Structure and Evolution of the Alternately Spliced Fast Troponin T Isoform Gene*

(Received for publication, February 2, 1999, and in revised form, April 6, 1999)

Elizabeth A. Bucher, Gurtej K. Dhoot, Mark M. Emerson, Margaret Ober, and Charles P. Emerson, Jr.

From the Department of Cell and Developmental Biology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6058

The vertebrate fast skeletal muscle troponin T gene, TnTf, produces a complexity of isoforms through differential mRNA splicing. The mechanisms that regulate splicing and the physiological significance of TnTf isoforms are poorly understood. To investigate these questions, we have determined the complete sequence structure of the quail TnTf gene, and we have characterized the developmental expression of alternatively spliced TnTf mRNAs in quail embryonic muscles. We report the following: 1) the quail TnTf gene is significantly larger than the rat TnTf gene and has 8 non-homologous exons, including a pectoral muscle-specific set of alternatively spliced exons; 2) specific sequences are implicated in regulated exon splicing; 3) a 900-base pair sequence element, composed primarily of intron sequence flanking the pectoral muscle-specific exons, is tandemly repeated 4 times and once partially, providing direct evidence that the pectoral-specific TnT exon domain arose by intragenic duplications; 4) a chicken repeat 1 retrotransposon element resides upstream of this repeated intronic/pectoral exon sequence domain and is implicated in transposition of this element into an ancestral genome; and 5) a large set of novel isoforms, produced by regulated exon splicing, is expressed in quail muscles, providing insights into the developmental regulation, physiological function, and evolution of the vertebrate TnTf isoforms.

The troponin (Tn)1 complex subunit proteins, troponin T (TnT), troponin I (TnI), and troponin C (TnC), interact with tropomyosin (Tm) and actin in the thin filament and are the Ca2+-sensitive switch for muscle contraction (for review see Refs. 1–3). TnTf, TnI, and TnC are each encoded by small gene families that encode functionally related isoforms. Vertebrate TnT, which is the focus of this report, is encoded by three genes that are differentially expressed in fast, slow, and heart muscles (2). Alternative mRNA splicing generates additional isoforms encoded by each of these three genes (4–7). TnTf isoforms resulting from alternative mRNA splicing are differentially expressed during development and in different muscle types, indicating that these forms have specialized functions in muscle contraction.

Although TnT tethers the Tn complex to the thin filament, the regulatory functions of TnTf are arguably the least understood of the thin filament proteins. Physiological and biochemical studies (see, for example, Refs. 8–14) and the discovery of TnTf mutations in various organisms, including in humans, suggest previously unanticipated functions for TnT (15–17). These include regulation of the Ca2+ responsiveness of contraction, sarcomere assembly, and actin-myosin cross-bridge kinetics (18). That different TnTf protein domains also contribute to the functional diversity of contractile regulation in vivo is indicated by observations by ourselves and others (4–7, 16, 19, 20) of a remarkable number of TnTf isoforms. The fast TnTf isoform gene produces alternatively spliced variants that alter the length and acidity of the N-terminal domain, with largely unknown functional consequences (for reviews see Refs. 2, 3, and 21). Recent studies of the human hypertrophic cardiomyopathy mutation, I79N, close to residues that are hypervariable among TnTf isoforms, suggests that N-terminal isoform heterogeneity influences myosin-actin kinetics (18). Alternative exons located near the C terminus of the protein encode a domain that interacts with TnC, TnI, and Tm, providing further evidence that alternatively spliced exons encode TnTf domains that modulate its function in specific muscles (9, 11, 22, 23).

In this study, we have determined the complete sequence structure of the quail (Coturnix coturnix japonica) TnTf gene and a large set of TnTf cDNAs. These data provide a basis to investigate TnTf isoform diversity and to undertake analysis of the regulation and functions of specific TnTf exons during fast skeletal muscle development. Comparison of the quail and rat TnTf gene structures has provided new information on the functional diversity of TnTf isoforms, on the evolutionary origin of alternatively spliced exons, and on splice junction sequences that are hypothesized to regulate TnTf alternative exon splicing.

EXPERIMENTAL PROCEDURES

RNA Isolation and cDNA Analysis of TnTf Isoforms—RNA from day 10 embryonic leg, day 7 post-hatch pectoral, and 5-week post-hatch leg
and pectoral muscles was purified by acid guanidinium thiocyanate and phenol/chloroform extraction (24) and subjected to reverse transcription PCR as follows. RNA from the different tissue sources was reverse-transcribed into cDNA using the oligo(dT) primer (5'-GACTGAGTC- GACATCGATTCTTTTTTTTTTTTTTTTTT-3'). TnTf cDNAs were then amplified from the chain reaction using a sense primer having sequences that would anneal to exon 2/3 sequences that are upstream of the 5'-differentially spliced exons (5'-GGGAAGCTTGATA- TCACCGAGGAGGACGACCGG-3') and an antisense primer that would anneal to exon 11/12 sequences that are downstream of the differentially spliced exons (5'-GGGCTGAGAAGTCGCTGCTGCT- TCTAACCTCCCTC-3'). Restriction endonuclease sites (in italics) for the sense primer were HII (and EcoRI) and antisense (PstI and BamHI) primers included for subcloning into pBluescript KS+ (26). Recombinant clones representative of the differently sized classes were selected for DNA sequence analysis (27). Each of the cDNA sequences was submitted to GenBank™ and have the following accession numbers: AF139109 (qLeg2), AF139117 (qLeg2a), AF139118 (qLeg3), AF139119 (Pec1), AF139120 (qPec2), AF139121 (qPec3), AF139122 (qPec4), AF139123 (qPec5), AF139124 (qPec6), AF139125 (qPec7), AF139126 (qPec8), and AF139127 (qPec9).

Results

Sequence Analysis of a Large Set of Quail TnT Isoform cDNAs—Many TnT isoforms are generated by alternative exon splicing of avian and mammalian mRNA transcripts. We utilized reverse transcription PCR to examine the quail TnTf N-terminal domain diversity to 1) define differential usage of alternatively spliced exons in functionally specialized muscles; 2) compare avian and mammalian isoform expression as a basis to understand the evolution and function of isoform diversity; and 3) define exon structure in relation to TnTf mRNAs generated by alternative mRNA splicing.

We generated 5' end cDNA clones from day 10 embryonic leg, day 7 post-hatch leg, and 5-week post-hatch leg and pectoral muscles (see “Experimental Procedures”; Fig. 1) to identify a set of 5' cDNA sequences that would be representative of the N-terminal diversity found in different muscle types. Fifty four independent qTnTf cDNAs were isolated, sequenced, and compared with sequences of three quail fast TnTf isoform cDNAs that we had reported previously (6). This analysis established that these primers amplified TnTf isoform cDNAs and identified isoforms not previously reported.

Comparison of Quail, Chicken, and Mammalian Fast TnT Isoforms—Cloning of the qTnTf cDNAs enabled detailed comparison of quail TnT predicted amino acid sequences to chicken, rabbit, and rat fast isoform sequences. Comparison of the exon structures and the amino acids encoded by different exon sequences highlights that most TnTf constitutively spliced exons (i.e., spliced into all mature mRNAs) are highly conserved (exons 1, 2, 5, 9–15, 18). Differentially spliced exons 4, 7, and 17 are also highly conserved between rat and quail, whereas the differentially spliced exons (exons w, p1–5, y, 8, 16, and f) are divergent in their sequences (Fig. 2). The highly conserved exons encode domains of TnTf for which some biochemical functions have been ascribed, whereas the exact functions of the N terminus and the biochemical or physiological consequences of the alterations in charge and length of the N terminus are not well understood (1–3, 34).

Quail TnTf N-terminal Variant Isoforms—Comparisons of the 57 cDNA sequences established that they represent 14 different mRNA splice forms, including 8 novel forms. These sequences also enabled the unambiguous determination of the genomic intron/exon structure. The predicted N-terminal genomic sequence of each cDNA form is shown in Table I. Sequence analysis of 33 cDNAs recovered from pectoral mRNA demonstrates that 31 of these RNA encode a histidine-rich peptide (AHHHEE) repeated four times and a fifth exon with the sequence AHAHEE in quail, chicken, and mammalian mRNA segments.

Sequence analysis of 33 cDNAs recovered from pectoral RNA strongly supported the existence of these pectoral isoforms in quail; however, such clones were not recovered from pectoral cDNA libraries (6). Consistent with earlier protein expression studies and studies of mRNA expression from ourselves and others (“Discussion”), we designate these isoforms pectoral (qPec) (see below).

The pectoral cDNAs fall into 11 different cDNA classes, 9 having the His-rich peptide (qPec1–9) and 2 being more similar to the leg-type sequences. Thus, these latter two forms were called qLeg2 and qLeg2a (Table I). Although qLeg2 is not represented in our limited quail leg cDNA set, it is equivalent to a reported chicken leg form (38). The nine qPec forms vary by the inclusion or exclusion of peptides encoded by exons 4, w, 7, and 8. The 31
pectoral cDNAs that bear the His-rich peptide were identical for inclusion of p1–5. This is significant since Schachat and colleagues (38) reported chicken pectoral isoforms that have variant numbers of repeats of the His-rich peptide including a form that has two more His peptide repeats than we find in the qTnTf gene (38). Finally, consistent with protein studies (39), leg forms are found in our pectoral cDNA sample, but these were recovered at a relatively low frequency (3/33). The postembryonic pectoral forms show a transition to forms lacking exon 4, which is considered an embryonic exon (38).

We also examined 11 cDNA sequences recovered from embryonic day 10, and 9 cDNA sequences recovered from postembryonic leg mRNA. These 20 leg sequences fell into four different classes. All these leg clones have in common that they lack the His-rich pectoral exons, and most lack exon w, but include alternatively spliced exon 4.

Clones cC501 and cC605 were previously recovered from embryonic and adult leg cDNA libraries, respectively (6). Interestingly, of the 12 sequences recovered from embryonic leg, 10 are the C501 N-terminal form. This suggests that this is the predominant form expressed in leg mRNA. Similarly, of the 10 cDNAs sequenced from 5-week posthatch leg, 5 are identical to cC605, suggesting that this is the predominant form at this stage.

**Developmental Expression of qTnTf Isoforms**—We compared our set of clones with a set of 40 chicken cDNAs (20 adult pectoral and 20 adult gastrocnemius) that was reported by Ogut and Jin (40). The set of chicken cDNAs from Schachat and co-workers (38) was made from mRNA isolated from fetal (16 and 19 day embryos), perinatal (day 5), and adult chicken pectoral muscles and fell into 16 independent, alternatively spliced sequences that the authors placed into two broad classes based on length (Table I). Class I sequences were shorter in length, since they encoded neither the His-rich peptide sequence nor exon y, and were predominant in fetal and neonatal muscles. In contrast, class II sequences contained either of two length variants of the His-rich peptide and always encoded exon w and sometimes exon y. Exons 7 and 4 varied in either class cDNA. Class I sequences represented 50% of the forms in fetal pectoral muscle, with decreasing representation in neonatal and adult muscle samples (20%). In contrast, class II sequences have lower relative representation in fetal and neonatal muscles but are predominant in adult pectoral muscles. Exon y was only identified in class II fetal samples.

Similar to their findings, we identified the His-rich peptide sequences in cDNAs isolated from pectoral mRNA (36, 40); however, we observed no variation in the size of the His-rich peptide included, suggesting a difference between the coding potentials of the chicken and quail TnTf genes (38). Since the larger His-rich peptide was observed in chicken fetal muscle, it is possible that a larger variant exists in the quail and that
TABLE I
Summary of the qTnTf N-terminal isoform variants and comparison to known chicken isoforms

The abbreviations used are as follows: post-h (posthatch) columns represent posthatching stage; chick equiv. (equivalent); pred. indicates predominant. Columns 2–8 indicate exons in the combinatorial region examined; the next 4 columns indicate the stage mRNA was isolated, and the last column indicates whether an equivalent chick isoform has been reported. The rows indicate the name of the isoform represented. For each isoform, plus indicates that an exon is included, and minus indicates that an exon is excluded. Numbers represent how many times a particular isoform sequence was represented in the mRNA sample. Add is an adductor form based on primer extension sequences (6). cC501 and cC605 are cDNAs isolated from embryonic cultures and adult leg cDNA libraries (6). qLeg2f, qLeg2fa, and qLeg3 are quail forms that resemble the exon usage of the predominant mammalian forms TnT2f, TnT2fa, and TnT3 (66). Six isoforms are homologous to other chicken cDNAs (38,40), and eight are new TnT isoforms.

| Isoform name | 4 | 5 | 6 | 7 | 8 | Emb leg | Pect 7 day | Pect 5-wk | Leg 5-wk | Chick equiv. |
|--------------|---|---|---|---|---|---------|------------|-----------|----------|--------------|
| qPec1        | + | + | + | + | - | +        |            |           |          | No           |
| qPec2        | + | + | + | + | - | +        |            |           |          | Pred. fetal  |
| qPec3        | + | + | + | + | - | +        |            |           |          | No           |
| qPec4        | + | + | + | + | - | +        |            |           |          | No           |
| qPec5        | + | + | + | + | - | +        |            |           |          | No           |
| qPec6        | + | + | + | + | - | +        |            |           |          | No           |
| qPec7        | + | + | + | + | - | +        |            |           |          | 1-Neonatal   |
| qPec8        | + | + | + | + | - | +        |            |           |          | Pred. adult  |
| qPec9        | + | + | + | + | - | +        |            |           |          | pect.        |
| qLeg2f       | + | + | - | + | - | +        |            |           |          | 1 neonate adult gastroc. |
| qLeg2fa      | + | + | - | + | - | +        |            |           |          | 1 fetal      |
| qLeg3        | + | + | - | + | - | +        |            |           |          | 1 adult gastroc. 1 at all stages |
| cC605        | + | + | - | + | - | +        |            |           |          | No           |
| cC501        | + | + | - | + | - | +        |            |           |          | 9 + cC501    |

Fig. 2. Amino acid sequence alignment of avian and mammalian TnTf sequences. All known TnTf exonic sequences of rat, rabbit, human, chicken, and quail were aligned by the program MacBoxShade version 2.0.1. Black shaded regions are identical in all sequences. Gray shaded residues of uppercase letters are identical in at least 3 of the 5 sequences. Gray shaded residues in lowercase letters are conservative substitutions. Non-conservative substitutions are not highlighted. Exons are marked by slashes and numbered as described in the legend for Fig. 1, and the boundaries are based on the rat and quail genomic sequences. Exon boundaries are identical for rat and quail TnTf genes, except for exons that are unique to each species, and these species-specific exons are highlighted by the subscript m (mammalian) or q (quail), respectively. Although the rabbit cDNAs have sequences that look related to the avian fetal exon y (66), these sequences have not been identified in the rat genome, and it is not clear if the VHVP peptide is part of exon 7 or represents a homologous exon y. Thus we designate exon 7 as q and r (for rat), to represent that the exon boundaries are clear only for the quail and rat. This comparison highlights the substantial heterogeneity among mammalian and avian TnTf sequences in the N terminus and the high conservation in the remainder of the protein.

a Isoforms with exons (p1–5) that are essentially exclusive to pectoral muscle (qPec).

b Isoforms without p1–5 exons.
analysis of fetal quail muscle may identify these additional exons. However, as described below, our analysis of the genomic sequence has not identified additional His-rich encoding exons.

We did not identify a cDNA bearing exon y (Table I) that was found in the fetal chicken class II cDNA set (36); however, the quail gene has a predicted exon located between exons 7 and 8 that corresponds to y in position and sequence content (see below). Our failure to recover an isoform bearing exon y in any of our samples likely reflects our analysis of mRNAs from muscles of day 10 quail embryos, whereas other studies that identified exon y isoforms had analyzed later fetal stages.

In the N-terminal region, we found that phylogenetically conserved exons 4–5 showed similar splicing patterns between birds and mammals, consistent with the findings of Schachat et al. (38). Most significantly, exon 4 is present in mammalian perinatal TnT (35/40 quail and 29/31 chicken perinatal mRNAs have exon 4). In addition, we found three leg forms that appear similar to predominant vertebrate leg forms. Of the 25 exons, only one of which, represents the putative fetal exon y, has not been confirmed in a quail cDNA. Of the 25 exons, 13 of these exons are differentially expressed in fetal quail muscle.

Unexpectedly, we observed a lack of significant overlap among TnTf isoforms sequenced from the chicken and quail cDNA sets. Only three forms were represented both in the quail and the chicken sets. This difference, in part, is due to the lack of y exon inclusion in any of our forms and thus likely reflects the difference in stage of muscle analyzed (38). However, both chicken and quail studies identified the same predominant adult pectoral form (see Table I). Exon usage supports the inclusion of exon 4 in perinatal muscles, and His-rich pectoral peptides are predominant in the avian pectoral muscles. These data support the conclusion that many different splice forms are produced during pectoral muscle development and provide further compelling evidence that the developmental regulation of splicing that generates TnTf isoforms is sufficient to account for accumulation of specific TnTf proteins.

Intron/Exon Organization of the qTnTf Gene—To unambiguously define exon splicing patterns that give rise to TnTf isoforms, we determined the DNA sequence of a contiguous region of the qTnTf gene (see “Experimental Procedures” and Fig. 3). Clone gC1067 is a 13,876-nt clone that encodes the most 5′ sequences of the clone. Exon 1 starts at position 1178 in the sequenced regions of gC1067. qT1 bridges the genomic phage clones gC1067 and gC106 from position 13,563 to position 15,928 in the consensus sequence, ending 315 nt into gC106. gC106 is a 17,105-nt clone that begins at 15,625 and ends at 32,715 in the consensus sequence. qT2 begins at 32,363 in the consensus sequence and ends at 33,128 in exon 18. The remaining 3′ untranslated TnTf sequences are derived from the TnTf mRNA sequences up until the poly(A) tail at nucleotide 33,434. The structure, based on these DNA sequences and comparisons to qTnTf cDNAs, is represented in Fig. 3. We identified 25 exons, only one of which, the putative fetal exon y, has not been confirmed in a quail cDNA. Of the 25 exons, 13 of these exons are differentially spliced from the qTnTf mRNA transcript.

Identification of a Repeated Sequence Element Containing the Pectoral-specific Exons—In our sequence analysis of the genomic sequences encoding the pectoral exon, we discovered a remarkable degree of sequence identity that included and flanked each of the five pectoral-specific exons. These exons (4/5p exons are 15 nt in length) are encompassed within an approximately 900-bp sequence element that is repeated four times and one time partially (150 nucleotides), having 74–82% sequence identity in pairwise comparisons (Fig. 4).

Identification of a CR1 Retrotransposon Located 5′ of the Pectoral-specific Exon, p1—We identified a sequence related to the chicken CR1 transposable element upstream of the first pectoral repeat. A family of chicken middle repetitive sequences, termed CR1 for chicken repeat 1, have been described for the chicken genome. CR1 is a member of the non-long terminal repeat class of retrotransposons (41–43). CR1 elements are flanked by imperfect direct repeats of an octamer.
sequence having the consensus (CATTCTRT) (GATTCTRT). The sequence CAATTCT GATCTTCT in the quail intron 1.5 kb upstream of exon p1 was identified, as well as some flanking sequences of approximately 90 bp having similarity to CR1 sequences (nucleotides 8990–9081 in the submitted sequences), but we did not identify transposase encoding sequences associated with this element. This CR1 element apparently transposed into the TnTf locus, and its proximity to the pectoral repeat may be relevant to the introduction of these sequences into the avian genome (see “Discussion”).

Analysis of Splice Site Consensus Sequences—We compared splice acceptor and donor sequences flanking constitutive and alternatively spliced exons to assess if there are significant variations and similarities between the quail and rat splice junctions that could identify sequences important for regulating splicing (Table II and legend). The analysis shows that exons p1–5, 6, 7, and y have non-consensus (at 23 position) 39 SS; instead of the consensus YA(G/G), the site is AA(G/G). This variant acceptor sequence is highly exceptional, as the (23) position conforms to the consensus in 96% of acceptor junctions, suggesting that these acceptor sequences function in

FIG. 4. Alignment of the genomic repeat sequence element that includes the pectoral exons. Analysis of sequences surrounding the pectoral His repeat exons reveals remarkable, highly conserved repeated elements that are 74–82% identical in pairwise comparisons. The repeat is referred to based on the His repeat exon encoded within it (p1–5, 5’ to 3’). C represents the consensus sequence for all repeats based on identity in 4/5 or 3/4 in the regions where the p5 repeat has ended; minus indicates a lack of consensus. * indicates nucleotide gap(s) in the best fit alignment made, and in some cases this is also the consensus. The PTB-binding site is underlined and the non-conserved 3’ SS (6–3) is in bold (also see Table III).

sequence having the consensus (CATTCTRT) (GATTCTRT). The sequence CAATTCT GATCTTCT in the quail intron 1.5 kb upstream of exon p1 was identified, as well as some flanking sequences of approximately 90 bp having similarity to CR1 sequences (nucleotides 8990–9081 in the submitted sequences), but we did not identify transposase encoding sequences associated with this element. This CR1 element apparently transposed into the TnTf locus, and its proximity to the pectoral repeat may be relevant to the introduction of these sequences into the avian genome (see “Discussion”). The difference in the alternative exon coding potential between these two genes resides in the N-terminal hypervariable regions of the protein, with the remainder of the gene structure (exons 9–18) being essentially identical in organization between the genes. This high conservation is also reflected in the nearly complete identity of the amino acid sequences in the C-terminal coding region (Fig. 2).

Comparison of the Rat and Quail TnTf Genes Structure—The quail TnT gene is encoded in greater than 33 kb (Fig. 3), whereas the rat TnT gene spans approximately 16 kb (7, 44). The quail protein is encoded in 25 exons, 12 constitutively spliced exons, and 13 alternatively spliced exons, whereas the rat gene has 19 exons, with 11 exons predicted as constitutively spliced, and 8 exons predicted to be alternatively spliced (including exon 5, which is more likely constitutively spliced; see “Discussion”). The analysis of alternative exon coding potential between these two genes resides in the N-terminal hypervariable regions of the protein, with the remainder of the gene structure (exons 9–18) being essentially identical in organization between the genes. This high conservation is also reflected in the nearly complete identity of the amino acid sequences in the C-terminal coding region (Fig. 2).
splicing regulation, as has been proposed for alternatively regulated exons of some other genes (e.g., Ref. 45). The rat fetal exon (f) also shows a similar non-consensus 3' SS and the chicken gene shows alternative splicing of the pectoral region in fetal muscle. Thus, this 3’ SS sequence may have special significance for fetal splicing regulation. Finally, it is noted that two poly-pyrimidine sequences, with PTB consensus binding sites, are located immediately 5' of the splice acceptors of the tandemly repeated, pectoral specific exons p1-p5 (Fig. 4). PTB consensus binding sites have been implicated in the positive and negative regulation of alternative exon splicing (46-48).

Comparison of the rat and quail with nucleotide sequences flanking the alternatively spliced, mutually exclusive exons 16 and 17 (Table III) reveals a pattern of conserved purines (5 for 16 and 7 for 17) in the pyrimidine-rich region of the 3’ SS. Furthermore, the extent of sequence identity was high in the 3’ SS consensus. Variations in the sequence from consensus sequences are indicated by boldface letters. Sequence comparisons of other intronic regions of the quail and rat genes did not reveal any significant sequence homology, supporting this possibility.

### DISCUSSION

Here, we report the complete structure of the quail TnTf gene and a detailed analysis of the expression of TnT isoforms produced by alternative splicing of its N-terminal exons. These structural and expression data provide significant new information on the evolution, function, and developmental isomorph regulation of TnTf genes in birds and mammals. We show that the qTnTf gene has 25 exons. The reported sequences include 1178 nt of sequences upstream of the transcription start site and exon sequences that comprise only 4% of the total genomic sequence. The amino acids encoded by constitutively spliced exons are highly conserved between quail and rat TnTf, with the exception of mini-exon 3. The rat and quail exons share identical split codons for all homologous exons. This conservation suggests that the differential splicing of exons in this gene evolved prior to mammalian/avian divergence. As previously discussed (7, 44), distribution of split codons requires that exons 3 and 9 always be spliced into mRNAs to maintain the translational reading frame, whereas different combinations of combinatorial exons between 3 and 9 can be included. The quail TnTf gene, however, is nearly twice the size of the rat gene. This size difference may be due in part to the introduction of the avian-specific exons and to the exon/intron duplications of the pectoral exon region; however, the intron sizes in general are larger in the quail than in the rat (see Fig. 4). The species-specific (p, y, fetal) exons likely serve specialized functions in the TnTf protein (see below). Finally, although the splicing and organization of exons 16 and 17 show similar alternative splicing and similar organization in the quail and rat gene, however, is nearly twice the size of the rat gene. This size difference may be due in part to the introduction of the avian-specific exons and to the exon/intron duplications of the pectoral exon region; however, the intron sizes in general are larger in the quail than in the rat (see Fig. 4). The species-specific (p, y, fetal) exons likely serve specialized functions in the TnTf protein (see below). Finally, although the splicing and organization of exons 16 and 17 show similar alternative splicing and similar organization in the quail and
rat, exon 17 encodes divergent protein sequences, suggesting specialized functions for this domain in avian and mammalian muscles (23).

Several significant features of the quail TnTf gene were revealed through these structural and exon expression analyses and comparisons to the homologous rat TnTf gene (7, 44). These features include the following: 1) identification of 13 constitutively spliced, 11 (4, w, p1–p5, 6, 7, y, and 8) combinatorially spliced, and 2 (16 and 17) alternatively spliced exons; 2) the combinatorially spliced exons encode N-terminal TnT sequences and are subject to developmental and muscle-specific splicing regulation in the skeletal muscles of the quail embryo; 3) 5 of these N-terminal exons, which encode the pectoral-specific, His-rich domain that is unique to the pectoral muscle TnTfs of Galliformes and Craciformes, are encoded within a novel, 900-bp intronic sequence that is tandemly repeated; 4) a CR1 transposable element sequence located immediately 5′ of this tandemly repeated domain is implicated in transposition of these pectoral muscle exons and associated intron splicing in the N-terminal TnTf region of additional combinatorially spliced exons; 5) non-consensus splice acceptor sequences and adjacent polyzygotic tracks and PTB-binding sites that are implicated in tissue-specific splicing regulation of the pectoral muscle-specific exons (46–48); and 6) conserved, non-consensus splice acceptor sequences associated with the mutually exclusive, alternatively spliced exons 16 and 17 in the C-terminal domain, further implicating splice acceptor sequences in alternative exon splicing regulation.

Evolution of the Pectoral-specific His Repeat Exon Domain—Intronic sequences that surround the five pectoral-specific exons (p1–5) are highly conserved and tandemly repeated 4 and 1/8 times, resulting in four identical mini-exons (p1–4) bearing the His repeat motif, AHHEE, and one divergent mini-exon bearing the His motif, AHAEE. Since this motif is amplified to variable amounts in different species of birds, this demonstrates that recombination is actively occurring within this domain during the speciation within the Galliformes and Craciformes orders (49). Although a similar recombination mechanism has been proposed for the origin of the N-terminal combinatorial exons and exons 16 and 17 of the TnTf gene based on their related exon sequences (7, 44), no related intronic sequences were observed in the regions flanking these alternative exons. To our knowledge, the conservation of the repeat sequences of the quail pectoral-specific exons and surrounding intronic sequences is novel, providing direct evidence of intragenic recombination, generating duplicated exons. Indeed, conservation of the intron and exon sequences of the His repeat region of quail TnTf, is striking; comparisons of other intronic sequences among chick and quail genes reveal only 25% similarity,2 the level expected for random drift of “non-functional” sequences after divergence of the chick and quail lineages. This indicates that there are strong selective pressures to maintain the conservation of the intron and exon sequences around exons p1–5. These selective pressures are likely to be multiple, including protein structure and functional requirements to maintain and amplify the His repeat motif within the TnTf protein for specialized pectoral muscle functions and requirements to conserve splicing regulatory sequences that restrict the splicing of these exons to the pectoral muscles. In this regard, it is notable that the chicken TnTf pectoral exon has two additional His repeat motifs as compared with the quail (38). Together these data suggest a role for recombination mechanisms to amplify to maintain homology in these intronic and exon sequences. Such mechanisms have been discussed for maintenance of repeated gene families (50–52). It will be of considerable interest to determine the genomic organization of the chick TnTf intron/exon domain encoding these pectoral exons and to compare these with the quail pectoral exon domains, as an approach to the analysis of the recombination mechanisms that lead to the conservation of these intronic sequences, and the sequences responsible for pectoral exon splicing regulation.

Two DNA transposition hypotheses can now be considered for the introduction of the His repeat exon domain and its repeated intron sequences into the ancestral genome of Galliformes and Craciformes. We have identified a CR1 retrotransposon element (41) 1.5 kb upstream of the tandemly repeated intronic domain region, suggesting that a retrotransposon mechanism could have inserted and duplicated the pectoral exons. In its current state, the CR1 repeat in the qTnTf gene is imperfect as its second direct repeat is 6 and not 8 nt, and no transposase open reading frame is associated with this CR1 element; however, there are approximately 90 bp upstream having CR1-related sequences, verifying its relationship to CR1. Although sequences associated with the pectoral repeat have not been otherwise associated with CR1 elements, it is possible that a mobile CR1 picked up these sequences at another genomic site prior to insertion into TnTf. Evidence that CR1 retrotranspositional machinery can be usurped and facilitate movement and insertion of non-CR1 sequences has been presented by others (42, 53). As a test of this hypothesis, it will be of interest to examine and compare the TnTf genomic sequences of the chick and other Galliformes and Craciformes for the presence and structure of this CR1 element and its relationship to the His repeat domain.

An alternative hypothesis for the origin of the pectoral exons is based on the observation that the pectoral exons are 54% related to portions of the histidine/alanine-rich proteins present in the avian malarial parasite Plasmodium lophurae (49), suggesting the possibility that this exon sequence was acquired by a recombinational event between the malarial parasite genome and the genome of the ancestor of the Galliformes and Craciformes. Since the sequences encoded within a single avian exon (AHHEE) do not correspond exactly to this metal-binding motif (HXXXH) (see Figs. 1 and 2), this motif would have been split into exons, according to this model, either prior to or after introduction into the ancestral genome.

Physiological Functions of the Pectoral His Repeat Sequences—The pectoral muscles of Galliformes and Craciformes are distinctive in that they function in explosive but short-lived flight, and expression of TnTf isoforms with the His repeat could reflect a specialized functional adaptation related to muscle metabolism and function. Interestingly, bacterial and vertebrate metal-binding proteins have HXXXH motifs typical of the quail pectoral exon. These sequences, in α-helical form (such as predicted for the avian motif), generate high affinity binding sites for transition metal ions (Cu2+, Ni2+, Zn2+, and Co2+), and TnTf isoforms bearing this domain are specifically retained on metal affinity columns (49). Birds having the HXXXH motif also have lower Zn2+ concentration in their pectoral muscle, suggesting that this domain could regulate the free metal ion concentrations in pectoral muscle. Consistent with a possible role in muscle protein function, binding of metal ions to the N terminus has been shown to generate conformational changes in a TnT N-terminal peptides by stabilizing α-helical structure and influencing binding to Tm (54). Alternatively, the His repeat domain may be sensitive to alterations in pH that occur during anaerobic work, e.g. a change from cellular pH of 7.1 to a pH of 6.5 (55, 56). Histidine-imidazole groups are unusual in that they have a pK close to 7 at 25 °C.

2 D. Pinney and C. P. Emerson, Jr., unpublished data.
At pH values typical of biological systems, these groups are approximately half-protonated and thus are well suited to function in reversible protonation/deprotonation events. For instance, a pH-dependent effect of histidine-imidazole groups in enzyme function is well documented for lactate dehydrogenase in muscle (reviewed in Ref. 56). Consistent with a specialized function of the His repeat motif in anaerobic pectoral muscle, recent studies show that the N-terminal region of the pectoral TnTf isoform shows pH-dependent binding affinities for Tm and TnTf to Tm binding more stable at lower pH (40). Therefore, specialized functions of the His exon protein domain in pectoral TnTf could control the conformation of this domain and interaction with Tm and thus shift the contraction force, PκK dynamics, in the muscle in response to the change in the protonation state of the histidine residues. These mechanisms could provide dynamic ways to alter the activation state, dependent on the pH of the muscle. Although the functional contributions of the His domain in vivo are unclear, the N-terminal variable domain of TnT, which includes the pectoral His domain, is physiologically distinct for fast skeletal muscle types (10, 14, 57), and the importance of this region has been highlighted by the discovery of mutations near this N-terminal domain that disrupt TnT function and result in human disease (18). We are currently testing these hypotheses of TnT domain function through in vivo expression studies of TnTf isoforms characterized in this study.

mRNA and Protein Analysis of TnT Isoform Expression—Through combinatorial and alternative exon splicing, the TnTf gene has the potential to produce a large number of different mRNAs for production of isoforms with variations of protein structure in the N- and C-terminal domains. Our quail data in combination with chicken TnTf isoform data (36, 40) establish that there is great diversity of TnTf mRNAs and that their expression is highly regulated in different avian muscles and during development. It also is of interest that there is very little overlap in the isoforms that we have identified with the isoforms reported in chick (36). The findings of extensive TnTf mRNA isoform diversity, as assayed by reverse transcription PCR at the mRNA level, are consistent with results from studies of TnT protein isoforms (for example see Refs. 35, 39, and 58–61). It is currently difficult to compare the muscle specificity of avian TnT isoforms to the characterized mammalian isoforms, although it is notable that qTnT2a, which is abundant in adductor muscle, is homologous to the mammalian TnT2a form in fast oxidative muscles (tongue), which are similar in fiber type to the avian adductor. In addition, although a remarkable number of isoforms (at mRNA and protein level) are made in avian muscles, some studies suggest that the number of abundant cDNAs and proteins made in mammals may be more restricted (62). Furthermore, our data and the data of Briggs and Schachat (62) suggest that exon 5 is most likely a constitutively splice exon, rather than a combinatorially spliced exon, since all cDNAs sequenced to date from chicken, quail, and mammals include this exon.

Non-consensus Splice Acceptor Sequences of Alternatively Spliced TnTf Exons—The combinatorially spliced TnTf exons, p1–5, 6, 7, and y, all have a non-consensus purine nucleotide at the −3 position in the 3′ SS consensus. The rat TnTf fetal exon is also non-consensus in this position. This variant purine in the 3′ SS consensus likely contributes to the regulated splicing of these exons, as supported by genetic studies of the Drosophila myosin heavy chain gene, which also has alternative exons with variant splice acceptors (45, 63). It also is notable that there are perfect polypyrimidine sequences containing PTB-binding sequences immediately 5′ of the non-consensus splice acceptors of the five pectoral exons (p1–5). Such polypyrimidine/PTB sequences participate in the regulated alternative exon splicing of muscle tropomyosin and other tissue-restricted mRNA splicing (46–48). In addition to regulatory sequences associated with the splice acceptor of TnTf combinatorial exons, acceptor sequences immediately 5′ of alternatively spliced exons 16 and 17 in the rat and quail TnTf genes have nearly identical purine substitutions in what should be the pyrimidine-rich domain, and a variant lariat branch sequence, which does not match the consensus in the region upstream of exon 17, is displaced toward the donor. The conserved pattern of purine substitutions may control splice site selection by splicing regulatory factors (64), and displacement of the lariat may result in steric hindrance of lariat formation. It is also possible that splicing enhancers within exon sequences contribute to regulated splicing mechanism, as has been found for the chicken heart TnTf isoform gene (65). These findings, together with the availability of the complete TnTf gene structure, make possible directed experiments to test the specific regulatory functions of the splicing acceptor sequences of both combinatorial and alternatively spliced exons and to identify other candidate splicing regulatory sequences as well as the muscle-specific splicing factors that interact with these sequences and likely regulate their splicing.

Acknowledgments—We thank L. Hong for assistance with figure preparation; P. Benx for technical assistance; and J. Burch, J. Marden, L. Sweeney, and D. Standiford for helpful discussions. E.A.B. also thanks M. Buckingham and the Pasteur Institute for providing library and computer support during the preparation of this manuscript.

REFERENCES

1. Leavis, P. C., and Gergely, J. (1984) CRC Crit. Rev. Biochem. 16, 235–305
2. Perry, S. V. (1998) J. Muscle Res. Cell Motil. 19, 575–602
3. Tobacman, L. S. (1996) Annu. Rev. Physiol. 58, 447–481
4. Cooper, T. A., and Ordahl, C. P. (1985) J. Biol. Chem. 260, 11140–11148
5. Giermann, R., Truitt, A. F., Fyn, P., Gunning, P., and Redes, L. (1987) J. Biol. Chem. 262, 16122–16126
6. Bucher, E. A., de Broussse, F. C., and Emerson, C. P., Jr. (1989) J. Biol. Chem. 264, 12462–12491
7. Breitbart, R. E., Nguyen, H. T., Medford, R. M., Destree, A. T., Mahdavi, V., and Nadal-Ginard, B. (1985) Cell 41, 67–82
8. Nassar, R., Malouf, N. N., Kelly, M. B., Oakely, A. E., and Anderson, P. A. (1991) Circ. Res. 69, 1470–1475
9. Pan, B. S., and Potter, J. D. (1992) J. Biol. Chem. 267, 23052–23056
10. Tobacman, L. S. (1988) J. Biol. Chem. 263, 16688–16722
11. Wu, L. J., Jha, P. K., Du, Y., Leavis, P. C., and Sarkar, S. (1995) Gene (Amst.) 155, 225–230
12. L’Ecuyer, T. J., Schulte, D., and Lin, J. J. (1991) Pediatr. Res. 30, 232–238
13. Reiter, P. J., Greaser, M. L., and Moss, R. L. (1992) J. Physiol. (Lond.) 449, 573–588
14. Schachat, F. H., Diamond, M. S., and Brandt, P. W. (1987) J. Mol. Biol. 198, 551–554
15. Myers, C. D., Goh, P. Y., Allen, T. S., Bucher, E. A., and Bogen, G. T. (1996) J. Cell Biol. 132, 1061–1077
16. Fyryberg, E., Fyryberg, C., Beall, C., and Saville, D. L. (1990) J. Mol. Biol. 216, 657–675
17. Thierfelder, L., Watkins, H., MacRae, C., Lamas, R., McKenna, W., Vosberg, H. P., Seidman, J. G., and Seidman, C. E. (1994) J. Clin. Invest. 97, 2842–2848
18. Jin, J. P., Huang, Q. Q., Yeh, H. I., and Lin, J. J. (1992) J. Mol. Biol. 227, 1269–1276
19. Schiaffino, S., and Reggiani, C. (1996) Physiol. Rev. 76, 371–423
20. Zot, A. S., and Potter, J. D. (1987) Annu. Rev. Biophys. Biophys. Chem. 16, 535–559
21. Medford, R. M., Nguyen, H. T., Destree, A. T., Summers, E., and Nadal-Ginard, B. (1984) Cell 38, 409–421
22. Hastings, K. E., Bucher, E. A., and Emerson, C. P., Jr. (1985) J. Biol. Chem. 260, 13699–13703
23. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
24. Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. (1990) PCR Protocols: A Guide to Methods and Applications, Academic Press, San Diego
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
27. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499–560
28. Henikoff, S. (1984) Gene (Amst.) 28, 351–359
29. Staden, R. (1984) Nucleic Acids Res. 12, 505–519
30. Staden, R. (1984) Biochem. Soc. Trans. 12, 1005–1008
31. Pearson, W. R., and Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 2444–2448
Quail TnTf Gene Expression, Structure, and Evolution

33. Lipman, D. J., and Pearson, W. R. (1985) Science 227, 1435–1441
34. Zot, H. G., and Potter, J. D. (1982) J. Biol. Chem. 257, 7678–7683
35. Bird, I. M., Dhoot, G. K., and Wilkinson, J. M. (1985) Eur. J. Biochem. 150, 517–525
36. Schachat, F., Schmidt, J. M., Maready, M., and Briggs, M. M. (1995) Dev. Biol. 171, 233–239
37. Smillie, L. F., Golosinska, K., and Reinach, F. C. (1988) J. Biol. Chem. 263, 18816–18820
38. Schachat, F., Schmidt, J. M., Maready, M., and Briggs, M. M. (1995) Dev. Biol. 171, 233–239
39. Dhoot, G. K. (1988) J. Muscle Res. Cell Motil. 9, 446–455
40. Ogut, O., and Jin, J.-P. (1998) J. Biol. Chem. 273, 27858–27866
41. Vandergon, T. L., and Reitman, M. (1994) Mol. Biol. Evol. 11, 886–898
42. Haas, N. B., Grabowski, J. M., Sivitz, A. B., and Burch, J. B. (1997) Gene (Amst.) 197, 305–309
43. Silva, R., and Burch, J. B. (1989) Mol. Cell. Biol. 9, 3563–3566
44. Breitbart, R. E., and Nadal-Ginard, B. (1986) J. Mol. Biol. 188, 313–324
45. Standiford, D. M., Davis, M. B., Sun, W., and Emerson, C. P., Jr. (1997) Genetics 147, 725–741
46. Achiya, M., and Grabowski, P. J. (1997) RNA (NY) 3, 996–1015
47. Perez, I., Lin, C. H., McAfee, J. G., and Patton, J. G. (1997) RNA (NY) 3, 764–778
48. Valcarcel, J., and Gebauer, F. (1997) Curr. Biol. 7, R705–R708
49. Jin, J. P., and Smillie, L. B. (1994) FEBS Lett. 341, 135–140
50. Stark, G. R., and Wahl, G. M. (1984) Annu. Rev. Biochem. 53, 447–491
51. Holland, S. K., and Blake, C. C. (1987) Biosystems 20, 181–206
52. Gilbert, W. (1978) Nature 271, 501
53. Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1996) Mol. Cell. Biol. 16, 3756–3764
54. Ogut, O., and Jin, J. P. (1996) Biochemistry 35, 16581–16590
55. McGilvery, R. W. (1985) in Metabolic Adaptation to Prolonged Physical Exercise (Howald, H., and Poortmans, J. R., eds) Birkhauser Verlag, Basel
56. Hochachaka, P. W., and Somero, G. N. (1984) Biochemical Adaptation, Ph.D. thesis, Princeton University Press, Princeton
57. Briggs, M. M., McGinnis, H. D., and Schachat, F. (1990) Dev. Biol. 140, 253–260
58. Abe, H., Komiya, T., and Ohnata, T. (1986) Dev. Biol. 118, 42–51
59. Imai, H., Hirai, S., Hiroto, H., and Hirabayashi, T. (1986) J. Biochem. (Tokyo) 99, 923–930
60. Wilkinson, J. M., Moir, A. J., and Waterfield, M. D. (1984) Eur. J. Biochem. 143, 47–56
61. Yao, Y., Nakamura, M., Miyazaki, J. I., Kirinoki, M., and Hirabayashi, T. (1992) Dev. Biol. 151, 531–549
62. Briggs, M. M., and Schachat, F. (1996) Am. J. Physiol. 270, C298–C305
63. Davis, M. B., Dietz, J., Standiford, D. M., and Emerson, C. P., Jr. (1998) Genetics 150, 1105–1114
64. Breitbart, R. E., and Nadal-Ginard, B. (1987) Cell 49, 793–803
65. Cooper, T. A., Cardone, M. H., and Ordahl, C. P. (1988) Nucleic Acids Res. 16, 8443–8465
66. Briggs, M. M., and Schachat, F. (1989) J. Mol. Biol. 206, 204–249
67. Shapiro, M. B., and Senapathy, P. (1987) Nucleic Acids Res. 15, 7155–7174
68. Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472