ARTICLE

Micro-exons of the cardiac myosin binding protein C gene: flanking introns contain a disproportionately large number of hypertrophic cardiomyopathy mutations

Rune Frank-Hansen[1,3], Stephen P Page[2], Petros Syrris[2], William J McKenna[2], Michael Christiansen[1] and Paal Skytt Andersen*,[1]

[1]Department of Clinical Biochemistry, Statens Serum Institut, Copenhagen, Denmark; [2]The Heart Hospital, University College London, London, UK

Hypertrophic cardiomyopathy is primarily caused by mutations in genes encoding cardiac sarcomere proteins. Large screening studies identify mutations in 35–65% of the diagnosed patients and 15–30% of these are discovered within the *MYBPC3* gene encoding the cardiac myosin binding protein C. The aim of this study is to determine whether intronic variation flanking the three micro-exons in *MYBPC3* is disease-causing. Two hundred and fifty unrelated patients with hypertrophic cardiomyopathy were genotyped in *MYBPC3*, using automated single-strand conformation polymorphism, and sequenced for confirmation. Mutations located in the flanking introns of the *MYBPC3* micro-exons were examined using *in silico* methods. Ectopic expression of mRNA in blood leukocytes in the respective patients was examined using reverse transcription-PCR. A total of seven mutations were discovered in the introns flanking the two micro-exons 10 and 14, but none were found in introns flanking exon 11. Functional studies together with co-segregation analysis indicate that four mutations are associated with HCM, in the respective patients. All four mutations result in premature termination codons, which suggests that haploinsufficiency is a pathogenic mechanism of this type of mutation. It is demonstrated that the use of *in silico* methods together with RNA studies on peripheral blood leukocytes is a useful tool to evaluate the potential effects of mutations on pre-mRNA splicing.

Introduction

Mutations in the gene encoding myosin binding protein C (*MYBPC3*) are one of the most common causes of hypertrophic cardiomyopathy (HCM), generally accounting for 15–30% of all identified HCM mutations.[1–3]

All recent studies have concluded that large-scale mutation screenings of HCM patients, where eight or nine sarcomeric genes are screened, identify mutations in 35–60% of diagnosed patients.[2] One reason for the relatively low detection rate could be the presence of a number of novel non-sarcomeric HCM disease genes. Another reason could be presence of several unidentified disease-causing mutations in the known disease genes.

Correspondence: Dr PS Andersen, Department of Clinical Biochemistry, Statens Serum Institut, 5 Artillerivej, Copenhagen S DK-2300, Denmark.
Tel: +45 3268 8190; Fax: +45 3268 3878;
E-mail: psa@ssi.dk
[3]Present address: Section of Forensic Genetics, Department of Forensic Medicine, University of Copenhagen, Copenhagen, Denmark

Missense mutations are the most common type of mutation in HCM, but in *MYBPC3* the majority of the more than 150 known mutations are insertions, deletions, splice site and nonsense mutations and most of them lead to premature termination codons (PTCs). This suggests that these mutations could constitute a separate disease-causing mechanism in *MYBPC3*-associated HCM, and therefore make *MYBPC3* a candidate for the discovery of novel domains and mutations that could be involved in generating aberrant transcripts. The reason for difference in the type of mutation is not known, but we and others have speculated that these mutations lead to nonsense-mediated decay (NMD) and subsequent haploinsufficiency rather than malfunctioning truncated proteins.[1,4]

One way of generating transcripts with frameshifts and ultimately PTCs is through aberrant splicing. Several examples of mutations of both acceptor and donor splice sites in *MYBPC3* that can produce aberrant transcripts have already been demonstrated in HCM.[1,4-7]

The difficulty with intronic mutations is, however, to determine whether they are disease-causing. In cases where it is possible to study the co-segregation of the genotype with the HCM phenotype in a family and proper controls are available, the task is alleviated; however, the precise effect of the mutation or the severity of the splicing defect remain undetected. Several factors determine the inclusion rate of an exon and therefore mutations can have a varying effect on the resulting pre-mRNA splicing and the effect may be difficult to detect.

The distribution of exon lengths suggests functional constraints, with 65% of exons having sizes between 68 and 208 nt.[8] Optimal splicing occurs at sizes between 50 and 500 nt with limits being dependent on the size of the adjacent introns.[9-11] Nevertheless, it is estimated that 10% of all exons are shorter than 60 nt,[8] and some genes carry exons below 25 nt, also termed micro-exons.[12] It is well known that short exons have intrinsically poor splice signals and require strong accessory signals from the flanking intron region.[13,14] The presence of micro-exons has been demonstrated in several genes in different organisms such as potato,[15] Drosophila[8] and humans.[16] Of the cardiac sarcomere protein genes, the troponin genes (*TNNT2* and *TNNI3*) carry micro-exons. Micro-exons have also been described for the nuclear cell adhesion molecule L1 (*NCAML1*)[17] and one of the GABA receptors (*GABABR1*).[18] In *GABABR1*, the micro exon 4 has been identified as a hotspot for regulated alternative splicing,[18] whereas in other genes the micro-exons appear constitutive. The micro-exons appear to be well conserved among vertebrates, at least in the case of *MYBPC3* and *TNNT2*.

Owing to the unusual small size of micro-exons, steric hindrance prevents proper exon recognition and splicing by the spliceosome.[11] Expression studies of micro-exons within heterologous minigene constructs have identified the need of both intronic context, in the form of *cis*-acting

splice signals,[13] and in some cases the presence of the native upstream exon.[10] It is believed that splicing is achieved by the spliceosome recognizing the micro exon together with the upstream intron and exon as a single unit, thus cleaving the downstream intron first. Mutations in the introns flanking micro-exons could, therefore, be expected to generate unpredictable transcripts.

The *MYBPC3* gene contains three micro-exons, exons 10, 11 and 14. Exon 10 and 14 are conserved in murine and canine sequences and both consist of 3 nt. In the course of routine mutation-screening of 250 unrelated HCM patients, we have discovered a disproportionately high number of mutations in the flanking introns of exons 10 and 14. No mutations were discovered in the flanking introns of exon 11, which corresponds to previous reports of HCM mutations.[1-3]

The purpose of this study was to use *in silico* methods together with the study of ectopic mRNA expression in peripheral blood leukocytes to determine the precise effect of micro exon mutations on pre-mRNA splicing.

Methods
Patients
Two hundred and fifty patients diagnosed with HCM from The Heart Hospital, London, UK, were included in the study and blood samples were collected for purification of genomic DNA. Patients who were carrying micro exon mutations were recalled and asked to provide a second blood sample for isolation of peripheral blood leukocytes. DNA samples from 192 Caucasians (cat. nos. HCR-1 and -2, European Collection of Cell Cultures (ECACC), Porton Down, UK) were used as controls. RNA purified from peripheral blood leukocytes provided by a single donor was used as control in all reverse transcription-PCRs. Total RNA purified from the heart of a single donor was used as control in all reverse transcription-PCRs (Human Heart Total RNA; (cat. no. 540011, Stratagene, La Jolla, CA, USA).

Genotyping
Genomic DNA was extracted from EDTA-blood using the QiaAmp® DNA Blood Mini Kit (Qiagen, Germany). The coding regions of *MYBPC3* were amplified using intronic primers according to Niimura *et al*[19] and Andersen *et al*.[1] Primer sequences and PCR conditions are available on request (psa@ssi.dk). Mutation detection of amplicons was performed by automated single-strand conformation polymorphism on an Applied Biosystems ABI PRISM® 3100[20,21] and the presence of mutations confirmed by automated cycle sequencing (Applied Biosystems, Foster City, CA, USA). All patients carrying intron 9, 10, 13 or 14 mutations were screened in nine other known HCM disease genes (*MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *MYL2*, *MYL3*, *ACTC*, *PLN* and *CSRP3*) by automated single-strand conformation polymorphism and DNA sequencing as described.[22]

In silico analysis

The potential effect of the mutations was evaluated by calculating Shapiro–Senapathy (S&S) splice site scores[23,24] (http://www.genet.sickkids.on.ca/~ali/splicesitescore.html). Neural network predictions of splice sites were performed using NetGene2 (www.cbs.dtu.dk/netgene2/)[25,26] and Alternative Splice Site Predictor (ASSP).[27]

RNA analysis

In the case where more than one index patient carried a specific intron mutation, only one proband was recalled to provide a second blood sample for RNA. Peripheral blood leukocytes were isolated from a fresh blood sample and stabilized in RNAlater™ (Ambion, Austin, TX, USA) and RNA was extracted using the Ambions RiboPure™ Blood Kit. Reverse transcription-PCR was performed using Qiagen® One-Step RT-PCR protocol with the following primers: exon 10: forward primer (exons 7 and 8) 5′-CCTT CCGCCGCACGAG-3′, reverse primer (exons 11 and 12) 5′-GCTTCGAGTCCCTCGG-3′, exon 14: forward primer (exon 12) 5′-GGGCATGAGGCGCGATGAGA-3′, reverse primer (exon 16) 5′-TCCAAGGGGCGCGTGATGAG-3′. Reverse transcription-PCR products were visualized on 2% agarose gels. All products were sequenced. All reverse transcription-PCR products that included exon 10 were cloned using Invitrogen™ (Carlsbad, CA, USA) TOPO TA cloning® according to the manufacturer's protocols and sequenced for confirmation.

Clinical characterization

The cohort consisted of 250 consecutive, unrelated patients who fulfilled diagnostic criteria for HCM based upon a maximal wall thickness of $\geqslant 15$ mm, in the absence of a hemodynamic cause.[28,29] All patients underwent full clinical evaluation including 12-lead electrocardiogram (ECG) and transthoracic echocardiogram (ECHO).

Sequences

MYBPC3 genomic reference sequence (GenBank accession no. U91629) and *MYBPC3* cDNA reference sequence (GenBank accession no. NM_000256.2) were used.

Results

In 250 unrelated HCM patients a total of seven mutations were discovered near the two micro-exons 10 (AGA) and 14 (CAA) (g.6651–53 and g.10387–9, respectively). Five of these were novel and four unique for the patient in whom they were found (Table 1). The c.1226 + 49C > T variant is a common polymorphism seen in both normal controls and patients at equal frequencies. None of the six other mutations were seen in UK Caucasian (ECACC HCR-1 and -2) control groups. One of the mutations was identified in three unrelated index patients (c.1224−19G > A) of a different ethnic background (Greek Cypriot, Indian and British, respectively) in the study. Only one of these was contacted for subsequent RNA studies. All patients were screened for nine other HCM disease genes without finding any other disease associated mutations. The c.1224−2A > G mutation has previously been published.[2] No mutations were discovered in the intronic area flanking exon 11.

Splice site prediction

The potential effect of the discovered mutations was assessed using three different methods for prediction of splice sites and the scores obtained are presented in Table 2. The reference sequence for *MYBPC3* (GenBank acc. no. U91629) was altered at the respective mutated sites and the same mutation data were analyzed with NetGene2 and ASSP as well as with an online algorithm for the calculation of S&S scores. Predictions were calculated for the authentic splice sites and the effect of the mutation on the particular splice site (mutated splice site) as well as for activated

Table 1 Allelic frequencies and transcriptional consequence of variants flanking micro-exons in *MYBPC3*

Pedigree ID	Allelic variant[a]	Allele	FHC patients Frequency	Caucasians Frequency	RNA[a] Effect	Transcriptional Consequence
Exon10						
730	c.906−36G > A	A	1/500	0/384	r.906−34_906-1ins	Frameshift, PTC
	c.906−15G > C	C	1/500	0/384	None	—
917	c.906−1G > C	C	1/500	0/384	r.906_907del	Frameshift, PTC
	c.908+39G > A	A	1/500	0/384	None	—
Exon14						
896	c.1224−19G > A[b]	A	3/500	0/384	r.1224−17_1224-1ins	Frameshift, PTC
447	c.1224−2A > G[c]	G	1/500	0/384	r.1224_27del	Frameshift, PTC
	c.1226+49C > T	T	2/500	2/384	None	—

[a]Numbering according to GenBank Acc. NM_000256.2 (ATG = 1).
[b]Discovered in three unrelated index patients.
[c]Previously published by Richard *et al*, 2003.

Table 2 Splice site predictions for exons 10, 11 and 14 and flanking aberrant splice sites

Location	Mutation	Splice site	Sequence[a]	S&S score	ASSP	NetGene2
Exon 10		Authentic 3'ss	ctgctgtggcccag|AGA|gtaaga	78.7	*2166*	NI
		Authentic 5'ss	ctgctgtggcccag|AGA|**gt**aaga	74.1	8532	0.45
	c.906-1G>C	Mutated 3'ss	ctgctgtggccc<u>a</u>c|AGA|gtaaga	62.6	NI	NI
		Cryptic 3'ss	ctgctgtggccca**c**ag|A|gtaaga	81	5095	NI
		Mutated 5'ss	ctgctgtggccca<u>c</u>|AGA|**gt**aaga	74.1	9172	0.46
	c.906-36G>A	De Novo 3'ss	tccctctcccccc<u>ag</u>|CCCCTCC	86	9923	0.95
Exon 11		Authentic 3'ss	cccctggagc**ag**|CAGTTT	70.6	*0.241*	NI
		Authentic 5'ss	TCCGGACCCCGAG|**gt**gagt	96.7	15196	0.82
Exon 14		Authentic 3'ss	cttctgttctac**ag**|CAA|gtaagt	89.6	9936	NI
		Authentic 5'ss	cttctgttctacag|CAA|**gt**aagt	87.6	13567	0.86
	c.1224-2A>G	Mutated 3'ss	cttctgttcta**gg**|CAA|gtaagt	54.8	*-1026*	NI
		Mutated 5'ss	cttctgttctac<u>g</u>|CAA|**gt**aagt	87.6	13136	0.93
	c.1224-19G>A	De Novo 3'ss	gaacacttcaac**ag**|CCCCTTC	70.8	*-1438*	NI

[a]Mutated nucleotide positions are underlined in sequence, calculated values are for splice sites in bold.
ASSP values in italics are below default cutoff values, 3'-ss = 2.2 and 5'-ss = 4.5.
NI: Not identified.
Scores are given for the authentic splice site (ss) and the effect the mutation has on the particular site (mutated splice site). The value is also given for the cryptic splice site that can then be activated. The *de novo* splice site scores are given for comparison to the competing authentic splice site.

cryptic splice sites and *de novo* splice site for comparison with the competing authentic splice site.

Exon 10 NetGene2 was unable to detect the authentic acceptor splice sites of exon 10, and subsequently unable to detect any change brought on by the splice site mutation c.906−1G>C. Both ASSP and the S&S scores identify the correct acceptor splice site. The ASSP prediction of the authentic exon 10 acceptor splice site is marginally below the default cutoff provided. ASSP predicts a 1 nt exon 10 as the consequence of the c.906−1G>C mutation as does the S&S score. A *de novo* acceptor splice site is the predicted outcome of c.906−36G>A by all models, tentatively extending the 5′-end of exon 10, thus including a large portion of intron 9 in the mRNA (34 nt). The c.906−15G>C and c.908+39G>A mutations did not change the likelihood of induced cryptic splicing and did not affect any known splicing mechanisms (data not shown).

Exon 11 All models correctly identify the donor splice site of exon 11. Both ASSP and S&S scores detect the acceptor splice site, with ASSP providing a score below the normal default. NetGene2 is unable to detect the acceptor splice site.

Exon 14 NetGene2 was unable to detect the acceptor splice site of exon 14 or any change brought on by the splice site mutation c.1224−2A>G. ASSP and S&S scores identify exon 14 and predict the disruption of the acceptor splice site (c.1224−2A>G). None of the three models predict activation of a cryptic splice site. As the only model, S&S scores predict the formation of a *de novo* splice

site as the consequence of c.1224−19G>A the outcome of which is a 17 nt extension of the 5′-end of exon 14.

RNA analysis
In the absence of cardiac biopsies, peripheral blood leukocytes from the respective patients were used for the detection of aberrant splice variants. In all reverse transcription-PCRs, RNA purified from control blood leukocytes and total RNA from human heart were included to ensure that inherent differences in leukocyte splicing mechanisms compared to that of cardiac tissue would not be interpreted as false positive effects of the discovered mutations. The mRNA from the patients carrying seven different mutations was examined by reverse transcription-PCR. The results are summarized in Table 1.

Of the four mutations flanking exon 10, c.906−36G>A and c.906−1G>C were predicted to cause aberrant splicing by two of the prediction models. The assay demonstrated that the upstream G>A transition created a *de novo* acceptor splice site ((g>a)g|CCCC) and resulted in the inclusion of 34 nt of intron 9 into exon 10 (Figure 1). Exon 10 maintained the correct donor splice site, thus resulting in a new exon of 37 nt. This inclusion produced a frameshift and eventually resulted in a PTC in exon 12 (c.993_995). The c.906−1G>C mutation disrupted the existing 3′ splice site and activated the neighboring cryptic 3′ splice site positioned 2 nt downstream of the existing splice site. Reverse transcription-PCR and subsequent cloning demonstrated that the mutation resulted in exclusion of the first two bases of exon 10, thus yielding an exon 10 of only one base (A) (Figure 1). This change produced a frameshift and eventually resulted in a PTC in
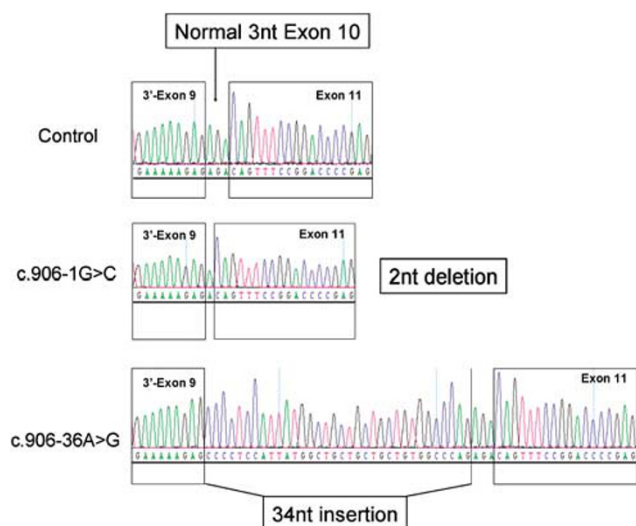
Figure 1 Exon 10 RT-PCR results. cDNA from a control and two patients was cloned and sequenced. The 3′-end of exon 9 and all of exon 11 are represented by boxes. Exon 10 is seen to be spliced correctly in the normal control. The c.906−1G>C mutation results in skipping of 2 nt of exon 10. c.906−36A>G activates a cryptic splice-site resulting in a 34 nt intron inclusion.

exon 12 (c.993_995). No change in the splicing of exon 10 was detected in reverse transcription-PCR as a result of c.906−15G>C or c.908+39G>A mutations.

Three mutations were found near exon 14, one of which is a known polymorphism (c.1226+49A>T). This polymorphism was not predicted to produce any change in the splicing of exon 14 and this was confirmed by reverse transcription-PCR. The upstream c.1224−19G>A transition produced a *de novo* acceptor splice site and extended the transcript by 17 nt, thus introducing a frameshift and ultimately a PTC in exon 15 (c.1294_1296). The splice site mutation c.1224−2A>G was demonstrated to disrupt the authentic splice site of exon 14, but with no cryptic splice site activation as seen in exon 10. Instead the mutation generates a 4 nt deletion (r.1224_1227del) in the transcript, comprised of exon 14 (3 nt) and the most 5′-nucleotide (G) of exon 15, resulting in a frameshift and PTC in exon 15 (c.1294_1296).

Clinical characterization
Family 917 The c.906−1G>C mutation was discovered in a 57-year-old female (III.5) presenting asymmetrical septal hypertrophy (ASH) with systolic anterior motion (SAM). Further genetic analysis revealed that the patient was compound heterozygous with a second mutation in *MYBPC3*, a V1125M missense mutation. The only other family member to carry both mutations, the 64-year-old brother, presented with ASH, right bundle branch block (RBBB) and left atrium dilatation. Two family members (II.2, IV.2), mother and son of the index patient, carried only the c.906-1G>C mutation. The 94-year-old mother
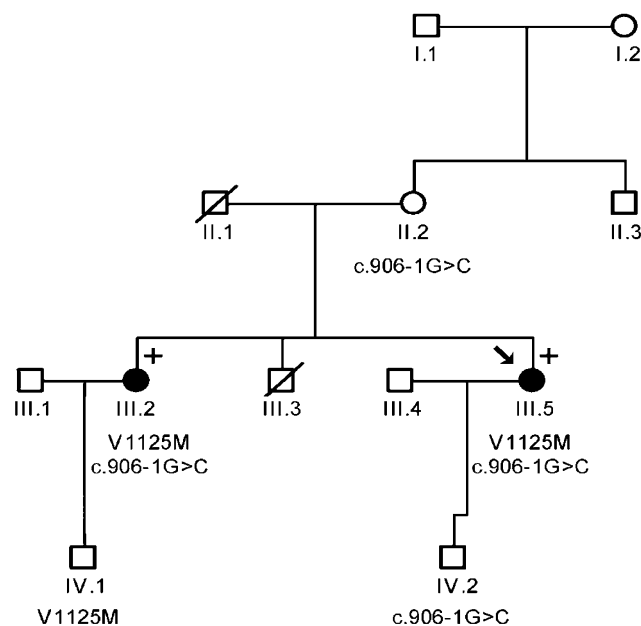

Figure 2 Pedigree of family 917. (+) indicates clinical disease status. Arrow indicates proband. The proband III5 and her sister III.2 carry both the c.906−1G>C and the V1125M mutation. II.2 and IV.2 carry only the c.906−1G>C mutation, IV.1 carries only the V1125M mutation.

(II.2) had a borderline diagnosis with an abnormal ECG, more specifically T wave inversion in the lateral leads and abnormal Q waves in the high lateral leads (I and aVL), but normal ECHO. The son (26 years, IV.2) of the index patient was unaffected with normal ECG and ECHO. The nephew of the index patient, having a borderline diagnosis of hypertrophy, was the only person carrying the V1125M missense mutation by itself. See Figure 2 for pedigree.

Family 730 The c.906−36G>A mutation was discovered in a 64-year-old female presenting left ventricular hypertrophy (LVH), ASH, chest pain on exertion as well as provocable left ventricular outflow tract obstruction (LVOTO). Two daughters were found to be carriers, but were both asymptomatic at ages 34 and 25, respectively.

Family 896 The c.1224−19G>A mutation was discovered in a 32-year-old male of Greek Cypriot descent who had mild hypertrophy and severe LVOTO treated with left ventricular myomectomy. Blood samples from the proband as well as family members were examined. This mutation was also identified in two other unrelated index patients of different ethnic background (Indian and British, respectively) in the study, both of whom had a relatively mild form of the disease.

Family 447 The c.1224−2A>G mutation was discovered in a 70-year-old male with LVH and ASH with LVOTO. The patient has a permanent pacemaker, and has experienced

transient ischemic attacks and paroxysmal atrial fibrillation. A son and daughter were found to be asymptomatic carriers at ages 36 and 39, respectively.

Discussion

The potential effects of mutations located in the flanking introns of the two *MYBPC3* micro-exons, exons 10 and 14, were examined using three different *in silico* methods. NetGene2 was not useful in evaluating the potential effect of mutations in these regions, only predicting the effect of the c.906−36G>A substitution, while not identifying any of the authentic acceptor splice sites. This is most likely due to the fact that micro-exons have almost no statistical signal that allows either software or neural networks such as NetGene2 to recognize their presence.[12] With the exception of NetGene2, the acceptor splice sites of all three micro-exons of MYBPC3 were recognized. All three models correctly predicted the donor splice sites. Comparison of the predictions and the examination of RNA extracted from peripheral blood leukocytes revealed that S&S scores and predictions from ASSP were a reliable way to predict the outcome of a potential mutation. The S&S algorithm only provides a score for the sequence of choice, and cutoff values or confidence intervals are not given. This leaves a significant amount of decisions for personal evaluation, which is not useful for larger studies. Of the three models, only the S&S scores gave any prediction of the *de novo* splice site in intron 13, in spite of it appearing as a functional acceptor splice site in the reverse transcription-PCR. This may reflect a weak *de novo* splice site and reflect that only a fraction of the pre-mRNA will be incorrectly spliced. Nevertheless, the combined result of reverse transcription-PCR and the fact that the mutation was discovered in three unrelated index patients and not seen in any controls suggest that c.1224−19G>A is a potential disease-causing mutation.

The c.1224−2A>G is predicted to disrupt the exon 14 acceptor splice site; however, it results in a 4 nt deletion of the mRNA and thereby a frameshift and eventual PTC in exon 15. The result raises interesting questions regarding splicing within this region of *MYBPC3*. Inspection of the adjacent exon 13 donor and exon 15 acceptor splice sites revealed that the deletion must be composed of exon 14 (3 nt) and the first nucleotide of exon 15; however, this requires that the downstream intron 14 must be spliced before intron 13 (Figure 3). Intron 14 splicing should proceed by utilizing the proper intron 14 splice sites, and in the absence of the authentic exon 14 3′ splice site, the subsequent exon–exon junction between exons 14 and 15 activates a cryptic acceptor splice site resulting in cleavage of the most 5′ nucleotide of exon 15.

The splicing is, therefore, not sequential and proceeds in a manner previously described for the Troponin I micro
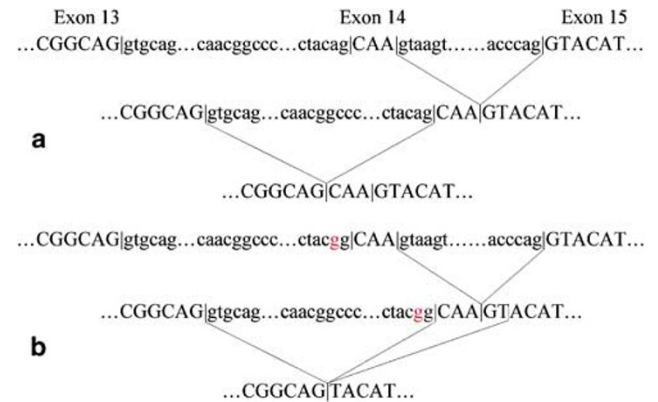

Figure 3 Exon 14 splicing. (**a**) Normal micro-exon 14 splicing. Exon 13, intron 13 and exon 14 are recognized as a single exon structure with cleavage of intron 14 proceeding first. The resulting joining of exons 14 and 15 then facilitates the recognition of intron 13, which is subsequently cleaved producing the correct mRNA. (**b**) The c.1224−2A>G mutation (marked in red) does not interfere with the intron 14 cleavage. Owing to the mutation, there is no longer a true intron 13 3′-ss. The nearest cryptic splice site is the one produced by the joining of the 3′-end of exon 14 and the 5′-end of exon 15 (AG|TACA) (marked in red). This is then used to produce the new mRNA with a 4 nt deletion.

exon.[10] Also described in the latter paper was the retention of the upstream intron as a result of disruption of the upstream 5′ splice site, instead of the expected exon skipping. The consequence of the 5′ splice site disruption is therefore that the initial recognition of the exon–intron–exon unit is maintained after splicing of the downstream intron. The mechanism of splicing was regulated by a downstream intronic sequence enhancing the recognition of the 5′ splice site of the micro exon.[13] This appears to be a sound premise as the enhancer sequence will provide a preference for the micro exon 5′ splice site *versus* the upstream 5′-splice site. This enhancer element is removed as a result of the initial splicing of the downstream intron, whereas an upstream splicing regulation would be maintained post-splicing.

The mode of micro exon splicing requires that two adjacent 5′-splice sites are utilized in reverse order. This observation fits the dynamic model of exon definition, where the phosphorylated C-terminal domain (CTD) of RNA polymerase II (RNAPII) is believed to bind U1 snRNPs through a complex with the human ortholog of yeast Prp40p and is capable of tethering multiple 5′-splice sites as they emerge from the RNA channel.[30] With an enhancer present to ensure the use of the micro exon 5′-splice site, the downstream intron is spliced first, leaving the upstream donor splice site tethered to the CTD. With the downstream enhancer and associated U1 snRNP removed, U2AF can associate with the acceptor splice site of the micro exon, completing splicing of the upstream intron.

The use of peripheral blood leukocytes in the genetic diagnosis of HCM has been in use since Rosenzweig *et al*[31]

first reported the presence of β-myosin heavy chain transcripts in circulating blood lymphocytes and showed its use in preclinical diagnosis. In the initial publications of *MYBPC3*-associated HCM, both normal and mutated transcripts were observed in blood lymphocytes of patients with *MYBPC3* mutations[5,6] and comparison of the transcripts present in both blood and cardiac tissue of a patient with an exon 31 donor splice site mutation of *MYBPC3* revealed that the same mRNA species could be observed.[4] The c.906−36G>A, c.906−1G>C, c.1224−19G>A and c.1224−2A>G mutations were all demonstrated to cause aberrant transcripts, and no use of alternative splice sites was detected in the examined exons 10 and 14 of the two controls.

All mutations in the studied regions seemed to be associated with mild and late-onset hypertrophy, and only in the case of c.906−1G>C together with the missense mutation V1125M, in itself seemingly benign, did the patient appear to present with severe hypertrophy. Cardiac biopsies were not available, and therefore the outcomes of the mutations both at the RNA and protein levels remain speculative. A recent report by Sarikas *et al*[32] demonstrated the impairment of the ubiquitin-protease system (UPS) of truncated cardiac MYBPC potentially resulting in competitive inhibition of other UPS substrates. It should also be noted that NMD should remove transcripts with PTCs that are upstream of the NMD boundary, which is located 50–55 bases 5′ of the last exon–exon junction.[33] There are, however, examples of different levels of degradation of nonsense mRNA in various cell types,[34] making it unclear to what extent nonsense transcripts are degraded as well as the amount of residual mRNA. Both mechanisms should, by the way of degradation of either mRNA or protein, result in haploinsufficiency, and of any residual truncated cardiac MYBPC only the exon 14 mutations would retain a complete MYBPC motif (S2-binding). The MYBPC3 N-terminus interaction with the S2 region of myosin (K_d of 5 μM)[35] is such that low levels of truncated protein should be insufficient to compete with normal filament-associated protein.[36] Therefore, any damage of a potential poisonous peptide would be limited.

In a clinical genetic setting, the discovery of a mutation prompts three questions: (1) Is the mutation known? (2) Does it co-segregate with the disease in the family? and (3) Is it a conserved residue? We have demonstrated that the use of *in silico* methods together with the stabilization and long-term storage of peripheral blood leukocytes for later use in RNA studies is useful tool to evaluate the potential effects of mutations on pre-mRNA splicing. Therefore, the practice of drawing blood in RNA-stabilizing solutions should be at least considered when performing genetic screening or beginning larger screening studies. The presence of disease-causing mutations at positions up to 36 nt from the coding regions of *MYBPC3* has only been discovered due to the current positioning of PCR primers,

and this warrants further examination of intronic variation and has implications for the general approach to genetic screening. Characterization of the regulation involved in micro exon splicing in *MYBPC3*, as well as *TNNT2* and *TNNI3*, could aid future mutation studies in the correct identification of potential disease-causing mutations. Particularly, identifying to what extent downstream intronic elements or adjacent exons regulate micro exon splicing in *MYBPC3* would provide further insights. An understanding of this mechanism could also aid in the improved *in silico* prediction of micro-exons and provide further knowledge on the occurrence of micro-exons in other genes. In *MYBPC3*-associated HCM, we suspect that haploinsufficiency is one of the main causes of disease. Therefore, a detailed study of the almost 17 kb of intronic sequence of *MYBPC3* seems a natural step in the genetic diagnosis of HCM as opposed to the screening of novel candidate genes.

References
1 Andersen PS, Havndrup O, Bundgaard H *et al*: Genetic and phenotypic characterization of mutations in myosin-binding protein C (MYBPC3) in 81 families with familial hypertrophic cardiomyopathy: total or partial haploinsufficiency. *Eur J Hum Genet* 2004; **12**: 673–677.
2 Richard P, Charron P, Carrier L *et al*: Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. *Circulation* 2003; **107**: 2227–2232.
3 Van Driest SL, Vasile VC, Ommen SR *et al*: Myosin binding protein C mutations and compound heterozygosity in hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2004; **44**: 1903–1910.
4 Rottbauer W, Gautel M, Zehelein J *et al*: Novel splice donor site mutation in the cardiac myosin-binding protein-C gene in familial hypertrophic cardiomyopathy. Characterization Of cardiac transcript and protein. *J Clin Invest* 1997; **100**: 475–482.
5 Watkins H, Conner D, Thierfelder L *et al*: Mutations in the cardiac myosin binding protein-C gene on chromosome 11 cause familial hypertrophic cardiomyopathy. *Nat Genet* 1995; **11**: 434–437.
6 Bonne G, Carrier L, Bercovici J *et al*: Cardiac myosin binding protein-C gene splice acceptor site mutation is associated with familial hypertrophic cardiomyopathy. *Nat Genet* 1995; **11**: 438–440.
7 Erdmann J, Raible J, Maki-Abadi J *et al*: Spectrum of clinical phenotypes and gene variants in cardiac myosin-binding protein C mutation carriers with hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2001; **38**: 322–330.
8 Mount SM: Genomic sequence, splicing, and gene annotation. *Am J Hum Genet* 2000; **67**: 788–792.
9 Sterner DA, Carlo T, Berget SM: Architectural limits on split genes. *Proc Natl Acad Sci USA* 1996; **93**: 15081–15085.
10 Sterner DA, Berget SM: *In vivo* recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol Cell Biol* 1993; **13**: 2677–2687.
11 Dominski Z, Kole R: Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* 1991; **11**: 6075–6083.

12 Volfovsky N, Haas BJ, Salzberg SL: Computational discovery of internal micro-exons. *Genome Res* 2003; **13**: 1216–1221.

13 Carlo T, Sterner DA, Berget SM: An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA* 1996; **2**: 342–353.

14 Carlo T, Sierra R, Berget SM: A 5′ splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon. *Mol Cell Biol* 2000; **20**: 3988–3995.

15 Simpson CG, Hedley PE, Watters JA *et al*: Requirements for mini-exon inclusion in potato invertase mRNAs provides evidence for exon-scanning interactions in plants. *RNA* 2000; **6**: 422–433.

16 Carrier L, Bonne G, Bahrend E *et al*: Organization and sequence of human cardiac myosin binding protein C gene (MYBPC3) and identification of mutations predicted to produce truncated proteins in familial hypertrophic cardiomyopathy. *Circ Res* 1997; **80**: 427–434.

17 De Angelis E, Brummendorf T, Cheng L, Lemmon V, Kenwrick S: Alternative use of a mini exon of the L1 gene affects L1 binding to neural ligands. *J BiolChem* 2001; **276**: 32738–32742.

18 Holter J, Davies J, Leresche N, Crunelli V, Carter DA: Identification of two further splice variants of GABABR1 characterizes the conserved micro-exon 4 as a hot spot for regulated splicing in the rat brain. *J Mol Neurosci* 2005; **26**: 99–108.

19 Niimura H, Bachinski LL, Sangwatanaroj S *et al*: Mutations in the gene for cardiac myosin-binding protein C and late-onset familial hypertrophic cardiomyopathy. *N Engl J Med* 1998; **338**: 1248–1257.

20 Larsen LA, Christiansen M, Vuust J, Andersen PS: High-throughput single-strand conformation polymorphism analysis by automated capillary electrophoresis: robust multiplex analysis and pattern-based identification of allelic variants. *Hum Mutat* 1999; **13**: 318–327.

21 Andersen PS, Jespersgaard C, Vuust J, Christiansen M, Larsen LA: High-throughput single strand conformation polymorphism mutation detection by automated capillary array electrophoresis: validation of the method. *Hum Mutat* 2003; **21**: 116–122.

22 Hougs L, Havndrup O, Bundgaard H *et al*: One third of Danish hypertrophic cardiomyopathy patients with MYH7 mutations have mutations [corrected] in MYH7 rod region. *Eur J Hum Genet* 2005; **13**: 161–165.

23 Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987; **15**: 7155–7174.

24 Senapathy P, Shapiro MB, Harris NL: Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol* 1990; **183**: 252–278.

25 Brunak S, Engelbrecht J, Knudsen S: Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J Mol Biol* 1991; **220**: 49–65.

26 Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 1996; **24**: 3439–3452.

27 Wang M, Marin A: Characterization and prediction of alternative splice sites. *Gene* 2006; **366**: 219–227.

28 McKenna WJ, Spirito P, Desnos M, Dubourg O, Komajda M: Experience from clinical genetics in hypertrophic cardiomyopathy: proposal for new diagnostic criteria in adult members of affected families. *Heart* 1997; **77**: 130–132.

29 Maron BJ, McKenna WJ, Danielson GK *et al*: American College of Cardiology/European Society of Cardiology clinical expert consensus document on hypertrophic cardiomyopathy. A report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents and the European Society of Cardiology Committee for Practice Guidelines. *J Am Coll Cardiol* 2003; **42**: 1687–1713.

30 Goldstrohm AC, Greenleaf AL, Garcia-Blanco MA: Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* 2001; **277**: 31–47.

31 Rosenzweig A, Watkins H, Hwang DS *et al*: Preclinical diagnosis of familial hypertrophic cardiomyopathy by genetic analysis of blood lymphocytes. *N Engl J Med* 1991; **325**: 1753–1760.

32 Sarikas A, Carrier L, Schenke C *et al*: Impairment of the ubiquitin-proteasome system by truncated cardiac myosin binding protein C mutants. *Cardiovasc Res* 2005; **66**: 33–44.

33 Nagy E, Maquat LE: A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 1998; **23**: 198–199.

34 Bateman JF, Freddi S, Nattrass G, Savarirayan R: Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage. *Hum Mol Genet* 2003; **12**: 217–225.

35 Gruen M, Gautel M: Mutations in beta-myosin S2 that cause familial hypertrophic cardiomyopathy (FHC) abolish the interaction with the regulatory domain of myosin-binding protein-C. *J Mol Biol* 1999; **286**: 933–949.

36 Moolman JA, Reith S, Uhl K *et al*: A newly created splice donor site in exon 25 of the MyBP-C gene is responsible for inherited hypertrophic cardiomyopathy with incomplete disease penetrance. *Circulation* 2000; **101**: 1396–1402.