Mutation Research 642 (2008) 86–89

Contents lists available at ScienceDirect

Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

Short communication

Evaluation of the flanking nucleotide sequences of sarcomeric hypertrophic cardiomyopathy substitution mutations

Kathryn M. Meurs*, Katrina L. Mealey

Department of Veterinary Clinical Sciences, Washington State University College of Veterinary Medicine, Grimes Street, Pullman, WA 99164, United States

ARTICLE INFO

Article history:
Received 14 February 2008
Received in revised form 15 April 2008
Accepted 16 April 2008
Available online 24 April 2008

Keywords:
Hypertrophic cardiomyopathy
Mutation
Methylation
Deamination

ABSTRACT

Hypertrophic cardiomyopathy (HCM) is a familial myocardial disease with a prevalence of 1 in 500. More than 400 causative mutations have been identified in 13 sarcomeric and myofilament related genes, 350 of these are substitution mutations within eight sarcomeric genes. Within a population, examples of recurring identical disease causing mutations that appear to have arisen independently have been noted as well as those that appear to have been inherited from a common ancestor. The large number of novel HCM mutations could suggest a mechanism of increased mutability within the sarcomeric genes. The objective of this study was to evaluate the most commonly reported HCM genes, beta myosin heavy chain (MYH7), myosin binding protein C, troponin I, troponin T, cardiac regulatory myosin light chain, cardiac essential myosin light chain, alpha tropomyosin and cardiac alpha-actin for sequence patterns surrounding the substitution mutations that may suggest a mechanism of increased mutability. The mutations as well as the 10 flanking nucleotides were evaluated for frequency of di-, tri- and tetranucleotides containing the mutation as well as for the presence of certain tri- and tetranculeotide motifs. The most common substitutions were guanine (G) to adenine (A) and cytosine (C) to thymidine (T). The CG dinucleotide had a significantly higher relative mutability than any other dinucleotide ($p < 0.05$). The relative mutability of each possible trinucleotide and tetranucleotide sequence containing the mutation was calculated; none were at a statistically higher frequency than the others. The large number of G to A and C to T mutations as well as the relative mutability of CG may suggest that deamination of methylated CpG is an important mechanism for mutation development in at least some of these cardiac genes.

1. Introduction

Hypertrophic cardiomyopathy (HCM) is a primary myocardial disease characterized by increased left ventricular mass and wall thickness in the absence of a pressure overload or metabolic stimulus [1]. It has a prevalence of 1 in 500 and is believed to be familial in at least 60% of the cases, usually inherited as an autosomal dominant trait [2]. More than 400 causative mutations have been identified in 13 sarcomeric and myofilament related genes, however, the majority of the mutations are substitution mutations found in the coding regions of eight genes that encode for sarcomeric proteins [2,3]. Within a population, examples of recurring identical disease causing mutations that appear to have arisen independently have been noted as well as those that appear to have been inherited from a common ancestor [4–7].

The large number of novel substitution mutations observed within these eight sarcomeric genes could suggest a mechanism of increased gene mutability. The most commonly cited mechanisms for development of endogenous mutations are chemical (deamination of CpG regions), physical (DNA slippage) or enzymatic (post-replication repair) and the efficiency of all of these processes appears to be dependent on the local DNA sequence environment [8]. Specifically, evidence from sequence analysis and mutation rates of human DNA sequences and from early studies of the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/ac/index.php) suggests that the identity of neighboring bases can have an influence on both the type and rate of mutation events that occur at specific positions within a gene [9,10].

The objective of this study was to evaluate the eight most commonly reported hypertrophic cardiomyopathy genes, beta myosin heavy chain (MYH7), myosin binding protein C (MYBPC3), troponin I (TNNI3), troponin T (TNNT2), cardiac regulatory myosin light chain (MYL2), cardiac essential myosin light chain (MYL3), alpha tropomyosin (TPM1) and cardiac alpha-actin (ACTC) for sequence

* Corresponding author. Tel.: +1 509 335 0817; fax: +1 509 335 0880.
E-mail address: Meurs@vetmed.wsu.edu (K.M. Meurs).

0027-5107/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrfmmm.2008.04.005

Table 1
Nucleotide altered within the coding region of each gene

	ACTC	MYH7	MYBPC3	TNNI3	TNNT2	TPM1	MYL2	MYL3
Transcript size (bp)	1134	5925	4304	708	882	1105	811	868
Number of mutations	8	190	71	24	31	11	10	5
A	2	71	7	4	8	3	3	2
C	2	33	19	8	7	1	2	1
G	4	37	40	10	11	3	4	2
T	0	49	5	2	5	4	1	0

Table 2
Total number of each type of substitution mutation for the eight cardiac sarcomeric genes

G > A	113
C > T	56
A > G	37
G > C	30
T > C	23
G > T	21
C > G	18
A > T	12
C > A	13
A > C	12
T > G	8
T > A	7

patterns surrounding the reported substitution mutations to identify a pattern for increased mutability.

2. Materials and methods

The number of known substitution mutations and the coding sequences for the MYBPC3, MYH7, TNNI3, TNNT2, TPM1, ACTC, MYL2 and MYL3 genes were acquired from the CardioGenomics (http://cardiogenomics.med.harvard.edu/home) and UCSC Genome Bioinformatics data bases (http://genome.ucsc.edu/) [11,12]. The number and type of mutation was sorted according to the nucleotide mutated, the type of substitution and the di-, tri-, and tetranucleotides containing each mutation. The number of each type of di-, tri- and tetranucleotide containing the mutation was divided by the number of times that the same sequence appeared in the coding sequence of the gene to calculate the relative mutability for that sequence and multiplied by 100 to present it as a percentage. The mutability of each gene was determined by dividing the number of substitution mutations in each gene by the size of the transcript in nucleotides and multiplying by 100 to present it as a percentage.

The potential methylation status of each CG dinucleotide was assessed by scanning each gene for CpG islands that suggest unmethylated sites using the UCSC Genome Bioinformatics data base (http://genome.ucsc.edu/) [12].

Finally, the 10 nucleotides immediately preceding and following the mutation were evaluated for the presence of previously reported trinucleotide and tetranucleotide motifs that are thought to predispose to mutation development including TTT, CTT, TGA, TTG, CTTT, TCTT and TTTG [9]. The frequency of each of these was divided by the total number of mutations (350) and multiplied by 100 to determine the frequency (in percentage) for which each of these was observed to be associated with a substitution mutation.

A one-way ANOVA was used to assess statistical significance of the relative mutability for each dinucleotide, trincucleotide and tetranucleotide. A Pearsons correlation was used to assess the correlation between the relative mutability of each gene and its transcript size, as well as the relative mutability of each gene and the relative CG mutability. A Chi squared analysis was performed to determine if the observed frequency of the seven tri- and tetranucleotides motifs within the flanking regions of each mutation occurred at a greater frequency than expected. An alpha of $p \leq 0.05$ was considered to be significant.

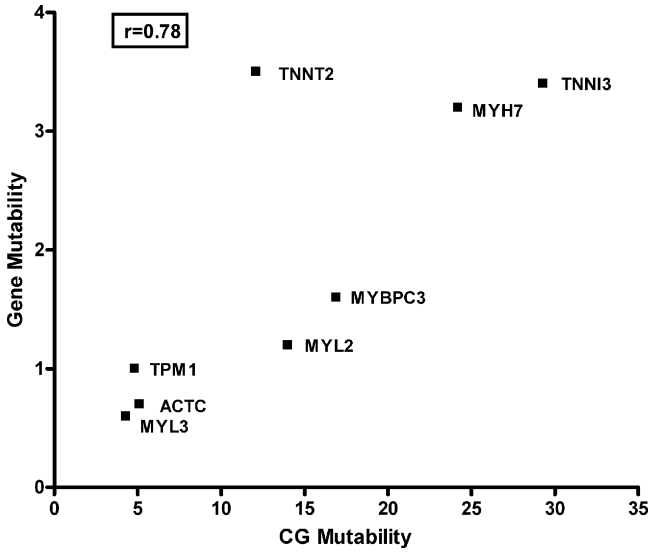

Fig. 1. The mutability of each gene was determined by dividing the number of substitution mutations in each gene by the size of the transcript in nucleotides. The relative mutability of the CG dinucleotide was calculated by dividing the number of CGs in each gene that contained a mutation by the number of times that dinucleotide appears in the coding region of each gene. A Pearson's correlation was performed to determine if the gene mutability correlated with CG mutability, and an alpha of < 0.05 was considered to be significant. The relative mutability of the individual genes was found to correlate with CG mutability ($p < 0.05$, $r = 0.78$).

3. Results

A total of 350 substitution mutations were evaluated including 190 in MYH7, 71 in MYBPC3, 31 in TNNT2, 24 in TNNI3, 11 in TPM1, 10 in MYL2, 8 in ACTC and 5 in MYL3 (Table 1). The most common substitutions were G > A (113) and C > T (56) (Table 2).

The CG dinucleotide had a significantly higher relative mutability than any other dinucleotide ($p < 0.05$) and had the highest mutability in the TNNI3 gene (Table 3). Relative mutability of the individual genes correlated with CG mutability ($p < 0.05$, $r = 0.78$), but not the size of transcript (NS, $r = 0.27$) (Figs. 1 and 2). CpG islands suggesting unmethyled CG sites were identified at exon 27 of MYH7, exon 24 in the MYBPC3 gene and exons 3–5 in TNNI3. A total of three C > T or G > A mutations occurred in these regions. The MYL3, MYL2, ACTC, TNNT2 and TPM1 genes did not have any CpG islands in the exonic regions.

The relative mutability of the trinucleotides and tetranucleotides containing the mutation were calculated; none were observed to occur statistically more frequently than the others.

Of the previously reported trinucleotide or tetranucleotide motifs that have been associated with mutation development when

Table 3
Relative gene mutability and relative CG mutability for each gene

Gene	Transcript size (bp)	Number of mutations	Gene mutability	Number of CG mutations	Number of CG sites in transcript	CG mutability
TNNT2	882	31	3.5	4	33	12.1
TNNI3	708	24	3.4	12	41	29.3
MYH7	5925	190	3.2	58	240	24.2
MYBPC3	4304	71	1.6	33	195	16.9
MYL2	811	10	1.2	4	29	14.0
ACTC	1134	8	0.7	2	39	5.1
TPM1	1105	11	1.0	2	41	4.8
MYL3	868	5	0.6	1	23	4.3