J Mol Evol (1988) 27:228–235

Journal of Molecular Evolution


# Evolution of Tropomyosin Functional Domains: Differential Splicing and Genomic Constraints

S. Colote,[1] J. Sri Widada,[1] C. Ferraz,[1] F. Bonhomme,[2] J. Marti,[3] and J.-P. Liautard[1]

[1] INSERM U-249, CRBM du CNRS, Université de Montpellier I, Route de Mende, B.P. 5051, 34033 Montpellier, France
[2] Institut des Sciences de l'Évolution, CNRS UA 327, USTL pl. E. Bataillon, 34060 Montpellier, France
[3] INSERM U-65, Institut de Biologie Bd Henri IV, 34000 Montpellier, France

**Summary.** We have cloned and determined the nucleotide sequence of a complementary DNA (cDNA) encoded by a newly isolated human tropomyosin gene and expressed in liver. Using the least-square method of Fitch and Margoliash, we investigated the nucleotide divergences of this sequence and those published in the literature, which allowed us to clarify the classification and evolution of the tropomyosin genes expressed in vertebrates. Tropomyosin undergoes alternative splicing on three of its nine exons. Analysis of the exons not involved in differential splicing showed that the four human tropomyosin genes resulted from a duplication that probably occurred early, at the time of the amphibian radiation. The study of the sequences obtained from rat and chicken allowed a classification of these genes as one of the types identified for humans.

The divergence of exons 6 and 9 indicates that functional pressure was exerted on these sequences, probably by an interaction with proteins in skeletal muscle and perhaps also in smooth muscle; such a constraint was not detected in the sequences obtained from nonmuscle cells. These results have led us to postulate the existence of a protein in smooth muscle that may be the counterpart of skeletal muscle troponin.

We show that different kinds of functional pressure were exerted on a single gene, resulting in different evolutionary rates and different convergences in some regions of the same molecule.

Codon usage analysis indicates that there is no strict relationship between tissue types (and hence the tRNA precursor pool) and codon usage. G+C content is characteristic of a gene and does not change significantly during evolution. These results are in good agreement with an isochore composition of the genome, and thus suggest a similar chromosomal environment in chicken, rat, and human.

**Key words:** Tropomyosin — Differential splicing — Evolution — Isochore — Codon usage — Sequence convergence — Functional constraints

## Introduction

Tropomyosins are a family of highly related proteins present not only in muscle (skeletal, cardiac, and smooth; Smillie 1979) but in all types of tissue (Stossel et al. 1985), although different forms of the protein are characteristic of particular cell types (Giometti and Anderson 1984). Each type of cell contains multiple isoforms; in mammals, at least nine isoforms have been identified. Skeletal muscle tropomyosins have been found in nonmuscle cells and nonmuscle tropomyosins in embryonic muscle (Giometti and Anderson 1984).

Tropomyosins are a fundamental component of muscle. They bind to actin microfilaments (Smillie 1979). In skeletal muscle their function is clearly established. They are involved in calcium regulation of muscle contraction. They bind troponins (a protein involved in calcium modulation of muscle contraction) in the presence of calcium and allow myosin to interact with microfilaments (Smillie 1979). In smooth muscle and in other tissues where no troponin has been detected, the actual function of

*Offprint requests to:* J.-P. Liautard

tropomyosin is unknown, although it was found to stabilize actin microfilaments against the severing effect of gelsolin (Berstein and Bambourg 1982). Moreover, tropomyosin expression is closely related to various morphological stages of the cell. This phenomenon was observed during differentiation of embryonal carcinoma cells (Paulin et al. 1979) and following cell transformation by viruses or oncogene injection. In the latter cases the expression was considerably reduced and the isoform composition was profoundly altered (Hendricks and Weintraub 1981; Cooper et al. 1985).

The different tropomyosins share common structures. They consist totally of alpha helix (Smillie 1979), which forms coiled-coil duplexes (McLachlan and Stewart 1975) and head-to-tail assemblies giving rise to long-range structures (Smillie 1979). In vitro, paracrystals can be obtained from these types of interactions. Furthermore, tropomyosins must interact with actin filaments and, at least in the case of skeletal muscle, with troponins. All these structural properties and functional requirements subject tropomyosins to strong evolutionary constraints.

Elucidation of the relationships between the isoform types and tissue specificity should help explain the function, structure, and evolution of this protein class. Molecular biological analysis has shown that isoform production results from many genes, which undergo alternative splicing (MacLeod et al. 1985). This mechanism appears to play an important role in the tissue-specific expression of the isoforms (Reinach and MacLeod 1986), but nothing is known about the actual role of this type of splicing. Furthermore, the genes are very difficult to isolate due to the large number of pseudogenes.

Until now, the gene sequences were not known, and the only possibility for characterizing the number of genes and their relationships was to analyze the complementary DNA (cDNA) whose sequences were known. Furthermore, we considered that evolutionary analysis of the tropomyosin sequences could help clarify the relationships among the different genes. An analysis of the structures conserved through evolution could explain the functions of the different parts of the molecule and the reasons why they undergo differential splicing.

We isolated cDNA clones of tropomyosin and compared their sequences with those published in the literature to determine the structures retained through evolution, to understand the mechanisms of the evolutionary constraints, and to elucidate the role of differential splicing. Our analysis shows that tropomyosins are susceptible to two kinds of pressure: functional pressure due to interactions with other specific proteins and another pressure probably exerted by genomic structural requirements.

## Materials and Methods

*Screening the Libraries.* Human liver cDNA library in λgt11 was obtained from Genofit (Switzerland) and screened with the following oligonucleotide: GCTGAAGGAGGCTGAGACCCG. Three rounds of screening [37°C; 5× standard saline phosphate (SSP)] were performed to purify the responding colonies. The inserts digested with the appropriate restriction enzymes were subcloned into M13 and pUR222 and then sequenced. Other libraries were also analyzed, and the same clones were found.

*Sequencing Methods.* We used two methods to sequence the clones: the dideoxy chain terminal procedure (Sanger et al. 1977) after subcloning in M13 and the chemical method (Maxam and Gilbert 1980). A sequencing kit from Amersham was used for the dideoxy method according to the manufacturer's recommendations.

*Sequence Analysis.* The sequences were aligned using the Diagon procedure (Staden 1982) on an IBM-PC computer. A mutation divergences matrix was constructed containing all the sequences analyzed. In some cases, deletions were included to obtain good similarity in the conserved exon part (see Results for justification).

Evolutionary trees were constructed with the Fitch-Margoliash least-square method proposed in J. Felsenstein's PHYLIP program package (Phylogeny Inference Package version 2.8; Department of Genetics, University of Washington, Seattle, Washington 98195, USA).

## Results

### *Sequence of a New Tropomyosin Gene*

We have screened a human liver library with a synthetic probe (GCTGAAGGAGGCTGAGACCCG) whose sequence was deduced from an analysis of the known tropomyosin sequences. Two clones were retained and sequenced as described in Materials and Methods. One of these clones was the product of a human tropomyosin gene not yet described (Fig. 1). This clone, though incomplete, contains the coding sequence from amino acid 60 onward and a large part of the 3′ noncoding region. Further attempts to isolate the entire coding sequences were not fruitful. We also looked for this cDNA in a human embryonic muscle library (a generous gift of Dr. F.S. Walsh). Although numerous clones were found and sequenced, none was larger than the one found in liver. Furthermore, several clones terminate at exactly the same point in the two libraries. This is probably due to a special structure of the messenger. These findings mean that this gene is expressed in liver and in embryonic muscles. In preliminary studies we did not detect the corresponding mRNA in transformed cells (HeLa and HL 60).

The nucleotide sequence of the clone is shown in Fig. 1. The cDNA coding region and the 3′ noncoding region are very similar to rat smooth muscle tropomyosin (Ruiz-Opazo et al. 1985), which led us to conclude that the clone we isolated is of the smooth

```
         10         20         30         40         50         60         70
CTGCAGGCTCCGGATCTTCCTGCGCACGGCCTCCAGCGAGCATCTCCACGAGAGGAAGCTGAGGGAGACC
  C  R  L  R  I  F  L  R  T  A  S  S  E  H  L  H  E  R  K  L  R  E  T

         80         90        100        110        120        130        140
GCTGAAGCCGACGTAGCTTCTCTGAACAGACGCATCCAGCTGGTTGAGGAAGAGTTGGATCGTGCCCAGG
 A  E  A  D  V  A  S  L  N  R  R  I  Q  L  V  E  E  E  L  D  R  A  Q

        150        160        170        180        190        200        210
AGCGTCTGGCAACAGTTTTGCAGAAGCTGGAGGAAGCTGAGAAGGCAGCAGATGAGAGTGAGAGAGGCAT
E  R  L  A  T  V  L  Q  K  L  E  E  A  E  K  A  A  D  E  S  E  R  G  M

        220        230        240        250        260        270        280
GAAAGTCATTGAGAGTCGAGCCCAAAAAGATGAAGAAAAAATGGAAATTCAGGAGATCCAACTGAAAGAG
  K  V  I  E  S  R  A  Q  K  D  E  E  K  M  E  I  Q  E  I  Q  L  K  E

        290        300        310        320        330        340        350
GCCAAGCACATTGCTGAAGATGCCGACCGCAAATATGAAGAGGTGGCCCGTAAGCTGGTCATCATTGAGA
 A  K  H  I  A  E  D  A  D  R  K  Y  E  E  V  A  R  K  L  V  I  I  E

        360        370        380        390        400        410        420
GCGACCTGGAACGTGCAGAGGAGCGGGCTGAGCTCTCAGAAGGCCAAGTCCGACAGCTGGAAGAACAATT
S  D  L  E  R  A  E  E  R  A  E  L  S  E  G  Q  V  R  Q  L  E  E  Q  L

        430        440        450        460        470        480        490
AAGAATAATGGATTCAGACCTTGAAAGCATTAATGCTGCAGAGGATAAGTACTCGCAGAAGGAAGACAGA
  R  I  M  D  S  D  L  E  S  I  N  A  A  E  D  K  Y  S  Q  K  E  D  R

        500        510        520        530        540        550        560
TATGAGGAAGAGATCAAGGTCCTTTCCGACAAGCTGAAGGAGGCTGAGACTCGGGCTGAGTTTGCGGAGA
 Y  E  E  E  I  K  V  L  S  D  K  L  K  E  A  E  T  R  A  E  F  A  E

        570        580        590        600        610        620        630
GGTCAGTAACTAAATTGGAGAAAAGCATTGATGACTTAGAAGAGAAAGTGGCTCATGCCAAAGAAGAAAA
R  S  V  T  K  L  E  K  S  I  D  D  L  E  E  K  V  A  H  A  K  E  E  N

        640        650        660        670        680        690        700
CCTTAGTATGCATCAGATGCTGGATCAGACTTTACTGGAGTTAAACAACATGTGAAAACCTTCTTAGCTG
  L  S  M  H  Q  M  L  D  Q  T  L  L  E  L  N  N  M  *

        710        720        730        740        750        760        770
CGACCACATTCTTTCATTTTGTTTTGTTTTGTTTTGTTTTTAACACTGCTTACCCTAAATGCAATTTATT

        780        790        800        810        820        830        840
TACTTTTACCATGTCACAGAAACACCCACAAGATACCACTAGGTCAGGGGTGGGAAAACACATACAAAAA

        850        860        870        880        890        900        910
GGCAAGCCCATGTCAGGCGATCCTGGTCAAATGTGCCATTTCCCGGGTGATGCTGCCACCACTTGTAGAG

         920        930        940        950        960        970        980
AGTTAGCAACACAGTGTGCTTAGTCAGTGTAGGAATCCTCACTAAAGCAGAAGAAGTTCCATTCAAAGTC

        990       1000       1010       1020       1030       1040       1050
CCAATGATAGAGTCAACAGGAAGGTTAATGTTGGAAACACAATCAGGTGTGGATTGGTGCTACCTTGAAC

       1060       1070       1080       1090       1100       1110       1120
AAAAGGTCCCCCTGTGGTCTTTTGTTCAACATTGTACAATGTAGAACTCTGTCCAACACTAATTTATTTT

       1130       1140       1150       1160
GTCTTGAGTTTTACTACAAGATGAGACTATGGATCCCGCATGC
```

**Fig. 1.** Sequence of a new human tropomyosin cDNA isolated from liver. One of the clones we isolated was totally sequenced. We used a Pst I restriction site at nucleotide number 457 to split the DNA into two subclones, which were sequenced with the method of Maxam and Gilbert (1980) or the M13 dideoxy method (Sanger et al. 1977). The bases are numbered from the beginning of the sequence, but the first amino acid (cysteine) corresponds to amino acid 58 of muscle tropomyosin. Due to its tissue origin and the similarity of exons 6 and 9 with those of smooth muscle tropomyosin, we classified this cDNA as a smooth muscle type. The 3′ noncoding sequence was not complete, and we found neither Poly-A nor polyadenylation signal.

muscle type. It corresponds to a new gene not yet described in humans, and it is yet not clear which liver cell expresses it. However, it was a missing link in the family of the tropomyosins described until now, and its discovery has allowed us to complete a comparative analysis of these molecules.

*Alignment of the Tropomyosin Sequences*

In this study we used many sequences described in the literature. To avoid confusion, we provide a nomenclature for all the sequences used (Table 1).

The tropomyosin isoforms have been classified into three categories:

1) One class is called skeletal tropomyosins (SK), because it constitutes the major tropomyosin component of skeletal muscle. At least two isoforms (alpha and beta) of 284 amino acids (aa) have been identified (Giometti and Anderson 1984). This class of tropomyosins has also been detected in other cell types (MacLeod et al. 1985).

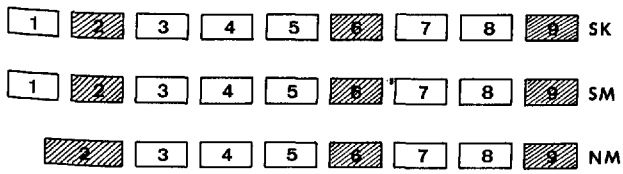


Fig. 2. Schematic representation of the exon composition of the three types of tropomyosin. Muscle tropomyosins have 284 amino acids, whereas nonmuscle tropomyosins have 248, but the different isoforms can be obtained by alternative splicing. Consequently, there are extensive similarities in some regions of the molecules. To clarify the correspondences, the exons are numbered as shown above. Exons subject to alternative splicing are dashed and the others are shown as open boxes. Since nonmuscle tropomyosins probably consist of eight exons, exon 2 of nonmuscle tropomyosin is aligned with exon 3 of muscle-type tropomyosins. Consequently, what is actually exon 1 in nonmuscle tropomyosins is called exon 2 in this paper, the actual exon 2 is called exon 3, and so on.

2) The second category comprises smooth muscle tropomyosins (SM), of 284 aa, which are the major tropomyosin isoforms of smooth muscle. They have also been found in fibroblasts and other epithelial cells (Giometti and Anderson 1984).
3) The nonmuscle tropomyosin isoforms (NM) have 247 aa and are major components of circulating cells (Cote et al. 1978; Giometti and Anderson 1984) and transformed cells (Leavitt et al. 1986). They are also found in fibroblasts and other cells (Giometti and Anderson 1984).

To align the coding sequences determined on cDNAs of different sizes, we used exon separation suggested by the work of MacLeod et al. (1985). We divided the muscle tropomyosins into nine exons and the nonmuscle tropomyosins into eight exons (see Fig. 2 for a schematic representation). The amino acid sequence of exon 3 in the muscle isoforms resembles that of exon 2 in the nonmuscle types. For clarity, we here designate as exon 2 what is in fact exon 1 of nonmuscle tropomyosin. The tropomyosin sequences can be aligned from the beginning of exon 3. They share a high percentage of amino acids and are of exactly the same size. Three tropomyosin exons are subject to differential splicing (see MacLeod et al. 1985; Reinach and MacLeod 1986; and this paper), and we numbered them 2, 6, and 9.

### *Mean Divergence of Individual Exons*

A divergence matrix was constructed for each of the nine tropomyosin exons. When the whole sequence was not known, only the exons with complete sequences were used. We define mean divergence as the sum of all the terms of the matrix normalized to the number of comparisons. Figure 3 shows a plot of mean divergence as a function of the exons. For exons 1, 3, 4, and 8 the mean divergence corresponds to a variation on the third base of the codon. Other exons exhibit a larger mean divergence, implying changes in the amino acid composition. It can be seen clearly that the three exons that undergo differential splicing diverge more than the others.

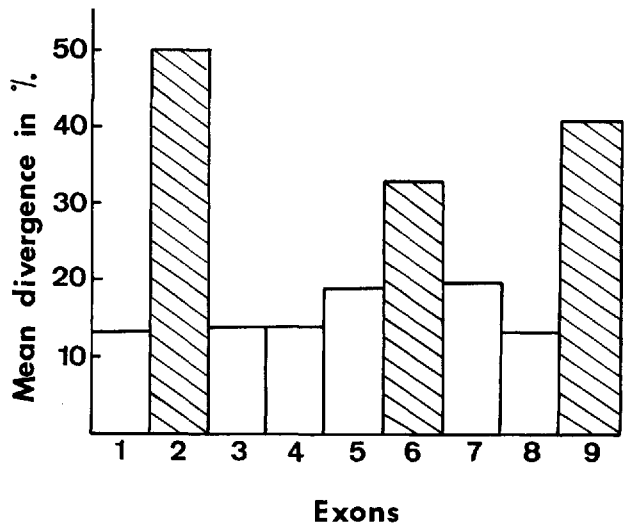


Fig. 3. Analysis of the mean divergence of the different exons. A divergence matrix was calculated for each exon. Mean divergence was calculated as the sum of divergence percentage divided by the number of comparisons. The mean divergence obtained was plotted versus the exon number.

The divergence of exons 2 and 9 is higher than that of exon 6. There are two possible explanations for the greater divergence observed in the exons that undergo differential splicing. First, these exons may not be very important for the function they support. Second, a different function may be supported by these exons. Moreover, both explanations may apply to these phenomena, each on a part of the exon, which would explain why exon 6 differs from exons 2 and 9. To resolve this problem we analyzed the results in more detail.

### *Comparison of Exons Not Subject to Alternative Splicing*

The evolution of the more conserved exons, which are also those not involved in differential splicing, can form the basis for an identification and classification of the different genes. This is important because not all the genes have been identified.

Because exons 4, 5, 7, and 8 are conserved and because their sequences are known (Table 1), we used these sequences to construct a divergence matrix, which gives the evolutionary distance between the sequences (Fitch and Margoliash 1967). We then constructed evolutionary trees. The tree presented in Fig. 4 is rooted, implying that constant evolution

**Table 1.** Nomenclature of the sequences coding for tropomyosins

| Name | Species | Reference | Isoform[a] | Gene A | Gene B |
|---|---|---|---|---|---|
| H1 | human | MacLeod et al. (1985) | SM | $\beta$ | $\beta$ |
| H2 | human | this paper | SM | — | $\alpha$ |
| H3 | human | Reinach and MacLeod (1986) | SK | ? | $\gamma$ |
| H4 | human | MacLeod et al. (1985) | SK | $\beta$ | $\beta$ |
| H5 | human | MacLeod et al. (1986) | NM | $\alpha$ | $\gamma$ |
| H6 | human | MacLeod et al. (1987) | NM | ? | $\delta$ |
| R10 | rat | Ruiz-Opazo et al. (1985) | SK | $\alpha$ | $\alpha$ |
| R11 | rat | Ruiz-Opazo et al. (1985) | SM | $\alpha$ | $\alpha$ |
| R12 | rat | Helfman et al. (1986) | SK | $\beta$ | $\beta$ |
| R13 | rat | Helfman et al. (1986) | SM | $\beta$ | $\beta$ |
| P22 | chicken | Helfman et al. (1984) | SM | $\alpha$ | $\beta$ |
| P23 | quail | Flach et al. (1986) | SK | $\alpha$ | $\alpha$ |
| P24 | chicken | MacLeod (1982) | SK | $\alpha$ | $\alpha$ |

This table summarizes the data known on the tropomyosin sequences, the isotype classification, and the origin of the different isoforms. The first isotype classification is that proposed by the authors who published the sequence (gene A). The second is the isotype classification we have deduced from the evolutionary analysis presented in the Results (gene B)

[a] SK, skeletal muscle; SM, smooth muscle; NM, nonmuscular. Hn, Rn, Pn are the abbreviations used in the figures

took place along its branches. However, relaxing this constraint yielded trees of exactly the same topology.

The results show clearly that the sequences determined until now correspond to four human genes, and some of them have counterparts in the rat and chicken. Therefore, the divergence of the human genes took place before the separation of birds and mammals, and is consequently more than 300 Myr old. The same kind of result was obtained with another muscle/cytoskeletal component, actin (Alonso et al. 1986). The sequence divergence occurs almost exclusively on the third base of the codon and is in 90% of conserved amino acids. Given the high percentage of third-position sites with nucleotide changes among the four defined gene isotypes, it is very likely that absolute estimates of divergences are strongly biased by multiple substitutions and that therefore the precise timing of the duplications yielding the present forms from a single ancestral gene remains hypothetical.

Moreover, Fig. 4 shows clearly that at least four tropomyosin genes are expressed in vertebrates, and it also allows a better classification among the three species studied. We have used this result to reexamine the nomenclature of tropomyosins, which was not totally clear in the literature. We divided tropomyosin genes into four isotypes: $\alpha$, $\beta$, $\gamma$, $\delta$, as shown. This is summarized in Table 1.

With the isotypes defined, it was then possible to analyze the relationships between differential splicing and the functional domains of tropomyosins.

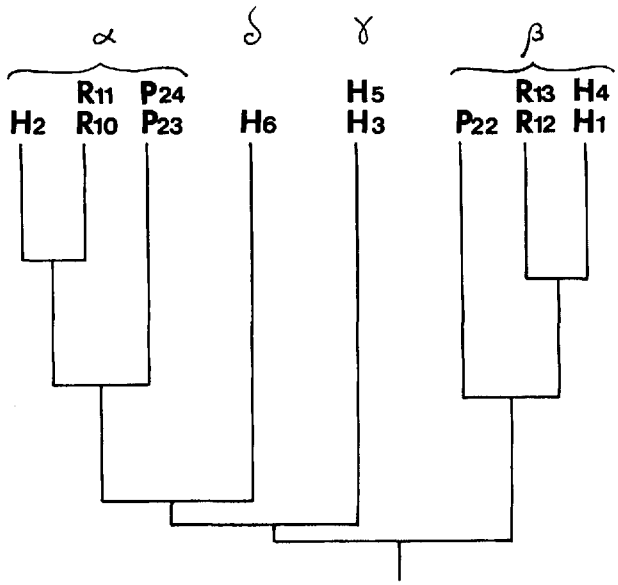


**Fig. 4.** Evolutionary tree of tropomyosin deduced from conserved exons. Exons 4, 5, 7, and 8 (as defined in Fig. 2) were used to calculate the evolutionary tree. The divergence matrix was analyzed with the PHYLIP program to produce phylogenetic trees that were rooted or unrooted. The mean square deviations between the two were very close, hence we present rooted trees for clarity. The length of the branches is proportional to the divergence.

### *Comparison of Exons Subject to Alternative Splicing*

A divergence matrix and evolutionary trees were constructed for exons 6 and 9. We did not perform the same analysis for exon 2 because only part of the known sequences reach this exon. However, the results seem to be comparable.

There is a striking discrepancy between the results shown in Fig. 4 for conserved exons and those of Fig. 5 for exons 6 and 9. For instance, cDNA H1 and cCNA H4, which were together in Fig. 4 because they are transcribed from the same gene, are at the two extremities of the tree in Fig. 5. The same kinds of results were also obtained with H1 and R12. The conserved exons allow discrimination between genes

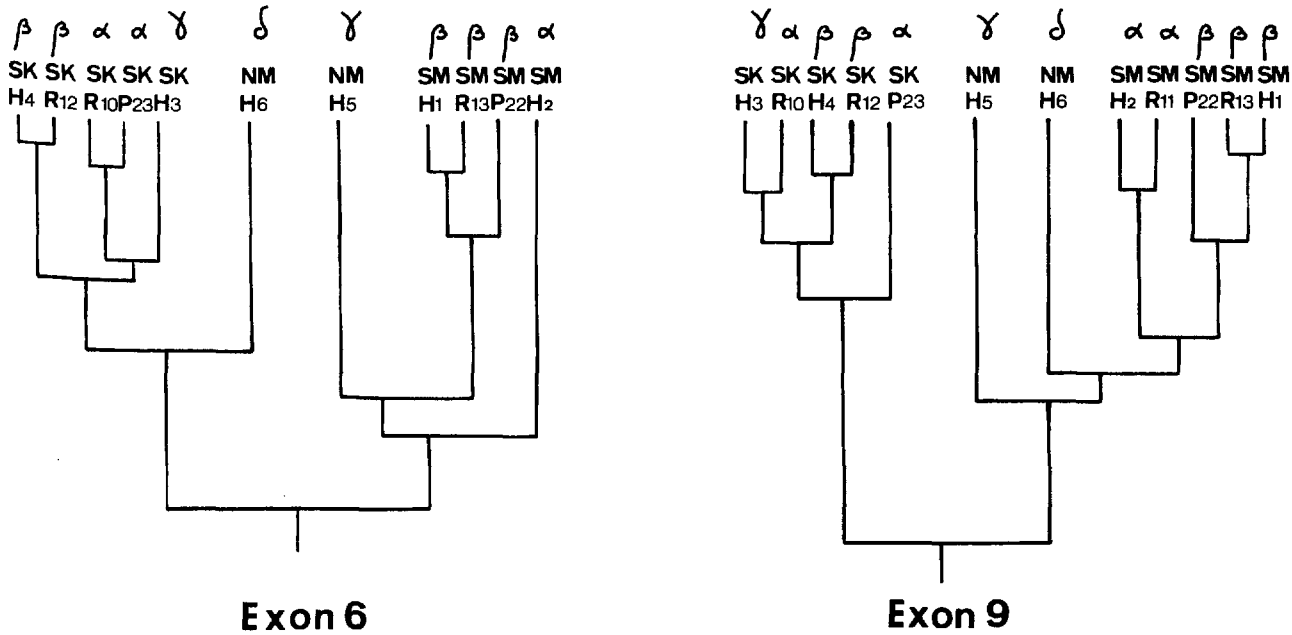


**Fig. 5.** Evolutionary tree computed for exons 6 and 9 in tropomyosin. The phylogenetic trees were calculated as described in Fig. 4.

(Fig. 4); on the other hand, analysis of exons 6 and 9 associates the skeletal muscle tropomyosins (i.e., they are very similar) and to a lesser extent the smooth muscle tropomyosins. For example, exon 9 of human H4 and of rat R12 are identical at the protein level; human H3 and H4 exhibit 94% similarity at the nucleic acid level although they are coded by distant genes (see Fig. 4); and rat R10 and R12 share 22 of their 26 amino acids although they are transcribed from two different genes. On the contrary, still in exon 9, rat R10 and R11, belonging to the same gene (see Fig. 4), have in common only 4 amino acids out of 26. These results argue in favor of a convergence mechanism or a coevolution phenomenon driven by a constraint generated by interaction with another protein. This protein is known in skeletal muscle as the troponin complex. Furthermore, it is known that troponin interacts with the COOH region of tropomyosin (i.e., exon 9) and with cysteine 191 contained in exon 6. However, troponin is present only in skeletal muscle and cannot be responsible for the similarities observed between smooth muscle isoforms. In this case, the phenomenon could be due to a similar mechanism, i.e., interaction with another protein. However, because there are two relatively different types of sequences in exon 9, the mechanism would have to involve different proteins. These proteins have not yet been identified. In exon 6, the difference between the sequences appears to exclude interaction with a unique specific protein. Furthermore, in rat tropomyosins (R10 and R11) the smooth muscle and skeletal muscle isoforms are identical with respect to exon 6. A gene conversion between sequences could explain the proximity of the smooth muscle-type tropomyosin sequences. Because their exons are relatively short, such a conversion would be detectable in the putative introns. The hypothesis should be confirmed or invalidated when the genes are sequenced.

On the other hand, the nonmuscular tropomyosins are less similar (only 58% similarity at the nucleic acid level for H6 and H5). The role of exon 9 in nonmuscle tropomyosin is not clear. Its apparent variability seems to minimize the function of this exon in these isoforms. However, because only two nucleotide sequences of this type are known, it is possible that they reflect two functionally different classes of nonmuscular tropomyosins.

### *Comparison of Codon Usage and G+C Content*

Tropomyosins are expressed in different tissues from different genes, but tissue specificity is also provided by differential splicing. Because of this it was of interest to examine the respective roles of the tissue-specific tRNA pool and the genomic organization of the sequences (Bernardi and Bernardi 1985) in codon usage.

To analyze the influence of codon usage on base composition we calculated the percentages of G+C content in the coding sequences and in the third base of the codon. Figure 6 shows a linear relationship between these two values. Two conclusions can be drawn from this result. First because the proteins are, at least in part, highly conserved, there is only one possible explanation for the increase in the G+C content: a change in the third base of the codon. Indeed the G+C content of the two first bases of the codon lies between 45% and 49% for the se-

quences analyzed. However, another interesting conclusion can be drawn: differential splicing does not dramatically change the G+C content in the coding sequences, although this splicing is characteristic of the tissue and thus the tRNA pool. This indicates that the G+C content does not depend mainly on the tRNA pool, at least in the case of tropomyosins. On the contrary, it suggests that the pressure on the base composition originates from another phenomenon.

Another aspect of the question of codon usage concerns the genomic organization of the sequences. Genes located in the different isochores (Bernardi et al. 1985) are composed of long stretches of DNA with a relatively uniform and distinct base composition. The results presented in Fig. 6 suggest that the alpha and gamma genes are in the same class, while beta genes are in an isochore of high G+C content. There is a general tendency for G+C content to be conserved throughout evolution. The genes we assumed to be of the same origin fall into the same isochore whether they are from human, chicken, or rat.

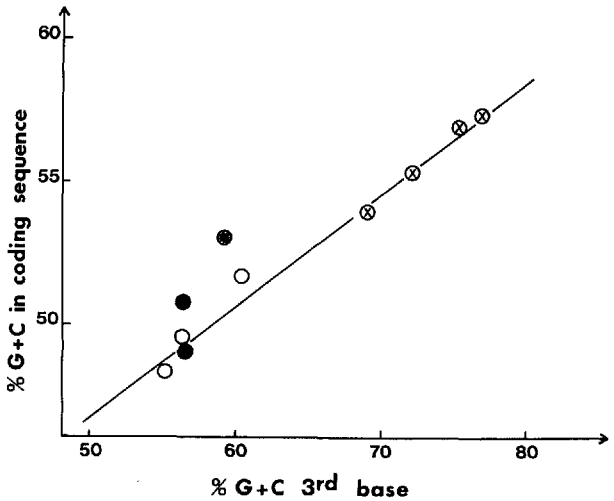


**Fig. 6.** Relation between the G+C content of the coding sequence and of the third base of the codon. G+C content was calculated either for the whole coding sequence or for the third base of the codon. The percentage of G+C obtained for the coding sequence is plotted versus the percentage of G+C of the third base of the codon. ○ Alpha genes; ⊗ beta genes; ● gamma genes; ◒ delta genes.

## Discussion

This comparison of tropomyosin sequences leads to some interesting conclusions. First, the examination of sequences from the standpoint of their evolutionary relationships enabled us to determine the genetic origin of the different sequences published. Before this study, this origin was not clear, because without further data it was difficult to know whether a paralogous or orthologous comparison was appropriate. We showed that there are at least four tropomyosin genes in humans and that all the nonhuman sequences analyzed correspond to one of the genetic categories found in the human. Least-square divergence analysis indicated that the different human genes originated from a genetic duplication that occurred before the separation of mammals and birds, perhaps at the time of the early amphibian radiation. Comparable results have been obtained with another muscle protein, actin (Alonso et al. 1986).

Tropomyosin RNAs undergo differential splicing, which is at the origin of their tissue specificity (MacLeod et al. 1985). We show here that specificity of function led to a convergence phenomenon in skeletal muscle in the case of exons 6 and 9. This phenomenon was probably due to the interaction of troponin with the protein sequences coded by exons 6 and 9 of the tropomyosin genes (Smillie 1979). To explain the same phenomenon in smooth muscle tropomyosin, a similar protein can be assumed to exist in these types of cells. Smooth muscle is regulated by $Ca^{2+}$, but troponin has not been found in it. Direct interaction with $Ca^{2+}$ has been proposed (Cavadore et al. 1985), but this is unlikely because the tropomyosins consist totally of alpha helix. Alternatively, the convergence observed in the SM isoforms could be explained by a conversion phenomenon. The gene must be isolated before this hypothesis can be tested. Our analysis shows that, due to the generation of isoforms by alternative splicing, a molecule can be subject to different constraints on different parts, giving rise to differing convergences in the same molecule.

The analysis of G+C content indicates that tropomyosin base composition is distorted not by the tRNA pool but by its location in the genome. Our results are well explained by the hypothesis of isochore compartmentalization developed by Bernardi and coworkers (Bernardi et al. 1985; Bernardi and Bernardi 1986). These observations suggest that tropomyosin has been subjected to two types of pressure during evolution. Functional pressure has been very important because the tropomyosin molecule interacts with different proteins and is totally composed of alpha helix. However, this pressure is exerted on the protein and thus on the first two bases of the codon. In contrast, the pressure on the third base of the codon seems to arise mainly from genomic compositional constraints such as those described for isochores of various other genes.

*Acknowledgments.* This work was supported by the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, and a grant from the Association pour la Recherche sur le Cancer. We thank B. Jory for expert editorial assistance.

## References


Alonso S, Minty A, Bourlet Y, Buckingham M (1986) Comparison of three actin-sequences in the mouse; evolutionary relationships between the actin genes of warm-blooded vertebrates. J Mol Evol 23:11–22

Bernardi G, Bernardi G (1985) Codon usage and genome composition. J Mol Biol 22:363–365

Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. J Mol Evol 24:1–11

Bernardi G, Olofsson B, Filipski J, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958

Berstein B, Bambourg J (1982) Tropomyosin binding to F-actin protects the F-actin from disassembly by actin depolymerising factor. Cell Motil 2:1–8

Cavadore JC, Berta P, Axelrud-Cavadore C, Haiech J (1985) Calcium binding of arterial tropomyosin: involvement in the thin filament regulation of smooth muscle. Biochemistry 24: 5216–5221

Cooper H, Feurstein N, Noda M, Bassin R (1985) Suppression of tropomyosin synthesis, a common biochemical feature of oncogenesis by structurally diverse retroviral oncogenes. Mol Cell Biol 5:972–983

Cote G, Lewis W, Pato M, Smillie L (1978) Platelet tropomyosin: lack of binding to skeletal muscle troponin and correlation with sequence. FEBS Lett 94:131–135

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Flach J, Lidquester G, Berish S, Hickam K, Devlin R (1986) Analysis of tropomyosin cDNAs isolated from skeletal and smooth muscle mRNA. Nucleic Acids Res 14:9193–9211

Giometti C, Anderson L (1984) Tropomyosin heterogeneity in human cells. J Biol Chem 259:14113–14120

Helfman D, Cheley S, Kuismanen E, Finn L, Yamawaki-Kataoka Y (1986) Non muscle and muscle tropomyosin isoforms are expressed from a single gene by alternative RNA splicing and polyadenylation. Mol Cell Biol 6:3582–3595

Helfman D, Feramisco J, Ricci W, Hughes S (1984) Isolation and sequence of a cDNA clone that contains the entire coding region for chicken smooth muscle tropomyosin. J Biol Chem 259:14136–14143

Hendricks M, Weintraub H (1981) Tropomyosin is decreased in transformed cells. Proc Natl Acad Sci USA 78:5633–5637

Leavitt J, Latter G, Lutomski L, Goldstein D, Burbeck S (1986) Tropomyosin isoform switching in tumorigenic human fibroblast. Mol Cell Biol 6:2721–2726

MacLeod A (1982) Distinct $\alpha$-tropomyosin mRNA sequences in chicken skeletal muscle. Eur J Biochem 126:293–297

MacLeod A, Houlker C, Reinach F, Smillie L, Talbot K, Modi G, Walsh F (1985) A muscle-type tropomyosin in human fibroblasts: evidence for expression by an alternative RNA splicing mechanism. Proc Natl Acad Sci USA 82:7835–7839

MacLeod A, Houlker C, Reinach F, Talbot K (1986) The mRNA and RNA-copy pseudogenes encoding TM30 nm a human cytoskeletal tropomyosin. Nucleic Acids Res 14:8413–8426

MacLeod A, Talbot K, Smillie L, Houlker C (1987) Characterisation of a cDNA defining a gene family encoding TM30 pl a human fibroblast tropomyosin. J Mol Biol 194:1–10

Maxam AM, Gilbert W (1980) Sequencing of end-labelled DNA with base specific cleavage reactions. Methods Enzymol 65: 499–560

McLachlan A, Stewart M (1975) Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. J Mol Biol 98:293–304

Paulin D, Perreau J, Jakob H, Jacob F, Yaniv M (1979) Tropomyosin synthesis accompanies formation of actin filaments in embryonal carcinoma cells induced to differentiate by hexamethylene bisacetamide. Proc Natl Acad Sci USA 76:1891–1895

Reinach F, MacLeod A (1986) Tissue-specific expression of the human tropomyosin gene involved in the generation of the trk oncogene. Nature 322:648–650

Ruiz-Opazo N, Weinberger J, Nadal-Ginard B (1985) Comparison of $\alpha$-tropomyosin sequences from smooth and striated muscle. Nature 315:67–70

Sanger F, Niclen S, Coulson A (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Smillie L (1979) Structure and function of tropomyosins from muscle and non-muscle sources. Trends Biochem Sci 4:151–155

Staden R (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res 10:2951–2960

Stossel T, Chaponnier C, Ezzell R, Hartwig J, Janmey P, Kwiatkosky D, Lind S, Smith D, Southwick, Yin H, Zaner K (1985) Nonmuscle actin-binding proteins. Annu Rev Cell Biol 1:353–402


Received August 1, 1987/Revised and accepted November 1, 1987