 to IVS + 6 positions. In general, the IVS bases where mutation has been observed to occur create sites with lower values than what would occur at the bases where no mutation has been seen. For example, changes in IVS1 would still have very good scores and are not observed as might be expected. However, IVS2 changes would lower values significantly but have not been ob-

Table 22
Intronic acceptor splice site mutations not at the invariant ag

Site	Normal Sequence	Senapathy Score	R_i value in bits	Mutation*	Senapathy Score	R_i value in bits	Splicing Result
ex 7	tttgtaattaacag C	81.1	4.6	tttgtaattaagag C	69.3	< 0	ex 7 loss, does not use new ag (score 76.5, R_i 1.3 bits)
				ttagtaattaatag C	74.1	2.0	ex 7 loss, does not use new ag (score 69.9, R_i 1.3 bits)
ex 8	atgattctttttag T	77.1	2.8	aggattctttttag T	71.4	0.3	ex 8 loss, does not use new ag (score 54.2, $R_i < 0$)
				atgagtctttttag T	71.1	0.6	ex 8 loss, does not use new ag (score 57.3, $R_i < 0$)
				atgattcttgttag T	73.7	< 0	ex 8 loss
				atgattcttattag T	74.1	< 0	ex 8 loss
				atgattcttttaag T	74.1	0.2	ex 8 loss
				gaaatggcttatag TT > G at IVS7-96	NC	NC	ex 8 loss, does not use new ag (score 60.8, $R_i < 0$)
ex 9	attttttttatag C	85.3	8.9	attttttttag C	81.9	4.5	splicing moves to new ag (score 81.5, R_i 7.4 bits)
				atttttttaaag C	82.2	6.3	Surprisingly, splicing moves to usual cryptic (score 72.0, R_i 2.7 bits)
				agatttttttttatag C	85.3	8.9	Surprisingly, splicing moves to new ag (score 69.8, R_i 3.5 bits)

*Mutated base is underlined and change is shown. If an ag is formed by the mutation, it is shaded.

Table 23
Donor sites IVSX + 3 to IVSX + 6 where mutations have been observed

Site	Normal Sequence	Senapathy Score	R_i value in bits	Mutation*	Senapathy Score	R_i value in bits	Splicing Result
ex 1	TG gtgagc	83.0	7.6	TG gtgaac	69.7	4.1	Uses cryptic at +49 [scores 67.2 (5.0 bits)]
				TG gtgatc	69.7	3.7	Uses cryptic at +49 [scores 67.2 (5.0 bits)]
ex 5	AA gtaagt	87.6	8.0	AA gtaaat	73.2	4.5	Uses cryptic at +67 [scores 94.2, 11.0 bits]
				AA gtaatt	73.2	4.1	exon 4–5 exclusion And use of cryptic
ex 7	CT gtaagt	79.0	7.6	CT gtaact	64.8	3.7	ex 7 loss
				CT gtaatt	64.6	3.7	ex 7 loss
				CT gtaaat	64.6	4.1	ex 7 loss
				CT gtaagc	73.2	6.2	ex 7 loss
				CT gttagt	69.0	3.1	ex 7 loss
ex 8	AT gtaagt	87.2	9.2	AT gttagt	77.2	4.8	ex 8 loss
				AT gtaggt	76.5	6.7	ex 8 loss
				AT gtaact	73.0	5.3	ex 8 loss
				AT gtaaat	72.8	5.7	ex 8 loss
				AT gtaagg	82.7	8.0	ex 8 loss
				AT gtaagc	81.4	7.8	ex 8 loss
				AT gtaaga	81.6	7.7	ex 8 loss

*Mutated base is underlined and change is shown.

Table 24
Donor sites IVSX + 3 to IVSX + 6 where mutations have not been observed to date

Site	Normal	Senapathy Score	Mutation site (underlined)	Senapathy Score with worst base at site
ex 1	TG gtgagc	83.0	TG gtgagc	76.3
			TG gtgagc	71.5
			TG gtgagc	**
ex 2	AG gtaagt	100.0	AG gtaagt	90.0
			AG gtaagt	88.5
			AG gtaagt	85.6
			AG gtaagt	94.2
ex 3	GT gtgagt	76.1	GT gtgagt	69.3
			GT gtgagt	64.6
			GT gtgagt	61.7
			GT gtgagt	70.3
ex 4	AG gtatgt	88.7	AG gtatgt	78.7
			AG gtatgt	**
			AG gtatgt	74.3
			AG gtatgt	82.9
ex 5	AA gtaagt	87.6	AA gtaagt	77.6
			AA gtaagt	76.1
			AA gtaagt	81.8
ex 6	AG gtatgt	88.7	AG gtatgt	78.7
			AG gtatgt	**
			AG gtatgt	74.3
			AG gtatgt	82.9
ex 7	CT gtaagt	79.0	CT gtaagt	67.5

** All mutations at that site improve or do not change the score—therefore, it would not be expected to find mutations.

served (see further discussion on exon 3 splice site mutations below). Of note also is that for mutations at IV1: + 5, the values are above that of the cryptic but the cryptic is utilized. Two observations should be noted: first, that mutation at some bases bases where a non-consensus nucleotide is present can actually improve values and second, that just because a mutation has not yet been observed at a position does not mean that one will not be seen in the future.

Exon 3 is especially intriguing in that only mutations at the invariant + 1g are known; this is surprising in light of the fact that its Senapathy score and R_i value is so low. While this paper summarizes 277 splicing mutants; there are an additional 341 exon exclusion mutations in the database where the genomic mutation has not been determined. These include 26 (7.6%) exon 3 exclusion mutations which could contain missing IVS3 + 2–5 mutations (some IVS3 + 1 mutations cause exon 2–3 exclusion), this percentage is still lower than expected. For example, there are 27 exon 2 exclusion and 58 exon 4 exclusion mutants in the uncharacterized data set.

Mutations have been observed at all 6 intronic donor positions for exon 8. This is interesting in that, unlike its acceptor site, this donor site is quite strong (Senapathy score 87.2, R_i value of 14.7 bits). Exon 8 is also unique in that there are a number of missense mutations within the exon that cause either exon 8 loss or mixed cDNAs of exon 8 loss and full length. Steingrimsdottir et al. [79] hypothesized a hairpin structure for exon 8 and that these exonic missense mutations disrupted this hairpin. However, the placement of the mutated bases from all the database mutations does not clearly support that model, although the unusual mutations observed in and around exon 8 do suggest that DNA secondary structure may be important for splicing. Some additional unusual mutations have also been found in this exon. A deletion of IVS7-45 to -22 causes loss of exon 8 from the mRNA. This is outside of the

conventional consensus acceptor and it does not create an ag. A possible explanation could be that it removes the branch site. The mutations ΔIVS7-10 to -8 (Senapathy score of 72.8, R_i value of 1.1 bits) and ΔIVS7-9 to -8 (Senapathy score of 77.6, R_i values of 1.7 bits) also cause exon loss. The Senapathy score for the latter is above those of that normal site although the information theory value is less and would predict no splicing, as observed, as the value is below the 2.36 bit cutoff.

Other exons do have internal mutations that at first glance appear to affect splicing. However, these are either the exon 3 mutation at 209G → T that creates a new donor site, or nonsense mutations. As discussed in the introduction, nonsense mutations reduce the stability of the full length mRNA and alternatively spliced mRNAs that are always present but at very low levels may be preferentially amplified. Often these can be mRNAs the in which the exon containing the nonsense codon is spliced out (if that exon loss is in frame) or other in frame exon losses. For *HPRT*, exons 2–3 loss appears to be a normal, low level mRNA seen with a number of different splicing mutations. This has been more carefully studied in the CHO *hprt* gene where it has been demonstrated that only nonsense mutants in the central exons cause reduction in full length message and the 'appearance' of exon excluded mRNAs (92–94). Nonsense mutations in the first 2 exons may allow reinitiation and nonsense mutations in the last few exons do not appear to affect mRNA stability [92]. It has also been shown that mRNAs in which exons have been lost are present at the same levels in nonsense mutants and wild type but are obscured by the high levels of full length product in wild type cells [93,94].

Exon 2–3 loss in cDNA is a common phenotype of human *HPRT* mutation for several reasons. First, genomic exon 2–3 deletions are common resulting both from V(D)J recombinase mediated events [77,78] and also because introns 1 and 3 are quite large and therefore quite large targets for deletion breakpoints. These cannot really be considered 'splicing' mutations. Secondly, some exon 2 acceptor and exon 3 donor or acceptor mutants give exon 2–3 loss phenotypes [exon 2 donor mutations give only exon 2 loss]. This, plus the fact that many mutants show a low level of exon 2–3 loss transcripts, suggests that exon 2 + exon 3 may be 'clustered'—i.e., spliced as a unit.

Because of its small size and the fact that 'better' exons could be spliced in the region, it might have been expected that exon 5 would require an enhancer to function properly. However, no mutations have been described which suggest the existence of an enhancer sequence (Table 20).

Also, surprisingly, no evidence for branch point mutations for any exon (except possibly a deletion of IVS7-45 to -22) was found. This suggests, that if the branch 'A' is mutated, a nearby alternative is utilized without affecting splicing. This implies that such alternatives must be common and may also explain why several exons do not have easily recognizable branch sites.

Another conclusion that can be drawn from the *HPRT* mutants is that the acceptor is longer than the 14 bases in the Senapathy consensus [31] as recently proposed by Stephens and Schneider [35]. This becomes clear from analysis of some of the cryptic sites utilized. As discussed in the results section, several of these cryptic acceptors have quite poor Senapathy scores and R_i values, yet they are spliced. They all have in common, however, a stretch of 6 or more pyrimidines 5′ of the usual acceptor site. As discussed in the introduction, alternative splice sites appear to have a pyrimidine stretch upstream which is associated with the branchpoint; perhaps some of the unusual *HPRT* sites are spliced using an 'alternative site' mechanism. There is currently no consensus or information theory models to predict such sites. The exon 8 acceptor also has a stretch of 8 pyrimidines (ttttcttt) at IVS7-31 to -37 which may strengthen the site. Since exon 8 has no easily identifiable branch site; it might require this pyrimidine stretch.

The above underscores the limitiations of consensus and information-based models of splice sites. While the R_i value does appear to explain several *HPRT* splicing situations better than the Senapathy score, there are still instances of very low R_i sites being spliced (i.e., those < 2.36). This implies the involvement of other factors (enhancers?) which the model does not, as of yet at least, include. However, lastly, in defense of the models, there is the possibility that descriptions of some mutants may be incorrect leading to misplaced blame on the models.

The CHO *hprt* gene is also used for mutagenicity studies and a number of splice mutants have been identified [92–98]. The genes have quite similar intron/exon structure and are 95% identical in the coding region. Comparison of CHO and human mutants finds a number of similarities but also interesting differences. There appears to be an exon 2–3–4 clustering in CHO cells, as well as the exon 2–3 clustering seen in human mutations. The base 209G → T mutation creating a donor site in exon 3 is also seen in CHO mutants. The internal exon 8 mutations that affect splicing in human *HPRT* are not present in CHO mutations but internal exon 2 mutations affecting splicing are seen in CHO. A hairpin is postulated where mutations that strengthen the hairpin, decrease correct splicing and increase alternative mRNA products [96]. Internal exon 4 mutations have also been reported to affect splicing [97]. Interestingly, a dearth of exon 3 donor mutations is reported [98] which is similar to the low number of such mutations seen in the human gene.

In conclusion, the large number of *HPRT* splicing mutants has allowed insight into the relative importance of various splice site bases to spliceosome action. The unusual features of mutations affecting exon 8 splicing leads to a hypothesized secondary structure, while we still have no explanation for how exon 5, with its small size, is still spliced. The large number of splicing mutants described for which the genomic mutation has not been identified are a wealth of future material. The knowledge of *HPRT* gene sequence and structure combined with the splicing mutations present a valuable model for future biochemical and molecular studies of splicing, which heretofore have used nearly exclusively in vitro constructs. Such studies could include: splice mutation reversion assays, spliceosome protein binding assays and branch site determinations.

Acknowledgements

The research of the authors is supported by an American Cancer Society grant ACS CG-45 to J.A.N. and NIH R01 CA 68288 to J.P.O'N. This work was also supported, in whole or in part, by a grant from The Merck Genome Research Institute, to P.K.R. We are grateful to Marcia Dalton and Inge Gobel for their excellent and patient word processing, to Dr. David Dummit for calculating the Senapathy scores and to Dr. Carrie Valentine for an extensive and thought-provoking evaluation of this review.

Appendix A. Publications which comprise the *HPRT* database (release 6)

[1] Andersson, B., S. Fält, and B. Lambert (1992), Strand specificity for mutations induced by (+)-anti bpDE in the *hprt* gene in human T-lymphocytes, Mutation Res., 269, 129–140.

[2] Andersson, B., S.-M. Hou, and B. Lambert (1992), Mutations causing defective splicing in the human *hprt* gene, Environ. Molec. Mutagenesis, 20, 89–95.

[3] Bastlova, T., B. Andersson, B. Lambert, and A. Kilman (1993), Molecular analysis of ethylene oxide induced mutations at the *hprt* locus in human diploid fibroblasts, Mutation Res., 287, 283–292.

[4] Bhattacharyya, N.P., A. Ganesh, G. Phera, B. Richards, A. Skandalis, and M. Meuth (1995), Molecular analysis of mutations in mutator colorectal carcinoma cell lines, Hum. Molec. Genet., 4, 2057–2064.

[5] Bronstein, S.M., J.E. Cochrane, T.R. Craft, J.A. Swenberg, and T.R. Skopek (1991), Toxicity, Mutagenicity, and Mutational Spectra of *N*-Ethyl-*N*-nitrosourea in Human Cell Lines with Different DNA Repair Phenotypes, Cancer Res., 51, 5188–5197.

[6] Burkhart-Schultz, K., C.B. Thomas, C.L. Thompson, C.L. Strout, E. Brinson, and I.M. Jones (1993), Characterization of in vivo somatic mutations at the hypoxanthine phosphoribosyltransferase gene of a human control population, Environ. Health Perspect., 101, 68–74.

[7] Burkhart-Schultz, K.J., C.L. Thompson, and I.M. Jones (1996), Spectrum of somatic mutation at the hypoxanthine phosphoribosyltransferase (*hprt*) gene of healthy people, Carcinogenesis, 17, 1871–1883.

[8] Chiou, C.C., and J.L. Yang (1995), Mutagenicity and specific mutation spectrum induced by 8-methoxypsoralen plus a low dose of UVA in the *hprt* gene in diploid human fibroblasts, Carcinogenesis, 16, 1357–1362.

[9] Curry, J., G.T. Rowley, V. Saddi, D. Beare, J. Cole, and B.W. Glickman (1995), Determination of *hprt* mutant and mutation frequencies and the molecular characterization of human derived in vivo T-lymphocyte mutants, Environ. Molec. Mutagenesis, 25, 169–179.

[10] Denault, C.M., T.R. Skopek, and H.L. Liber (1993), The effects of hypoxia and cysteamine on X-ray mutagenesis in human cells, Rad. Res., 136, 271–279.

[11] Dorado, G., H. Steingrimsdottir, C.F. Arlett, and A.R. Lehmann (1991), Molecular analysis of ultraviolet induced mutations in a xeroderma pigmentosum cell line, J. Mol. Biol., 217, 217–222.

[12] Fujimori, S., T. Tagaya, N. Yamaoka, H. Saito, N. Kamatani, and I. Akaoka (1994), Direct evidence for a hot spot of germline mutation at *HPRT* locus, Adv. Exp. Med. Biol., 370, 679–682.

[13] Gibbs, R.A., P.N. Nguyen, A. Edwards, A.B. Civitello, and C.T. Caskey (1990), Multiplex DNA deletion detection and exon sequencing of the hypoxanthine phosphoribosyltransferase gene in Lesch–Nyhan families, Genomics, 7, 235–244.

[14] Giver, C.R., S.L. Nelson, and A.J. Grosovsky (1993), Spectrum of spontaneous *HPRT*$^{-}$ mutations in TK6 human lymphoblasts, Environ. Molec. Mutagenesis, 22, 138–146.

[15] Gordon, R.B., P.A. Dawson, D.G. Sculley, B.T. Emmerson, C.T. Caskey, and R.A. Gibbs (1991), The molecular characterisation of $HPRT_{chermside}$ and $HPRT_{coorparoo}$: two Lesch–Nyhan patients with reduced amounts of mRNA, Gene, 108, 299–304.

[16] Guillouf, C., A. Laquerbe, E. Moustacchi, and D. Papadopoulo (1993), Mutagenic processing of psoralen monoadducts differ in normal and Fanconi anemia cells, Mutagenesis, 8, 355–361.

[17] Hou, S.M., A.M. Steen, S. Falt, and B. Andersson (1993), Molecular spectrum of background mutation at the *hprt* locus in human T lymphocytes, Mutagenesis, 8, 43–49.

[18] Kagawa, Y., F. Yatagai, M. Suzuki, Y. Kase, A. Kobayashi, M. Hirano, T. Kato, M. Watanabe, and F. Hanaoka (1995), Analysis of mutations in the human HPRT gene induced by accelerated heavy-ion irradiation, J. Radiat. Res., 36, 185–195.

[19] Laquerbe, A., C. Guillout, E. Moustacchi, and D. Papadopoulo (1995), The mutagenic processing of psorlen photolesions leaves a highly specific signature at an endogenous human locus, J. Mol. Biol., 254, 38–49.

[20] Lichtenauer-Kaligis, E.G.R., J.C.P. Thijssen, H. den Dulk, P. van de Putte, M. Giphart-Gassler, and J.G. Tasseron-de Jong (1995), Spontaneous mutation spectrum in the *hprt* gene in human lymhoblastoid TK6 cells, Mutagenesis, 10, 137–143.

[21] Marcus, S., E. Christensen, and G. Malm (1993), Molecular analysis of the mutations in five unrelated patients with the Lesch–Nyhan syndrome, Hum. Mutat., 2, 473–477.

[22] Marcus, S., D. Hellgren, B. Lambert, S.P. F@llstr'm, and J. Wahlstr'm (1993), Duplication in the hypoxanthine phosphoribosyl-transferase gene caused by Alu–Alu recombination in a patient with Lesch Nyhan syndrome, Hum. Genet., 90, 477–482.

[23] Mattano, S.S., T.D. Palella, and B.S. Mitchell (1990), Mutations induced at the hypoxanthine guanine phosphoribosyltransferase locus of human T lymphocytes by perturbations of purine deoxyribonucleoside triphosphate pools, Cancer Res., 50, 4566–4571.

[24] McGregor, W.G., R.-H. Chen, L. Lukash, V.M. Maher, and J.J. McCormick (1991), Cell cycle-dependent strand bias for UV-induced mutations in the transcribed strand of excision repair-proficient human fibroblasts but not in repair-deficient cells, Molec. Cell. Biol., 11, 1927–1934.

[25] McGregor, W.G., V.M. Maher, and J.J. McCormick (1994), Kinds and locations of mutations induced in the hypoxanthine-guanine phosphoribosyl transferase gene of human T-lymphocytes by 1-nitrosopyrene, inducing those caused by V(D)J recombinase, Cancer Res., 54, 4207–4213.

[26] Monnat, R.J., Jr., T.A. Chiaverotti, A.F.M. Hackman, and G.A. Maresh (1992), Molecular structure and genetic stability of human hypoxantine phosphoribosyltransferase (hprt) gene duplications, Genomics, 13, 788–796.

[27] Morgenthaler, P.M.L., and D. Holzhauser (1995), Analysis of mutations induced by 1-amino-1-methyl-6-phenylimidazo[4,4-b]pyridine (PhIP) in human lymphoblastoid cells, Carcinogenesis, 16, 713–718.

[28] Nelson, S., C.R. Giver, and A.J. Grosovshy (1994), Spectrum of X-ray-induced mutations in the human *hprt* gene, Carcinogenesis, 15, 495–502.

[29] Papadopoulo, D., A. Laquerbe, C. Guillouf, and E. Moustacchi (1993), Molecular spectrum of mutations induced at the *hprt* locus by a cross linking agent in human cell lines with different repair capacities, Mutation Res., 294, 167–177.

[30] Recio, L., J. Cochrane, D. Simpson, T.R. Skopek, J.P. O'Neill, J.A. Nicklas, and R.J. Albertini (1990), DNA sequence analysis of in vivo *hprt* mutation in human T-lymphocytes, Mutagenesis, 5, 505–510.

[31] Recio, L., D. Simpson, J. Cochrane, H. Liber, and T.R. Skopek (1990), Molecular analysis of hprt mutants induced by 2-cyanoethylene oxide in human lymphoblastoid cells, Mutation Res., 242, 195–208.

[32] Rigaud, O., A. Laquerbe, and E. Moustacchi (1995), DNA sequence analysis of *HPRT* mutants induced in human lymphoblastoid cells adapted to ionizing radiation, Rad. Res., 144, 181–189.

[33] Rossi, A.M., A.D. Tates, A.A. van Zeeland, and H. Vrieling (1992), Molecular Analysis of Mutations Affecting *hprt* mRNA Splicing in Human T-lymphocytes in vivo, Environ. Molec. Mutagenesis, 19, 7–13.

[34] Sege-Peterson, L., J. Chambers, T. Page, O.W. Jones, and W.L. Nyhan (1992), Characterization of mutations in phenotypic variants of hypoxanthine phosphoribosyltransferase deficiency, Hum. Molec. Genet., 1, 427–442.

[35] Steingrimsdottir, H., G. Rowley, A. Waugh, D. Beare, I. Ceccherine, J. Cole, and A.R. Lehmann (1993), Molecular analysis of mutations in the *hprt* gene in circulating lymphocytes from normal and DNA repair deficient donors, Mutation Res., 294, 29–41.

[36] Vrieling, H., J.C.P. Thijssen, A.M. Rossi, F.J. van Dam, A.T. Natarajan, A.D. Tates, and A.A. van Zeeland (1992), Enhanced *hprt* mutant frequency but no significant difference in mutation spectrum between a smoking and a non-smoking human population, Carcinogenesis, 13, 1625–1631.

[37] Yamada, Y., H. Goto, K. Suzumori, A. Ritsuko, and N. Ogasawara (1992), Molecular analysis of five independent Japanese mutant genes responsible for hypoxanthine guanine phosphoribosyltransferase (*HPRT*) deficiency, Hum. Genet., 90, 379–384.

[38] Yamada, Y., H. Goto, S. Tamura, and N. Osasawara (1993), Molecular genetic study of a Japanese family with Lesch–Nyhan syndrome: a point mutation at the consensus region of RNA splicing (*hprt* Keio), Jpn. J. Human. Genet., 38, 413–419.

[39] Yang, J.L., M.C. Hu, and C.W. Wu (1991), Novel mutational spectrum induced by *N*-methyl-*N′*-nitro-*N*-nitrosoguanindine in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in diploid human fibroblasts, J. Mol. Biol., 221, 421–430.

[40] Yang, J.L., P.C. Lee, S.R. Lin, and J.G. Lin (1994), Comparison of mutation spectra induced by *N*-ethyl-*N*-nitrosourea in the *hprt* gene of Mer + and Mer − diploid fibroblasts, Carcinogenesis, 15, 939–945.

[41] Yang, J.L., J.G. Lin, M.C. Hu, and C.W. Wu (1993), Mutagenicity and mutational specificity of *N*-methyl-*N′*-nitro-*N*-nitrosguanidine in the *hprt* gene in G1-S and late S phase of diploid human fibroblasts, Cancer Res., 53, 2865–2873.

[42] Yang, S., J.G. Lin, C.C. Chiou, L.Y. Chen, and J.L. Yang (1994), Mutation specificity of 8-methoxypsoralen plus two doses of UVA irradiation in the *hprt* gene in diploid human fibroblasts, Carcinogenesis, 15, 201–207.

References

[1] A. Kramer, The structure and function of proteins involved in mammalian pre-mRNA splicing, Annu. Rev. Biochem. 65 (1996) 367–409.

[2] C. Will, R. Luhrmann, Protein functions in pre-mRNA splicing, Curr. Opin. Cell Biol. 9 (1997) 320–328.

[3] R. Reed, Initial splice-site recognition and pairing during pre-mRNA splicing, Curr. Opin. Genet. Dev. 6 (1996) 215–220.

[4] X.D. Fu, The superfamily of arginine/serine rich splicing factors, RNA 1 (1995) 663–680.

[5] M.D. Chiara, O. Gozani, M. Bennett, P. Champion-Arnaud, L. Palanfjian, R. Reed, Identification of proteins that interact with exon sequences, splice sites, and the branchpoint sequence during each stage of spliceosome assembly, Mol. Cell. Biol. 16 (1996) 3317–3326.

[6] G. Dreyfuss, M.J. Matunis, S. PiZol-Roma, C.G. Burd, hnRNP proteins and the biogenesis of mRNA, Annu. Rev. Biochem. 62 (1993) 289–321.

[7] X.-D. Fu, Specific commitment of different pre-mRNAs to splicing by single SR proteins, Nature 365 (1993) 82–85.

[8] S.M. Berget, Exon recognition in vertebrate splicing, J. Biol. Chem. 270 (1995) 2411–2414.

[9] J.-S. Sun, J.L. Manley, A novel U2–U6 snRNA structure is necessary for mammalian mRNA splicing, Genes Dev. 9 (1995) 843–854.
[10] E.J. Sontheimer, J.A. Steitz, The U5 and U6 small nuclear RNAs as active site components of the spliceosome, Science 262 (1993) 1989–1991.
[11] J.D. Crispino, B.J. Blencowe, P.A. Sharp, Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP, Science 265 (1994) 1866–1869.
[12] W.Y. Tarn, J.A. Steitz, Modulation of 5′ splice site choice in pre-messenger RNA by two distinct steps, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 2504–2508.
[13] J.F. Ca-ceres, S. Stamm, D.M. Helfman, A.R. Krainer, Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors, Science 265 (1994) 1706–1708.
[14] Q. Sun, A. Mayeda, R.K. Hampson, A.R. Krainer, F.M. Rottman, General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer, Genes Dev. 7 (1993) 2598–2608.
[15] R. Singh, J. Valc < rcel, M.R. Green, Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins, Science 268 (1995) 1173–1176.
[16] K. Nakai, H. Sakamoto, Construction of a novel database containing aberrant mutations of mammalian genes, Gene 141 (1994) 171–177.
[17] M. Talerico, S.M. Berget, Effect of 5′ splice site mutations on splicing of the preceding intron, Mol. Cell. Biol. 10 (1990) 6299–6305.
[18] S. Brunak, J. Engelbrecht, S. Knudsen, Prediction of human mRNA donor and acceptor sites from the DNA sequence, J. Mol. Biol. 220 (1991) 49–65.
[19] B.L. Robberson, G.J. Cote, S.M. Berget, Exon definition may facilitate splice site selection in RNAs with multiple exons, Mol. Cell. Biol. 10 (1990) 84–94.
[20] J.Y. Wu, T. Maniatis, Specific interactions between proteins implicated in splice site selection and regulated alternative splicing, Cell 75 (1993) 1061–1070.
[21] R. Reed, T. Maniatis, A role for exon sequences and splice-site proximity in splice-site selection, Cell 46 (1986) 681–690.
[22] A. Watakabe, K. Tanaka, Y. Shimura, The role of exon sequences in splice site selection, Genes Dev. 7 (1993) 407–418.
[23] R. Xu, J. Teng, T.A. Cooper, The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element, Mol. Cell. Biol. 13 (1993) 3660–3674.
[24] A. Laviguer, H. La Branche, A.R. Kornblihtt, B. Chabot, A splicing enhancer in the human fibronectin alternate exon ED1 exon interacts with SR proteins and stimulates U2 snRNP binding, Genes Dev. 7 (1993) 2405–2417.
[25] R.K. Hampson, L. La Follette, F.M. Rottman, Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences, Mol. Cell. Biol. 9 (1989) 1604–1610.
[26] G.S. Huh, R.O. Hynes, Regulation of alternative pre-mRNA splicing by a novel repeated hexanucleotide element, Genes Dev. 8 (1994) 1561–1574.
[27] T. Carlo, D.A. Sterner, S.M. Berget, An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon, RNA 2 (1996) 342–353.
[28] F. Del Gatto, M.-C. Gesnel, R. Breathnach, The exon sequence TAGG can inhibit splicing, Nucl. Acids Res. 24 (1996) 2017–2021.
[29] P.J. Furdon, R. Kole, The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing in vitro, Mol. Cell. Biol. 8 (1988) 860–866.
[30] K. Inoue, M. Ohno, Y. Shimura, Aspects of splice site selection in constitutive and alternative pre-mRNA splicing, Gene Express. 4 (1995) 177–182.
[31] P. Senapathy, M.B. Shapiro, N.L. Harris, Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project, Meth. Enzymol. 183 (1990) 252–278.
[32] N.L. Harris, P. Senapathy, Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis, Nucl. Acids Res. 18 (1990) 3015–3019.
[33] F.E. Penotti, Human pre-mRNA splicing signals, J. Theor. Biol. 150 (1991) 385–420.
[34] T.D. Schneider, Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences, Nucl. Acids Res. 25 (1997) 4408–4415.
[35] R.M. Stephens, T.D. Schneider, Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites, J. Mol. Biol. 24 (1992) 1124–1136.
[36] M.B. Shapiro, P. Senapathy, RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression, Nucl. Acids Res. 15 (1987) 7155–7174.
[37] B.M. Faux, T.D. Schneider, P.K. Rogan, Information analysis of human splice site mutations, Am. J. Hum. Genet. 59 (1996) A257.
[38] P.K. Rogan, T.D. Schneider, Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites, Hum. Mutat. 6 (1995) 74–76.
[39] P.K. Rogan, B.M. Faux, T.D. Schneider, Information analysis of human splice site mutations, Hum. Mut. 12 (1998) 153–171.
[40] R.F. Roscigno, M. Weiner, M.A. Garcia-Blanco, A mutational analysis of the polypyrimidine tract of introns, J. Biol. Chem. 268 (1993) 11222–11229.
[41] R. Reed, The organization of 3′ splice-site sequences in mammalian introns, Genes Dev. 3 (1989) 2113–2123.
[42] C.-S. Chu, B.C. Trapnell, S. Curristin, G.R. Cutting, R.G. Crystal, Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA, Nature Genet. 3 (1993) 151–156.
[43] K.K. Nelson, M.R. Green, Mammalian U2 snRNP has a sequence-specific RNA-binding activity, Genes Dev. 3 (1989) 1562–1571.
[44] R. Reed, T. Maniatis, Intron sequences involved in lariat formation during pre-mRNA splicing, Cell 41 (1985) 95–105.
[45] D.M. Helfman, W.M. Ricci, Branch point selection in alter-

native splicing of tropomyosin pre-mRNAs, Nucl. Acids Res. 17 (1989) 5633–5650.
[46] P.A. Estes, N.E. Cooke, S.A. Liebhaber, A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms, J. Biol. Chem. 267 (1992) 14902–14908.
[47] T.A. Cooper, In vitro splicing of cardiac troponin precursors, J. Biol. Chem. 267 (1992) 5330–5338.
[48] D. Libri, M. Goux-Pelletan, E. Brody, M.Y. Fiszman, Exon as well as intron sequences are *cis*-regulating elements for the mutually exclusive alternative splicing of the α tropomyosin gene, Mol. Cell. Biol. 10 (1990) 5036–5046.
[49] C.W.J. Smith, B. Nadal-Ginard, Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing, Cell 56 (1989) 749–758.
[50] A.M. Carothers, G. Urlaub, D. Grunberger, L.A. Chasin, Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells, Mol. Cell. Biol. 13 (1993) 5085–5098.
[51] W.J.T. Smith, C.T. Chu, B. Nadal-Ginard, Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns, Mol. Cell. Biol. 13 (1993) 4939–4952.
[52] B. Ruskin, J.M. Greene, M.R. Green, Cryptic branch point activation allows accurate in vitro splicing of human $-globin intron mutants, Cell 41 (1985) 833–844.
[53] Z. Dominiski, R. Kole, Selection of splice sites in pre-mRNAs with short internal exons, Mol. Cell. Biol. 11 (1991) 6075–6083.
[54] X. Liu, J.E. Mertz, Sequence of the polypyrimidine tract of the 3′-terminal 3′ splicing signal can affect intron-dependent pre-mRNA processing in vivo, Nucl. Acids Res. 24 (1996) 1765–1773.
[55] E. Izaurraide, J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, I.W. Mattaj, A nuclear cap binding protein complex involved in pre-mRNA splicing, Cell 78 (1994) 657–668.
[56] M. Niwa, S.D. Rose, S.M. Berget, In vitro polyadenylation is stimulated by the presence of an upstream intron, Genes Dev. 4 (1990) 1552–1559.
[57] B. Wieringa, E. Hofer, C. Weissman, A minimal intron length but no specific internal sequence is required for splicing the large rabbit $-globin intron, Cell 37 (1984) 915–925.
[58] J.F. Bateman, D. Chan, I. Moeller, M. Hannagan, W.G. Cole, A 5′ splice site mutation affecting the pre-mRNA splicing of two upstream exons in the collagen COL 1A1 gene, Biochem. J. 302 (1994) 729–735.
[59] D.A. Sterner, S.M. Berget, In vivo recognition of a vertebrate mini-exon as an exon–intron–exon unit, Mol. Cell. Biol. 13 (1993) 2677–2687.
[60] H. Ge, J. Noble, J. Colgan, J.L. Manley, Polyoma virus small tumor antigen pre-mRNA splicing requires cooperation between two 3′ splice sites, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 3338–3342.
[61] M. Aoshima, H. Nunoi, M. Shimazu, S. Shimizu, O. Tatsuzawa, R.T. Kenney, S. Kanegasaki, Two-exon skipping due to a point mutation in p67-*phox*-deficient chronic granulomatous disease, Blood 88 (1996) 1841–1845.
[62] M.J.L. Ligtenberg, A.M.C. Gennissen, H.L. Vos, J. Hilkens, A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episialin mRNA, Nucl. Acids Res. 19 (1991) 297–304.
[63] M. Krawczak, J. Reiss, D.N. Cooper, The mutational spectrum of single base-pair substitutions in the mRNA splice junctions of human genes: causes and consequences, Hum. Genet. 90 (1992) 41–54.
[64] H.C. Dietz, D. Valle, C.A. Francomano, R.J. Kendzior Jr., R.E. Pyeritz, G.R. Cutting, The skipping of constitutive exons in vivo induced by nonsense mutations, Science 259 (1993) 680–683.
[65] G. Urlaub, P.J. Mitchell, C.J. Ciudad, L.A. Chasin, Nonsense mutations in the dihydrofolate reductase gene affect RNA processing, Mol. Cell. Biol. 9 (1989) 2868–2880.
[66] D. Weil, M. D'Alessio, F. Ramirez, W. de Wet, W.G. Cole, D. Chan, J.F. Bateman, A base substitution in the exon of a collagen gene causes alternative splicing and generates a structurally abnormal polypeptide in a patient with Ehlers–Danlos syndrome Type VII, EMBO J. 8 (1989) 1705–1710.
[67] C.W. Fisher, C.R. Fisher, J.L. Chuang, K.S. Lau, D.T. Chuang, R.P. Cox, Occurrence of a 2-bp (AT) deletion allele and a nonsense (G-to-T) mutant allele at the E2 (DBT) locus of six patients with maple syrup urine disease: multiple-exon skipping as a secondary effect of the mutations, Am. J. Hum. Genet. 52 (1993) 414–424.
[68] M.H. Ricketts, M.J. Simons, J. Parma, L. Mercken, Q. Dong, G. Vassart, A nonsense mutation causes hereditary goitre in the Afrikaner cattle and unmasks alternative splicing of thyroglobulin transcripts, Proc. Natl. Acad. Sci. U.S.A. 84 (1987) 3181–3184.
[69] P.M. Smooker, J. Christodoulou, R.R. McInnes, R.G.H. Cotton, A mutation causing DHPR deficiency results in a frameshift and a secondary splicing defect, J. Med. Genet. 32 (1995) 220–223.
[70] S. Aoufouchi, J. Yjlamos, C. Milstein, Nonsense mutations inhibit RNA splicing in a cell-free system: recognition of mutant codon is independent of protein synthesis, Cell 85 (1996) 415–422.
[71] I. Santisteban, F.X. Arredondo-Vega, S. Kelly, M. Loubser, N. Meydan, C. Roifman, P.L. Howell, T. Bowen, K.I. Weinberg, M.L. Schroeder, M.S. Hershfield, Three new adenosine deaminase mutations that define a splicing enhancer and cause severe and partial phenotypes: implications for evolution of a CpG hotspot and expression of a transduced ADA cDNA, Hum. Mol. Genet. 4 (1995) 2081–2087.
[72] R. Pulak, P. Anderson, mRNA surveillance by the *Caenorhabditis elegans smg* genes, Genes Dev. 7 (1993) 1885–1897.
[73] L.E. Maquat, Defects in RNA splicing and the consequence of shortened translational reading frames, Am. J. Hum. Genet. 59 (1996) 279–286.
[74] A. Edwards, H. Voss, P. Rice, A. Civitello, J. Stegemann, C. Schwager, J. Zimmerman, H. Erfle, C.T. Caskey, W. An-

sorge, Automated DNA sequencing of the human HPRT locus, Genomics 6 (1990) 593–608.
[75] J.A. Nicklas, T.C. Hunter, J.P. O'Neill, R.J. Albertini, Fine structure mapping of the *hprt* gene region of the human X chromosome (Xq26), Am. J. Hum. Genet. 49 (1990) 267–278.
[76] N.F. Cariello, T.R. Craft, H. Vrieling, A.A. van Zeeland, T. Adams, T.R. Skopek, Human HPRT mutant database: software for data entry and retrieval, Environ. Mol. Mutagenesis 20 (1992) 81–83.
[77] J.C. Fuscoe, L.J. Zimmerman, M.L. Lippert, J.A. Nicklas, J.P. O'Neill, R.J. Albertini, V(D)J recombinase-like activity mediates *hprt* gene deletion in human fetal T-lymphocytes, Cancer Res. 51 (1991) 6001–6005.
[78] J.C. Fuscoe, L.J. Zimmerman, K. Harrington-Brock, L. Burnette, M.M. Moore, J.A. Nicklas, J.P. O'Neill, R.J. Albertini, V(D)J recombinase-mediated deletion of the *hprt* gene in T-lymphocytes from adult humans, Mutation Res. 283 (1992) 13–20.
[79] H. Steingrimsdottir, G. Rowley, G. Dorado, J. Cole, A.R. Lehmann, Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene, Nucl. Acids Res. 20 (1992) 1201–1208.
[80] C.R. Giver, S.L. Nelson, A.J. Grosovsky, Spectrum of spontaneous $HPRT^-$ mutations in TK6 human lymphoblasts, Environ. Mol. Mutagenesis 22 (1993) 138–146.
[81] C. Guillouf, A. Laquerbe, E. Moustacchi, D. Papadopoulo, Mutagenic processing of psoralen monoadducts differ in normal and Fanconi anemia cells, Mutagenesis 8 (1993) 355–361.
[82] Y. Yamada, H. Goto, K. Suzumori, A. Ritsuko, N. Ogasawara, Molecular analysis of five independent Japanese mutant genes responsible for hypoxanthine guanine phosphoribosyltransferase (HPRT) deficiency, Hum. Genet. 90 (1992) 379–384.
[83] S. Nelson, C.R. Giver, A.J. Grosovshy, Spectrum of X-ray-induced mutations in the human *hprt* gene, Carcinogenesis 15 (1994) 495–502.
[84] S. Marcus, D. Hellgren, B. Lambert, S.P. Fällström, J. Wahlström, Duplication in the hypoxanthine phosphoribosyl-transferase gene caused by Alu–Alu recombination in a patient with Lesch–Nyhan syndrome, Hum. Genet. 90 (1993) 477–482.
[85] S. Yang, J.G. Lin, C.C. Chiou, L.Y. Chen, J.L. Yang, Mutation specificity of 8-methoxypsoralen plus two doses of UVA irradiation in the *hprt* gene in diploid human fibroblasts, Carcinogenesis 15 (1994) 201–207.
[86] T. Bastlova, B. Andersson, B. Lambert, A. Kilman, Molecular analysis of ethylene oxide induced mutations at the *hprt* locus in human diploid fibroblasts, Mutation Res. 287 (1993) 283–292.
[87] A.M. Rossi, A.D. Tates, A.A. van Zeeland, H. Vrieling, Molecular Analysis of Mutations Affecting *hprt* mRNA Splicing in Human T-lymphocytes in vivo, Environ. Mol. Mutagenesis 19 (1992) 7–13.
[88] E.G.R. Lichtenauer-Kaligis, J.C.P. Thijssen, H. den Dulk, P. van de Putte, M. Giphart-Gassler, J.G. Tasseron-de Jong, Spontaneous mutation spectrum in the *hprt* gene in human lyphoblastoid TK6 cells, Mutagenesis 10 (1995) 137–143.
[89] L. Sege-Peterson, J. Chambers, T. Page, O.W. Jones, W.L. Nyhan, Characterization of mutations in phenotypic variants of hypoxanthine phosphoribosyltransferase deficiency, Hum. Mol. Genet. 1 (1992) 427–442.
[90] M.J. Lippert, I.R. Rainville, J.A. Nicklas, R.J. Albertini, Large deletions partially external to the human *hprt* gene result in chimeric transcripts, Mutagenesis 12 (1997) 185–190.
[91] M.L. Cleary, Oncogenic conversion of transcription factors by chromosomal translocations, Cell 66 (1991) 619–622.
[92] M.G. Manjanatha, L.A. Lindsey, R.A. Mittelstaedt, R.H. Heflich, Low hprt mRNA levels and multiple *hprt* mRNA species in 6-thioguanine-resistant Chinese hamster cell mutants possessing nonsense mutations, Mutation Res. 308 (1994) 65–75.
[93] C.R. Valentine, R.H. Heflich, The association of nonsense mutation with exon-skipping in *hprt* mRNA of Chinese hamster ovary cells results from an artifact of RT-PCR, RNA 3 (1997) 660–676.
[94] Valentine, C.R. (in press), Origin of exon-deleted mRNAs associated with nonsense mutations, Mutation Research.
[95] P. Menichini, A. Inga, G. Fronza, R. Iannone, P. Degan, P. Campomenosi, A. Abbondandolo, Defective splicing induced by 4NQO in the hamster *hprt* gene, Mutation Res. 323 (1994) 159–165.
[96] E.E. Hennig, A.H. Conney, S.-J.C. Wei, Characterization of *hprt* splicing mutations induced by the ultimate carcinogenic metabolite of benzo[a]pyrene in Chinese Hamster V-79 cells, Cancer Res. 55 (1995) 1550–1558.
[97] L.-H. Zhang, H. Vrieling, A.A. van Zeeland, D. Jenssen, Spectrum of spontaneously occurring mutations in the *hprt* gene of V79 Chinese hamster cells, J. Mol. Biol. 223 (1992) 627–635.
[98] C.R. Valentine, R.H. Heflich, Genomic DNA sequencing of mRNA splicing mutants in the *hprt* gene of Chinese Hamster ovary cells, Environ. Mol. Mutagenesis 25 (1995) 85–96.