

species from mouse muscle and PLAT-E cells as described previously [3]. The virus titers for NIH3T3 cells were $\sim 1.6 \times 10^7$ PFU/ml. In NIH3T3 cells infected with retrovirus particles, we confirmed expression of the transgenes by RT-PCR and by Western blot analysis using anti-human ColQ polyclonal antibody (Gene way).

AAV vectors

We inserted either human *COLQ* or human *COLQ-IRES-EGFP* downstream of a CMV promoter in the pAAV-MCS vector (Stratagene). Adenovirus-independent helper plasmid pDF6 carrying a serotype 6 capsid was a kind gift of Dr. Mark A. Kay, Stanford University.

We cotransfected HEKT293T cells with pAAV-*COLQ* and pDF6 by the calcium phosphate precipitation method [4] to generate rAAV6-*COLQ*. We infected AAV-HT1080 cells (Stratagene) with rAAV6-*COLQ*, and confirmed expression of *COLQ* by RT-PCR.

Results and discussion

3.1 Retrovirus-mediated expression of asymmetric AChE in cultured cells

PLAT-E cells were transfected with pSIREN-*ACHE-IRES-COLQ* to yield virus particles. In these cells, we observed production of both globular and asymmetric forms of AChE, as we can see in wild-type mouse skeletal muscles by sedimentation analysis in sucrose gradients (Fig.1).

We next infected NIH3T3 cells with the *ACHE-IRES-COLQ* retrovirus, or with both the *ACHE* and *COLQ* retroviruses. In both cells, we confirmed expression

of the transgene mRNAs and proteins. We also observed secretion of ColQ in culture medium by Western blotting (data not shown).

3.2 AAV-mediated expression of ColQ in cultured cells

rAAV6-COLQ harbors serotype 6, and should have tropism for skeletal muscles. In AAV-HT1080 cells infected with rAAV6-COLQ-IRES-EGFP, we observed expressions of COLQ mRNA by RT-PCR, and of EGFP by flow cytometry (data not shown). We administered $\sim 1 \times 10^8$ vector genomes in the tail vein of a ColQ-knockout mouse. We are currently observing if the AAV vector shows any beneficial effects on the treated mouse.

3.3 Future perspectives

In this study, we demonstrate that the recombinant retrovirus and AAV efficiently express transgenes in cultured cells. We are currently working on achieving further higher expression levels of transgenes in cultured cells, as well as in culture medium. Through the current strategy, we hope to ameliorate motor symptoms of the knockout mice [5] and of the patients.

Acknowledgments

The research was supported by Grants-in-Aid for Scientific Research from the Japan Foundation for Neuroscience and Mental Health.

References

[1] K. Ohno, A.G. Engel, J.M. Brengman, X.M. Shen, F. Heidenreich, A. Vincent,

M. Milone, E. Tan, M. Demirci, P. Walsh, S. Nakano, I. Akiguchi, The spectrum of mutations causing end-plate acetylcholinesterase deficiency, *Ann. Neurol.* 47 (2000) 162-170

[2] L.M. Kimbell, K. Ohno, A.G. Engel, R.L. Rotundo, C-terminal and heparin-binding domains of collagenic tail subunit are both essential for anchoring acetylcholinesterase at the synapse, *J. Biol. Chem.* 279 (2004) 10997-11005

[3] K. Ohno, J. Brendman, A. Tsujino, A.G. Engel, Human endplate acetylcholinesterase deficiency caused by mutations in the collagen-like tail subunit (ColQ) of the asymmetric enzyme, *Proc. Natl. Acad. Sci. USA* 95 (1998) 9654-9659

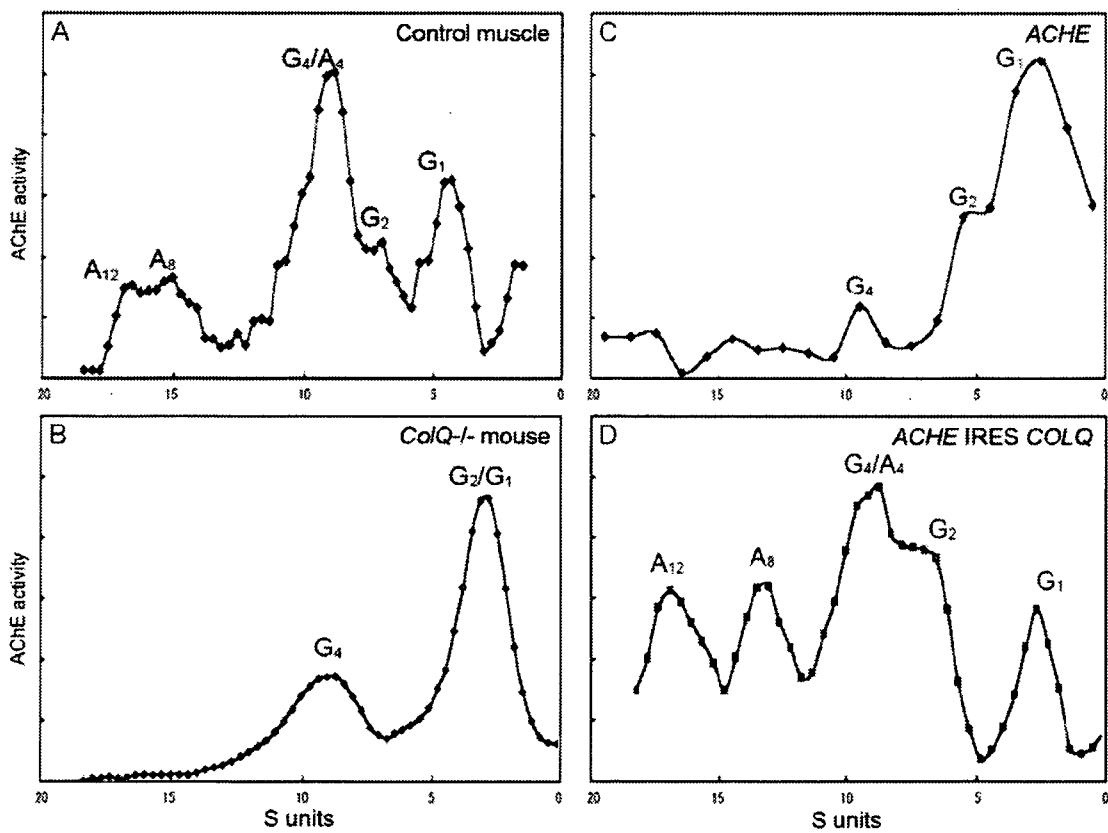
[4] T. Okada, T. Nomoto, K. Shimazaki, W. Lijun, Y. Lu, T. Maatsushita, H. Mizukami, M. Urabe, Y. Hanazono, A. Kume, S. Muramatsu, I. Nakano, K. Ozawa, Adeno-associated virus vectors for gene transfer to the brain, *Methods* 28(2) (2002) 237-247

[5] G. Feng, E. Krejci, J. Molgo, J.M. Cunningham, J. Massoulié, J.R. Sanes, Genetic analysis of collagen Q: Roles in acetylcholinesterase and butyrylcholinesterase assembly and in synaptic structure and function, *J. Biol. Chem.* 144 (1999) 1349-1360

Figure Legend

Fig. 1. Sedimentation profiles of AChE species. (A) In wild-type mouse, we observe both asymmetric and globular forms of AChE. (B) In ColQ-deficient mice [4], no asymmetric AChE species are formed. (C) In PLAT-E cells transfected with pSIREN-ACHE alone, only globular forms are generated. (D) In PLAT-E cells transfected with pSIREN-ACHE-IRES-COLQ, both asymmetric and globular forms are generated.

Fig. 1.



Long-range PCR for the diagnosis of spinocerebellar ataxia type 10

Tatsuaki Kurosaki · Tohru Matsuura · Kinji Ohno · Shintaroh Ueda

Received: 28 November 2007 / Accepted: 20 December 2007
© Springer-Verlag 2008

Sirs,
Spinocerebellar ataxia type 10 (SCA10) is an autosomal dominant neurodegenerative disorder characterized by a unique combination of progressive ataxia and seizures. The mutation causing SCA10 is an unstable and massive expansion (800–4500 repeats; unaffected range, 10–29) of an ATTCT pentanucleotide repeat in intron 9 of the ataxin-10 (ATXN10) gene on chromosome 22q13.3 [1]. In molecular analysis of SCA10, polymerase chain reaction (PCR) of the ATTCT repeat typically shows a single normal-sized allele since the expanded allele is non-amplifiable. Southern blot analysis is required to detect the repeat expansion and determine its size [1]. However, this method is laborious, costly, and not applicable to degraded samples or those with limited quantities of DNA. Ataxia patients are not often tested for SCA10 due to these limitations as well as insurance problems; a simple and cost-effective screening test would be preferred.

Recent development of repeat-primed PCR (RP) [2, 3] allows quick detection of the ATTCT expansion by PCR alone, even in degraded DNA samples; however, it cannot

determine the repeat length, merely the presence of expansion. Long-range PCR (LP) is effective for accurately and directly determining the size of expansions, requires little DNA, and can be carried out quickly. We recently succeeded in amplifying a 280-ATTCT expansion by LP [4]. Here, we show the improved LP protocol to accurately detect the full expansion ~4,000 repeats, which can completely replace the use of Southern blot analysis.

We used 100 normal controls and 20 patient-DNA samples in which the SCA10 diagnosis was confirmed and repeat sizes were determined by Southern analysis. LP was performed in the supplementary conditions. As shown in Fig. 1, we could detect variably expanded alleles ranging from 280 to 3,500 repeats. There were no false-positive results, i.e., no expansions in the 100 normal controls. A size estimate for each mutation was consistent with the results from Southern data, excluding the possibility of PCR artifact. We believe this protocol is a rapid and accurate method to both confirm the presence of an expansion mutation and determine its size without limitation of the mutation range, and samples positive for expansions by RP should be subjected to LP.

Wick et al. [5] reported similar efforts in the diagnosis of Friedreich ataxia and established a protocol to amplify ~1,700 GAA repeats by LP. Recent improvements in LP technology such as hot start and high-fidelity *Pfu* polymerase mixtures with 3'→5' proofreading activities are now allowing us to amplify longer, more complex genomic targets that were previously difficult to amplify. Our protocol will reduce time, labor, and cost requirements in clinical laboratory settings. It is superior to RP in molecular testing of SCA10 when the sizing of the expansion is critical (e.g., a family with reduced penetrance [4]) or homozygotes must be distinguished from heterozygotes. It will be also applicable to cloning of the expanded allele,

Electronic supplementary material The online version of this article (doi:10.1007/s10048-007-0117-x) contains supplementary material, which is available to authorized users.

T. Kurosaki · S. Ueda
Department of Biological Sciences, Graduate School of Science,
The University of Tokyo,
Tokyo, Japan

T. Kurosaki · T. Matsuura (✉) · K. Ohno
Division of Neurogenetics and Bioinformatics,
Center for Neurological Diseases and Cancer,
Nagoya University Graduate School of Medicine,
Nagoya 466-8550, Japan
e-mail: tohrum@med.nagoya-u.ac.jp

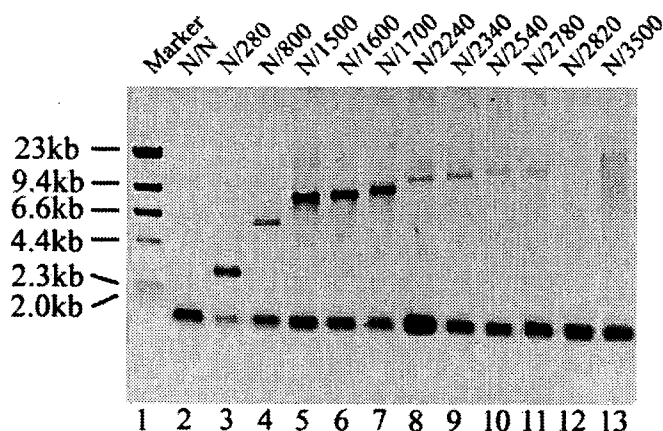


Fig. 1 Long-range PCR of the ATTCT expansion in spinocerebellar ataxia type 10. *Lane 1:* λ DNA/*Hind* III marker; *Lane 2:* normal individual; *Lanes 3–13:* individuals with SCA10 expansions. The number of ATTCTs of each individual's expanded allele is shown, with "N" representing a normal-range ATTCT repeat. *Lane 13* shows somatic mosaicism of SCA10 expanded allele. The length of normal allele is 1.5 kb

which can be used for studies of its repeat sequence and instability, eventually leading to understanding the disease mechanism [6] as well as genetic diagnosis and counseling in SCA10.

Acknowledgements The authors thank Dr. Tetsuo Ashizawa for providing SCA10 DNA samples. This study was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan (S.U.), National Ataxia Foundation, National Organization for Rare Disorders, Ichihara International Scholarship Foundation,

Japan, Takeda Science Foundation, Japan, The Nitto Foundation, Japan, Japan Brain Foundation, Japan, Grant-in-Aid for Scientific Research on Priority Areas—Research on Pathomechanisms of Brain Disorders—from MEXT, Japan, and grant from the Research Committee for Ataxic Diseases of the Ministry of Health, Labor and Welfare, Japan (T.M.).

References

1. Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall A, Davis CF, Zu L, Achari M, Pulst SM, Alonso E, Noebels JL, Nelson DL, Zoghbi HY, Ashizawa T (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet* 26:191–194
2. Matsuura T, Ashizawa T (2002) Polymerase chain reaction amplification of expanded ATTCT repeat in spinocerebellar ataxia type 10. *Ann Neurol* 51:271–272
3. Cagnoli C, Michielotto C, Matsuura T, Ashizawa T, Margolis RL, Holmes SE, Gellera C, Migone N, Brusco A (2004) Detection of large pathogenic expansion in FRDA1, SCA10, and SCA12 genes using a simple fluorescent repeat-primed PCR assay. *J Mol Diag* 6:96–100
4. Matsuura T, Fang P, Pearson CE, Jayakar P, Ashizawa T, Roa BB, Nelson DL (2006) Interruption in the expanded ATTCT repeat of spinocerebellar ataxia type 10: repeat purity as a disease modifier. *Am J Hum Genet* 78:125–129
5. Wick MJ, Matthias-Hagen VL, Day JW, Gomez CM, McGlennen RC (1998) Long range polymerase chain reaction-based diagnosis of Friedreich ataxia in the clinical molecular Diagnostics laboratory. *Mol Diagnosis* 3:3–9
6. Wakamiya M, Matsuura T, Liu Y, Shuster GC, Gao R, Xu W, Sarkar P, Lin X, Ashizawa T (2006) The role of ataxin 10 in the pathogenesis of spinocerebellar ataxia type 10. *Neurology* 67:607–613

Human branch point consensus sequence is yUnAy

Kaiping Gao, Akio Masuda, Tohru Matsuura and Kinji Ohno*

Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan

Received December 6, 2007; Revised January 17, 2008; Accepted February 5, 2008

ABSTRACT

Yeast carries a strictly conserved branch point sequence (BPS) of UACUAAC, whereas the human BPS is degenerative and is less well characterized. The human consensus BPS has never been extensively explored *in vitro* to date. Here, we sequenced 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes. Among the 367 clones, a misincorporated nucleotide at the branch point was observed in 181 clones, for which we can precisely pinpoint the branch point. The branch points were comprised of 92.3% A, 3.3% C, 1.7% G and 2.8% U. Our analysis revealed that the human consensus BPS is simply yUnAy, where the underlined is the branch point at position zero and the lowercase pyrimidines ('y') are not as well conserved as the uppercase U and A. We found that the branch points are located 21–34 nucleotides upstream of the 3' end of an intron in 83% clones. We also found that the polypyrimidine tract spans 4–24 nucleotides downstream of the branch point. Our analysis demonstrates that the human BPSs are more degenerative than we have expected and that the human BPSs are likely to be recognized in combination with the polypyrimidine tract and/or the other splicing *cis*-elements.

INTRODUCTION

In higher eukaryotes, pre-mRNA splicing is mediated by degenerative splicing *cis*-elements comprised of the branch point sequence (BPS), the polypyrimidine tract (PPT), the 5' and 3' splice sites and exonic/intronic splicing enhancers/silencers. Stepwise assembly of the spliceosome starts from recruitment of U1 snRNP, SF1, U2AF65 and U2AF35 to the 5' splice site, the branch site, the PPT and the 3' end of an intron, respectively (Complex E). SF1, a 75-kDa polypeptide, is a mammalian homolog of

yeast BBP (branch point-binding protein). U2AF65 and U2AF35 bring U2snRNP to the BPS in place of SF1 (1,2). The BPS establishes base pairing interactions with a stretch of 'GUAGUA' of U2 snRNA (3,4), which then bulges out the branch site nucleotide, usually an adenosine (Complex A) (5). Thereafter, pre-mRNAs are spliced in two sequential transesterification reactions mediated by the spliceosome. In the first step, the 2'-OH moiety of the branch site nucleotide carries out a nucleophilic attack against a phosphate at the 5' splice site, generating a free upstream exon, as well as a lariat carrying the intron and the downstream exon. In the second step, the 3'-OH moiety of the upstream exon attacks the 3' splice site leading to intron excision and ligation of the upstream and downstream exons (6). The branch site is thus involved in the first step of splicing, and potentially in the second step of splicing, although the detailed molecular mechanisms of contribution to the second step remain elusive (7).

The BPS is strictly conserved in yeast and has the sequence of UACUAAC, where the branch point adenosine is underlined. On the other hand, the human BPSs are degenerative. No extensive *in vitro* identification of human BPSs has been reported. Five communications address the mammalian consensus BPSs (Table 1). Three reports are based on 11–20 *in vitro* identified BPSs, and two are dependent on the *in silico* analysis of the human genome.

In an effort to establish the human consensus BPS based on *in vitro* experiments, we analyzed 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes. We found that the human consensus BPS is yUnAy. Our analysis demonstrates that the human BPSs are more degenerative than we have expected and that the BPS is likely recognized in combination with the PPT and/or the other splicing *cis*-elements.

MATERIALS AND METHODS

Lariat RT-PCR primers for human housekeeping genes

Among the 575 human housekeeping genes registered at http://www.compugen.co.il/supp_info/Housekeeping_genes.html (8), we excluded 82 genes, for which we

*To whom correspondence should be addressed. Tel: +81 52 744 2446; Fax: +81 52 744 2449; Email: ohnok@med.nagoya-u.ac.jp

Table 1. Previously reported consensus BPSs

Consensus BPS	Note	References
<i>S. cerevisiae</i> UACU <u>A</u> AC	Invariant BPS	(36,37)
Mammals YNYUR <u>A</u> Y	11 mammalian BPSs	(25)
YNCUR <u>A</u> C	20 mammalian BPSs	(38)
YNCUR <u>A</u> Y	15 BPSs of human HBB	(39)
CUR <u>A</u> Y	<i>In silico</i> homology search	(40)
YUV <u>A</u> Y ^a	<i>In silico</i> homology search	(41)
CUS <u>A</u> Y ^b		

Branch point 'A' is underlined.

Y, U or C; R, A or G; S, G or C; V, A, C or G.

^aLow GC% region.

^bHigh GC% region.

could not find entries in the EST profile viewer of the NCBI UniGene database (<http://www.ncbi.nlm.nih.gov/UniGene/>). Among 4188 introns of the remaining 493 human housekeeping genes, we excluded introns with a size of less than 300 nucleotides or with multiple repeated segments, because it was difficult to design appropriate PCR primers for such introns. We next sorted the 493 genes in the order of skin expression levels according to the EST profile viewer, and picked up the 20 best genes. We thus analyzed 52 introns of the 20 human housekeeping genes (Supplementary Table 1).

We placed the sense primers at least 100 nucleotides upstream of the 3' end of an intron, and the antisense primers at least 10 nucleotides downstream of the 5' end of an intron. The melting temperatures of the primers were designed to be 64–67°C according to the nearest neighbor method. Gene symbols, intron numbers and primer sequences are indicated in Supplementary Table 1.

Lariat RT-PCR to identify the branch point

We performed nested lariat RT-PCR to amplify a fragment spanning the 2'–5' phosphodiester bond at the branch point (9). We isolated total RNA from HEK293 cells grown to confluency in DMEM medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and penicillin–streptomycin (Invitrogen). First-strand cDNA was synthesized with SuperScript II reverse transcriptase (Invitrogen) using an intron-specific antisense primer C (Figure 1) located close to the 5' end of an intron. The first round of lariat RT-PCR was performed using primers C and D with *Taq* HS DNA polymerase (Takara) in 25 µl. The nested lariat RT-PCR was carried out with primers A and B using 0.2 µl of the first-round lariat RT-PCR product in 50 µl. The first-round PCR program was followed by an initial denaturation step at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min. For the nested PCR, we performed 35 cycles of amplification.

We purified the nested lariat RT-PCR products using the Wizard SV Gel and PCR Clean-up system (Promega), and cloned them into the pGEM-T-Easy vector (Promega). We sequenced 10 clones for each intron using the CEQ8000 genetic analyzer (Beckman Coulter).

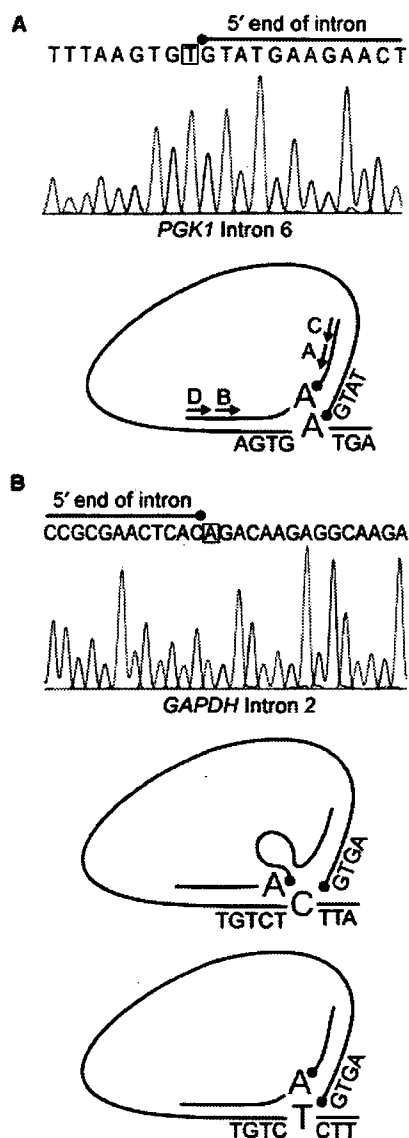


Figure 1. (A) Lariat RT-PCR of *PGK1* intron 6 indicates a misincorporated 'A' nucleotide at the branch point. We can pinpoint the branch point in this situation. The sequencing is performed with primer B. The small dots indicate the 5' end of an intron. (B) Lariat RT-PCR of *GAPDH* intron 2 exhibits no misincorporated nucleotide. The branch point can be either at 'C' or upstream 'T' depending on whether skipping of the reverse transcriptase occurs or not. We cannot locate the exact branch point in this case. The sequencing is performed with primer A.

Presentations of sequence motifs

Sequence motifs are presented using the Pictogram web server at <http://genes.mit.edu/pictogram.html> (10).

To indicate the amount of information conferred by each nucleotide at each position, we employed the WebLogo program at <http://weblogo.berkeley.edu/> (11,12). We also calculated the total amount of information content at each position using the following formula:

$$R_{\text{sequence}} = 2 + \sum Pi \log_2 Pi \quad (i = A, C, G \text{ and } U)$$

where P_i represents the probability of nucleotide i at each position. R_{sequence} represents the degree of conservation

Table 2. Previously identified mammalian and viral BPSs

Species	Gene	Intron	BPS	Position	Predicted BP ^a	BPS Score ^a	Reference
H. sapiens	<i>CALCA</i>	4	CACTC <u>A</u> C	-36 ^b	-36A	3.85	(42)
H. sapiens	<i>CALCA</i>	3	TACTG <u>T</u> C	-23 ^b	-42A	2.60	(21)
H. sapiens	<i>CALCA</i>	3	G <u>T</u> ACTGT	-24 ^b	-42A	2.60	(21)
H. sapiens	<i>CALCA</i>	3	GGTGC <u>A</u> T	-32 ^b	-42A	2.60	(21)
H. sapiens	<i>CSH1</i>	1	CCTCC <u>A</u> T	-23 ^b	-23A	2.75	(22)
H. sapiens	<i>DQB1</i>	3	CACAG <u>A</u> C	-21 ^c	-21A	3.25	(17)
H. sapiens	<i>GH1</i>	1	CTCTG <u>T</u> T	-22 ^b	na	na	(22)
H. sapiens	<i>GH1</i>	1	GGCTC <u>C</u> C	-28 ^b	na	na	(22)
H. sapiens	<i>GH1</i>	1	TGCTC <u>T</u> C	-36 ^b	na	na	(22)
H. sapiens	<i>GH1</i>	4	GCCTC <u>T</u> C	-24 ^b	-37A	2.80	(22)
H. sapiens	<i>GH1</i>	4	ACCCA <u>A</u> G	-37 ^b	-37A	2.80	(22)
H. sapiens	<i>GH1</i>	4	TACCC <u>A</u> A	-38 ^b	-37A	2.80	(22)
H. sapiens	<i>HBA</i>	1	CCCTC <u>A</u> C	-19 ^b	-37A	3.25	(25)
H. sapiens	<i>HBA</i>	2	CACTG <u>A</u> C	-18 ^b	-18A	3.95	(25)
H. sapiens	<i>HBB</i>	1	CACTG <u>A</u> C	-37 ^b	-37A	3.95	(25)
H. sapiens	<i>HBE1</i>	1	CTCTA <u>A</u> T	-31 ^b	-31A	3.45	(25)
H. sapiens	<i>HBG1</i>	1	TTCTG <u>A</u> C	-30 ^b	-30A	3.85	(25)
H. sapiens	<i>MYH10</i>	5	TGCTA <u>A</u> C	-31 ^b	na	na	(15)
H. sapiens	<i>XPC</i>	3	TGTTG <u>A</u> T	-4 ^b	na	na	(16)
H. sapiens	<i>XPC</i>	3	TACTG <u>A</u> T	-24 ^b	na	na	(16)
M. musculus	<i>Hbb-b2</i>	1	CACTA <u>A</u> C	-36 ^b	-36A	3.85	(25)
M. musculus	<i>Igh</i>	5	AATTC <u>A</u> C	-22 ^b	-22A	3.30	(14)
R. norvesicus	<i>Ins1</i>	1	CCTCA <u>A</u> C	-18 ^b	-18A	3.15	(25)
O. cuniculus	<i>Hbb</i>	1	TGCTG <u>A</u> C	-34 ^b	-34A	3.85	(25)
O. cuniculus	<i>Hbb</i>	2	TGCTA <u>A</u> C	-32 ^b	-32A	3.75	(28)
Adenovirus 5	<i>E1A</i>	1	GTTT <u>A</u> AA	-30 ^b	-30A	2.70	(25)
Adenovirus 2	<i>E2a-2</i>	1	GACTG <u>A</u> C	-26 ^b	-26A	3.70	(42)
Adenovirus 5	<i>Major Late</i>	1	TACTT <u>A</u> T	-24 ^b	-24A	3.05	(42)
SV40	<i>T antigen</i>	1	ATTCT <u>A</u> A	-19 ^b	-19A	2.00	(42)

^aPredicted BPs and BPS scores are according to the Branch-Site Analyzer at <http://ast.bioinfo.tau.ac.il/BranchSite.htm> (14). na, not available.

^bIdentified by the primer extension method.

^cIdentified by lariat RT-PCR.

of a sequence motif at a specific position (13). It becomes 2.0 when a single nucleotide is exclusively observed at a specific position, whereas it becomes 0.0 when four nucleotides are evenly observed.

RESULTS

Collation of previously identified mammalian BPSs

As shown in Table 1, five communications address the mammalian consensus BPSs. In order to understand the *in vitro* determined mammalian consensus BPS, we collated 29 previously reported BPSs comprised of 25 mammalian and four viral introns (Table 2). Viral introns should be spliced in the same way as the mammalian genes. The branch points are located between positions -38 to -4 (mean and SD, -26.9 ± 7.8). Nucleotides 'C', 'U', 'A' and 'Y' at positions -3, -2, 0 and +1 are observed at 21 (72.4%), 21 (72.4%), 23 (79.3%) and 25 sites (86.2%), respectively. The deduced consensus BPS thus becomes CUnAy at positions -3 to +1 (Figure 2A and B), when we arbitrarily assume that positions with the information contents above 0.45 are significant.

Lariat RT-PCR with or without a misincorporated nucleotide at the branch point

To further explore the human consensus BPS, we chose 52 introns of 20 human housekeeping genes. We performed nested lariat RT-PCR and cloned the amplified products.

We sequenced ten clones from each intron, and 367 clones carried available inserts, which represented 117 possible branch sites (Table 3). The remaining 153 clones carried either no inserts or PCR artifacts.

The 367 clones were divided into two classes: 181 clones carrying misincorporated nucleotides at the branch points, and 186 clones without misincorporated nucleotides. For those carrying misincorporated nucleotides, we could pinpoint the exact branch points (Figure 1A). On the other hand, for those carrying no misincorporated nucleotides, the reverse transcriptase might have skipped one or two nucleotides at the 2'-5' phosphodiester bond at the branch points (Figure 1B).

Among the 367 clones, we observed two or more possible branch sites in 36 of 52 introns. The 36 introns carried a total of 101 possible branch sites. Among the 101 sites, 25 were followed by an immediate downstream branch site, making 25 possible branch-site pairs. Among the 25 upstream branch sites, 19 carried no misincorporated nucleotides. In addition, 13 of the 19 upstream sites were followed by an 'A' nucleotide. Furthermore, when we simply deduced the consensus BPS from all 367 clones, the consensus BPS became more degenerative and less informative (data not shown). These findings suggest that the observed upstream branch points are likely due to skipping of a nucleotide in lariat RT-PCR. We thus employed the 181 clones carrying misincorporated nucleotides at the branch points in the following analyses unless otherwise stated.

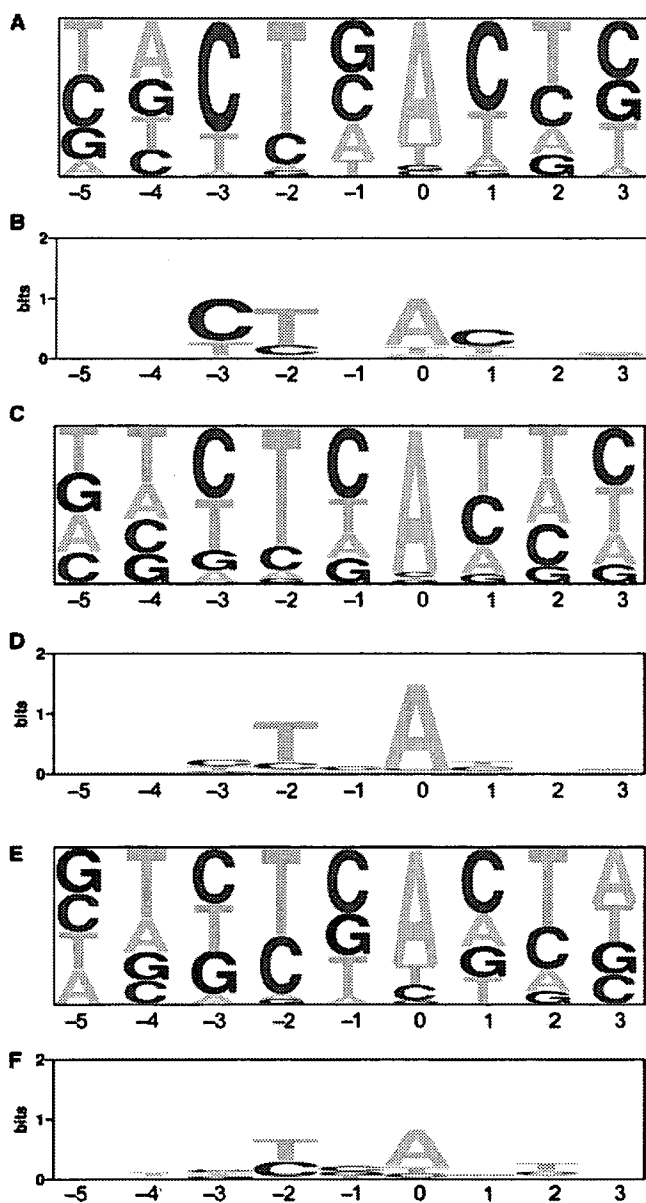


Figure 2. Pictogram (A, C and E) and WebLogo (B, D and F) presentations of mammalian BPSs. (A and B) Twenty-nine mammalian and viral BPSs identified by *in vitro* experiments (Table 2). (C and D) BPSs with a misincorporated nucleotide at the branch point in our studies. (E and F) BPSs without a misincorporated nucleotide at the branch point in our studies. We assume that 'A' residue one or two nucleotides downstream of the sequenced branch point is the actual branch point (see Figure 1B). Position 0 represents the branch point.

We counted each clone as a single occurrence of a branch point in order to weigh the preferred branch points. For example, in *PGKI* intron 6, eight clones mapped to 'A' at position -23 , whereas one clone pointed to 'A' at position -28 (Table 3). We assumed that the branch point at position -23 was eight times more frequently employed than that at position -28 . This analysis method might have overweighted introns that gave rise to more clones. An alternative analysis method would be to make the contribution of each intron equal regardless of the number of available clones. The alternative method, however, is also biased in favor of introns with fewer

clones. For example, *PGKI* intron 8 had a single available clone mapping to position -27 , whereas *EEF1A1* intron 1 had ten clones all mapping to position -23 . A single clone of *PGKI* might have arisen from one of many branch points, and we might have sequenced it by chance. On the other hand, it is likely that *EEF1A1* intron 1 indeed had a single branch point. We analyzed our data using both methods and obtained similar results (data not shown), except that the frequency of C at position -1 was slightly lower with the alternative method (44.8% versus 36.3%). In the current communication, we employed the former method, in which each clone was counted as a single occurrence of a branch site.

Positions and sequence motif of the branch points

Analysis of the 181 clones revealed that the positions of the branch points were from -50 to -5 , where position -1 represents the 3' end of an intron (Figure 3A). Among the 181 sites, 150 (83%) were at positions -34 to -21 .

We observed U at position -2 in 74.6% branch sites, and A at position 0 in 92.3% branch sites (Table 4). In addition, pyrimidines were observed at positions -3 and $+1$ in 79.0% and 75.1% branch sites, respectively (Table 4). We can thus conclude that the human consensus BPS is $yUnAy$ at positions -3 to $+1$ (Figure 2C), where the branch site is underlined and the less conserved nucleotides are indicated in lowercase letters. The information contents of 0.27 and 0.23 at positions -3 and $+1$, however, were not as high as those of 0.85 and 1.48 at positions -2 and 0 (Figure 2D), or 0.39 ± 0.12 (mean \pm SD) of the polypyrimidine tract at positions $+4$ to $+24$ (Figure 3C). Therefore, the consensus sequence alternatively becomes UnA according to the information contents (Figure 2D).

Among the 41 introns yielding the 181 clones, 14 introns carried multiple branch sites. In eight of the 14 introns (57%), the most downstream branch sites were most frequently used (Table 3). Although the ratio of 57% was not high, the downstream branch sites were four to eight times more frequently used than the upstream sites in four of the eight introns. We could not observe this magnitude of differential branch site usage in the remaining six introns, in which the downstream branch points were not overrepresented. Accordingly, when there are multiple branch points, downstream branch points are more likely to be employed than their upstream counterparts.

We also predicted BPSs of our housekeeping genes with the Branch Site Analyzer (14), and found that the actual branch sites matched to the predicted positions in 80 of the 181 sites (44.2%) (Table 3).

Alignment of polypyrimidine tract in respect to the branch point

We next aligned the PPT's in respect to the 181 branch points (Figure 3B and C). We observed a polypyrimidine stretch from position $+4$ down to position $+24$. The 'U' nucleotide was preferred over 'C' especially at positions $+4$ to $+12$ in the PPT. Alignment of the PPT in respect to the 3' end of an intron also demonstrated a stretch of pyrimidines from positions -20 to -4 (Figure 3D and E).

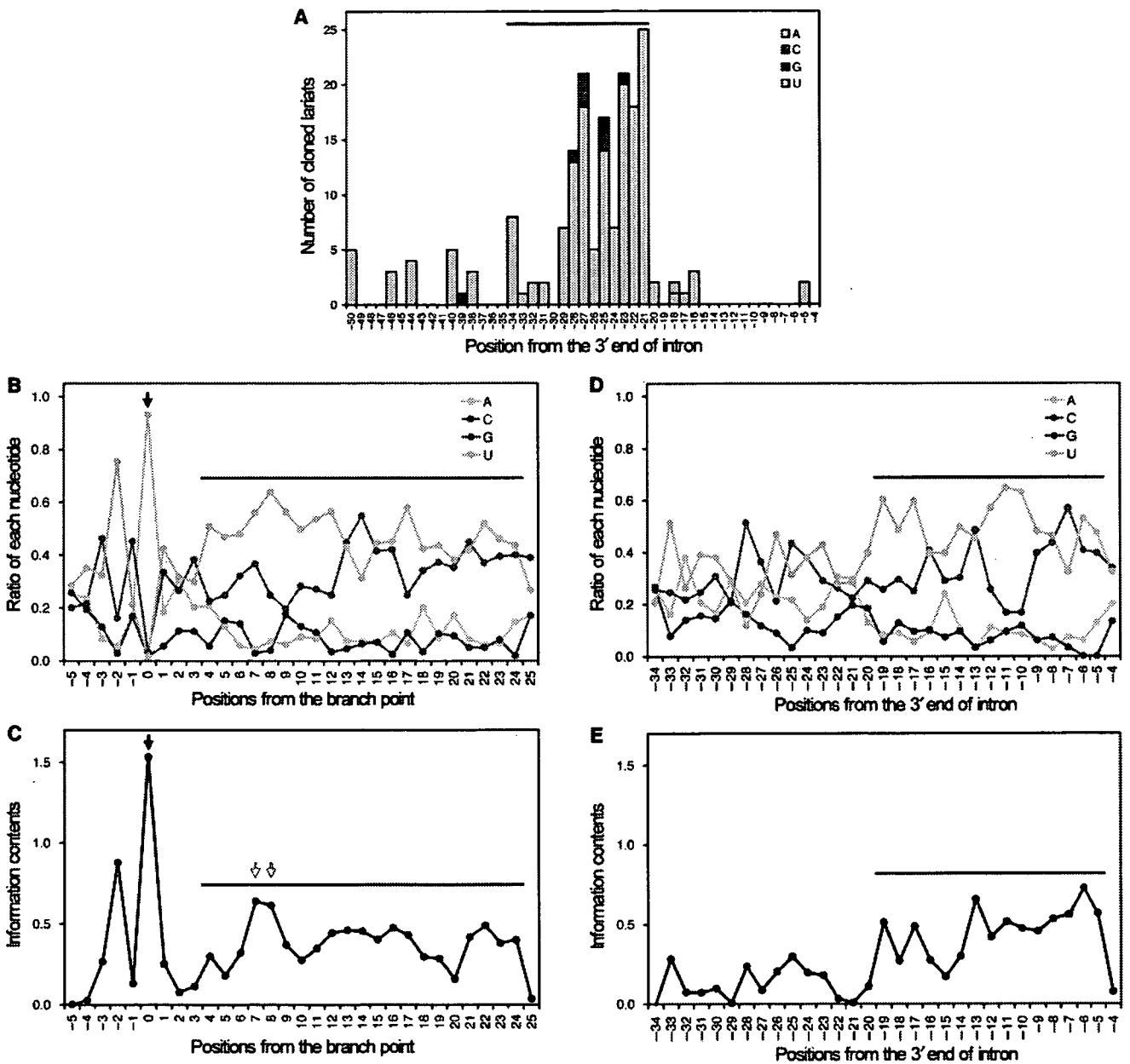


Figure 3. (A) Positions and nucleotides of 181 branch points with misincorporated nucleotides in our studies, where position -1 represents the 3' end of an intron. The median value of the branch points is -26 , and the mean and SD is -27.7 ± 7.6 . Among the 181 sites with misincorporated nucleotides at the branch points, 150 sites (83%) are at positions -34 to -21 (horizontal bar on top). Native nucleotides, not the misincorporated nucleotides, are indicated. Nucleotide preferences (**B** and **D**) and information contents (**C** and **E**) are deduced from 181 branch points. (**B** and **C**) Plots are aligned in respect to the branch point (closed arrows), which is designated as position 0. Open arrows point to peaks of information contents at positions $+7$ and $+8$. A polypyrimidine stretch starts from position $+4$ down to position $+24$ (bars). The plots are truncated at position $+25$, because the numbers of observations fall below 40 after position $+25$, and the plots become less informative and uneven. The last three nucleotides of introns are excluded from the plots. (**D** and **E**) Plots are aligned in respect to the 3' end of each intron, which is designated as position -1 . A polypyrimidine stretch spans positions -19 to -5 (bars).

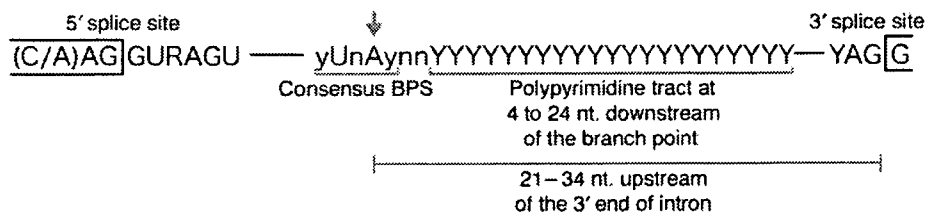


Figure 4. Representative composition of the branch point sequence (arrow) and the PPT deduced from our studies.

Table 3. Analyzed introns and observed branch points

Gene	Intron	Intron size (bp)	Predicted BP ^a	BPS Score ^a	Observed BP	Number of clones	Misincorporated nucleotide	Intronic sequence from BPS position -5 to the 3' end of an intron ^b
<i>ACTB</i>	3	441	-24A	3.10	-30A	10	-	TCCCCAGTGTGACATGGTGTATCTCTGCCTTACAG
<i>ALOD</i>	8	984	-24A	3.05	-26T	2	-	TGTCCTAATGTTGTTACCCTGACCCCAACAG
					-25A	2	T	GTCTTAATGTTGTTACCCTGACCCCAACAG
					-5A	1	-	ACCCCAACAG
<i>CCT3</i>	4	1069	-21A	3.30	-32A	3	-	TGAATAGTGTGAATTCAGTGTGATCTACC
								TTTTTAG
					-30T	1	-	AATAGTGTGAATTCAGTGTGATCTACCTTTTTAG
					-21A	2	T	AATTCAGTGTGATCTACCTTTTTAG
<i>CCT3</i>	11	845	-44A	2.95	-24A	6	T	GCTTCATAGTGTGTTGCTTCTCCAAG
					-10C	1	-	GTTTGCCTTCCAAG
<i>EEF1A1</i>	2	366	-24A	2.80	-28A	2	T	TAGTAAACCAAGTAACGACTCTTAATCCTTACAG
					-19C	1	-	AGTAAACGACTCTTAATCCTTACAG
<i>EEF1A1</i>	1	943	-33A	3.25	-23A	10	T	GGTTCAAAGTTTTTCTTCCATTTACAG
<i>ENO1</i>	2	2837	-33A	3.25	-27T	6	-	ATTGCTACTACATCTTTTTCTCTCATCCAG
					-25C	2	T	TGCTACTACATCTTTTTCTCTCATCCAG
<i>ENO1</i>	4	2394	-21A	3.45	-21A	10	T	CCCTCATTTCTCCCCCTCTCCCTCGTAG
<i>ENO1</i>	5	737	-32A	3.20	-27A	6	T	ACTTCAATCCACTCGGTTCTTCTGTTCTAG
					-24C	1	-	TCATTCAGTCTCGGTTCTTCTGTTCTAG
<i>ENO1</i>	6	615	-36A	2.85	-30T	2	G	CCCAGTGCATGCTTCTGCTCTGCTCTCCCCAG
					-27C	3	A	AGTGCCATGCTTCTGCTCTGCTCTCCCCAG
					-26A	3	-	GTGCCATGCTTCTGCTCTGCTCTCCCCAG
<i>ENO1</i>	7	796	-25A	3.05	-38A	3	T	TACCTACCTGTTTTCCAAACCTGTTGTCACCATC
								TCTTCCCAG
					-28C	1	-	TTTTCCAAACCTGTTGTCACCATCTCTTCCCAG
					-27A	2	T	TTTCCAAACCTGTTGTCACCATCTCTTCCCAG
					-26A	2	T	TTCCAAACCTGTTGTCACCATCTCTTCCCAG
<i>ENO1</i>	9	547	-30A	2.70	-5T	2	A	TGGCTTCCAG
<i>ENO1</i>	11	1457	-26A	2.95	-48T	4	-	AGGCTGACTTTTTCTTTTTCTCTCCCCATCTCTTACC
								TTTCTCCTTCCAAG
					-47G	2	-	GGTCTGACTTTTTCTTTTTCTCTCCCCATCTCTTT
								ACCTTTCTCCTTCCAAG
					-46A	3	T	GTCTGACTTTTTCTTTTTCTCTCCCCATCTCTTACC
								TTTCTCCTTCCAAG
<i>G22P1</i>	1	6031	-28A	2.90	-30A	2	-	AGGACAACATTTTCTTCCATTTTTTCCCCATAG
					-28A	4	T	GACAAAACATTTTCTTCCATTTTTTCCCCATAG
<i>G22P1</i>	8	3212	-26A	2.50	-31A	2	-	AAGTCAAATCAAAGAAAATTTATCTCCTTTCTTCAG
					-26A	1	T	AAATCAAAGAAAATTTATCTCCTTTCTTCAG
					-25A	1	T	AATCAAAGAAAATTTATCTCCTTTCTTCAG
<i>G22P1</i>	10	2978	-33A	3.60	-33A	10	-	GACTCACAGGCCACTCTCTGTGTTTGATT
								TCTAG
<i>GAPDH</i>	2	1633	-26A	3.35	-6T	10	-	TTGTCTCTTAG
<i>HSPA8</i>	1	734	-39A	3.15	-23A	1	T	TTTTAAACAGATTTTTCTTTTTTTTCAG
					-19A	1	-	AAACCAAGATTTTTCTTTTTTTTCAG
					-17A	1	C	ACCAGATTTTTCTTTTTTTTCAG
<i>HSPCB</i>	6	448	-18A	3.15	-22A	6	T	GTACCACTTATTTTGGTTTCTTTCAG
					-23C	1	-	TGTACCACTTATTTTGGTTTCTTTCAG
<i>HSPCB</i>	10	777	-22A	3.20	-24T	1	-	CAATCTAAGGCTTTTGTGATCGTCCACAG
					-22A	2	T	ATCTAAGGCTTTTGTGATCGTCCACAG
<i>HSPCB</i>	1	1434	-22A	2.75	-18A	1	T	AATTAATGAGATTTTATTTTAG
<i>LDHB</i>	2	7526	-22A	3.35	-24T	4	-	GGTTCTAATGCCTGTTTTGCGTTTACAG
					-23A	1	T	GTTCTAATGCCTGTTTTGCGTTTACAG
					-22A	6	T	TTCTAATGCCTGTTTTGCGTTTACAG
<i>PGK1</i>	1	5461	-28A	3.00	-29G	2	-	AAGTTGATCATGGTCTTGCATCTTTCTTTTTTAG
					-28A	4	T	AGTTGATCATGGTCTTGCATCTTTCTTTTTTAG
<i>PGK1</i>	2	3826	-35A	3.00	-26T	4	-	CATTCIGTTTGTGCTCTCTTTGGTTGCAG
<i>PGK1</i>	4	3151	-29A	3.35	-33C	1	-	GGAGCCATCACATTTCTGTTTTGTTTTCTCTA
								TAG
					-29A	5	T	CCATCACATTTCTGTTTTGTTTTCTCTATAG
<i>PGK1</i>	5	635	-32A	3.40	-29A	1	T	TGACTAGAATCTGAATGTCTTTGATCTTTCTAG

(Continued)

Table 3. Continued

Gene	Intron	Intron size (bp)	Predicted BP ^a	BPS Score ^a	Observed BP	Number of clones	Misincorporated nucleotide	Intronic sequence from BPS position -5 to the 3' end of an intron ^b
PGK1	6	4664	-27A	3.15	-22G	7	-	AATCTGAATGCTTTGATCTTTCTAG
					-21A	2	T	ATCTGAATGCTTTGATCTTTCTAG
					-28A	1	T	TCTTTAAGTGATGATTCTTGCTTTCTTTGTAG
PGK1	8	1499	-27A	3.20	-23A	8	T	AAGTGATGATTTCTTGCTTTCTTTGTAG
					-27A	10	T	AGCTCACTTCTCTTACCTCTACCCCTCAG
PGK1	10	364	-35A	2.75	-36A	10	-	ATAGTAATGCTGTCTATGTATGTGTGCTCTCTC AAAAACAG
PKM2	2	1443	-39A	3.05	-31A	2	T	AATTAATACTTGTGGCTTTAAAACCTTTTCCTAATAG
					-29A	1	T	TTAATACTTGTGGCTTTAAAACCTTTTCCTAATAG
					-25G	1	C	TACTTGTGGCTTTAAAACCTTTTCCTAATAG
					-23G	1	C	CTTGTGGCTTTAAAACCTTTTCCTAATAG
PKM2	3	6930	-21A	2.75	-25T	3	-	ACGCTTGTGTCATCTTCTCTTTTCCCCCAG
					-21A	4	T	TTGTCACTTCTCTTTTCCCCCAG
PKM2	4	487	-26A	2.85	-38T	1	-	TGGTGTCTCCAGTTGGACTCTTGCTTACTCTCTTGT CCCTAG
					-33A	1	T	TCTCCAGTTGGACTCTTGCTTACTCTCTTGTCC CTAG
					-23C	1	-	GGACTCTTGCTTACTCTTGTGCCCTAG
					-16A	1	T	TGCTTACTCTCTTGTGCCCTAG
					-8G	1	-	CTCTTGTCCCTAG
					-5C	1	-	TTGTCCCTAG
PKM2	5	781	na	na	-32T	1	-	CGTGCTGCTCCCTACTTACCCTTTTTCATACAG
					-31C	1	-	GTGCTCTGCTCCCTACTTACCCTTTTTCATACAG
					-28C	3	-	CTCTGCTCCCTACTTACCCTTTTTCATACAG
					-20A	1	T	CCCCACTTACCCTTTTTCATACAG
					-18T	1	A	CCTACTTACCCTTTTTCATACAG
					-16A	2	T	TACTTACCCTTTTTCATACAG
PKM2	6	1343	-29A	3.10	-39G	1	C	CCTCTGTTCTATATAACCTCTCTCCCCCAACTTTG TCCATCAG
					-34A	6	T	GTTCTATATAACCTCTCTCCCCCAACTTTG TCCATCAG
					-32A	2	T	TCTATATAACCTCTCTCCCCCAACTTTGTCCATCAG
					-65T	2	-	CCTTTTGTGACAAAGCTCTGACAAAGCTCTGTCCC CCTCTCGTCCCTCTGGACGGATGTTGCTCCCTAG
PKM2	8	4107	na	na	-52T	1	-	AGCTCTGACAAAGCTCTGTCCCCCTCTCGTCCCTC TGGACGGATGTTGCTCCCTAG
					-50A	5	T	CTCTGACAAAGCTCTGTCCCCCTCTCGTCCCTCTGGA CGGATGTTGCTCCCTAG
					-27T	1	-	TTTACTCACCAACCTCCCTTCTTCTCCCTCCAG
PKM2	10	717	-25A	3.75	-26C	2	-	TFACTCACCAACCTCCCTTCTTCTCCCTCCAG
					-25A	6	T	TACTCACCAACCTCCCTTCTTCTCCCTCCAG
					-44A	4	T	CTGTTATTCAGCCCAATATCCCCCATGGTTTTCC CCCAATCTCCCTAG
PSMB4	4	393	-40A	3.05	-40A	5	T	TATTCAGCCCAATATCCCCCATGGTTTTCCCCA ATCTCCCTAG
					-26A	1	C	ACCCCACTTAACTCTTCTCATTACCAACAG
RPL13	4	583	-21A	3.30	-23T	1	-	CCACTTAACTCTTCTCATTACCAACAG
					-22A	4	T	CACCTTAACTCTTCTCATTACCAACAG
					-22A	9	-	GTTTAAACAACCTGTCTTTCTTCTTAG
RPL13	5	492	-22A	3.30	-20A	1	T	TTAACACCTGTCTTTCTTCTTAG
					-22C	7	-	GAGTCCCTTTTGCCTTGTCTCCACAG
RPL13A	1	2205	-26A	3.55	-8T	1	-	TTGTCCTCCACAG
RPL3	2	770	-21A	3.70	-21A	4	T	GTCTGACTACTGCTTTTTTTTTGCGAG
					-19T	3	-	CTGACTACTGCTTTTTTTTTGCGAG
RPL3	4	1150	-22A	3.30	-24T	10	-	GGAGCTGAGCTGTGTCTACCTTCTCCTAG
RPL3	5	522	-22A	2.50	-29T	1	-	GGCGCTGAGGTGAAGTAATGTGTATCCATTCCAG
RPL3	6	477	-34A	3.45	-22T	3	-	AGCCTTACACCTTCTTGTTCATTAG
					-21A	3	T	GCCTTACACCTTCTTGTTCATTAG
RPL8	4	806	na	na	-23C	5	-	GTTCCCTGAGGTATCTGATCCCCTACAG
SLC25A3	2	1591	-30A	2.85	-31A	1	-	ATATTAAAATGCATGGTGTGCTTCTTACTACAG

(continued)

Table 3. Continued

Gene	Intron size (bp)	Predicted BP ^a	BPS Score ^a	Observed BP	Number of clones	Misincorporated nucleotide	Intronic sequence from BPS position -5 to the 3' end of an intron ^b
<i>SNRPB</i> 1	2929	-24A	3.20	-24A	1	T	GTCTC <u>A</u> TCCCTGTCCATTTCTCCTTGCAG
<i>SNRPB</i> 2	1787	-34A	3.85	-36T	2	-	ACCTC <u>A</u> AACACTTTTTTTGTTCCCTTCTAAAC CTCTCTTTAG
				-35A	5	-	CCTCT <u>A</u> AACACTTTTTTTGTTCCCTTCTAAACC TCTCTTTAG
				-34A	2	T	CTCTA <u>A</u> CACTTTTTTTGTTCCCTTCTAAAC CTCTCTTTAG
<i>SNRPB</i> 3	1808	-34A	3.10	-30G	2	-	CACTG <u>G</u> GATCAGAGCATATTTGTTTATTT TTCAG
				-29G	3	-	ACTG <u>G</u> GATCAGAGCATATTTGTTTATTT TTCAG
				-28C	1	G	CTG <u>G</u> GATCAGAGCATATTTGTTTATTTTTCAG
				-27A	2	-	TG <u>G</u> GATCAGAGCATATTTGTTTATTTTTCAG
<i>SNRPB</i> 4	519	-25A	3.75	-27T	7	-	TCTTCT <u>A</u> ACTCTTTCTTCTTATGTCCTCTTAG
				-26A	1	T	CTTCT <u>A</u> ACTCTTTCTTCTTATGTCCTCTTAG
				-25A	5	T	TTCTA <u>A</u> CTCTTTCTTCTTATGTCCTCTTAG
<i>SNRPB</i> 6	696	-25A	3.95	-27T	8	-	GGC <u>A</u> CTGACTAAACTTCTTACTCTTACTTCAG
<i>UBB</i> 1	717	-33A	3.45	-30T	6	-	TGAGG <u>T</u> GACACGCTTATGTTTTACTTTTAAA CTAG
				-29G	1	-	GAGG <u>T</u> GACACGCTTATGTTTTACTTTTAA ACTAG
				-28A	2	T	AGG <u>T</u> GACACGCTTATGTTTTACTTTTAAACTAG

^aPredicted BPs and BPS scores are according to the Branch-Site Analyzer at <http://ast.bioinfo.tau.ac.il/BranchSite.htm> (14).

^bObserved branch sites are underlined.

The information contents at the PPT's were similar between the two alignments. We observed peaks of information contents seven and eight nucleotides downstream of the branch point. The functional significance of these peaks, however, remains elusive.

Information obtained from lariat RT-PCR clones without misincorporated nucleotides

We next asked if we could exploit the 186 clones without misincorporated nucleotides at the branch point. If there was an 'A' nucleotide one or two nucleotides downstream of the sequenced branch point and the sequenced branch point was not 'A', we assumed that one or two nucleotides were skipped by the reverse transcriptase and that the particular downstream 'A' was the actual branch point. A similar assumption has also been applied to three other genes in previous reports (15–17). We aligned the branch points under this assumption, and plotted the nucleotide preferences and the information contents (Figure 2E and F). Compared to those of misincorporated nucleotides, the information contents were generally lower, but the Pictogram and WebLogo presentations resulted in similar patterns. These analyses suggest that one or two nucleotides were skipped when there were no misincorporated nucleotides, but definite experimental evidence is lacking to employ these clones to deduce the human consensus BPS.

DISCUSSION

Highly degenerative human BPS

We determined splicing branch points in 52 introns of 20 human housekeeping genes by lariat RT-PCR.

Our analysis disclosed the following features (Figure 4). First, 83% of the branch points are located 21–34 nucleotides upstream of the 3' end of an intron (Figure 3A). Second, a polypyrimidine stretch spans 4–24 nucleotides downstream of the branch point (Figure 3B and C). Third, the human branch point consensus sequence is yUnAy (Figure 2C and D). The first and the second features underscore the previous *in silico* observations (6,14), whereas the degeneracy of the human BPS is more than we have expected.

It is interesting to note that among the six consensus BPSs proposed for the mammalian branch points (Table 1), the shared nucleotides are yUnAy, which is identical to that determined by our analysis.

SF1 binds to BPS using its KH domain (18). NMR analysis of SF1 bound to the BPS revealed that a hydrophobic motif of Gly-Pro-Arg-Gly within the KH domain builds hydrogen bonds with 'UAA' at positions -2 to 0 of the yeast BPS, 'UACUAA' (19). Our analysis suggests that the binding of the KH domain to position -1 may enhance, but may be dispensable for, the recognition of the BPS. Berglund and colleagues (20) also demonstrate that, in 'UACUAA' at positions -5 to +1, nucleotide substitutions only at position -2 or 0, but not at the other positions, compromise the binding of SF1.

Non-'A' nucleotides at position 0

We observed an 'A' nucleotide at 92.3% of the branch points. Non-'A' nucleotides at the branch point have been reported in *CALCA1* (21) and *GHI* (22) (Table 2). The two reports demonstrate six such examples in

four introns. As these unusual branch points constitute 21% (6/29) of the previously reported *in vitro* determined branch points, the ratio of 'A' at the branch point is reduced to 79% (Table 2). Additionally, the potential observation bias posed by these unusual BPSs may account for the differences in the Pictogram and WebLogo patterns between the previously identified BPSs (Figure 2A and B) and our BPSs (Figure 2C and D).

Disease-causing mutations disrupting BPSs

According to the Human Gene Mutation Database (23), splicing mutations account for 13.7% (1768 of 12 879) of single nucleotide substitutions. Most splicing mutations, however, are at the splice donor or acceptor sites. To our knowledge, sixteen disease-causing mutations and a single polymorphism disrupt BPSs and give rise to aberrant splicings (Table 5). Nine variants are at position 0, and the other eight are at position -2. Among the nine variants affecting position 0, seven are A-to-G mutations, which supports the notion reported by Kralovicova and colleagues (24) that A-to-G transitions at position 0 are more deleterious than A-to-T or A-to-C transversions. For all the variants, aberrant splicings have been determined either in patients or minigenes. The actual branch points, however, have been identified only in two variants by lariat RT-PCR, whereas the remaining fourteen variants have been mapped to putative BPSs. Exclusive confinement of BPS-disrupting nucleotide changes at positions -2 and 0 also underscores our observation that the BPS consensus is $yUnAy$.

Conversely, mutations disrupting $yUnAy$ are not always deleterious. When the branch point 'A' is mutated or deleted, a neighboring cryptic 'A' residue is employed as a branch point (25–27), or the mutant 'C', 'G' or 'U' residue is used as a surrogate branch point (28). Additionally, we observed two or more branch sites in 15 of 41 introns (Table 3), which also implies that a mutation-harboring BPS can be readily substituted for by another BPS.

How is the highly degenerative BPS recognized?

It is hard to believe that SF1 simply recognizes $yUnAy$. We expect that SF1 recognizes the BPS along with the other *cis*-element(s) and their interacting *trans*-factor(s).

The SELEX screening of the yeast BBP binding motifs revealed a stem and loop structure immediately upstream of the BPS of 'UACUAAC' in 9 out of 48 selected motifs (29). A gel shift assay also showed preferential binding of human SF1 to 'UACUAAC' carrying an upstream stem and loop. Our BPSs, however, had no upstream stem and loop structures (data not shown). An upstream stem and

loop may help recognize highly degenerative mammalian BPSs for a subset of introns.

In the early step of the spliceosome assembly, SF1, U2AF65 and U2AF35 bind to the BPS, the PPT and AG at the 3' end of an intron, respectively, to form complex E (1,2). In *S. pombe*, SF1/BBP is tightly associated with U2AF59, a yeast homolog of mammalian U2AF65 recognizing the PPT, as well as with U2AF23, a yeast homolog of mammalian U2AF35 recognizing the 3' AG (30).

Table 5. Sixteen mutations and a single polymorphism disrupting BPSs

Gene and intron	Sequence	Consequence	Reference
<i>LCAT</i> intron4			
Wild-type	CCCTGAC		
Mutant	CCCCGAC	Intron retention ^a	(43,44)
Mutant	CCCGGAC	Intron retention ^b	(45)
Mutant	CCCAGAC	Intron retention ^b	(45)
<i>FBN2</i> intron30			
Wild-type	TACTAAG		
Mutant	TACGAAG	Exon skipping ^a	(46)
<i>COL5A1</i> intron32			
Wild-type	GACTGAC		
Mutant	GACGGAC	Exon skipping ^a	(47)
<i>ITGB4</i> intron31			
Wild-type	GGCTCAC		
Mutant	GGCACAC	Intron retention, ^a cryptic 3' splice site ^a	(48)
<i>TH</i> intron10			
Wild-type	GGCTGAT		
Mutant	GGCAGAT	Exon skipping, ^a cryptic 3' splice site ^a	(49)
<i>LICAM</i> intron18			
Wild-type	ATCCAAG		
Mutant	ATCCACG	cryptic 3' splice site ^a	(50)
<i>LIPC</i> intron1			
Wild-type	CCCCAAT		
Mutant	CCCCAGT	cryptic 3' splice site ^a	(51)
<i>FBN2</i> intron28			
Wild-type	TTGCAAT		
Mutant	TTGCAGT	Exon skipping ^a	(52)
<i>HEXB</i> intron10			
Wild-type	TTGCAAT		
Mutant	TTGCAGT	Cryptic 3' splice site ^a	(53)
<i>NF2</i> intron5			
Wild-type	TTCTAGC		
Mutant	TTCTAAC	Intron retention ^a	(54)
<i>TSC2</i> intron38			
Wild-type	GCGTGAC		
Mutant	GCGTGGC	Cryptic 3' splice site, ^a intron retention ^a	(55)
<i>XPC</i> intron3 ^c			
Wild-type	TACTGAT		
Mutant	TACTGGT	Exon skipping ^a	(16)
<i>NPC1</i> intron6			
Wild-type	CACTAAT		
Mutant	CACTAGT	Exon skipping ^a	(56)
<i>F9</i> intron 2			
Wild-type	CGTTAAT		
Mutant	CGTTAGT	Exon skipping ^b	(24,57)
<i>DQB1</i> intron 3 ^{c,d}			
Genotype A	CACAGAC	Exon skipping ^b	(17)
Genotype U	CACUGAC	Exon inclusion ^b	(17)

Mutations or a polymorphism are underlined.

Aberrant splicings have been determined in patients^a or minigenes.^b

^cBranch points have been identified by lariat RT-PCR. Others are putative BPSs lacking *in vitro* evidence.

^dPolymorphism.

Table 4. Nucleotide frequencies at the 181 branch sites

Position	-5	-4	-3	-2	-1	0	1	2	3
A	0.254	0.232	0.083	0.066	0.166	0.923	0.182	0.302	0.201
C	0.210	0.227	0.470	0.160	0.448	0.033	0.331	0.274	0.391
G	0.254	0.193	0.127	0.028	0.177	0.017	0.066	0.112	0.112
U	0.282	0.348	0.320	0.746	0.210	0.028	0.420	0.313	0.296

In mammals, the association between SF1 and U2AF65 is mediated by the 28 N-terminal amino acids of the KH domain of SF1(31) and by the third RBD of U2AF65 (32). Wang and colleagues determined that Ser20 in the N-terminal region of the KH domain is essential for binding to U2AF65 and that phosphorylation of Ser20 inhibits its binding and formation of complex A (33). Berglund and colleagues also report that the SF1-U2AF65 interaction promotes cooperative binding to the BPS and the PPT (32). Our analysis also demonstrates positional association of the BPS and the PPT (Figure 3B and C). On the other hand, Kent and colleagues demonstrate that U2AF65 and U2AF35 are dispensable for the binding of SF1 to the BPS (34). Sharma and colleagues similarly show that complex H includes SF1 in the absence of U2AF65 and U2AF35 (35). Although the exact order of the SF1, U2AF65 and U2AF35 assembly remains elusive, the BPS is possibly recognized along with the PPT and the 3' AG. Alternatively, SF1 is bound to any yUnAy sequences in complex H, and a particular SF1 that successfully associates with the U2AF heterodimer exclusively survives to form complex E.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We appreciate Jun Shinmi and Keiko Itano for technical assistance. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, as well as from the Ministry of Health, Labor, and Welfare of Japan. Funding to pay the Open Access publication charges for this article was provided by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832–835.
2. Zorio, D.A. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature*, **402**, 835–838.
3. Arning, S., Gruter, P., Bilbe, G. and Kramer, A. (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA*, **2**, 794–810.
4. Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, **89**, 403–412.
5. Query, C.C., Moore, M.J. and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, **8**, 587–597.
6. Will, C.L. and Lührmann, R. (2006) Spliceosome structure and function. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press Plainview, NY, pp. 369–400.
7. Query, C.C., Strobel, S.A. and Sharp, P.A. (1996) Three recognition events at the branch-site adenine. *EMBO J.*, **15**, 1392–1402.
8. Eisenberg, E. and Levanon, E.Y. (2003) Human housekeeping genes are compact. *Trends Genet.*, **19**, 362–365.
9. Vogel, J., Hess, W.R. and Borner, T. (1997) Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res.*, **25**, 2030–2031.
10. Burge, C.B., Tuschl, T. and Sharp, P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 525–560.
11. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
12. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
13. Rogan, P.K. and Schneider, T.D. (1995) Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum. Mutat.*, **6**, 74–76.
14. Kol, G., Lev-Maor, G. and Ast, G. (2005) Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum. Mol. Genet.*, **14**, 1559–1568.
15. Guo, N. and Kawamoto, S. (2000) An intronic downstream enhancer promotes 3' splice site usage of a neural cell-specific exon. *J. Biol. Chem.*, **275**, 33641–33649.
16. Khan, S.G., Metin, A., Gozukara, E., Inui, H., Shahnavi, T., Muniz-Medina, V., Baker, C.C., Ueda, T., Aiken, J.R., Schneider, T.D. *et al.* (2004) Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum. Mol. Genet.*, **13**, 343–352.
17. Kralovicova, J., Houngrinou-Molango, S., Kramer, A. and Vorechovsky, I. (2004) Branch site haplotypes that control alternative splicing. *Hum. Mol. Genet.*, **13**, 3189–3202.
18. Berglund, J.A., Fleming, M.L. and Rosbash, M. (1998) The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA*, **4**, 998–1006.
19. Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngrinou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. and Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, **294**, 1098–1102.
20. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, **89**, 781–787.
21. Adema, G.J., Bovenberg, R.A., Jansz, H.S. and Baas, P.D. (1988) Unusual branch point selection involved in splicing of the alternatively processed Calcitonin/CGRP-I pre-mRNA. *Nucleic Acids Res.*, **16**, 9513–9526.
22. Hartmuth, K. and Barta, A. (1988) Unusual branch point selection in processing of human growth hormone pre-mRNA. *Mol. Cell. Biol.*, **8**, 2011–2020.
23. Krawczak, M., Ball, E.V., Fenton, I., Stenson, P.D., Abeyasinghe, S., Thomas, N. and Cooper, D.N. (2000) Human gene mutation database—a biomedical information and research resource. *Hum. Mutat.*, **15**, 45–51.
24. Kralovicova, J., Lei, H. and Vorechovsky, I. (2006) Phenotypic consequences of branch point substitutions. *Hum. Mutat.*, **27**, 803–813.
25. Reed, R. and Maniatis, T. (1985) Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell*, **41**, 95–105.
26. Ruskin, B., Greene, J.M. and Green, M.R. (1985) Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell*, **41**, 833–844.
27. Padgett, R.A., Konarska, M.M., Aebi, M., Hornig, H., Weissmann, C. and Sharp, P.A. (1985) Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. *Proc. Natl Acad. Sci. USA*, **82**, 8349–8353.
28. Hornig, H., Aebi, M. and Weissmann, C. (1986) Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing in vitro. *Nature*, **324**, 589–591.
29. Garrey, S.M., Voelker, R. and Berglund, J.A. (2006) An extended RNA binding site for the yeast branch point-binding protein and

- the role of its zinc knuckle domains in RNA binding. *J. Biol. Chem.*, **281**, 27443–27453.
30. Huang, T., Vilardell, J. and Query, C.C. (2002) Pre-spliceosome formation in *S.pombe* requires a stable complex of SF1-U2AF(59)-U2AF(23). *EMBO J.*, **21**, 5516–5526.
 31. Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA*, **4**, 551–565.
 32. Berglund, J.A., Abovich, N. and Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.*, **12**, 858–867.
 33. Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P.J., Kramer, A. and Robinson, P.J. (1999) Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J.*, **18**, 4549–4559.
 34. Kent, O.A., Ritchie, D.B. and Macmillan, A.M. (2005) Characterization of a U2AF-independent commitment complex (E') of the mammalian spliceosome assembly pathway. *Mol. Cell. Biol.*, **25**, 233–240.
 35. Sharma, S., Falick, A.M. and Black, D.L. (2005) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell*, **19**, 485–496.
 36. Langford, C.J. and Gallwitz, D. (1983) Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. *Cell*, **33**, 519–527.
 37. Pikielny, C.W., Teem, J.L. and Rosbash, M. (1983) Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for UI RNA and metazoan mRNA splicing. *Cell*, **34**, 395–403.
 38. Green, M.R. (1986) Pre-mRNA splicing. *Annu. Rev. Genet.*, **20**, 671–708.
 39. Zhang, Y., Goldstein, A.M. and Weiner, A.M. (1989) UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc. Natl Acad. Sci. USA*, **86**, 2752–2756.
 40. Harris, N.L. and Senapathy, P. (1990) Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.*, **18**, 3015–3019.
 41. Zhang, M.Q. (1998) Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.*, **7**, 919–932.
 42. Nelson, K.K. and Green, M.R. (1989) Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev.*, **3**, 1562–1571.
 43. Kuivenhoven, J.A., Weibusch, H., Pritchard, P.H., Funke, H., Benne, R., Assmann, G. and Kastelein, J.J. (1996) An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). *J. Clin. Invest.*, **98**, 358–364.
 44. Li, M. and Pritchard, P.H. (2000) Characterization of the effects of mutations in the putative branchpoint sequence of intron 4 on the splicing within the human lecithin:cholesterol acyltransferase gene. *J. Biol. Chem.*, **275**, 18079–18084.
 45. Li, M., Kuivenhoven, J.A., Ayyobi, A.F. and Pritchard, P.H. (1998) T→G or T→A mutation introduced in the branchpoint consensus sequence of intron 4 of lecithin:cholesterol acyltransferase (LCAT) gene: intron retention causing LCAT deficiency. *Biochim. Biophys. Acta*, **1391**, 256–264.
 46. Maslen, C., Babcock, D., Raghunath, M. and Steinmann, B. (1997) A rare branch-point mutation is associated with missplicing of fibrillin-2 in a large family with congenital contractural arachnoidactyly. *Am. J. Hum. Genet.*, **60**, 1389–1398.
 47. Burrows, N.P., Nicholls, A.C., Richards, A.J., Luccarini, C., Harrison, J.B., Yates, J.R. and Pope, F.M. (1998) A point mutation in an intronic branch site results in aberrant splicing of COL5A1 and in Ehlers-Danlos syndrome type II in two British families. *Am. J. Hum. Genet.*, **63**, 390–398.
 48. Chavanas, S., Gache, Y., Vailly, J., Kanitakis, J., Pulkkinen, L., Uitto, J., Ortonne, J. and Meneguzzi, G. (1999) Splicing modulation of integrin beta4 pre-mRNA carrying a branch point mutation underlies epidermolysis bullosa with pyloric atresia undergoing spontaneous amelioration with ageing. *Hum. Mol. Genet.*, **8**, 2097–2105.
 49. Janssen, R.J., Wevers, R.A., Haussler, M., Luyten, J.A., Steenbergen-Spanjers, G.C., Hoffmann, G.F., Nagatsu, T. and Van den Heuvel, L.P. (2000) A branch site mutation leading to aberrant splicing of the human tyrosine hydroxylase gene in a child with a severe extrapyramidal movement disorder. *Ann. Hum. Genet.*, **64**, 375–382.
 50. Rosenthal, A., Jouet, M. and Kenwrick, S. (1992) Aberrant splicing of neural cell adhesion molecule L1 mRNA in a family with X-linked hydrocephalus. *Nat. Genet.*, **2**, 107–112.
 51. Brand, K., Dugi, K.A., Brunzell, J.D., Nevin, D.N. and Santamarina-Fojo, S. (1996) A novel A→G mutation in intron 1 of the hepatic lipase gene leads to alternative splicing resulting in enzyme deficiency. *J. Lipid Res.*, **37**, 1213–1223.
 52. Putnam, E.A., Park, E.S., Aalfs, C.M., Hennekam, R.C. and Milewicz, D.M. (1997) Parental somatic and germ-line mosaicism for a FBN2 mutation and analysis of FBN2 transcript levels in dermal fibroblasts. *Am. J. Hum. Genet.*, **60**, 818–827.
 53. Fujimaru, M., Tanaka, A., Choeh, K., Wakamatsu, N., Sakuraba, H. and Isshiki, G. (1998) Two mutations remote from an exon/intron junction in the beta-hexosaminidase beta-subunit gene affect 3'-splice site selection and cause Sandhoff disease. *Hum. Genet.*, **103**, 462–469.
 54. De Klein, A., Riegman, P.H., Bijlsma, E.K., Helder, A., Muijtjens, M., den Bakker, M.A., Avezaat, C.J. and Zwarthoff, E.C. (1998) A G→A transition creates a branch point sequence and activation of a cryptic exon, resulting in the hereditary disorder neurofibromatosis 2. *Hum. Mol. Genet.*, **7**, 393–398.
 55. Mayer, K., Ballhausen, W., Leistner, W. and Rott, H. (2000) Three novel types of splicing aberrations in the tuberous sclerosis TSC2 gene caused by mutations apart from splice consensus sequences. *Biochim. Biophys. Acta*, **1502**, 495–507.
 56. Di Leo, E., Panico, F., Tarugi, P., Battisti, C., Federico, A. and Calandra, S. (2004) A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in Niemann-Pick type C disease. *Hum. Mutat.*, **24**, 440.
 57. Ketterling, R.P., Drost, J.B., Scaringe, W.A., Liao, D.Z., Liu, J.Z., Kasper, C.K. and Sommer, S.S. (1999) Reported in vivo splice-site mutations in the factor IX gene: severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. *Hum. Mutat.*, **13**, 221–231.

Novel AChR δ subunit mutation causing myasthenia hinders intersubunit link essential for channel gating

Xin-Ming Shen,¹ Taku Fukuda,¹ Kinji Ohno,¹ Steven M. Sine,² and Andrew G. Engel¹

¹Muscle Research Laboratory, Department of Neurology; and ²Receptor Biology Laboratory, Department of Physiology and Biophysics, Mayo Clinic, Rochester, MN 55905

Dr. Fukuda's current address is First Department of Internal Medicine, Nagasaki University Hospital, Nagasaki, Japan; Dr. Ohno's current address is Department of Neurogenetics and Bioinformatics, Nagoya University Graduate School of Medicine, Nagoya, Japan.

Address correspondence to Andrew G. Engel, Department of Neurology, Mayo Clinic, Rochester MN 55905. Phone: 507-284-5102; Fax: 507-284-5831; E-mail: age@mayo.edu.

Nonstandard abbreviations used in text: CMS, congenital myasthenic syndrome, EP, endplate; EPP, endplate potential; HEK, human embryonic kidney; MEPC, miniature EP current; MEPP miniature EP potential.

Abstract

We trace the cause of a congenital myasthenic syndrome associated with endplate AChR deficiency and low amplitude synaptic potentials to the hetero-allelic mutations δ L42P and δ I58K. We find that δ I58K abolishes AChR expression by impairing association of α - and δ -subunits, whereas δ L42P enables low expression but severely attenuates channel gating, thus determining the phenotype. AChRs containing δ L42P exhibit shortened bursts of channel openings elicited by a low ACh concentration, predicting accelerated decay of the synaptic response, and a reduced probability the di-liganded AChR channel will open, accounting for reduced response amplitude. Substitutions of Gly, Lys, and Asp for δ L42, or of Pro for δ N41, δ I43, and δ S44 similarly attenuate channel gating. Mutation of residues equivalent to δ L42 in the ϵ , α , and β subunits impede, do not significantly affect, and enhance the response to ACh, indicating the mutated region contributes to channel gating in a subunit specific manner. Thermodynamic mutant cycle analyses reveal strong energetic coupling between δ L42 and Y127 from the opposing α -subunit, which forms an inter-subunit linkage with δ N41 known to be crucial for channel gating. Our overall findings indicate δ L42P perturbs its constituent β -strand of the δ -subunit, disrupting the inter-subunit linkage between δ N41 and α Y127 required for rapid and efficient channel gating.

Introduction

Congenital myasthenic syndromes (CMSs) are heterogeneous disorders caused by defects in presynaptic, synaptic, or postsynaptic proteins (1). Targets for CMS mutations include choline acetyltransferase (2), the collagenic tail subunit of the end-plate species of acetylcholinesterase (3, 4), rapsyn (5), the $\text{Na}_v1.4$ sodium channel (6), the muscle-specific receptor tyrosine kinase (MuSK) (7), Dok-7, the muscle-intrinsic activator of MuSK (8), and the nicotinic acetylcholine receptor (AChR) (9). Most postsynaptic CMS are caused by mutations in different subunits of AChR that alter its kinetics of activation by ACh, reduce its expression, or both.

AChR is a ligand-gated ion channel with a pentameric structure consisting of four homologous subunits with a stoichiometry $\alpha_2\beta\delta\gamma$ in the fetus or $\alpha_2\beta\delta\varepsilon$ in the adult. Each AChR subunit has a large N-terminal extracellular domain, four transmembrane domains, an extracellular linker and two intracellular domains (10). The extracellular domain of each subunit consists largely of β -strands that twist to form a β -sandwich, creating a conserved hydrophobic core within each subunit and highly complementary subunit interfaces, two of which form the ligand binding sites (11). The β_1 through β_6 strands compose the inner sheets of the β -sandwich, with the β_7 through β_{10} strands forming the outer sheets. A disulfide bond of the signature cysteine loop links the two sheets together (11, 12). At interfaces that form the agonist binding sites, the β_1 strand from the inner sheets of the δ - and ε -subunits approaches the junction of the inner and outer sheets of the neighboring α -subunit. At this inter-subunit interface, Asn 41 in the δ subunit and the equivalent Asn 39 in the ε subunit couple energetically with Tyr 127 in the α -subunit to form an inter-subunit link essential for channel gating (13). However, whether other residues in β_1 strands of the δ and ε subunits contribute to the inter-subunit link is not known, and to date no naturally occurring mutations have been detected in any AChR β_1 strand.