A COMPARATIVE ANALYSIS OF GENOMIC DNA SEQUENCE FROM DIFFERENTIALLY EXPRESSED ACTIN GENES IN THE SEA STAR *Pisaster ochraceus*

by

David John Kowbel
B.Sc., Carleton University, 1977

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science
in the Department
of
Biological Sciences

Simon Fraser University
November, 1988

APPROVAL

Name: David Kowbel

Degree: Master of Science

Title of Thesis:

A COMPARATIVE ANALYSIS OF GENOMIC DNA SEQUENCE FROM DIFFERENTIALLY EXPRESSED ACTIN GENES IN THE SEA STAR PISASTER OCHRACEUS

Examining Committee:

Chairman: Dr. C.L. Kemp, Associate Professor

Dr. M.J. Smith, Professor, Senior Supervisor

Dr. D.L. Baillie, Professor

Dr. B.M. Honda, Associate Professor

Dr. K.C. Eastwell, Visiting Assistant Professor, Biological Sciences, Simon Fraser University, Public Examiner

Date Approved Dec. 5th, 1988

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

A comparative analysis of genomic DNA sequence from differentially expressed actin genes in the sea star Pisaster ochraceus

Author: _____
(signature)

David Kowbel
(name)

Dec 6/88
(date)

ABSTRACT

This thesis is a comparative analysis of the DNA sequence of two actin genes in the sea star Pisaster ochraceus. Two clones isolated from a sea star genomic library contain actin genes that are differentially expressed. The cytoplasmic (Cy) actin gene is expressed in eggs and early development. The muscle (M) actin gene is expressed in tube feet and testes. Analysis of DNA sequence from both genes is used to illustrate potential functional and structural constraints at the protein and DNA level.

Both genes contain an 1125 nucleotide coding region interrupted by three introns at codons 41, 121 and 204. The M actin gene contains two additional introns at codons 150 and 267. The intron position at codon 150, although present in higher vertebrates, has not been reported in actin genes from invertebrates. The comparison of introns between genes reveals only a 37% sequence identity.

The M gene coding region has 89.5% nucleotide homology to the Cy actin and differs from the Cy actin gene in 13 of 375 amino acids. The distribution of replacement substitutions is non-uniform with 11 of 13 amino acid changes in the carboxy terminal half of the gene. The overall amino acid sequence of both genes show an actin isotype common to vertebrate cytoskeletal actins and not to vertebrate muscle specific isoforms. A comparison of DNA sequence in coding regions

reveals that the amino terminal region has a significantly low substitution ratio at 4-fold degenerate sites. The codon usage is biased for C in the 3rd position of codons from both genes. Selective pressure on messenger RNA secondary structure may explain some restrictions on codon usage.

Potential promoter sequences are located in the 5' flanking regions of both genes. The Cy gene contains one site that has sequence identity to the core of a transcription-factor-binding site found in other actin genes. The comparison of nucleotide sequence upstream to the coding start site suggests that both genes may have an intron in the 5' untranslated region.

Acknowledgements

I would like to express my sincere thanks to all those who have given me guidance throughout this research. I am extremely grateful to Dr. M.J. Smith, who has given me invaluable advice and help and to Dr. D. Baillie and Dr. B. Honda who critically supported my research. I also acknowledge J. Boom, D. Clarke, S. Prasad, E. Cabot, K. Thomas and L. Brown who gave me encouragement from the beginning of it all. I also wish to thank the Big Man, who made sea stars in His own image.

TABLE OF CONTENTS

Approval.....ii

Abstract.....iii

Acknowledgements.....v

List of Tables.....viii

List of Figures.....ix

Introduction.....1

Materials and Methods.....6

Isolation of genomic clones.....6

Preparation of DNA.....7

Sequencing of subclones.....9

Analysis of sequence.....9

Results.....10

Overall gene structure of Cy and M genes.....10

5' flanking regions.....16

Introns.....21

3' flanking regions.....22

Potential transcript sizes.....23

Coding region.....23

Discussion.....33

Sequence analyses.....33

Post translational modifications.....35

Actin in oocytes and tube feet.....38

Regulation of Actin Expression.....41

The untranslated regions.....44

Developmental expression of actin genes.....46

Actin gene evolution.....51

Conclusions.....56

References.....58

LIST OF TABLES

Table 1. Intron positions, lengths and splice junctions...18

Table 2. Comparison of nucleotide substitutions.....26

Table 3. Codon usage in sea star genes.....32

Table 4. Intron positions from actin genes.....55

LIST OF FIGURES

Figure 1. Overall gene structure of Cy and M actin genes..12

Figure 2. Genomic sequence of Cy actin gene.....14

Figure 3. Genomic sequence of M actin gene.....15

Figure 4. Sequence comparison of 5' and 3' flanking DNA...17

Figure 5. Comparison of coding regions of Cy and M actin..24

Figure 6. Amino acid comparison of variable region in sea star and sea urchin actin genes.....27

Figure 7. Comparison of nucleotide differences and silent substitution ratios across the coding region....30

Figure 8. Comparison of amino acid variability and protein binding sites to sea star actin coding region...42

INTRODUCTION

Actin is a ubiquitous cellular protein in eukaryotes that is a major constituent of the cytoskeletal matrix. In non-muscle cells actin is involved in an array of cellular processes such as cell motility, cytokinesis and cytoplasmic streaming (Korn, 1982). It is also a major component of thin filaments that associate with myosin in differentiated muscle tissue. In vertebrates, actin is a 44 Kd protein that can assemble into microfilaments. The assembly of actin is partially co-ordinated by the interactions with various proteins in a cell cycle or cell specific manner (Pollard 1986; Matsudaira and Jamney, 1988).

The actin monomer is a highly conserved protein that exists in various isoforms. In mammals and most vertebrates there are two major cytoskeletal isotypes designated beta and gamma, and one major muscle isotype labeled alpha. These can be separated by iso-focusing electrophoresis on two dimensional acrylamide gels, and identified by the variable N-terminal amino acid sequences (Vandekerckhove and Weber, 1978; 1979; 1984). The recent isolation of a third cytoskeletal isoform in chickens suggests that more isoforms are yet to be identified (Bergsma et al., 1985). The primary structure of cytoskeletal and muscle actin protein from various invertebrate species are similar to mammalian cytoskeletal isoforms (Vandekerckhove and Weber, 1984). This has lead to the hypothesis that muscle specific isotypes evolved as a duplication event of ancestral

cytoskeletal actin genes.

The actin proteins are encoded by multiple genes. Within eukaryotes the number of actin genes range from 2-4 in Trypanosoma brucei to over 20 in humans (Amar et al., 1988; Buckingham and Minty, 1983). Both the yeast Saccharomyces and the ciliated protozoan Tetrahymena are exceptions to other eukaryotes since they have a single actin gene within the genome (Shortle et al., 1982; Hirono et al., 1987). The actin gene organization is different between taxonomic groups (Buckingham and Minty, 1983). The six actin genes of Drosophila melanogaster are highly dispersed. Chromosomes 3R and 2R each contain two genes while chromosomes 3L and X each bear a single actin gene (Loukas and Kafatos, 1986). In the nematode C. elegans 3 of 4 actin genes are within 12 kb of each other and are located on linkage group V (Files et al., 1983 Landel et al., 1984). A fourth actin gene is on the X chromosome (Krause and Hirsh, 1986). The sea urchin Strongylocentrotus purpuratus has 8 actin genes (Lee et al., 1984). The cytoplasmic actin genes are tightly linked in two clusters. These clusters exhibit polymorphisms which have been used as molecular markers for genetic studies. The actin gene clusters segregate genetically from each other and from a single muscle actin gene within the sea urchin genome (Minor et al., 1987).

The gene structure of actin has been analysed in many organisms (Firtel, 1981; Buckingham and Minty, 1983). The

coding regions are interrupted by a limited number of introns. Generally the number and position of introns are more variable between species than within a species. The coding regions are flanked by 5' and 3' transcribed but untranslated sequences (UTR). Within an individual actin gene family, the 5' UTRs are more homologous than intron sequences. The 3' UTRs may be gene specific within a species but may show an increased homology between the same isoform in different species (Eldridge et al., 1985). Blocks of sequence identity are noted in the 3' UTRs which may have a structural or regulatory function in actin gene expression (Ordahl and Cooper, 1983; Deponti-Zilli et al., 1988). AATAAA polyadenylation signals in 3' UTRs may differ between genes of different species.

Upstream to the transcriptional start are CAAT and TATAAT like sequences similar to the promoter region of other eukaryotic genes (Buckingham and Minty, 1983). Sequence comparisons between similarily expressed actin genes have also revealed other evolutionarily conserved blocks 5' to the promoter regions (Minty and Kedes, 1986; Carroll et al., 1988).

Actin multigene families are expressed in a developmental and/or tissue specific manner (Fyrberg et al., 1983; Mohun et al, 1984; Cox et al., 1986). The use of 3' UTR gene specific probes hybridized to RNA from staged embryos of Xenopus, mice, sea urchin and Drosophila have shown that cytoskeletal isoforms are expressed throughout embryogenesis. Muscle specific actins are expressed in mesodermal germ layer derivatives

leading to myogenic lineages. The microinjection of chimeric gene fusion products into staged oocytes or embryos from sea urchin or mice reveal that proper temporal and spatial expression requires sequences 5' flanking to the gene regions (Shani, 1986; Katula et al., 1987). Temporal and spatial expression may be uncoupled in a Xenopus muscle actin gene. Cell type specific regulation is modulated by sequences within the transcribed region while correct temporal expression requires the 5' flanking segment of the gene (Steinbeisser et al., 1988).

Actin genes may be used as a model system for understanding the evolution of duplicate genes as well as for understanding the regulatory regions that are involved in differential expression. The goal of this thesis is to elucidate the gene structure of two differentially expressed actin genes from the sea star Pisaster ochraceus. Sea stars (Phylum: Echinodermata Class: Asteroidea) display a wide range of phyletic diversity. The echinoderms are also representative of early deuterostomes and therefore are phylogenetically linked to vertebrates (Raff and Kaufman, 1983; Field et al., 1988).

There are five to six actin genes in Pisaster ochraceus as defined by restriction map differences in genomic clones and their differing patterns of expression in ontogeny (Kovesdi et al., 1984; Kovesdi and Smith, 1985a; 1985b). Gene specific 3' UTR cDNA probes hybridized to staged embryo polyA+ RNA indicate that a single gene is expressed as a 2.3 kb transcript in

unfertilized oocytes and through early development (Kovesdi and Smith, 1985b). This is designated the Cy (cytoplasmic) actin gene. A second category of actin genes of which there are two copies per haploid genome (Kovesdi and Smith, 1985a) is designated the M (muscle) class. Expression of M genes is first detected in tube foot polyA+ RNA as a 2.2 Kb transcript. Both Cy and M genes are expressed during spermatogenesis with the M gene product as the less abundant of the two messages (Boom and Smith, in press).

In contrast to Pisaster ochraceus, the sea urchin Strongylocentrotus purpuratus (Class:Echinoidea) has eight actin genes of which six display a temporal and cell-lineage pattern of expression throughout development (Cox et al., 1986; Lee et al., 1986). In the sea star each egg contains over 2 X 10^5 actin transcripts while an egg from S. purpuratus has only 3-5 X 10^3 maternal transcripts from two actin genes (Kovesdi and Smith, 1985b; Lee et al., 1986). Both organisms also have contrasting developmental expression strategies for histone genes (Raff et al., 1984; Banfield et al., 1988). In S. purpuratus an abundant maternal RNA pool for histones is stored in the nucleus while in sea stars the histones have less than 10^2 transcripts/egg (Banfield et al., 1988). The histone pattern of expression in sea stars resembles that of primitive sea urchins where a low abundance of maternal histone transcripts have also been observed (Raff et al., 1984).

Although the expression pattern of sea star actin genes has

been described there is little information on their gene structure. This thesis is a comparative analysis of two sea star actin genes at the DNA sequence level. The gene structure of both genes is compared to actin genes from different species. An analysis of the DNA sequence will also be used to illustrate potential functional and structural constraints upon the actin protein and regulatory regions.

MATERIALS AND METHODS

A) Isolation of genomic clones

The creation of a Charon 4 genomic library from sperm DNA and isolation of genomic clones used in these studies have been previously published (Kovesdi et al., 1984). The genomic library was initially screened with an actin coding probe pDMA2 from Drosophila melanogaster (a gift from Dr. N. Davidson). After a third screen, eighteen clones were categorized into three classes according to restriction map differences. The localization and orientation of the actin coding regions were identified within 5 genomic clones by independent hybridization with pDMA2 and pSpG2, a 3' actin coding probe from the sea urchin S. purpuratus.

The genomic clone λPoA145 is representative of a class that is single copy within the haploid genome (Kovesdi and Smith, 1985a). Hybridization of a cDNA actin probe pPoTA9 from Pisaster ochraceus to PstI digests of individual sea star genomic DNA or to λPoA145 reveal a 4.1 kb positive band common

to both digests. This fragment was subcloned into the plasmid vector pUC19 for further sequencing. The genomic clone λPoA160 is representative of a gene class that does not contain internal EcoRI, BamHI or SstI sites characteristic of the other two actin gene classes (Kovesdi et al., 1984). The class contains 2 non-allelic genes, as shown in dot blot and a Southern blot of an EcoRI digest of individual sea star genomic DNA using a 3' UTR probe pPoTA13/210 (Kovesdi and Smith, 1985a). The actin coding region is confined to an 8.3 kb HindIII fragment of λPoA160. This fragment was cloned into pUC19 for sequencing.

B) Preparation of DNA

The subclones pPoA145 and pPoA160 were digested with restriction enzymes that cut within the inserted 4.2 Kb PstI or 8.3 Kb HindIII fragments. The restriction enzymes used in the digestion of plasmids were purchased from Bethesda Research Labs or Pharmacia and used under the manufacturers specifications with minor modifications. Essentially for 1 μg of DNA, 3-5 units of enzyme was used. The final enzyme volume did not exceed 10% of the final digestion volume. The digestions were usually at 37°C for 1.5 to 2.0 hours.

Restriction fragments were subcloned into recombinant plasmids pUC19 (Messing and Vieira, 1982) or BlueScript (Stratagene) for sequencing. Plasmid DNA isolation by alkaline lysis method and subcloning procedures are described by Maniatis (Maniatis et al., 1982). Approximately 10 μg of

plasmid DNA was electrophoresed on 0.8% low melting point agarose (BRL) at 1.0 V/cm in 1x TAE buffer pH 8.1 (40 mM TrisOH, 2 mM EDTA, 20 mM acetic acid), with 0.125 mM EtBr. Supercoiled DNA was isolated from the gel under UV light at 375 nm and heated to 70°C. One times TE (10 mM Tris, 1 mM EDTA pH 8.0) was added to a final volume of 600 ul and the DNA was extracted with 2 X 300 ul of TE saturated phenol at 70°C and centrifuged at 12,000 rpm for 10 min. The aqueous layer was extracted with 1X volume cold Sevag (4% v/v isoamyl alcohol in chlorofom -20°C) and the DNA precipitated, after adding 1/3 volume of 10 M NH_4OAC, with 2.5 X volume 95% EtOH at -70°C. The DNA was pelleted at 12,000 rpm for 15 minutes at 4°C, washed with 70% EtOH, dried under vacuum for 10 minutes and resuspended in TE. ExoIII/S1 generated unidirectional deletions were done by the method of Henikoff (1987). Five to ten micrograms of supercoiled DNA were used in each set of serial deletions. The clones were doubly digested with restriction enzymes that cut in the polylinker but not in the inserted fragment and left a 5' overhang and a 3' overhang. The 5' overhang restriction site digest was always proximal to the insert. Serial deletions were generated in a 3' to 5' direction by digestion with 250-300 Units of Exonuclease III (BRL) at 37°C. Aliquots were taken every 30 sec and blunt-ended with S1 nuclease and Klenow enzymes (Pharmacia) as described (Henikoff, 1987). The time points were ligated overnight at 14°C with T4 ligase (Pharmacia). Competent E.

coli JM83 cells were used for all transformations.

C) Sequencing of subclones

The preparation of double stranded plasmids for sequencing was modified from the PEG precipitation method (Hattori and Sakaki, 1986). Approximately 2-5 μg of supercoiled template was gel purified and resuspended into 16 μl of sterile H_2O. The templates were denatured with 4 μl of 1 N NaOH, 1 mM EDTA for 5 min. at room temperature and precipitated with 2 μl of 2 M NH_4OAC pH 5.4 and 55 μl of 95% EtOH at -70^{o}C for 15 min. The collapsed pellets were washed and dessicated as above and stored at -20^{o}C. DNA prepared by this method can be stored at least one week without degradation of the template. The plasmids were sequenced by the dideoxy method (Sanger et al., 1977) using a commercially available sequencing kit (Pharmacia) and 800 Ci/mmol[α-^{32}P]dATP (Amersham).

D) Analysis of sequence

The sequence was analysed using the SEQNCE 2 program by Delaney. DNA sequence alignments were done with the ESEE program designed by E. Cabot. The silent site substitutions were calculated by a 2-parameter method which estimates transitional and transversional mutations separately in 1, 2, or 4-fold degenerate sites (Li et al. 1985). All 3-fold degenerate sites were put into the 2-fold category so that the program could be also used for mitochondrial DNA analysis. Four-fold degenerate substitutions/4-fold silent site in 75 nucleotide blocks were normalized by an angular transformation

and means for each block and 95% confidence limits were determined.

RESULTS

Overall gene structure of Cy actin and M actin

The partial restriction maps of pPoA145 and pPoA160 and sequencing strategies are shown in Figure 1. The localization of the actin coding region and flanking regions are also shown on Figure 1. The subclone pPoA145 contains a 4172 nucleotide PstI fragment. The ends of pPoA145 were initially sequenced to orient further subcloned fragments within the PstI insert. A 2.8 kb PstI/EcoRI fragment was inserted into the vector Bluescript (pPoA2.8) to provide a substrate to generate unidirectional deletions from both strands. A 230 bp EcoR1/Sst1 fragment and a 900 bp SstI fragment were both subcloned into pUC19, pPoA230 and pPoA900 respectively. The insert of pPoA230 was completely sequenced while deletions were generated along the anti-sense strand of pPoA900. The remaining 200 bp SstI/PstI fragment from the original pPoA145 clone was completely sequenced.

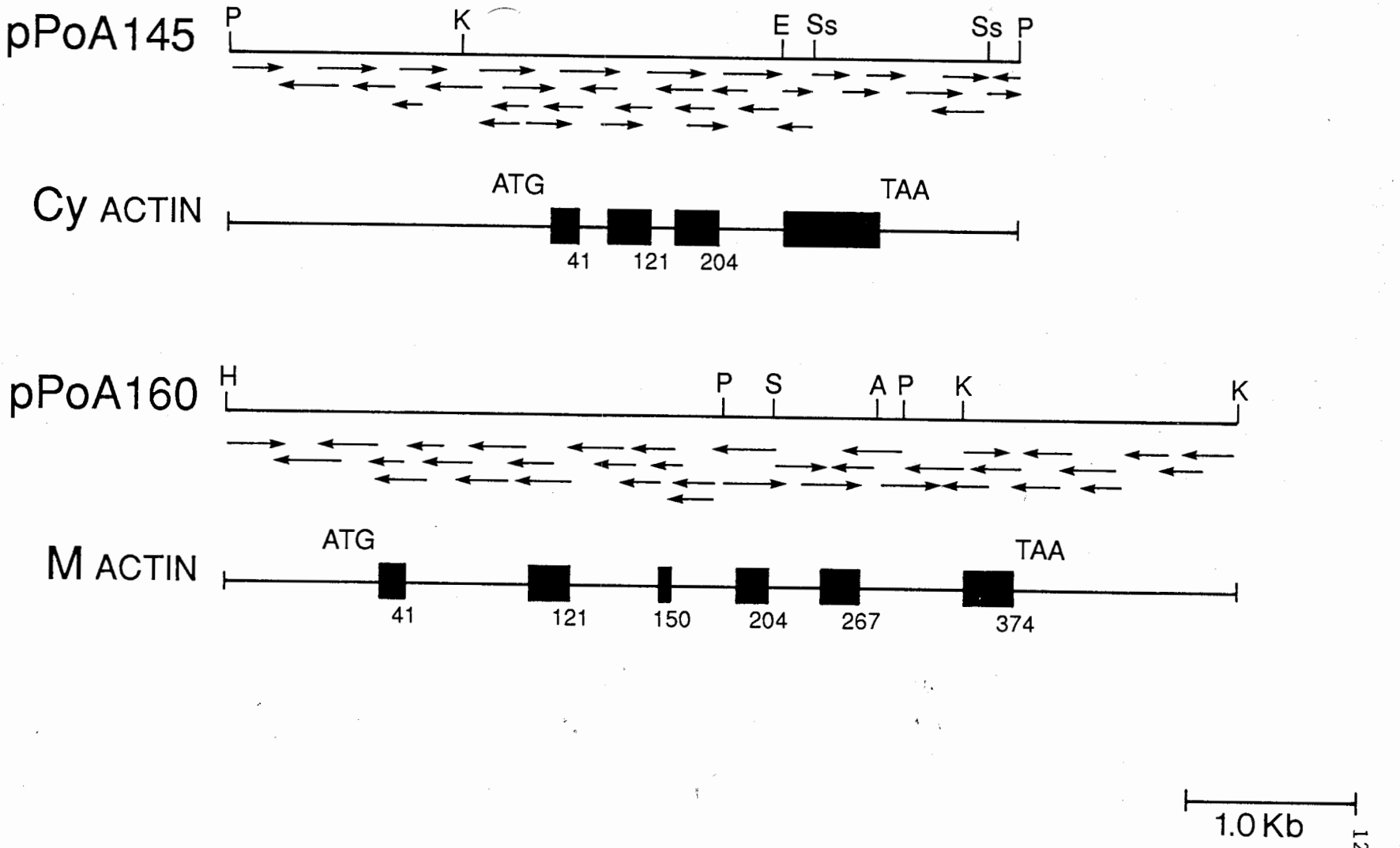
The coding regions of the Cy actin gene were identified by comparison of the DNA sequence to actin genes derived from the sea urchins Strongylocentrotus purpuratus and Lytechinus pictus. The coding regions within the Cy actin gene are shown in Figure 1. The gene contains four exons totaling 1125 nucleotides which translate into a 375 amino acid actin

molecule. The coding region is confined by an ATG translational start codon and terminates with a TAA stop codon. There are three introns located at the junction of codons 41/42, 121/122 and within codon 204. Regions flanking the coding region, 1755 nucleotides 5' to the start codon and 715 nucleotides 3' to the stop codon, were sequenced.

The M gene coding region is contained within an 8.3 kb HindIII fragment that has one internal Sal1 site. pPoA160 was digested with HindIII and SalI, and the 3.3 kb and 5.0 kb insert fragments were ligated into the HindIII/SalI site of Bluescript vector. These subclones are named pPoM3.3 and pPoM5.0 respectively. A 1.0 kb PstI fragment that hybridizes to the actin coding probe pPoA650 was originally subcloned into pUC13 (M. Smith, personal communication) and named pPoA160/1000. The DNA sequence generated from the ends of the 1.0 kb insert showed that the SalI site is located between the Pst1 sites. The sequence generated from the SalI site of pPoM3.3 complements actin coding region indicating that the start of the actin gene is near the HindIII end of pPoM3.3.

The subclone pPoM5.0 was further digested and cloned as smaller fragments to facilitate the generation of deletions and for accurate sequencing. The digestion of pPoM5.0 with ApaI results in four bands that migrate at 4.0 kb, 2.5 kb, 0.8 kb and 0.7 kb. Digestion of the 4.0 kb fragment with HindIII gives two bands of 2.9 kb and 1.1 kb in length. The original 4.0 kb band contained the vector Bluescript and a 1.1 kb

Figure 1.: Overall gene structure of Cy and M actin genes from Pisaster ochraceus. The clone pPoA145 contains the 4.2 kb Pst1 fragment in pUC19. The 8.3 kb HindIII fragment of pPoA160 was restricted with SalI to give a 3.3 kb fragment containing the amino terminal end of the M gene and a 5.0 kb fragment containing the 3' end plus approximately 3.0 kb of flanking DNA. Black boxes represent coding regions which are interrupted by introns at designated codon positions. The arrows represent the extent of sequence obtained in individual reactions. Restriction enzymes used for sequencing are as follows: P (PstI), K(KpnI), E(EcoRI), Ss(SstI), H(HindIII), A(ApaI), S(SalI).


pPoA145
P
K
E
Ss
Ss
P
Cy ACTIN
ATG
TAA
41
121
204
pPoA160
H
P
S
A
P
K
K
M ACTIN
ATG
TAA
41
121
150
204
267
374
1.0 Kb

HindIII/ApaI insert fragment that is located 3' to the coding region of the M actin gene. The 2.5 kb, 0.8 kb and 0.7 kb fragments were subcloned into Bluescript and the ends were sequenced. One end of pPoM2.5 complements part of the pPoA160/1000 and shows that the Apa1 site is internal to the 1.0kb Pst1 fragment by 145 bp. The 0.7 kb Apa1 fragment has one end that is identical to the sequence generated from the SalI site of pPoM5.0 plus polylinker DNA to the ApaI site within the Bluescript vector. The other end is the ApaI site that is internal to the 1.0 kb PstI fragment. The last ApaI fragment is therefore located 3' to the gene and joins the 2.5 kb ApaI fragment to the 1.1 kb ApaI/HindIII fragment.

Unidirectional deletions were generated from pPoM2.5, pPoM3.3 and pPoA160/1000. A 1.5 kb KpnI fragment was known to contain 3' coding region by hybridization of the probe pSpG2 from S. purpuratus (Kovesdi et al., 1984). The 1.5 kb KpnI was ligated into pUC19 (pPoM1.5) and the insert was localized internal to the 2.5 kb ApaI fragment, by comparison of the sequence generated from the ends of pPoM1.5 to that of deletions from pPoM2.5.

The overall structure of the M actin gene was defined by comparison of genomic sequence to the coding regions of the Cy actin. As shown in Figure 1, the whole coding region is defined by six exons which are interrupted by five introns between codons 41/42, 121/122, and within the codons of 150, 204 and 267. The coding region is preceded by the canonical

Figure 2.: The genomic sequence of the Cy actin gene. The start codon is designated +1. The amino acids are located above the codons. Introns are designated by lower case letters. Putative CAAT, TATA boxes, polyadenylation addition signals and transcriptional binding site (TBF) are underlined. Potential 5'UTR intron splice junctions (s.j.) are located at nucleotides -1372 and -52, respectively.

```
CTGCAGTGGAGAATACGCCCACATTTCTTCCTGCATAATGACCTTCTCTAAAACTCAAAAATTCAGAGGAAATATAATCTACCCTATGACCCCCGTTGGACTCATCGAGTGGACACGCCC (-1635)
 Pstl                                                                                CAAT                              TBF
TGAGGATCCCATGGTCCAATAAGAATCAACCAGGCTGCATATTATTGAGAAAAGCTGAACAAAGCTGTGCATGGTTATTCAAATGAACCCCAGCCAGCCTCTAGAATGTGACCATATTTG (-1515)
   BamHI                              TATA
GTAATGTCTCGGTCGCTCGATCACGTGGTACGATGGCCGCGCATATATTAAGGCCTGCATGGCAGACACCTGTTATTCAAGCAGTCATCTCTCAGACTAGTGGAACTCTGAACGCGTAGC (-1395)
                        5's.j.
TCTCAGTAGCTTTGGAACTGTGGTGAGTTAAACATTTTTCTGTTGGGATCTTCAATCAAGAAATAGCTAGTTTTCTATTTAGTAAGCAAGCATTGCTCACCCTACCAGTATTGAGATCGT (-1275)

TCATTACACTTCAAGTGATGTATTTTGTATGTTGCTTGTATCATTGCAGAAACTAACCATTTTAAAATCGAGTGAGATAGAAATGTAGGCTATAAAATCCCAAACAGTGTGTTACTATTA (-1155)

TATTGTTAGAAATGCTAAATTTATTTGACAACTGCATTCAAATGTTTTCGATGCACTTCTATTTTTGAGTCGGGCCTGTCACCCAAGTTGTTAATCATTTAGAAATGTTTATCATGTACA (-1035)

AAATGAGTCACTGGGAGCTTACTTGACGGGTGAAAGATTATGAGTGATTCACCGAACAAGCTGACTCATTGCACTTGCATATAAAAATGGGTCTCTATTAATTATAAGTATTAGTATGCA (-915)

ATTGATTACAATTCTGGACTTTTGGCGAGCAAACGCAATATTTTGAGGGTAAATTCATTCTTTCTAGAATTAAGATGCAATGATGATGTAAAGACTGCTGAAACTTTTTGCCGAATTTCA (-795)

TTTCCGTCTGTTTGAAGGAAATATTCATTATAATTATATGACAATGATGACGAAATATATCGATATTCTTAAATAATTGTATTGCTCTGAATTGTAGGACCTTAAGAAACATTCTACATT (-675)
                                                                                                                Kpnl
TAATTATTTAGGTTCTAACGGCTTTGTAGTATACTGGATTGTTGTAATTTGCAAAAAAAATGTATATTTCCTTGTCAAATTCACCGCGCCAACGCCCCTTTTTGATGGTACCCCAAGCCC (-555)
                                                                                                                    CAAT
ACGCTGCGATCAGGCCACCTGCTTGATACTCGTGCACCGGCCGGCTTTCCCCACACGTGGCTTTCCAGCTATCGAGGCACGCCCCATTACCTTGCCCTGTGGTGGCTCGCCATTGGTCAA (-435)
                      TATA
CAGCTGCATCATTATTAGCATAATAAAATATTCATCAAGGGGGGCGTTGGTACTACCATATTTGGTCATATTTTCTGCTTGGTGTTCGCCTGGCTGACTTCCTCGTTCCATATAGCGTGG (-315)

TCCCAAGTCAACAAAAGACTTGGGGATTTGCATAACATTATTGCGTTGGGTTTGCTTCCTGTTTGCAAAGAAGCGGATATTGTTCGATTTATCTCGCCCTCTTCAAAATTATCTCCTCCT (-195)

TTTCACGAAGTTTATCTTGGATTTTTTAAGGACGGAACAGGTAGTGTACTGTTATATTACGCGAACAAATTTCAACTCTTTTACGTGGAATACGTATACAAATGAATGTTGTTTTATTT (-75)
                                                                                        1
                        3's.j.                                                         M  C  D  E  D  V  A  A  L  V  V  D  N  G  S
ATTCGGCAATTTTTGTTCTTTTTCTTTGCAGACTTTTTCTATTCAACGAGTGAAAAAAAACTACAATCTATCAAAATGTGTGATGAAGATGTTGCCGCTCTGGTCGTAGACAATGGCTCC (45)
               20                                                                  (+1) 41
 G  M  C  K  A  G  F  A  G  D  D  A  P  R  A  V  F  P  S  I  V  G  R  P  R  H  Q
GGTATGTGCAAAGCCGGATTTGCCGGAGACGACGCTCCAAGAGCTGTCTTCCCGTCCATCGTCGGGAGACCCCGCCACCAGgtaaactcaattcccattcttctacaagaaattacttga (165)
attattctgccaatttcaacttgaattggaactgaaacctgttgatttcttttcttttcaaccccagctgtagttgtgttttctatttaaattaacagttcttatattgtgggtttttt (285)
                                                                  60
           G  V  M  V  G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  E  H  G  I  V  T  N  W
cacgtagGGTGTGATGGTTGGTATGGGTCAGAAGGACAGCTACGTCGGAGATGAAGCCCAGAGCAAGAGAGGTATCCTGACTCTCAAGTACCCCATTGAGCACGGTATCGTTACTAACTG (405)
 80                                                               100
 D  D  M  E  K  I  W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  N  P  K  A  N  R  E  K  M
GGACGATATGGAGAAGATTTGGCACCACACCTTCTACAATGAACTGCGTGTGGCACCAGAGGAACACCCCGTGCTCCTGACTGAAGCCCCCCTTAACCCCAAAGCCAACAGAGAAAAGAT (525)
120
 T  Q
GACACAGgtgaatatcttgacaaatgtttctatcagcatgtcactaataaaagtcccgccacgttgtttctgaagttgaactgactctgaacttaattgaactgacatgattgtaatcgc (645)
                                                                              140
                           I  M  F  E  T  F  N  T  P  A  M  Y  V  A  I  Q  A  V  L  S  L  Y  A  S  G  R  T  T  G  I  V  F
tctcctttaattgtttgttgcagATCATGTTCGAGACCTTCAACACACCCGCCATGTACGTTGCCATCCAGGCTGTACTCTCACTGTACGCCTCTGGTCGTACCACCGGAATCGTCTTCG (765)
                    160                                                       180
D  S  G  D  G  V  S  H  T  V  P  I  Y  E  G  Y  A  L  P  H  A  I  L  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  I  L
ACTCTGGTGACGGTGTGTCTCACACTGTGCCCATCTACGAGGGATACGCCCTCCCCCACGCCATCCTCCGTCTGGACTTGGCCGGTCGTGACTTGACCGACTACCTGATGAAGATCCTCA (885)
                    200
T  E  R  G  Y  S  F  T  T  T
CAGAGCGTGGCTACTCCTTCACAACAACTGgtgagtctcaactaaataaaaaagcaatttttttcaatcaaatcctcgtatttcttcttttttttcacacaaaatggccaactcgtgac (1005)
tggaacagtgagagttcaaacaaagttgtaatgttaaactgaagaagacaccaaaccagtcgtgtttccagctaactttccacctcgaattcttcttgaccttgatcaaacttgccaacc (1125)
                                                                                       EcoRI           220
                                                                A  E  R  E  I  V  R  D  I  K  E  K  L  C  Y  V  A  L  D  F  E
ttgcaattactgtaggtgcatgtatctaaatcaattatttatttcttcattttgcagCCGAGCGTGAAATCGTTCGTGACATCAAGGAGAAACTCTGCTACGTTGCCCTCGACTTCGAG (1245)
                              234a             240                                                           260
 Q  E  M  Q  T  A  A  S  S  S  S  L  E  K  S  Y  E  L  P  D  G  Q  V  I  T  I  G  N  E  R  F  R  C  P  E  A  L  F  Q  P
CAGGAGATGCAGACCGCTGCCTCAAGCTCCTCCCTGGAGAAGAGCTACGAGCTCCCCGACGGTCAGGTCATCACCATTGGCAACGAGCGTTTCCGTTGCCCCGAGGCTCTCTTCCAGCCA (1365)
                                                Sstl                                                         300
 S  F  L  G  M  E  S  A  G  I  H  E  T  T  Y  N  S  I  M  K  C  D  V  D  I  R  K  D  L  Y  A  N  T  V  L  S  G  G  S  T
TCCTTCTTGGGAATGGAAAGCGCCGGCATCCACGAGACCACATACAACAGCATCATGAAGTGCGACGTGGACATCCGTAAGGACCTGTACGCCAACACTGTCCTCTCCGGTGGCTCCACC (1485)
                                               320                                                           340
 M  F  P  G  I  A  D  R  M  Q  K  E  V  T  A  L  A  P  P  T  M  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I  G  G  S
ATGTTCCCCGGCATCGCCGACAGGATGCAGAAGGAGGTCACAGCCCTTGCCCCACCCACAATGAAGATCAAGATCATCGCTCCCCCAGAGCGCAAGTACTCCGTATGGATCGGTGGCTCC (1605)
                                               360
 I  L  A  S  L  S  T  F  Q  Q  M  W  I  S  K  Q  E  Y  D  E  S  G  P  S  I  V  H  R  K  C  F  *
ATCCTTGCCTCCCTGTCCACCTTCCAACAGATGTGGATCAGCAAGCAGGAGTACGATGAGTCTGGCCCCAGCATTGTTCACCGCAAGTGCTTCTAAACAAACTGTTAAAAGCATAACCAA (1725)

AGAACTGAACTTGTGGTTCCCTAAGTCTCGATTCTAACACCAAACTGAGACTCCTTACTAACCTTTTACTAAACTTGCTATGATCACATCAACTCAATCTGTGCCTGATATTTTGTTCAC (1845)

TGCACTATAGAAACCTCCCCGTTTTGTTGGCGTATCGGTCTGTTCTATAAAAAAAATAAAATTATAACACACAGGCTAAGAAGTATTTGCTGCCTGACTTGTTAACTTATACTTTAAAAA (1965)

AAAACACGGCTCTCTGAGTTGTCTAGTTACATTGGTACTAATTACAAACTGAAATGTCGTAAAAACAAACTTGAAAAAAAAGACCTTTTAACCGACATTTCATTCGCGGCTCGCAAAGAG (2085)

TTATTCTGATAAAATTGGCTTCAGAAAAACAAAATGTTTAAATTAAAAAAACAAGTGTTTGCATTATCTGTCTATAGTTGGCTGTAAACACGGTCACTATGGGATTATTTTGAAGGAAGA (2205)
                                                                                                                  Hairpin
TTTGATGGATCAGCTCTGAGCTCCGCGAAAAAAAAATTTGTATAGTTTACATTGGTACTAATTTACAAACTGAAATGTCGTGAAAACAAACTTGCAAAACACTTTCTAACCGACATTTAT (2325)
Loop                 Sstl   poly(A)                           poly(A)
GTCGCGGCTTGCGAAGAGTTGCTCTGATAAAATGGCTTCGGAAAAAAAAAAATGGTTAAATTAAAAAACGTTCTTGCAATAGCTGTCCTGCAG                         (2417)
                                                                                          Pstl
```

Figure 3.: The genomic sequence of the M gene. Amino acids are designated above the nucleotide sequence. Introns are shown in lower case. Putative CAAT, TATA, polyadenylation addition signals and 5' UTR intron splice junctions (s.j.) are underlined.

```
AAGCTTGATGTTTTCATTTGTGCGCCATGTTGTCATTTTCACTTTTCAGATTTTACAGTCAGACCAAATACTGTCAGTGCTCTTCAGTTATTTCTGCAGTATTTTCAGAGGCCATTTTAA (-745)
HindIII
AACAGAGAAAGTGTGTGTCCCATAAGCATGCCCATCTGTATAATGAGCATTCCTATAATGACCGCCTAATTTGGAAATCCAGAAGCTGAAACCTTCAACCAATCGCGAGAACTGTTATGT (-625)

CGTTAAGCCAACGGACCAATTAGAGCGCTCGATCTTCTGGTATTTTGACCGCATTCGTCTCCACTCCAAACAAGGTCATGGTGAATACATGCGCTGATGTGGAATGTAGCCGACTATGCC (-505)

GAAGTGTGTGTTACACATTACACTTATTATTTTTATGGGTTAGACATGCAAAGTTGATAAAGACCTAAAACGTGTGAGAAAAAAAAAAATGGGGATCTAACTGGGGTTTATTTTCGGTGGT (-385)

GCTAATACAATCGCAGGAAATGCCGCTTCGTGCTGTTATGTTGTTGAGACAGTTAGTAATGTAAAAAGGATTTGGATGTCAAACACAGATAAGCATTTTATTTGACACAATTTCGCAGTT (-265)
                CAAT                                    CAAT                    TATA
TGTATTTATTTATTTGTTCAATAATTTTTTTTTCAGAATTGGATTTGAACTCTTTCCATCAATTTGACATTTAGGGAGAGAGATAAGAATTTAAACGGACAGGAAAGTTATCGTGCAGCT (-145)
                                                                                                3's.j.
CTCAGCTTTTCCAGGAATGAATTTCGCCTCTCGATCGCCTCAATCTAAACAATTTTGTATACTATCTCAAGTTGGTGTATTTTGTTTACTTTTACCACTTTTCCAGGATTTGTTTGACGT (-25)
                          1                                                  20
                        M  C  D  E  D  V  A  A  L  V  V  D  N  G  S  G  M  C  K  A  G  F  A  G  D  D  A  P  R  A  V  F
GAAGGGAACTAGGTAATCATTCACCATGTGTGACGAGGATGTTGCCGCTTTGGTCGTAGACAATGGCTCCGGTATGTGCAAGGCCGGATTCGCCGGAGACGATGCCCCAAGGGCTGTGTT (95)
                        (+1) 41
  P  S  I  V  G  R  P  R  H  Q
CCCATCCATTGTAGGGAGACCACGTCACCAGgtttgtgaaatttgtctactgttcccatgcatttaagggagaaacgtcaacatattcctattaattgccacctcaacgaaagtcaaaca (215)
aaaagactgtaactcaccatcatgcatgcactcttctctgtgttttccctggcaacaagactggaagaaactgtgaaacatgtgacattgagtatcaaagcaacgtgacattttcgtcat (335)
caaatttcccatttggtcagtctacactacctgcaactctttgcacaacatggagcatgattacaaaaaatgccaattcaacaatgtactgggttgagggtgaaccaaccgtttcataag (455)
acagcccccatttgccatgccagatgtgtggcagagggagtctcccaacaaaaatgttcataggcctacactgtaccacaatcttgtaaggaagttcagccgaaatcatctaaaggattc (575)
ctgagattcccagaacacaatgggggcgggggggggtagaccataactcactgagcaccaaatgtcgcaccggattgtaagatgtcttaatccggatggccgggttaattgttcaaaat (695)
aacaagaacacaggagcaagactattttgaatagtactacaaaacctcaaaccaaaccaagccaaacaaatacttcttcatgaaactttgattcatgactagaatttgatttgtctttat (815)
                                                    60
          G  V  M  V  G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P  I  E  H  G  I  V  T  N
ccatatagGGTGTGATGGTAGGTATGGGTCAGAAGGACAGCTACGTAGGAGACGAAGCCCAGAGCAAGAGAGGTATCCTGACTCTCAAGTACCCCATTGAGCACGGTATCGTCACTAACT (935)
  80                                    100
W  D  D  M  E  K  I  W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  L  L  T  E  A  P  L  N  P  K  A  N  R  E  K
GGGACGATATGGAGAAGATCTGGCATCACACCTTCTACAATGAACTCCGTGTGGCACCAGAGGAGCACCCCGTGCTCTTGACTGAAGCCCCCCTCAACCCCAAAGCCAACAGAGAAAAGA (1055)
 120
M  T  Q
TGACACAGgtatgataatgctggtttgtttgtttgtttgttgttgagtcgttggggtgttggggtgtttatttgtttgtttgtttatttgttttttgggggcggttgatataaggagg (1175)
cattctttgtcctttacacctcagtttgattttaatacactgacacgagtttgatacaaaatggccgcagcatgattacggccctttggcgataacagttgaatcgaqcaqatqttaaaq (1295)
aacagctcttacaaaattatacgtatttgtcaagaatattgtttgaagaattgcatttttatttccactcactataagaagggcatgtatccattatgccgttgtaaattcacttttgt (1415)
ggggatcaagcgtgatctctcttctaagaaagtgaacattcagaagatcttgtattctttcacctaaacgccacacaaatcttttaggaaataagataacaatctggtgttattattgc (1535)
                                                    140
            I  M  F  E  T  F  N  S  P  A  M  Y  V  A  I  Q  A  V  L  S  L  Y  A  S  G  R  T  T
tctctccaaacagATCATGTTCGAGACCTTCAACTCCCCCGCCATGTACGTCGCCATCCAGGCTGTGCTCTCCCTCTACGCTTCCGGTCGTACCACCGgtaaatattctcttctacttttt (1655)
ctaaagaaaaattaaaataaacaacaactgaaaaaatggtgtaacgtttgtgcttgatgggggggaaaaaccacacacacaaatacacacagacgatcgttactctcaaagttttcattte (1775)
cataaattggagaaattctagctccgaaggcgacacttcatcgttagttggttcgttcgtattttgaaagtgcgtataaaaccaacttataaactatcttcttttaaaatgatcctcttg (1895)
                                                                                                                    160
                                                                                                G  I  V  F  D  T  G  D  G  V  S
gttcaatgcagtttgattgaaataatcttcctcgaaacattacagtcatgccaagacttaagtacatcttaaaaatcttctctgcagGTATTGTCTTCGATACCGGCGATGGTGTCTCC (2015)
                                                    180                     PstI                                    200
 H  T  V  P  I  Y  E  G  Y  A  L  P  H  A  I  L  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  I  L  T  E  R  G  Y  S  F
CACACCGTCCCAATCTACGAAGGTTACGCTCTCCCCCACGCCATCCTCCGTCTGGACTTGGCTGGTCGTGACTTGACCGACTACTTGATGAAGATCCTTACCGAGCGTGGCTACTCTTTC (2135)

 T  T  T
ACAACCACGGgtaggtttgtctgtttgttcgtaatgtacacttttataaaaacaaacttttagtgcaacaagtagaagtcgacgtgaaaagttaacgagaaagtcttccaaacttgtcc (2255)
                                                                    SalI
cgttcctcttctgtttttagttgtttatgttggattgccccaattaaatggaaattacattatggcttgtataagtcttcaatacacctcacccattcgtctaggctttttgacaatttc (2375)
                                                                                                    220
                                                A  E  R  E  I  V  R  D  I  K  E  K  L  C  Y  T  A  L  D  F  E  Q  E  M  Q  T
tgtacaaacaattgaaaacgctctctttattttccccctaaagCCGAGCGTGAAATCGTGCGTGACATCAAGGAGAAACTTTGCTACACCGCCTTGGACTTCGAGCAGGAGATGCAGACC (2495)
                234a                    240                                                 260
 A  S  S  S  S  S  L  E  K  S  Y  E  L  P  D  G  Q  V  I  T  I  G  N  E  R  F  R  C  P  E  T  L  F  Q  P  A  F  I
GCTTCATCCAGCTCCTCCTTGGAGAAGAGCTACGAGTTGCCCGACGGTCAGGTCATCACCATCGGAAACGAGCGCTTCAGGTGCCCCGAGACCCTCTTCCAGCCCGCTTTCATTGgtaag (2615)
agaagctgccccccccccatgatccttctatttttctatcaatatcgtggaagttgtagttttacaatataaagttcccacaaaaagaaatatcagattaagagaaaatgtatgcttttctgtt (2735)
caactctatatctcgtggaaaaacagaacacaaacaaatcatatacaggatgattacagcttgaccctgagtgggcataatgtgttgtcttttagatgggagctgtagggcccgtgtaca (2855)
ctaggcctatgattggtagtgcacacacaatgacccagaacacttagcgtggaagagttaatgccggtgtttctggttcacatataccaagcacgcttgttaacaggttctctctggctt (2975)
gctatatctgcagtgattgctttcgcttaatatcagacattcttggtttcaataatactagaaatatgtataaaatataataagtatacggatttcattggatctgatgctttgcattgt (3095)
        PstI                                                                                                G  M  E  S  A  G
tgggttttaaaagtagaagtaccggcccttagacgacccgaaaacaaagttattgagaagttgtgtttactattgtggtgctttctctattctcggtacccagGTATGGAATCTGCCGGT (3215)
                    280                                                 300                 KpnI
 I  H  E  T  T  Y  N  S  I  M  K  C  D  I  D  I  R  K  D  L  Y  A  N  T  V  L  S  G  G  T  S  M  Y  P  G  I  A  D  R  M
ATCCACGAAACTACCTACAACAGCATCATGAAGTGCGACATTGATATCCGTAAGGATCTGTACGCCAACACCGTCCTCTCCGGAGGTACCAGCATGTACCCAGGTATCGCTGACCGTATG (3335)
                    320                                                 340
 Q  K  E  I  Q  A  L  A  P  P  T  M  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I  G  G  S  I  L  A  S  L  S  T  F  Q
CAGAAGGAAATCCAGGCCCTCGCTCCACCCACGATGAAGATCAAGATCATCGCTCCCCCAGAGAGAAAGTACTCCGTATGGATCGGTGGCTCCATCCTTGCCTCTCTGTCCACCTTCCAA (3455)
                    360
 Q  M  W  I  S  K  Q  E  Y  D  E  S  G  P  S  I  V  H  R  K  C  F  *
CAGATGTGGATCAGCAAGCAGGAATACGACGAGTCTGGCCCCAGCATCGTCCACCGCAAATGCTTCTAAACTGTTTGCTGGAAATATTCAGACTGTATTTTTTTTTCAATTATTGTGCAT (3575)

TTTATTGTGCATTTTTTTTGTTTCATCACCAGCGTTGACATGGAAACCGAATGAGTTCACCATGGATTATTATTAATATTATTTATTTGTTCTCAGATTCTACTAATTTCAAAACAGTGC (3695)

AATTGTGGTCAAACTGAATGAAGTTGTCACAGCAGGCCAACCATGTGCAAAAGAAGAAAAAAAAAAAAAAACTACGATACTCTCAGCCTTGAGTCATTCAAACTGAAATTATTCCCTTTTT (3815)

TGGAGTCGCCGTTTTTATTTAGGTCTTTGTTTTGCAAAAAAGTAAGAACTGGCTTCGGTCTAATGAGAGAAGTTAATTCACTGATTGTCTTGAGTGCCTTTGATGGATTGGCCGTAAAAA (3935)
                                                                poly(A)                                         poly
ATCTTGAAGTAAGACCAAGCGAGCAAACTAAGATGACTTTTTTTCCCCCTTTGCTGNACCAAATAAAACAGCTTCAACAGCAACAACAAAATCTCGAAAAAAACCAAACTAGAAAAAATA (4055)
(A)
AAAGGTAGAGCTAGATTTATGATAGAAATAGAGACATTGTAAAAGCGAAAGTTGAATGGGCTGATCGATCTATATTGTTCAAATCAGATACATTTTTATTCTGGTCCGTGCGACTTACGA (4175)

TAGAGGGCGGGACACAAGTTGCAATATCAT                                                                                  (4205)
```

ATG start codon and terminated with an TAA codon. There is 865bp. of continuous flanking sequence 5' to the start codon and 680 nucleotides sequenced 3' to the stop codon. Total DNA sequence of the M gene region is 5170 nucleotides. The complete nucleotide sequences of the Cy and M gene are shown in Figures 2 and 3 respectively.

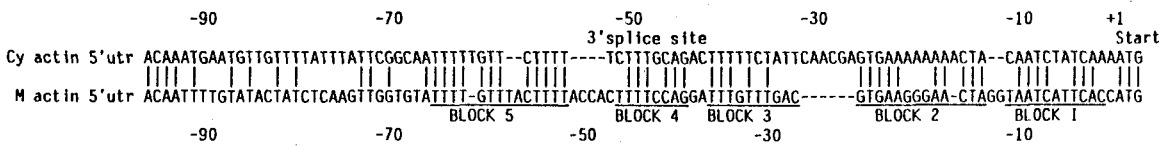
5' Flanking Regions

The nucleotide sequences upstream to the coding regions in both genes were compared to locate 5' UTR and regulatory regions. As shown in other studies the 5' UTR is usually of higher sequence identity than that of non-transcribed regions between actin genes that encode for similar isoforms (Cross et al., 1988; Carroll et al., 1986).

The sea star M gene clone pPoA160 has 865 nucleotides 5' to the start codon. This sequence was compared to an equivalent region from the Cy gene. These two sequences are 45.6% conserved. A segment of 70 bases from -1 to -70 is 67% matched. Within the 70 base pair comparison there are five blocks of identity (Figure 4). Block 1 is a 11 nucleotide sequence "YAATC(T/A)(T/A)CA(C/A)" which is similar to the CAATCATC site that immediately precedes the start codon in sea urchin Cy actin genes (Cooper and Crain, 1982; Akhurst et al., 1987). Block 2 is a purine rich "GTG$(R)_n$CTA" sequence which is followed by a pyrimidine rich sequence "TTT(T/G)TYTR(T/A)Y". This is underlined as block 3 in the Figure. Block 4

Figure 4.: Sequence comparison of actin flanking regions. 5' is the comparison of 5' leader sequences from Cy and M genes for the first 97 nucleotides upstream to start codon (shown on right side of sequence). The blocks of homology are underlined and described in text. 3' is the comparison of 3' flanking regions between the Cy and M genes. The stop codon and polyadenylation addition signals are underlined. The arrows represent inverted repeats that may form stem-loop structures. A putative hairpin loop region in the Cy actin sequence may precede the poly A addition signal.

5′


-90
-70
-50
-30
-10
+1
3'splice site
Start
Cy actin 5'utr ACAAATGAATGTTGTTTTATTTATTCGGCAATTTTTGTT--CTTTT----TCTTTGCAGACTTTTTCTATTCAACGAGTGAAAAAAAACTA--CAATCTATCAAAATG
M actin 5'utr ACAATTTTGTATACTATCTCAAGTTGGTGTATTTT-GTTTACTTTTACCACTTTTCCAGGATTTGTTTGAC------GTGAAGGGAA-CTAGGTAATCATTCACCATG
BLOCK 5
BLOCK 4
BLOCK 3
BLOCK 2
BLOCK I
-90
-70
-50
-30
-10

3′

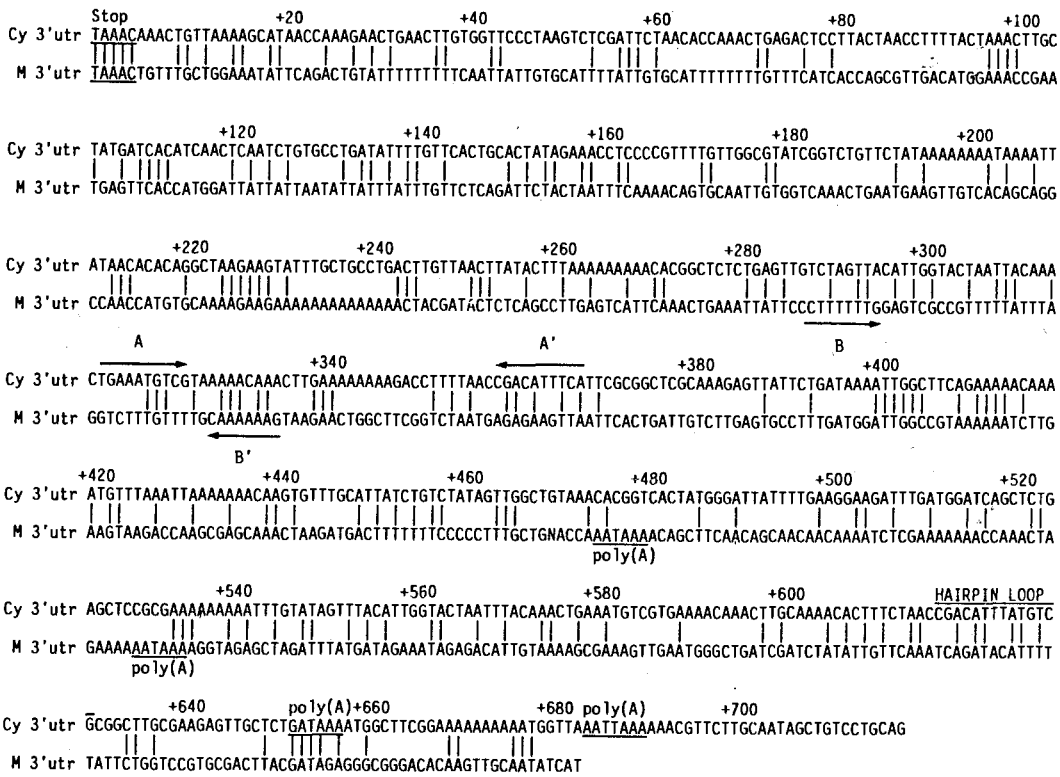

Stop
+20
+40
+60
+80
+100
Cy 3'utr TAAACAAACTGTTAAAAGCATAACCAAAGAACTGAACTTGTGGTTCCCTAAGTCTCGATTCTAACACCAAACTGAGACTCCTTACTAACCTTTTACTAAACTTGC
M 3'utr TAAACTGTTTGCTGGAAATATTCAGACTGTATTTTTTTTTCAATTATTGTGCATTTTATTGTGCATTTTTTTTGTTTCATCACCAGCGTTGACATGGAAACCGAA
+120
+140
+160
+180
+200
Cy 3'utr TATGATCACATCAACTCAATCTGTGCCTGATATTTTGTTCACTGCACTATAGAAACCTCCCCGTTTTGTTGGCGTATCGGTCTGTTCTATAAAAAAAATAAAATT
M 3'utr TGAGTTCACCATGGATTATTATTAATATTATTTATTTGTTCTCAGATTCTACTAATTTCAAAACAGTGCAATTGTGGTCAAACTGAATGAAGTTGTCACAGCAGG
+220
+240
+260
+280
+300
Cy 3'utr ATAACACACAGGCTAAGAAGTATTTGCTGCCTGACTTGTTAACTTATACTTTAAAAAAAAACACGGCTCTCTGAGTTGTCTAGTTACATTGGTACTAATTACAAA
M 3'utr CCAACCATGTGCAAAAGAAGAAAAAAAAAAAAAAACTACGATACTCTCAGCCTTGAGTCATTCAAACTGAAATTATTCCCTTTTTTGGAGTCGCCGTTTTTATTTA
A
A'
B
+340
+380
+400
Cy 3'utr CTGAAATGTCGTAAAAACAAACTTGAAAAAAAAGACCTTTTAACCGACATTTCATTCGCGGCTCGCAAAGAGTTATTCTGATAAAATTGGCTTCAGAAAAACAAA
M 3'utr GGTCTTTGTTTTGCAAAAAAGTAAGAACTGGCTTCGGTCTAATGAGAGAAGTTAATTCACTGATTGTCTTGAGTGCCTTTGATGGATTGGCCGTAAAAAATCTTG
B'
+420
+440
+460
+480
+500
+520
Cy 3'utr ATGTTTAAATTAAAAAAACAAGTGTTTGCATTATCTGTCTATAGTTGGCTGTAAACACGGTCACTATGGGATTATTTTGAAGGAAGATTTGATGGATCAGCTCTG
M 3'utr AAGTAAGACCAAGCGAGCAAACTAAGATGACTTTTTTTCCCCCTTTGCTGNACCAAATAAAACAGCTTCAACAGCAACAACAAAATCTCGAAAAAAACCAAACTA
poly(A)
+540
+560
+580
+600
HAIRPIN LOOP
Cy 3'utr AGCTCCGCGAAAAAAAAATTTGTATAGTTTACATTGGTACTAATTTACAAACTGAAATGTCGTGAAAACAAACTTGCAAAACACTTTCTAACCGACATTTATGTC
M 3'utr GAAAAAATAAAAGGTAGAGCTAGATTTATGATAGAAATAGAGACATTGTAAAAGCGAAAGTTGAATGGGCTGATCGATCTATATTGTTCAAATCAGATACATTTT
poly(A)
+640
poly(A)+660
+680 poly(A)
+700
Cy 3'utr GCGGCTTGCGAAGAGTTGCTCTGATAAAATGGCTTCGGAAAAAAAAAATGGTTAAATTAAAAAACGTTCTTGCAATAGCTGTCCTGCAG
M 3'utr TATTCTGGTCCGTGCGACTTACGATAGAGGGCGGGACACAAGTTGCAATATCAT

Table 1. Introns positions, lengths, and splice junctions of Cy and M actin genes

GENE	POSITION (Codon)	LENGTH (nt.)	5'SPLICE SITE	3'SPLICE SITE
M	-47 nt[a]		NOT FOUND	TTTTCCAG/G
	41	597	G/GTTTGT	CCATATAG/G
	121	485	G/GTATGA	CCAAACAG/A
	150	350	G/GTAAAT	CTCTGCAG/A
	204	273	G/GTAGGT	CCCTAAAG/C
	267	588	G/GTAAGA	GTACCCAG/G
Cy	41	166	G/GTAAAC	TCACGTAG/G
	121	136	G/GTGAAT	TGTTGCAG/A
	204	268	G/GTGAGT	TTTTGCAG/C
	-1373 nt[a]	1329	G/GTGAGT	CTTTGCAG/A

a Position of potential 5' UTR intron of Cy and M genes

"TTT(G/C)CAG" has 100% identity to the 3' acceptor splice site of intron 3 in the Cy actin gene (Figure 4 and Table 1). Block 5 is also a pyrimidine rich sequence "$(T)_n$GTT(TA)CTTTT". The remaining 798 bases that were compared are 45.5% matched.

A total of 1755 nucleotides upstream to the start codon were sequenced in the Cy actin gene. This region was searched for CAAT and TATA like sequences that are representative of promoter regions. Both consensus sequences are observed at six sites. Only CAAT and TATA boxes were considered which showed the evolutionarily conserved spacing between these sequences noted in other actin genes (Buckingham and Minty, 1983). Two pairs of sites seemed to meet the spacing criteria. The first of these pairs is a CAAAT sequence located at nt -1555. This is part of a larger sequence "GGTTATTCAAATGAACC" that may form a hairpin loop structure. A "TATATT" box is located at - 1471. These two consensus sequences are 80 nucleotides apart which is similar to the distance between these sites found in other actin genes (Mohun and Garrett, 1987; Buckingham and Minty, 1983). The start of transcription in most eukaryotic genes is an A nucleotide flanked by pyrimidines (Breathnach and Chambon, 1981). In the Cy gene the sequence "TCATC" is located at -1432 (Figure 2). This sequence is similar to the cap site found in the Cy1 gene of S. purpuratus and to that in X. laevis type 5 and 8 cytoplasmic actin genes (Katula et al., 1987; Mohun and Garrett, 1987). This would place the start of transcription at -1430 which is 36 nucleotides downstream from the putative TATA

box. A second site which has potential CAAT and TATA boxes includes a "CATTGG" site from nt -443 to -439 in the Cy actin gene. Twenty four nts. downstream from this sequence is a "TAATAAAAT" element that may represent a variant TATA box. A putative transcriptional start site is located at nt -380 which is 25 nucleotides from the TATA box.

In the 5' flanking region of the M gene, the CCAAT variant TCAAT occurs at nucleotides -247 and -206 (Figure 3). At nucleotide -181 the sequence TAAGA may be a TATA variant as was suggested for the H3 histones in sea stars (Cool et al., 1988). A TATA variant containing a G, ATAGAA has also been reported in a chicken actin gene (Bergsma et al., 1985).

The Cy actin gene was searched for other regulatory sequences involved in transcription. A sequence of imperfect dyad symetry "ACCATATTTGGT" at nt -1524 to -1513 is located between the putative CAAT and TATA box consensus sequences (Figure 2). This element shows strong homology (9/10 nucleotides) to a "CCATATTAGG" sequence that is reported to be conserved in the promoter region of various actin genes (Mohun and Garrett, 1987). This is part of a larger sequence "GATGTCCATATTAGGACATC" that is a transcription-factor-binding site located 300 bases upstream to a human c-fos proto-oncogene cap site (Hayes et al., 1987).

INTRONS

Intervening sequences were easily identified and localized as a disruption of the primary amino acid sequence. The intron splice junctions were also identified by the GT/AG rule that appears to be universally conserved throughout eukaryotic genes (Mount, 1982). The intron position, lengths, 5' donor and 3' acceptor splice junctions are shown on Table 2. The 5' splice site is always a GT preceded by a G from the adjacent codon. The 3' splice site terminates with a conserved AG sequence and is preceeded by pyrimidine rich blocks of variable length (Figures 2, 3 and Table 1). These are very similar to the consensus sequences for splice junctions as assigned by Mount (1982). The nucleotide 3' to the AG site in the actin gene introns may be a G, A, or C. Thymidine was not observed in that position.

The sequence identity within the introns of each gene was determined. Within the Cy gene, the introns are 50.0% to 57.4% homologous. The sequence identity within the M gene introns are 40.7% to 49.5%. All the introns have a high AT content (58 to 69 %). Some blocks of random homology are observed within these introns mostly in the form of A or T rich blocks. The total intron length of the M gene is 2393 nt. which is 4.2 times that of the combined introns of the Cy actin gene. There is 37% sequence identity between the combined intron sequences from both genes.

The comparison of 5' flanking regions reveals a block of

strong identity that resembles a 3' acceptor splice site at intron 204 (Table 1, Figure 4). This raises the possibility that there is an intron interrupting the 5' UTR in both genes. The sequence 5' to the coding region in the Cy gene was searched for consensus donor splice junctions that are observed in known introns from both the Cy and M genes. A "G/GTGAG" sequence located at -1372 to -1368 has 100% identity to the 5' splice site in the intron that splits codon 204 in the Cy actin gene.

3'flanking regions

There are potential variant poly A addition signals located 3' to the coding region in the Cy gene (Figure 4). Of these, two are located at +651 nt and +683 nt from the stop codon. These two sites are consistent with the 2.3 kb transcript that is expressed by the Cy actin gene (Kovesdi and Smith, 1985b; Boom and Smith, 1988). In the M gene two AATAAA sequences have been located +462 and +527 nucleotides from the stop codon. Both untranslated regions contain inverted repeats that could form imperfect stem-loop structures. A sequence 56 bases upsteam to the AATTAAA site, CGACATTTATGTCG, can form a perfect hairpin structure in the Cy gene. A comparison of the 3' UTRs between genes reveals only a 40% identity, which is a measure of the gene specificity of this region (Kovesdi and Smith, 1985b).

Potential transcript sizes

In the Cy gene, possible transcriptional start sites have been positioned at -382 and at -1430 nucleotides from the start codon. In the former case the processed message would have a 382 nt 5' UTR. The aggregate length of the transcript extending to the putative poly A addition signal would be 2248 nt. This is close to the 2.3 kb length of the Cy transcripts seen in embryos and testes of P. ochraceus (Kovesdi and Smith, 1985b; Boom and Smith, 1988). If the cap site is at -1430, a 111 nt 5' UTR is split by an intron 1329 nt in length. The size of the processed transcript would be 2077 nt long which is approximately 0.2 kb shorter than the values reported by Kovesdi and Smith (1985b). In the M gene the cap site is in the region of -155. This would give, with the 3' UTR and polyA tail, a transcript of 1976 nt in length which is significantly shorter than the 2.1-2.2 kb transcript seen in Northern blots of testes and embryos (Kovesdi and Smith, 1985b; Boom, J., 1988).

Coding Region

A comparison of the 1125 nt coding region from both genes reveals that it is highly conserved, having an 89.5% identity at the DNA sequence level (Figure 5). There are 118 nucleotide substitutions, 73 of which are transitional mutations and 45 are transversional mutations (Table 2). Of the 118

Figure 5.: Comparison of coding region between the Cy and M gene. The amino acid and nucleotide sequences of the Cy coding region are shown on the upper two lines starting with a cysteine codon. The nucleotide and amino acid differences are shown on lines 3 and 4.

```
          C  D  E  D  V  A  A  L  V  V  D  N  G  S  G  M  C  K  A  G  F  A  G  D  D  A  P  R  A  V  F  P  S  I  V   35
Cy Actin TGTGATGAAGATGTTGCCGCTCTGGTCGTAGACAATGGCTCCGGTATGTGCAAAGCCGGATTTGCCGGAGACGACGCTCCAAGAGCTGTCTTCCCGTCCATCGTC (105)
 M Actin .....C..G.....T.....G.....C.....T..C.....G.....G.....A.....T..A

          G  R  P  R  H  Q  G  V  M  V  G  M  G  Q  K  D  S  Y  V  G  D  E  A  Q  S  K  R  G  I  L  T  L  K  Y  P   70
Cy Actin GGGAGACCCCGCCACCAGGGTGTGATGGTTGGTATGGGTCAGAAGGACAGCTACGTCGGAGATGAAGCCCAGAGCAAGAGAGGTATCCTGACTCTCAAGTACCCC (210)
 M Actin ........A..T.....A.....A.....C.....

          I  E  H  G  I  V  T  N  W  D  D  M  E  K  I  W  H  H  T  F  Y  N  E  L  R  V  A  P  E  E  H  P  V  L  L  105
Cy Actin ATTGAGCACGGTATCGTTACTAACTGGGACGATATGGAGAAGATTTGGCACCACACCTTCTACAATGAACTGCGTGTGGCACCAGAGGAACACCCCGTGCTCCTG (315)
 M Actin .....C.....C.....T.....C.....G.....T..

          T  E  A  P  L  N  P  K  A  N  R  E  K  M  T  Q  I  M  F  E  T  F  N  T  P  A  M  Y  V  A  I  Q  A  V  L  140
Cy Actin ACTGAAGCCCCCCTTAACCCCAAAGCCAACAGAGAAAAGATGACACAGATCATGTTCGAGACCTTCAACACACCCGCCATGTACGTTGCCATCCAGGCTGTACTC (420)
 M Actin .....C.....T.C.....C.....G...
                                                                                S

          S  L  Y  A  S  G  R  T  T  G  I  V  F  D  S  G  D  G  V  S  H  T  V  P  I  Y  E  G  Y  A  L  P  H  A  I  175
Cy Actin TCACTGTACGCCTCTGGTCGTACCACCGGAATCGTCTTCGACTCTGGTGACGGTGTGTCTCACACTGTGCCCATCTACGAGGGATACGCCCTCCCCCACGCCATC (525)
 M Actin ..C..C.....T..C.....T..T.....TA.C..C..T.....C..C.....C..C..A.....A..T.....T.....
                                          T

          L  R  L  D  L  A  G  R  D  L  T  D  Y  L  M  K  I  L  T  E  R  G  Y  S  F  T  T  T  A  E  R  E  I  V  R  210
Cy Actin CTCCGTCTGGACTTGGCCGGTCGTGACTTGACCGACTACCTGATGAAGATCCTCACAGAGCGTGGCTACTCCTTCACAACAACTGCCGAGCGTGAAATCGTTCGT (630)
 M Actin .....T.....T.....T..C.....T.....C..G.....G...

                                                                                234a
          D  I  K  E  K  L  C  Y  V  A  L  D  F  E  Q  E  M  Q  T  A  A  S  S  S  S  L  E  K  S  Y  E  L  P  D  G  244
Cy Actin GACATCAAGGAGAAACTCTGCTACGTTGCCCTCGACTTCGAGCAGGAGATGCAGACCGCTGCCTCAAGCTCCTCCCTGGAGAAGAGCTACGAGCTCCCCGACGGT (735)
 M Actin .....T.....ACC...T.G.....T.A..C.....T.....T.G.....
                           T                                    S

          Q  V  I  T  I  G  N  E  R  F  R  C  P  E  A  L  F  Q  P  S  F  L  G  M  E  S  A  G  I  H  E  T  T  Y  N  279
Cy Actin CAGGTCATCACCATTGGCAACGAGCGTTTCCGTTGCCCCGAGGCTCTCTTCCAGCCATCCTTCTTGGGAATGGAAAGCGCCGGCATCCACGAGACCACATACAAC (840)
 M Actin .....C..A.....C...A.G.....A.C.....CG.T...A.T..T.....TCT.....T.....A..T..C.....
                                          T           A    I

          S  I  M  K  C  D  V  D  I  R  K  D  L  Y  A  N  T  V  L  S  G  G  S  T  M  F  P  G  I  A  D  R  M  Q  K  314
Cy Actin AGCATCATGAAGTGCGACGTGGACATCCGTAAGGACCTGTACGCCAACACTGTCCTCTCCGGTGGCTCCACCATGTTCCCCGGCATCGCCGACAGGATGCAGAAG (945)
 M Actin .....A.T..T.....T.....C.....A..TA...G.....A...A..T.....T...C.T.....
                      I                                    T  S    Y

          E  V  T  A  L  A  P  P  T  M  K  I  K  I  I  A  P  P  E  R  K  Y  S  V  W  I  G  G  S  I  L  A  S  L  S  349
Cy Actin GAGGTCACAGCCCTTGCCCCACCCACAATGAAGATCAAGATCATCGCTCCCCCAGAGCGCAAGTACTCCGTATGGATCGGTGGCTCCATCCTTGCCTCCCTGTCC (1050)
 M Actin ..AA..CAG.....C..T.....G.....A.A.....T.....
             I  Q

          T  F  Q  Q  M  W  I  S  K  Q  E  Y  D  E  S  G  P  S  I  V  H  R  K  C  F                              374
Cy Actin ACCTTCCAACAGATGTGGATCAGCAAGCAGGAGTACGATGAGTCTGGCCCCAGCATTGTTCACCGCAAGTGCTTC.                              (1125)
 M Actin .....A.....C.....C..C.....A.....
```

substitutions observed 16 are replacement site substitutions resulting in 13 amino acid changes. All of the amino acid changes are conservative, e.g. codon 129 is a threonine to serine.

Both genes start with a Met-Cys followed by a string of acidic residues Asp-Glu-Asp common to many actin genes (Fornwald et al., 1982; Crain et al., 1987). When the overall amino acid sequences are considered, both sea star genes show an isotype that is common to mammalian cytoskeletal actin proteins as do the sea urchin actin genes (Cooper and Crain, 1982; Akhurst et al., 1987). Among the sites that distinguish muscle and cytoskeletal isotypes both Cy and M genes show cytoskeletal isotypes at codon 5, 6, 10, 17, 76, 103, 162, 176, 201, 225, 296, 298, 357 and 364 (Vandekerckhove and Weber, 1978; 1979). At residues 259, 266 and 286 the Cy gene has cytoskeletal isotype while the M gene has muscle-like residues. Both genes have a muscle isotype at codons 271 and 278. A comparison of the M gene region between codons 259 and 278 to the SpCIIIa and SpG28 actins from S. purpuratus (Akhurst et al., 1987; Crain et al., 1987) indicate isotype similarity while the SpCy1 gene (Cooper and Crain, 1982) is more like the Pisaster Cy gene (Figure 6).

TABLE 2 The comparison of nucleotide substitutions between Cy and M actin genes

	DEGENERATE SITES		
	0-fold	2-fold	4-fold
Average # of Sites	751	199	175
Transitions	5.0	34.0	34.0
Transversions	11.5	1.8	31.7
Trns/Trvns	0.43	18.9	1.07
Substitutions/Site	0.02+0.005	0.22+0.04	0.53+0.08
Silent Substitution Ratio (Ks) 0.56+0.07			
Replacement Substitution Ratio (Ka) 0.02+0.005			

Figure 6.: An amino acid comparison of variable regions within genes from Pisaster ochraceus and Strongylocentrotus purpuratus. The regions compared are between codons 231 and 320. The first three sequences are cytoplasmic actins while sequences four and five are expressed in muscle.

	231										240								
PoCyact	A	S	S	S	S	L	E	K	S	Y	E	L	P	D	G	Q	V	I	T	I
SpCy1act	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SpCy3act	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PoMact	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SpMact	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	250										260								
PoCyact	G	N	E	R	F	R	C	P	E	A	L	F	Q	P	S	F	L	G	M	E
SpCy1act	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
SpCy3act	.	.	.	.	.	.	A	S	.	T	.	.	.	.	.	.	I	.	.	.
PoMact	.	.	.	.	.	.	.	.	.	T	.	.	.	.	A	.	I	.	.	.
SpMact	.	.	.	.	.	.	.	.	.	T	.	.	.	.	A	.	I	.	.	.

	270										280								
PoCyact	S	A	G	I	H	E	T	T	Y	N	S	I	M	K	C	D	V	D	I	R
SpCy1act	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	G	.	.	.
SpCy3act	.	.	.	.	.		.	C	.	.	R	.	.	.		.	.	.	.	.
PoMact	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.
SpMact	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.

	290										300								
PoCyact	K	D	L	Y	A	N	T	V	L	S	G	G	S	T	M	F	P	G	I	A
SpCy1act	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SpCy3act	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
PoMact	.	.	.	.	.	.	.	.	.	.	.	.	T	S	.	Y	.	.	.	.
SpMact	.	.	.	.	.	.	.	.	.	.	.	.	T	S	.	Y	.	.	.	.

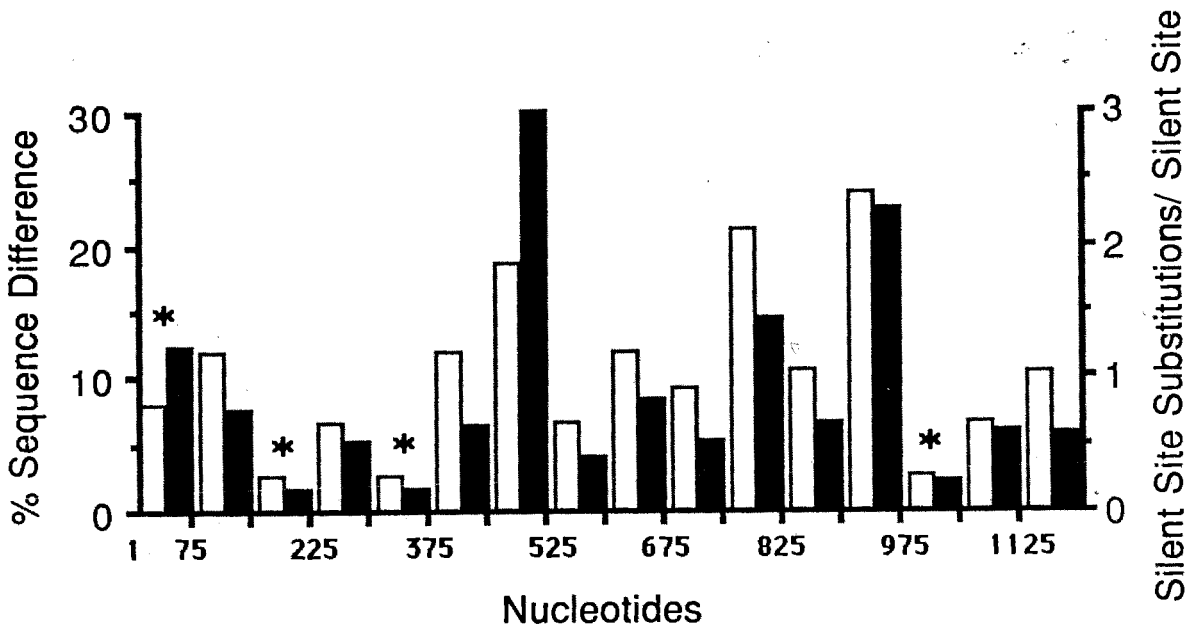
	310										320								
PoCyact	D	R	M	Q	K	E	V	T	A	L	A								
SpCy1act	.	.	.	.	.	.	I	.	.	.	.								
SpCy3act	.	.	.	.	.	.	I	S	.	.	.								
PoMact	.	.	.	.	.	.	I	Q	.	.	.								
SpMact	.	.	.	.	.	.	I	.	.	.	.								

Two methods were initially used to calculate the amount of overall sequence divergence between both genes. The percentage corrected divergence (PCD) method assumes that substitutions occur randomly between the 4 nts and equally weighs pathways that lead to either silent or replacement substitutions. Potential synonymous (silent) sites are classified into three categories according to whether these sites can afford 1,2 or 3 silent changes. Once calculated, each observed silent site is scored one point for each category to which it conforms. Each category is then corrected for multiple hits (Perler, et al., 1980). With this method the overall corrected percentage of silent substitutions was 106%. This suggests that the two genes are saturated for multiple hits at the silent sites.

The second analysis method used is similar to the PCD method except that transitional and transversional substitutions are treated separately and categorized into zero-fold, two-fold, or four-fold degenerate sites. Four-fold degenerate sites are the same as category 3 for silent substitutions in the PCD method. The weighted pathways taken by this procedure correspond to the available codon degeneracy in over thirty mammalian genes (Li et al., 1985). By this method the corrected silent substitutions/ silent site (Ks) is 0.56 (Table 2). The substitution ratios in 4-fold degenerate sites are essentially neutral since changes in these sites do not result in an amino acid change. When both methods are compared they both gave a corrected $Ks_{4\text{-fold}}$ value of 0.53.

There is a non-uniform distribution of nucleotide and amino acid substitutions within the coding region of the genes, with 33 nucleotide sustitutions and one amino acid change in the first 150 codons (Figure 5). To assess the amount of sequence variability across the coding region, the 1125 nucleotides were divided into 15 blocks of 75 nucleotides. The percent sequence difference was calculated manually and the silent site subsitutions/silent sites (Ks) were calculated within each block (Figure 7) by the method of Li (Li et al., 1985). Both percent sequence differences and Ks values are non-uniformly distributed across the coding region. The highest percent differences are seen between nucleotides 450-525, 750-825 and 900-975 and are 18.7%, 21.3% and 24.0% respectively. These blocks are saturated for silent substitutions with Ks values of 2.98, 1.43, and 2.25 respectively. Regions with a percent sequence difference below 10.5% are found between nucleotides 1-75, 151-225, 226-300, 301-375, 526-600, 676-750, 976-1050, and 1051-1125. All but one of these blocks have Ks values below 0.56 e.g., blocks 3, 4,and 5 have Ks values of 0.14, 0.49, and 0.14. The first block has a percent sequence difference of 8.0% but a Ks value of 1.21. To test if the reduced Ks values are due to constraints on either the DNA or protein level, the 4-fold substitutions/4-fold sites were calculated manually in each 75 nucleotide block. The values were normalised by angular transformation and the mean calculated as 0.33 + 0.15. Four blocks were below the mean

Figure 7.: A comparison of percent sequence differences (white boxes) and silent site substitution ratios (black boxes) between the Cy and M actin coding regions. The coding region was broken into 75 nucleotide blocks and comparisons were done as outlined in Materials and Methods. The asterisks (*) represent regions that have 4-fold degenerate substitutions/ 4-fold sites that are significantly below the mean of 0.33 + 0.15 for the whole coding region. The last pair of bars represent the percent sequence difference (10.5%) and average silent substitution ratio (0.56) over the entire coding region.


% Sequence Difference
30
20
10
0
3
2
1
0
Silent Site Substitutions/ Silent Site
*
*
*
*
1
75
225
375
525
675
825
975
1125
Nucleotides

value ($p < 0.05$). Blocks 1, 3,5 and 14 have uncorrected $Ks_{4\text{-fold}}$ values of 0.0, 0.11, 0.10. Blocks that have a reduced Ks value but a normal $Ks_{4\text{-fold}}$ value may be due to constraints on the amino acid level. Examples of these are blocks 2,4,12 and 15 which have $Ks_{4\text{-fold}}$ values of 0.47, 0.25, 0.39 and 0.18 respectively.

The codon usage for each gene was studied to determine if a codon bias may contribute to the observed increased DNA homogeneity seen across the coding regions of the two genes. As seen in Table 3, the Cy gene uses only 49 of the 61 available codons while the M gene uses 50 of the sense codons. Neither gene uses the codons TTA(Leu), CTA(Leu), ATA(Ile), TCG(Ser), CCT(Pro), GCG(Ala), TAT(Tyr), CGA(Arg), CGG(Arg) and AGT(Ser). Table 3 also shows the codon usage of a histone H3 gene from Pisaster ochraceus (Cool et al., 1988). The H3 gene uses only 46 of the 61 sense codons. The comparison of the codon table for all three genes reveals that CCT(Pro), TTA(Leu) and CTA(Leu) are the only codons not utilized by the actin or histone genes.

The base composition of the actin genes shows that the 3rd position of the codons is 48.8% cytosine in the Cy actin gene and 48.3% cytosine in the M actin gene. Two codons from both genes do not exhibit this bias. The codon CGT(Arg) is prevalently used over CGC(Arg) and GGT(Gly) is prevalently used over the other GGN codons in both genes.

Table 3. Codon usage in sea star genes

		Cy	M	H3			Cy	M	H3
Arg	CGT	10	10	4	Val	GTT	7	1	0
	CGC	3	2	3		GTC	8	9	1
	CGA	0	0	7		GTA	3	5	0
	CGG	0	0	2		GTG	6	6	5
	AGA	4	4	2					
	AGG	1	2	0	Ile	ATT	4	5	2
						ATC	22	24	4
Leu	CTT	3	3	3		ATA	0	0	1
	CTC	11	11	5					
	CTA	0	0	0	Lys	AAA	3	3	5
	CTG	10	4	3		AAG	16	16	16
	TTA	0	0	0					
	TTG	3	8	1	Asn	AAT	2	2	1
						AAC	7	7	0
Ser	TCT	4	4	0					
	TCC	12	13	0	Gln	CAA	1	1	1
	TCA	2	1	0		CAG	12	13	5
	TCG	0	0	1					
	AGT	0	0	3	His	CAT	0	1	0
	AGC	8	8	2		CAC	9	8	2
Thr	ACT	6	4	3	Glu	GAA	8	10	5
	ACC	10	17	5		GAG	19	17	3
	ACA	8	2	2					
	ACG	0	2	0	Asp	GAT	5	7	0
						GAC	17	15	3
Pro	CCT	0	0	0					
	CCC	14	12	3	Tyr	TAT	0	0	1
	CCA	5	8	1		TAC	15	16	2
	CCG	1	0	2					
					Cys	TGT	1	1	0
Ala	GCT	7	11	4		TGC	5	5	1
	GCC	21	16	7					
	GCA	1	1	3	Phe	TTT	1	0	0
	GCG	0	0	4		TTC	13	13	4
Gly	GGT	13	17	2					
	GGC	8	5	2					
	GGA	6	5	2					
	GGG	1	1	1					

DISCUSSION

Sequence Analyses

The initial analysis of the two sea star actin genes revealed that the genes are highly conserved at the nucleotide and protein level. To understand if the observed DNA identity was due to functional constraints at the protein level, extensive analysis was done on the DNA sequence. The 2 parameter method (Li et al., 1985) was used for two reasons. First, the 2 parameter method treats transitions and transversions separately in the calculations of substitution ratios between two sequences. Randomly there are twice as many possible substitution pathways for transversions than for transitions at any site (Kimura, 1980). It has been well documented that homologous sequences between closely related species have higher transition/transversion ratios than sequences from more diverged species (Li et al., 1985; Jukes, T., 1988). Second, deviations from this ratio do occur within the individual codon positions (Kimura, 1980). It is essential therefore to understand the pattern of distribution of transversions and transitions if constraints on the DNA or protein are to be assessed. To calculate the pattern of transition/transversion, the codon positions were categorized into non-degenerate, 2-fold degenerate and 4-fold degenerate sites. The 3^{rd} position of 32 of all the 61 sense codons are 4-fold degenerate. Substitutions in these sites are synonymous. Twenty four of the 3^{rd} positions are 2-fold

degenerate. Substitutions in these sites are synonymous if transitions occur. Transversions in these sites are non-synonymous. The 2nd position of all the sense codons are non-synonymous. The 1st position of 4 codons are 2-fold degenerate and the remainder are non-degenerate. The analyses included the third position of isoleucine codons as 2-fold degenerate although they are in fact 3-fold degenerate in nuclear genes. This change allowed the program to be also used on mitochondrial genes, where they are two-fold degenerate. Generally the 3rd position of codons can tolerate more substitutions than the 1st or 2nd position of codons without a change in the amino acid sequence.

When the 2 parameter method is applied to the 2 actin genes, it is seen that the ratio of transition/transversions is 73:45 (Table 2). The transitions are equally distributed over 2-fold and and 4-fold degenerate sites, yet the ratio of transitions/transversions in 4-fold degenerate sites is nearly 1:1. The bias towards transitions in 4-fold degenerate sites has been seen in many genes (Li et al., 1984; Jukes, 1987). Transitional biases are correlated with mispairs which, due to rare tautomerisms within a base, tend to favour transitions more than transversions (Topal and Fresco, 1976). Transversions would tend to accumulate more as the two sequences diverge. Transitions also predominate in 2-fold degenerate sites where the transition to transversion ratios are 19:1. The lowest substitution ratios are seen in the

nondegenerate sites where changes are replacement site substitutions. The ratio of transitions to transversions is 1:2.3. Thus transversions predominate in nondegenerate sites. Since any substitution in these sites cause amino acid changes, there is a definite bias towards transversions in the non-degenerate sites that are not observed in the other two categories. This contradicts the findings in mammalian genes where transition to transversion ratios are near unity in these sites (Li et al., 1985). This cannot be explained by the tissue specific isotypes at some of the codons. An example is GCT(Ala) to ACC(Thr) at codon 259 or GTG(Val) to ATT(Iso) at codon 286. The substitution in the first codon position is a transition in both cases. It is seen that multiple substitutions have occurred in most of the codons that result in a change in the amino acid. The amino acid changes also reside in blocks of coding sequence that have a percent nucleotide difference that is higher then 10.5% seen across the whole coding region (Figure 5). It is therefore probable that an increase in transversions may be in part due to multiple substitutions at nondegenerate sites.

Post Translational Modifications

The start of the coding region varies between actin genes. In invertebrates e.g. Drosophila, C. elegans, S. purpuratus, the start codon ATG is immediately followed by a Cys residue (Fyrberg et al., 1981; Cooper and Crain, 1982; Files et al.,

1983). The muscle actin genes of vertebrates also have a Cys residue following the start codon but the cytoplasmic isotypes of Xenopus, rats or humans do not. In these genes the Cys residue is substituted with an Ala residue as is seen in Xenopus (Cross et al., 1988) or, alternatively, is removed all together as seen in human cytoskeletal actins (Erba et al., 1986). The Ala residue has been observed in a third chicken cytoskeletal isoform (Bergsma et al., 1985) and is common amongst soybean and maize actin genes representative of plant genes (Shah et al., 1983). This suggests that Cys to Ala changes have occurred at this amino acid position throughout evolution. The ATG codon is ubiquitous as a start codon in nuclear eukaryotic genes. It differs from prokaryotic initiation codons by binding to a Met-tRNA that is not formylated (Moldave, 1985). It has been hypothesized that the AUG codon is recognised by the 40S ribosomal subunit to initiate translation by a scanning mechanism along the mRNA template. Factors that are involved in recognition of the start codon though are still obscure. In one study the sequence 5'(A/G)NNAUGG 3' is optimal for recognition by the 40S subunit. This is partially seen in both the Cy and M genes but the 3'G residue is replaced by a T (see Fig. 2 and 3). The sequences noted in Block 1 of Figure 4 may facilitate start codon recognition.

The first amino acid is shown to be post-translationally removed in all actin genes with the protein sequence starting

with an acetylated Asp or Glu residue (Vandekerckhove and Weber, 1984). The removal of a Cys or Ala residue from the genomic sequence of some actin genes may be an evolved modification of the post-translational process.

Within the sea star genes the Cys residue following the start codon uses the codon TGT. The other five cysteine residues use the codon TGC. This pattern of codon usage for Cys residues is also observed in sea urchin actins (Cooper and Crain, 1982), Drosophila actins (Fyrberg et al., 1981) and, in the muscle actin genes of vertebrates (Fornwald et al., 1982; Hamada et al., 1982; Hu et al., 1986). Actin genes that have an alanine instead of cysteine after the start site also display a bias for the codon GCA at that position. The cytoskeletal actin genes observed in Xenopus laevis or Xenopus borealis have the codon GCA(Ala) following the start site. This codon is only utilized in three of the other 28 Ala residues. The other 24 Ala residues use GC(Py) (Cross et al., 1988). One possible explanation for the codon bias seen at this site is that a particular secondary structure of the mRNA is needed for initiation of translation but does not extend past the first Cys or Ala residue. After translation of the message the residue is no longer needed and is subsequently cleaved. Alternatively, the cleavage of the amino-terminal residue may affect the half-life at either the protein or RNA level (Bachmair et al., 1986; Pachter et al., 1986). A change of the amino-terminal amino acid of β-galatosidase can alter

the stabilty of the protein when it is introduced into Saccharomyces cerevisiae (Bachmair et al., 1986). In Chinese hamster ovary cells tubulin expression is autoregulated at the translational level. Co-translational binding of tubulin subunits to the amino-terminal end of nascent polypeptides can destabilize the polysome bound mRNAs for tubulin (Pachter, et al., 1987). In the sea urchin Lytechinus pictus unpolymerized tubulin can destabilize mRNA and decrease the rate of tubulin synthesis (Gong and Brandhorst, 1988).

Actin in oocytes and tube feet

The maturation of sea star oocytes is triggered by the endogenous hormone 1-methyl adenine (1-MeAde) (Kanatani, 1969). The maturation process is necessary for the release of oocytes from meiotic prophase during spawning. In isolated oocytes from Pisaster ochraceus the hormone induces polyadenylation and translation of maternal RNA (Jeffery, 1977) as well as an array of cellular events including a cyclic assembly and depolymerization of microtubules and actin from the cortex during the cell cycle (Schroeder and Otto, 1984; Otto and Schroeder, 1984). The actin filaments form transient spike-like protrusions from the cell surface within 10 minutes of induction by 1-MeAde and disassemble again before germinal vesicle breakdown (Schroeder, 1981). Protein profiles of the cortex appear the same before and after spike formation which suggests that actin is recruited from existing cortical

constituents (Otto and Schroeder, 1984).

The regulation of actin polymerization is dependent on interactions with numerous proteins (Pollard, 1986; Hambly et al., 1986). The binding domains on the actin monomer are under selective contraints and should be conserved at the amino acid level. The known binding sites to actin are shown on Figure 8. In sea star oocytes some of the actin binding proteins have been identified. Depactin, from the sea star Asterias amurensis, inhibits F-actin assembly by binding actin monomers in a molar ratio of 0.63:1.0 (Mabuchi, 1983). This protein binds to the $-NH_2$ and carboxyl termini of actin monomers. Chemical cross-linking studies have shown that the binding sites are near residues 1-4, 11, 359, 361, 363, 364 and 373 of actin (Sutoh and Mabuchi, 1984). Myosin is also detected in oocytes from Pisaster ochraceus. Antibodies to sea urchin myosin do not appear to couple with the actin associated spikes and retain a uniform pattern of distribution throughout maturation (Otto and Schroeder, 1984). Myosin does inhibit depactin binding to the actin monomer in vitro and is highly dependent on ATP concentrations for competitive inhibition of depactin (Mabuchi, 1982). The myosin S1 subfragment binds to the N-terminus of actin between amino acids 1 and 11. The light chain subunits bind between residues 365-375 (Hambly et al., 1986).

The actin monomer appears to have 4 main binding sites for the polymerization of the right handed double helix that

constitutes F-actin (Hambly et al., 1986). These have been assigned by crosslinking of specific amino acids or by specific antibodies that inhibit the polymerization process. These appear to be located to specific amino acids within residues 40-69, 87-113, 168-226 and 283-291. These binding regions have also been confirmed by proteolytic studies which cleave at residues that are exposed to the surface of the monomer (Mornet and Ue, 1984).

The comparison of the actin binding domains to the nucleotide sequence of the sea star actin genes is depicted in Figure 8. Included with other actin binding proteins are tropomyosin and DNase I. These proteins although not studied in sea star systems, are associated with actin in all organisms that these proteins have been found. The variable amino acids are taken from the comparison of 30 actin genes that include all known isoforms (Hambly et al., 1986). The actin binding regions are from the accumulated information from two reviews (Pollard, 1986; Hambly et al., 1986). The major nucleotide sequence differences between the sea star Cy and M genes correlate well with regions that do not specify for protein binding domains. The domains which do not bind proteins are also highly variable in amino acid sequence. Protein binding domains have blocks of nucleotide differences reduced to, or below, the 10.5% average seen across the entire coding region, e.g. blocks 1, 4, 8, 9 and 15. These domains also have less variability in their primary amino acid structure than the non-

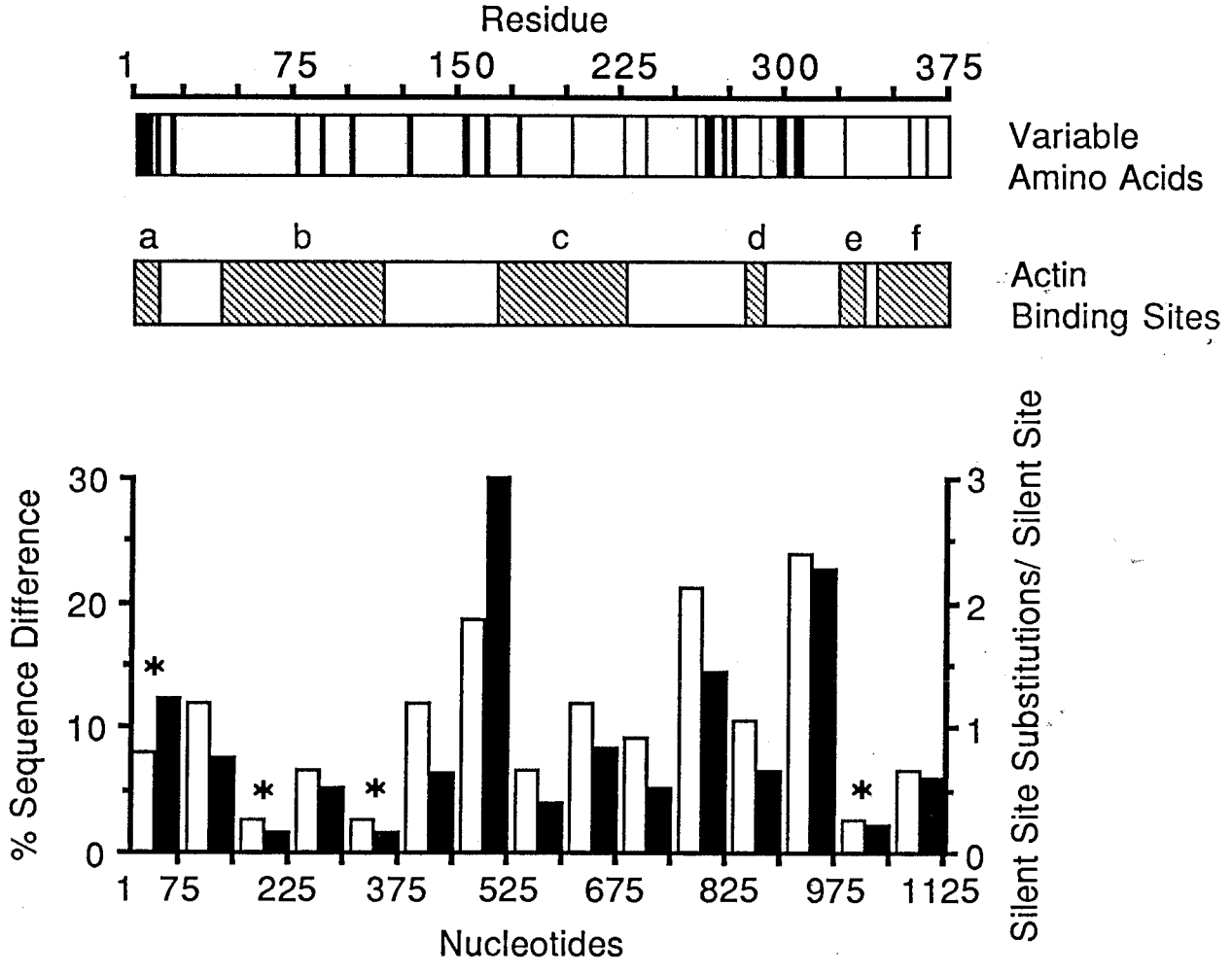
protein binding regions and represent functional constraints on the actin monomer. The only exception to this is the amino acid region covering the first 25 amino acids. The amino terminal end of the peptide is highly variable yet is known to bind to both myosin heavy chain and to depactin. This region is also used to isotype various actin proteins (Vandekerckhove and Weber, 1984).

Blocks of DNA sequence that have significantly reduced 4-fold degenerate silent substitution ratios (*) are all regions that have protein binding properties. Of these blocks, the N-terminal amino acid region (block 1) has a high silent substitution ratio. All substitutions are of the 2-fold degenerate category. The 16 available 4-fold degenerate sites are constrained from substitution between the Cy and M genes in this region (Figure 5). By definition 4-fold degenerate sites are under the least constraint to change (Li et al., 1985). Other factors such as RNA secondary structure and codon usage may be responsible for the increased DNA homogeneity. These will be discussed below.

Regulation of Actin Expression

Actin genes are differentially expressed throughout development. The sea star Pisaster ochraceus is no exception. The large maternal RNA pool for actin is transcribed from a single gene, Cy actin. Using 3' UTR specific DNA probes for the sea star actin genes, Cy actin is revealed as the sole gene

Figure 8.: Comparison of Cy and M genomic sequence to the amino acid variable regions and to functional binding sites to the actin monomer. The variable amino acid regions are from 30 actin primary amino acid structure including all known actin isotypes (Hambly et al., 1986). The actin binding sites are from reviews by Pollard (1986) and Hambly et al.(1986). Actin binding sites: (a) 1-12 Myosin, 1-4, 11 Depactin (b) 40-49, 87-113 Actin 48-82 DNase I 70-86 Tropomyosin (c) 168-226 Actin (d) 283-291 Actin (e) 326-328, 336 Tropomyosin (f) 340-375 Tropomyosin, 360-360 Myosin, 359-373 Depactin, 374-375 Profilin.


Residue
1
75
150
225
300
375
Variable
Amino Acids
a
b
c
d
e
f
Actin
Binding Sites
% Sequence Difference
30
20
10
0
1
75
225
375
525
675
825
975
1125
Nucleotides
Silent Site Substitutions/ Silent Site
3
2
1
0
*
*
*
*

that is transcribed throughout early development. The expression of maternal RNA is shown to be translationally contolled in fertilized sea star oocytes from Asterias forbesi (Rosenthal et al., 1982; Martindale and Brandhorst, 1984). Upon stimulation with 1-MeAde, the pattern of protein synthesis is qualitatively changed once germinal vesicle breakdown (GVBD) has occurred. During oocyte maturation the protein products translated by polysomal and non-polysomal RNA change once GVBD has occurred (Rosenthal et al., 1982). Fertilization of mature oocytes does not change the pattern of protein synthesis as analysed on 2-D electrophoretic gels. Therefore there is a selective recruitment of available maternal RNA onto polysomes. This recruitment onto polysomes is unaltered by fertilization nor is it dependent upon the germinal vesicle (Martindale and Brandhorst, 1984). In sea urchins there is also selective translation of available messages during oocyte maturation (Brandhorst, 1980). The RNA translated in the unfertilized egg is unstable. The increased protein synthesis in the fertilized egg is attributable to a recruitment of transcripts that were stable and untranslatable in the mature oocyte.

In Pisaster ochraceus the number of actin transcripts decreases from 2.9 - 4.3 x 10^5 in the immature oocyte to 1.2 x 10^5 in the early gastrula. By late gastrula the abundance of transcripts increases to 2.2 x 10^5 (Kovesdi and Smith, 1985b). If transcription of Cy actin does occur after fertilization it would be undetectable by RNA-driven hybridization, due to the

high abundance of actin transcripts in the egg. Nuclear run-off experiments from staged embryos would give a better representation of the time of embryonic transcription from actin genes. In the sea urchin S. purpuratus newly transcribed actin RNA is not detected until 7-9 hours post-fertilization as detected by nuclear run-off transcriptional assays and by RNA-RNA hybridization with specific actin gene probes (Lee et al., 1986; Hickey et al., 1987).

The untranslated regions

The function of the 5'and 3' untranslated regions in actin genes has to be inferred from the molecular analysis of other genes. The 3' UTR of actin genes within the same species have a lower degree of sequence identity than coding regions between different isotypes (Eldridge et al., 1985). This has allowed these regions to be used as molecular markers for identifying specific genes within the same family (Lee et al., 1984; Kovesdi et al., 1984). There is an increased sequence identity of the 3' UTR between species that encode for the same actin isotype (Eldridge et al., 1985; Erba et al., 1988). The untranslated region of eukaryotic genes have been associated with various regulatory processes (Brawerman, 1987). Histone genes have conserved boxes of homology in the 3' UTR that form stem-loop structures. These regions are involved in histone H2A RNA processing (Birchmeir et al., 1984). The stem-loop structures may interact with small nuclear ribonucleoproteins

to cleave the mRNA precursor at the 3' end of the transcript. Since histone transcripts generally do not contain polyadenylated tracts the stem-loop structure maybe an alternative RNA processing mechanism. The 5' UTR of human histone H3 genes are involved with the maintenance of stable mRNA transcripts during the S phase of the cell cycle (Morris et al., 1986). The inhibition of DNA synthesis with hydroxyurea causes the rapid destabilization of chimeric transcripts that contain 20 bases of the 5' leader histone sequence. The exogenous transcripts are stable though when the leader is replaced with a Drosophila hsp70 mRNA sequence (Morris et al., 1986). This suggests that the cell cycle dependence of endogenous histone H3 mRNA stability requires the 5' leader sequence. The coupling of H4 mRNA stability to the cell cycle requires sequences in the 3' UTR (Luscher. et al., 1985).

In yeast the GCN4 gene product regulates the co-ordinate derepression of genes involved in amino acid biosynthesis. During amino acid starvation GCN4 mRNA transcripts accumulate to the same levels as in cells growing under normal conditions, yet translation increases over 60X during starvation (Thireos et al., 1984). The generation of deletions in the transcribed region show that 280 bases in the 5' UTR is necessary for the increased translation of the gene product under amino acid starvation conditions (Thireos et al., 1984). Thus the untranslated regions can regulate mRNA stability, RNA

processing and translation of some genes.

In the case of one actin gene the 3' UTR is responsible for transcriptional regulation (Deponti-Zilli et al., 1988). During myogenesis there is a steady decrease in β-actin mRNA in all vertebrate myoblast cell lines and an increase in muscle actin gene transcription (Buckingham and Minty, 1983). Chimeric actin contructs that are injected into mouse myoblasts are expressed in parallel to the endogenous genes. These constructs are constitutively expressed when deletions are generated in the 3' UTR of the gene. Nuclear run-off experiments reveal that the endogenous β-actin gene is down regulated at the level of transcription. A 40 base pair transcriptional control sequence is located in the 3' UTR. It is conserved at the nucleotide level in all vertebrates that encode for the same cytoskeletal isotype (Deponti-Zilli et al., 1988).

The comparison of the two sea star actin genes shows a 40% nucleotide sequence identity between the 3' UTRs. Within the 3' UTR there are sites that can form putative secondary structures. The functional significance of these sites are yet to be determined. The abundance of Cy transcripts is reduced in tube feet (Kovesdi and Smith, 1985b) and may be regulated at the level of transcription.

Developmental Expression of Actin Genes

The expression of actin genes during development has been

examined in various species (Kovesdi and Smith, 1985b; Lee et al., 1986; Fyrberg et al., 1983). In Xenopus the cytoskeletal actin genes are expressed throughout embryogenesis, while muscle actin transcripts are first visible in the neurula stage (Mohun et al., 1984). The muscle specific genes are spatially regulated such that muscle transcripts are confined to cell lineages that lead to muscle tissue (Wilson et al., 1986). This is also observed in the highly determined ascidian embryo where muscle actin is expressed in blastomeres leading to tail muscle cell lineages (Tomlinson et al., 1987). Cytoskeletal actin genes are expressed in all non-muscle lineages, yet expression decreases in muscle cell lines (Mohun et al., 1984; Minty et al., 1986). In mammals the cytoskeletal isoforms are regulated differentially in non-muscle and muscle cell lines. The ratios of beta:gamma mRNA vary from 1.7:1.0 in testes and kidney to 114:1.0 in liver. In mammalian striated muscle the levels are 12:1 yet this accounts for less than 5% of the actin mRNA in this tissue (Erba et al., 1988). The variation of cytoskeletal isoforms in non-muscle tissue is also detected by immunoblotting tissue proteins with actin isotype specific antibodies (Otey et al., 1987).

In the echinoderm Strongylocentrotus purpuratus six of the eight actin genes are expressed in a cell specific pattern throughout development. The two prevalently expressed actin genes in the egg, SpCy1 and SpCyIIIa contribute a total of 3 X 10^3 transcripts to the maternal RNA pool (Lee et al., 1986).

The SpCy1 and SpCyIIIa genes are spatially expressed uniformly throughout in the 16 cell and 32 cell embryo (Cox et al., 1986). By 18 hours post fertilization the SpCy1 transcripts are confined to presumptive oral ectoderm blastomeres and the SpCyIIIa transcripts are confined to aboral ectodermal precursors. New transcription from the embryonic genome is shown to occur between the 32-64 cell stage of development as shown by RNA-RNA hybridization with specific actin gene probes (Lee et al., 1986; Hickey et al., 1987) or by incorporation of ^{32}P-UTP into isolated sea urchin nuclei (Hickey et al., 1987). Thus the differential expression of these genes results from differences in transcriptional regulation in the early embryo and not from selective degradation of specific mRNAs in the specific cell lineages. The Cy actin gene from Pisaster ochraceus is expressed in all stages of early development. The expression pattern of this gene was determined from RNA isolated from whole embryo preparations and not from specific cell-lineages. The possibility that some cell lineages may not express the Cy actin gene can not be excluded. The spatial pattern of actin gene expression resembles more of the cytoskeletal actin expression as observed in Xenopus embryos than that of S. purpuratus.

The sea star M actin gene transcripts are detected in tube feet, presumably in the myoepithelium that are retractor cells in this tissue (Wood and Cavey, 1981) and also in testes (Boom and Smith, in press). Although this gene is not expressed up

to 70 hr post-fertilization (late gastrula embryo) (Kovesdi and Smith, 1985b), the patterns of gene expression as assessed in other organisms suggest that M actin gene expression would commence in precursor cells determined to be myoepithelium.

A sequence in the Cy actin gene, CCATATTTGG, is located between putative promoter sequences as described in the results. It shows strong identity to the $CC(A/T)_6GG$ sequence found in promoter regions of cytoskeletal actins from humans, chickens, mice and amphibians (Erba et al., 1986; Mohun and Garrett, 1987). This region shares identity with a core sequence of a transcriptional factor binding site in a human c-fos proto-oncogene. Transient transcription of this gene can occur in tissue culture in response to growth factors or mitogens and is coincident with a factor binding to the observed site (Farmer, 1986; Hayes et al., 1987). There is also a transient increase of human cytoskeletal actin genes in response to growth factors (Farmer, 1986). Xenopus type 5 cytoskeletal actin can compete with the c-fos gene for factors isolated from nuclear extracts from Xenopus oocytes or from Hela cells. The $CC(A/T)_6GG$ sequence from the Xenopus actin gene is shown to be protected from DNase I digestion and can interfere with complex formation by methylation with DMS (Mohun et al., 1987). This sequence though, is not involved with tissue specific expression since the Xenopus type 8 actin is expressed only in muscle yet shares 100% identity to the $CC(A/T)_6GG$ sequence found in the type 5 cytoskeletal actin

(Mohun and Garrett, 1987). The competition for DNA binding factors common to both amphibians and humans suggests that this may be a common transcriptional factor in different organisms including sea stars.

Nucleotide sequences upstream to the transcriptional start site have implied regulatory functions for the differential expression of actin genes. In the sea star S. purpuratus, chimeric gene constructs with SpCy1 or SpCyIIIa require over 800 bases of 5' flanking DNA for proper spatial and temporal expression of these actin genes (Katula et al., 1987; Flytzanis et al., 1987). Upstream regions are also shown to dictate proper developmental expression in mouse embryos (Shani et al., 1986). Evolutionary conserved sequences in the flanking DNA are observed in the actin genes from Xenopus to man. In the case of a Xenopus alpha skeletal actin gene, cell-specific expression is not dependent upon 5' flanking DNA but probably involves sequences within the transcriptional unit (Steinbeisser et al., 1988). Temporal and spatial regulatory sequences are yet to be observed in the sea star actin genes. Chimeric constructs of Cy or M actin genes fused to reporter genes may be microinjected into sea star oocytes and their expression pattern followed throughout development, as shown in sea urchins (Flytzanis et al., 1987; Katula et al., 1987). It would be of interest to ascertain if evolutionary conserved sequences are observed in putative regulatory regions in sea star actin genes and if they are of functional significance.

Actin gene evolution

Actin is a highly conserved protein. The replacement substitutions observed usually result in conservative amino acid changes, which probably have little or no effect on the tertiary structure of the actin monomer. A noticeable exception to the canonical amino acid sequence is that of an actin gene from the ciliated protozoan Tetrahymena which differs from mammalian actin proteins in 90 of 375 amino acids (Hirono et al., 1987). In sea star actin genes the silent substitutions within the DNA sequence are non-random and suggests that constraints occur on 4-fold degenerate sites. Constraints on silent sites have also been observed in sea urchin and human actin genes (Erba et al., 1986; Crain et al., 1987). In sea urchin this constraint is attributed to a correction mechanism via a gene conversion between a muscle and cytoplasmic actin genes. Relative DNA homogeneity is confined to a region between amino acid 60 and 120 where the ratio of silent site substitutions is reduced relative to that of the rest of the coding region (Crain et al., 1987). In human actin, increased homogeneity occurs in the carboxyl region between two cytoskeletal actin genes (Erba et al., 1986). Thus correction mechanisms may not be confined to specific regions of actin genes. The maintenance of nucleotide sequence is noted in many multigene families including histones, globins and immunoglobulin genes (Maeda and Smithies, 1986). Within

these families the genes may or may not be physically linked. Actin genes may be clustered (sea urchin, Tetrahymena) or dispersed (Drosophila, humans). It is not known if the actin genes from Pisaster ochraceus show any linkage. The gene organization of other multigene families suggests that the constraints on nucleotide sequence may be independent of linkage.

In the case of the sea star genes, 4-fold degenerate sites are constrained in the 5' and 3' coding regions. The increased homogeneity is independent of protein binding domains. If a correction mechanism is acting upon these regions then it is operating upon discrete regions of the genes. Since the increased DNA homogeneity is seen only in coding regions it is interesting to speculate if a correction mechanism is mediated via an RNA template, as is suggested for 5S genes from Neurospora (Morzcyka- Wroblewska, 1985). In developmentally expressed globin genes, gene conversion covers translated and untranslated regions (Roninson and Ingram, 1982; Slightom et al., 1987). In the case of hominid fetal globin sequences, (T/G)n repeats flank the converted regions and are postulated to be hot spots for non-reciprocal recombination between genes (Slightom et al., 1987). The distribution of discrete blocks of increased DNA homogeneity in sea star actin genes preclude a pattern of simple gene conversion seen in other genes. Explanations for the constrained silent sites may include the codon usage by the genes. Four-fold degenerate sites are

prevalent in the third position of most sense codons. The 3rd positions of both sea star genes are over 40% Cytosine. Codon biases have been used to explain increased DNA homogeneity in three differentially expressed mouse actin genes, where it is also observed to be over 50% G/C in the 3rd position of codons (Alonso et al., 1986). The bias in codon usage is not dependent on tRNA availability since the comparison of sea star histone and actin genes reveal that all but two of the sense codons are used in Pisaster ochraceus (Table 3).

Another plausible explanation for increased DNA homogeneity may be secondary structural requirements of the RNA message. Nucleotide comparisons of eukaryotic genes show that the 5' and 3' untranslated regions have higher sequence conservation than that of intron regions (Miyata et al., 1980). The 5' UTR of both sea star genes display a 67% sequence identity for 50 bases upstream to the coding region. This region may have a functional importance in regulating RNA stability, or translation of actin genes. Constraints on coding nucleotide sequence may be also explained by RNA secondary structure. To test this possibility the entire coding region should be compared to multiple actin genes. Non-random substitutions in paralogous or orthologous comparisons of the coding regions would be suggestive evidence for constraints on the RNA structure.

The intron positions for Pisaster ochraceus actin genes were compared to actin genes from other organisms (Table 4).

There are a total of 16 different intron positions known among actin genes to date. Within the deuterostome phylogeny the intron positions appear to be conserved, accounting for 8 of the known introns positions. In protostomes, as represented by Caenorhabitis elegans and Drosophila melanogaster there is a broad variation of intron numbers and positions within species (Fyrberg et al., 1981). When actin genes from animals are compared to known actin genes from plants there are intron positions common to both kingdoms. The intron at codon 150 of soybean and maize actin genes is also seen in chicken, mouse, human and sea star actin genes. This is suggestive evidence for the possible hypothesis that actin genes from extant organisms evolved from a common primordial actin gene that contained all of the known intron positions (Doolittle, 1978). This has also been recently hypothesized for the triose phosphate isomerase gene (Marchionni and Gilbert, 1986). The recent genomic actin sequence from the myxomycete Physarum polycephalum has 4 introns not previously found in other actin genes. It has an isotype common to Dictyostelium, sea urchin, Drosophila and mammalian cytoskeletal actin genes (Gonzalez-y-Merchand and Cox, 1988). This suggests that an ancient actin gene probably contained a minimum battery of 18 introns which may have been randomly lost.

An intron is retained within the 5' UTR of all vertebrate actin genes and at least two sea urchin actin genes. An intron in this position may have a functional significance. The

Table 4. Intron position of actin genes from Deuterostomes

		INTRON POSITION							
Species	Type a	5'	41	84	121	150	204	267	327
P. ochraceus	Cy	?	X		X		X		
	M	?	X		X	X	X	X	
S. purpuratus	Cy1	X			X		X		
	CyIII	X			X		X		
	M	?	X		X		X	X	
X. laevis	Cy	X	X		X			X	X
X. borealis	Cy	X	X		X			X	X
Chicken	A	X	X	X	X	X	X	X	X
	A	X	X			X	X	X	X
	B-Cy	X	X		X			X	X
	5	X							
Mouse	A	X	X			X	X	X	X
Rat	A	X	X			X	X	X	X
	B-Cy	X	X		X			X	X
Human	G-Cy	X	X		X			X	X
	A	X	X	X	X	X	X	X	X

a actin isotypes are A(Alpha) B(Beta) G(Gamma) 5(Type 5)

actin 5C gene from Drosophila is alternately spliced in the 5' region such that the transcripts contain one of two non-homologous leader sequences (Bond and Davidson, 1986). The temporal expression patterns of both transcripts are the same during development, but the tissue specificity of the transcripts are as yet unknown. Alternative promoters in the 5' UTR introns can lead to different developmental expression as seen for the mouse alpha amylase gene and Drosophila ADH gene (Benyajti et al., 1983). In 3 of 4 actin genes from the nematode C. elegans a 22 base sequence is trans-spliced to a 3' acceptor site upstream to the translational start site (Krause and Hirsh, 1987). Alternative transcripts are also seen in a chicken muscle actin gene, but this may be due to utilization of alternative AATAAA sites in the 3' untranslated regions (Carroll et al., 1986). It is interesting to speculate if alternative splicing may lead to transcripts of differing stability.

Conclusion

The comparison of genomic sequence of two actin genes from the sea star Pisaster ochraceus has shown that the gene structure can be delimited by selective constraints on the gene. The least constrained overall are the intron sequences while the coding regions are highly constrained. The conserved DNA coding sequence can be partially explained by functional constraints on the protein, but the reduction of silent

substitutions in some parts of the genes suggests that constraints may occur at the DNA and/or RNA level. It has been hypothesized that mechanisms other than selection may be involved in the maintenance of some multigene families e.g. gene conversion. If true then the use of synonomous substitution rates for the analysis of divergence times between these genes has been perturbed and is therefore unreliable. It would be of interest to understand the gene structure of other actin genes in Pisaster ochraceus and study the extent of constraints on the DNA sequence.

The sequence differences in the untranslated regions of the two genes are intermediate to that of the coding and intron regions. While the 3' UTRs are gene specific, the 5' UTRs do share a 67% sequence identity for the first 50 nucleotides upstream to the translational start site. This may represent functional constraints for gene expression or RNA stability as assessed in other eukaryotic genes.

Future experiments are needed to understand the complete gene structure of both genes. The M gene needs to be sequenced further into the 5' flanking regions in order to compare both genes for functionally conserved sequences. Also the putative intron that interrupts the 5' UTR and transcriptional start sites to the Cy gene may be resolved by primer extension with a Cy specific probe. The formation of cDNAs specific to this gene may also be used to observe if alternative transcripts are generated from a single gene.

REFERENCES

Akhurst, R.J., Calzone, F.J., Lee, J.J., Britten, R.J. and Davidson, E.H.: Structure and organization of the CyIII actin gene sub family of the sea urchin Strongiocentrotus purpuratus. J. Mol. Biol. 194 (1987) 193-203.

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

Amar, M.F.B., Pays, A., Tebabi, P., Dero, B., Seebeck, T., Steinert, M. and Pays, E.: Structure and transcription of the actin gene of Trypanosoma brucei. Mol. Cell. Biol. 8 (1988) 2166-2176.

Bachmair, A., Finley, D. and Varshavsky, A.: In vivo half-life of a protein is a function of its amino-terminal residue. Science 234 (1986) 179-186.

Banfield, D.K., Boom, J.D.G., Honda, B.M. and Smith, M.J.: H3 histone RNA in eggs and embryos of the sea star Pisaster ochraceus. Biochem. Cell. Biol. (1988, in press).

Benyajti, C., Spoerel, N., Haymerie, H. and Ashburner, M.: The messenger RNA for alcohol dehydrogenase in Drosophila melanogaster differs at its 5'end in different developmental stages. Cell 33 (1983) 125-133.

Bergsma, D. J., Chang, K.S. and Schwartz, R.J.: Novel chicken actin gene: Third cytoplasmic isoform. Mol. Cell. Biol. 5 (1985) 1151-1162.

Birchmeier, C., Schumperli, D., Sconzo, G. and Birnstiel, M.L.: 3' editing of mRNA's: Sequence requirements and involvement of a 60-nucleotide RNA in maturation of histone mRNA precursors. Proc. Nat'l. Acad. Sci. U.S.A. 81 (1984) 1057-1061.

Bond, B.J. and Davidson, N.: The Drosophila melanogaster actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple mRNA species. Mol. Cell. Biol. 6 (1986) 2080-2088.

Boom, J.D.G.: Actin gene transcription during spermatogenesis in the sea star Pisaster ochraceus (Brandt). M.Sc Thesis, Simon Fraser University, British Columbia, Canada 1988.

Boom, J.D.G. and Smith, M.J.: Molecular analyses of gene expression during sea star spermatogenesis. J. Exp. Zoology (in press).

Brandhorst, B.P.: Simultaneous synthesis, translation and storage of mRNA including histone mRNA in sea urchin eggs. Dev. Biol. 79 (1980) 139-148.

Brawerman, G.: Determinants of messenger RNA stability.: Cell 48 (1987) 5-6.

Breathnach, R. and Chambon, P.: Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50 (1981) 349-383.

Buckingham, M.E. and Minty, A.J.: Contractile protein genes. In: Eukaryotic genes: Their structure, activity and regulation. Maclean, Gregory, S.P. and Flavell, R.A. (Eds), Butterworths and Co. Publishers 1983 PP. 365-395.

Carroll, S.L., Bergsma, D.L. and Schwartz, R.J.: Structure and complete nucleotide sequence of the chicken α-smooth muscle (aortic) actin gene. J. Biol. Chem. 19 (1986) 8965-8976.

Carroll, S.L., Bergsma, D.L. and Schwartz, R.J.: A 29-nucleotide segment containing an evolutionarily conserved motif is required in cis for cell-type -restricted repression of the chicken α -smooth muscle actin gene core promoter. Mol. Cell. Biol. 8 (1988) 241-250.

Cool, D., Banfield, D., Honda, B.M. and Smith, M.J.: Histone genes in three sea star species: Cluster arrangement, transcription polarity, and analyses of the flanking regions of H3 and H4 genes. J. Mol. Evol. 27 (1988) 36-44.

Cooper, A.D. and Crain, W.R.: Complete nucleotide sequence of a sea urchin actin gene. Nucleic Acid Res. 13 (1982) 4081-4092.

Cox, K.H., Angerer, L.M., Lee, J.J., Davidson, E.H. and Angerer, R.C.: Cell-lineage specific programs of expression of multiple actin genes during sea urchin embryogenesis. J. Mol. Biol. 188 (1986) 159-172.

Crain, W.R., Boshar, M.F., Cooper, A.D., Durica, D.S., Nagy, A. and Steffen, D.: The sequence of a sea urchin muscle actin gene suggests a gene conversion with a cytoskeletal actin gene. J. Mol. Evol. 25 (1987) 37-45.

Cross, G.S., Wilson, C., Erba, H.P. and Woodland, H.R.: Cytoskeletal actin gene families of Xenopus borealis and Xenopus laevis. J. Mol. Evol. 27 (1988) 17-28.

Deponti-Zilli, L., Seiler-Tuyns, A. and Patterson, B.M.: A 40-base pair sequence in the 3'end of the β-actin gene regulates β -actin mRNA transcription during myogenesis. Proc. Natl. Acad. Sci. USA 85 (1988) 1388-1393.

Doolittle, W.F.: Genes in pieces: Were they ever together? Nature 272 (1978) 581-582.

Eldridge, J., Zehner, Z. and Paterson, B.M.: Nucleotide sequence of the chicken cardiac alpha actin gene: absence of strong homologies in the promoter and 3'-untranslated regions with the skeletal alpha actin sequence.: Gene 36 (1985) 55-63.

Erba, H.P., Gunning, P. and Kedes, L.: Nucleotide sequence of the human γ cytoskeletal actin mRNA: anomalous evolution of vertebrate non-muscle actin genes. Nucleic Acids Res. 14 (1986) 5275-5294.

Erba, H.P., Eddy, R., Shows, T., Kedes, L. and Gunning, P.: Structure, chromosome location, and expression of the human β-actin gene: Differential evolution, location and expression of the cytoskeletal β- and γ-actin genes.: Mol. Cell. Biol. 8 (1988) 1775-1789.

Files, J.G., Carr, S. and Hirsh, D.: Actin gene family of Caenorhabditis elegans. J. Mol. Biol. 164 (1983) 355-375.

Firtel, R.L.: Multigene families encoding actin and tubulin. Cell 24 (1981) 6-7.

Flytzanis, C.N., Britten, R.J. and Davidson, E.H.: Ontogenic activation of a fusion gene introduced into sea urchin eggs. Proc. Nat'l. Acad. Sci. U.S.A. 84 (1987) 151-155.

Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P.: The complete nucleotide sequence of the chicken α-actin gene and its evolutionary relationships to the actin gene family. Nucleic Acids Res. 10 (1982) 3861-3876.

Fyrberg, E.A, Bond, B., Hershey, N., Mixter, K. and Davidson, N.: The actin genes of Drosophila: Protein coding genes are highly conserved but the intron positions are not. Cell 24 (1981) 107-116.

Fyrberg, E.A., Mahaffey, J.W., Bond, B.J. and Davidson, N.: Transcripts of the six Drosophila actin genes accumulate in a stage- and tissue-specific manner. Cell 30 (1983) 115-123.

Gonzalez-y-Merchand, J.A. and Cox, R.A.: Structure and Expression of an actin gene of Physarum polycephalum. J. Mol. Biol. 202 (1988) 161-168.

Gong, Z. and Brandhorst, B.: Autogenous regulation of tubulin synthesis via RNA stability during sea urchin embryogenesis. Development 102 (1988) 31-43.

Hamada, H., Petrino, M. and Kakunaga, T.: Molecular structure and evolutionary origin of human cardiac muscle actin gene. Proc. Nat'l. Acad. Sci. U.S.A. 79 (1982) 5901-5905.

Hambly, B.D., Barden, J.A., Miki, M. and Remedios, C.G.: Structural and functional domains on actin. BioEssays 4 (1986) 124-128.

Hattori, M. and Sakaki, Y.: Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. 152 (1986) 232-238.

Hayes, T.E., Kitchen, A.M. and Cochran, B.H.: Inducible binding of a factor to the c-fos regulatory region. Proc. Nat'l. Acad. Sci. U.S.A. 84 (1987) 1272-1276.

Henikoff, S.: Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods in Enzymology 155 (1987) 156-165.

Hickey, R.J., Boshar, M.F. and Crain, W.R.: Transcription of three actin genes and a repeated sequence in isolated nuclei of sea urchin embryos. Dev. Biol. 124 (1987) 215-227.

Hightower, R.C. and Meagher, R.B.: The molecular evolution of actin. Genetics 114 (1986) 315-332.

Hirono, M., Endoh, H., Okada, N., Numata, O. and Watanabe, Y.: Cloning and sequence of the Tetrahymena actin gene and identification of its gene product. J. Mol. Biol. 194 (1987) 181-192.

Hu, M., Sharp, S.B. and Davidson, N.: The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein - coding region. Mol. Cell. Biol. 6 (1986) 15-25.

Jeffery. W.R.: Polyadenylation of maternal and newly synthesized RNA during starfish oocyte maturation. Dev. Biol. 57 (1977) 98-108.

Jukes, T.H.: Transitions, transversions, and the molecular clock. J. Mol. Evol. 26 (1987) 87-98.

Kanatani, H.: Induction of spawning and oocyte maturation by L-methyl adenine in starfishes. Exp. Cell. Res. 57 (1969) 333-337.

Katula, K.S., Hough-Evans, B., Britten, R.J. and Davidson, E.H.: Ontogenic expression of a CyI actin fusion gene injected into sea urchin eggs. Development 101 (1987) 437-447.

Kimura, M.: A simple method for estimating evolutionary rates

of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (1980) 111-120.

Korn, E.D.: Actin polymerization and its regulation by proteins from non muscle cells. Physiological Reviews 62 (1985) 672-737.

Kovesdi, I., Preugschat, F., Stuerzl, M. and Smith, M.J.: Actin genes from the sea star Pisaster ochraceus. Biochimica et Biophysica Acta 782 (1984) 76-86.

Kovesdi, I. and Smith, M.J.: Actin gene number in the sea star Pisaster ochraceus. Can. J. Biochem. Cell Biol. 63 (1985a) 1145-1151.

Kovesdi, I. and Smith, M.J.: Quantitative assessment of actin transcript number in eggs, embryos and tube feet of the sea star Pisaster ochraceus. Mol. Cell. Biol. 5 (1985b) 3001-3008.

Krause, M. and Hirsh, D.: The actin genes of Caenorhabditis elegans. In: Cell and molecular biology of the cytoskeleton. Shay, J. (Ed), Plenum Press Publishers, New York 1986, pp.151-178.

Krause, M. and Hirsh, D.: A trans-spliced leader sequence on actin mRNAs in C. elegans. Cell 49 (1987) 753-761.

Landel, C.P., Krause, M., Waterston, R.H. and Hirsh, D.: DNA rearrangements of the actin gene cluster in Caenorhabditis elegans accompany reversion of three muscle mutants. J. Mol. Biol. 180 (1984) 497-513.

Lee, J.J., Shott, R.J., Rose, S.J., Thomas, T.L., Britten, R.J. and Davidson, E.H.: Sea urchin actin gene subtypes: Gene number, linkage and evolution. J. Mol. Biol. 172 (1984) 149-162.

Lee, J.J., Calzone, F.J., Britten, R.J., Angerer, R.C. and Davidson, E.H.: Measurement of transcript accumulation from five different genes in Strongylocentrotus purpuratus. J. Mol. Biol. 188 (1986) 173-183.

Li, W-H., Wu, C-I. and Luo, C-C.: A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative liklihood of nucleotide and codon changes. Mol. Biol. Evol. 2 (1985) 150-174.

Li, W-H., Luo, C-C. and Wu, C-I.: Evolution of DNA sequences. In MacIntyre, R.J. (Ed), Molecular Evolutionary Genetics, Plenum Press, New York, NY, 1986, pp.1-94.

Loukas, M. and Kafatos, F.C.: The actin loci in the genus Drosophila: Establishment of chromosomal homologies among

distantly related species by in situ hybridization. Chromosoma 94 (1986) 297-308.

Mabuchi, I.: Effects of muscle proteins on the interaction between actin and an actin-depolymerizing protein from starfish oocytes. J. Biochemistry 92 (1982) 1439-1447.

Mabuchi, I.: An actin-depolymerizing protein (depactin) from starfish oocytes: Properties and interaction with actin. J. Cell Biology 97 (1983) 1612-1621.

Maeda, N. and Smithies, O.: The evolution of multigene families. Ann. Rev. Genet. 20 (1986) 81-108.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1982.

Marchionni, M. and Gilbert, W.: The triosephosphate isomerase gene from maize: Introns antedate the plant-animal divergence. Cell 46 (1986) 133-141.

Martindale, M.Q. and Brandhorst, B.P.: Translational changes induced by 1-methyladenine in anucleate starfish oocytes. Dev. Biol. 101 (1984) 512-515.

Matsudaira, P. and Janmey: Pieces in the actin-severing protein puzzle. Cell 54 (1988) 139-140.

Messing, J. and Vierra, J.: A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. Gene 19 (1982) 269-272.

Minor, J.E, Lee, J.J., Akhurst, R.J., Leahy, P.S., Britten, R.J. and Davidson, E.H.: Sea urchin actin gene linkages determined by genetic segregation. Dev. Biol. 122 (1987) 291-295.

Minty, A.J., Hardemann, E., Gunning, P., Bains, W., Blau, H. and Kedes, L.: Regulation of actin gene expression in mammalian striated muscle. In UCLA symposia on molecular and cellular biology Vol 29, Emerson, C., Nadal-Ginard, B., Fischman, D., Siddiqui, M. (Eds) Alan R. Liss, Inc. 1986 pp. 507-519.

Minty, A. and Kedes, L.: Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: Presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. 6 (1986) 2125-2136.

Miyata, T., Yasunaga, T. and Nishida, T.: Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Nat'l. Acad. Sci. U.S.A. 77 (1980) 7328-7332.

Mohun, T.J., Garrett, N. and Treisman, R.: Xenopus cytoskeletal actin and human c-fos gene promoters share a conserved protein binding site. EMBO J. 5 (1987) 667-673.

Mohun, T.J. and Garrett, N.: An amphibian cytoskeletal-type actin gene is expressed exclusively in muscle tissue. Development 101 (1987) 393-402.

Mohun, T.J., Brennan, S., Dalthan, N., Fairman, S. and Gurdon, J.B.: Cell type-specific activation of actin genes in the early amphibian embryo. Nature 311 (1984) 716-721.

Moldave, K.: Eukaryotic protein synthesis. Ann. Rev. Biochem 54 (1985) 1109-1119.

Mornet, D. and Ue, K.: Proteolysis and structure of skeletal muscle actin. Proc. Natl. Acad. Sci. USA 78 (1981) 4319-4323.

Morris, T., Marashi, F., Weber, L., Hickey, E., Greenspan, D. and Bonner, J.: Involvement of the 5' -leader sequence in coupling the stability of a human H3 histone mRNA with DNA replication. Proc. Nat'l. Acad. Sci. U.S.A. 83 (1986) 981-985.

Morzycka-Wroblewska, E., Selker, E.U., Stevens, J.N. and Metzenberg, R.L.: Concerted evolution of dispersed Neurospora crassa 5S RNA genes: Pattern of sequence conservation between allelic and non-allelic genes. Mol. Cell. Biol. 5 (1985) 46-51.

Mount, S.M.: A catalogue of splice junction sequences. Nucleic Acids Research 10 (1982) 459-472.

Ordahl, C.P. and Cooper, T.A.: Strong homology in promoter and 3' untranslated regions of chick and rat α actin genes. Nature 303 (1983) 348-349.

Otey, C.A., Kalnoski, M.H. and Bulinnski, J.C.: Identification and quantification of actin isoforms in vertebrate cells and tissues. J. Cell. Biochem. 34 (1987) 113-124.

Otto, J.J. and Schroeder, T.E.: Assembly -disassembly of actin bundles in starfish oocytes: An analysis of actin-associated proteins in the isolated cortex. Dev. Biol. 101 (1984) 263-273.

Pachter, J.S., Yen, T.J. and Cleveland, D.W.: Autoregulation of tubulin expression is achieved through specific degradation of polysomal tubulin mRNAs. Cell 51 (1987) 283-292.

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J.: The evolution of genes - the chicken preproinsulin gene. Cell 20 (1980) 555-566.

Pollard, T.D.: Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. Ann. Rev. Biochem. 55 (1986) 987-1035.

Raff, R. and Kaufman, T.C.: Embryos, genes and evolution: The devolopmental-genetic basis of evolutionary change. Macmillan Publishers, New York, 1983.

Raff, R.A., Anstrom, J.A., Huffman, C.J., Leaf, D.S., Loo, J-H, Showman, R.M. and Wells, D.E.: Origin of gene regulatory mechanism in the evolution of echinoderms. Nature 310 (1984) 312-314.

Rosenthal, E., Brandhorst, B.P. and Ruderman, J.V.: Translationally mediated changes in patterns of protein synthesis during maturation of starfish oocytes. Dev. Biol. 91 (1982) 215-220.

Sanger, F., Nicklen, S. and Coulsen, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463-5467.

Shah, D.M., Hightower, R.C. and Meagher, R.B.: Genes encoding actin in higher plants: Intron positions are highly conserved but the coding sequences are not. J. Mol. Appl. Genet. 2 (1983) 111-126.

Shani, M: Tissue-specific and developmentally regulated expression of chimeric actin-globin gene in transgenic mice. Mol. Cell. Biol. 6 (1986) 2624-2631.

Shortle, D., Haber, J. and Botstein, D.: Lethal disruption of the yeast actin gene by integrative DNA transformation. Science 217 (1982) 371-373.

Slightom, J.L., Theisen, T.W., Koop, B.F. and Goodman, M.: Orangutan fetal globin genes. J. Biol. Chem. 16 (1987) 7472-7483.

Steinbeisser, H., Hofmann, A., Stutz, F. and Trendelenberg, M.F.: Different regulatory elements are required for cell-type and stage specific expression of the Xenopus laevis skeletal muscle actin gene upon injection in X. laevis oocytes and embryos. Nucleic Acids Research 16 (1988) 3223-3238.

Suck, D., Kabsch, W. and Mannherz, H.: Three-dimensional structure of the complex of skeletal muscle actin and pancreatic DNase I at 6Å resolution. Proc. Natl. Acad. Sci. USA 78 (1981) 4319-4323.

Sutoh, K. and Mabuchi, I.: N-terminal and C-terminal segments of actin participate in binding depactin, an actin-

depolymerizing protein from starfish oocytes. Biochemistry 23 (1984) 6757-6761.

Thireos, G., Penn, M.D. and Greer, H.: 5' untranslated sequences are required for the translational control of a yeast regulatory gene. Proc. Nat'l. Acad. Sci. 81 (1984) 5096-5100.

Tomlinson, C.R., Beach,R.L and Jeffery, W.R.: Differential expression of a muscle actin gene in muscle cell lineages of ascidian embryos. Development 101 (1987) 751-765.

Ueyama, H., Hamada, H., Battula, N. and Kakunaga, T.: Structure of human smooth muscle actin gene (aortic type) with a unique intron site. Mol. Cell Biol. 4 (1984) 1073-1078.

Vandekerckhove, J. and Weber, K.: Mammalian cytoplasmic actins are the products of at least two genes and differ in primary structure in at least 25 identified positions from skeletal muscle actins. Proc. Natl. Acad. Sci. USA. 75 (1978) 1106-1110.

Vandekerckhove, J. and Weber, K.: The complete amino acid sequence of actins from bovine aorta, bovine heart, bovine fast skeletal muscle and rabbit slow skeletal muscle. Differentiation 14 (1979) 123-133.

Vandekerckhove, J. and Weber, K.: Chordate muscle actins differ distinctly from invertebrate muscle actins. J. Mol. Biol. 179 (1984) 391-413.

Wilson, C., Cross, G.S. and Woodland, H.R.: Tissue-specific expression of actin genes injected into Xenopus embryos. Cell 47 (1986) 589-599.

Wood, R.L. and Cavey, M.J.: Ultrastructure of the coelomic lining in the podium of the starfish Stylasterias forreri. Cell Tissue Res. 218 (1981) 449-473.