A putative ancestral actin gene present in a thermophilic eukaryote: novel combination of intron positions

Alan G. Wildeman

Department of Molecular Biology and Genetics, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Received October 23, 1987; Revised and Accepted February 24, 1988

ABSTRACT

The gene encoding actin in the thermophilic fungus Thermomyces lanuginosus has been isolated and sequenced. It contains five introns, with three being at positions already known to be intron sites in actin genes from other eukaryotes. These three positions have not been found to occur simultaneously in any other organisms to date, suggesting that the actin gene in this fungus may more closely resemble an ancestral form of this highly conserved eukaryotic gene. The 5' flanking region of the gene contains a TATA-like sequence and two CCAAT motifs in positions almost identical to those in the yeast actin gene. Other features of the gene sequence, and possible adaptations to thermophily, are discussed.

INTRODUCTION

Actin proteins are found in all eukaryotic cells (1), and between even diverse species the amino acid sequence exhibits a high degree of conservation. The most divergent representatives known are those from ciliated protozoa such as Oxytricha and Tetrahymena, but even they retain at least 65% homology to yeast actin (2, 3). The roles of actin are thought to include cell motility, cytoplasmic movements, and cell surface organization (4, 5), in addition to its involvement in muscle cell contraction. Such a highly conserved macromolecule in eukaryotic cells would be expected to be indispensable and gene disruption experiments in yeast, which contains a single actin gene, have shown that loss of actin gene function is indeed lethal (6).

The structure of genes encoding actin is somewhat variable with respect to introns. The yeast actin gene contains a single intron, while those of ciliated protozoa (2, 3) and Dictyostelium contain no introns. Actin genes from multicellular eukaryotes contain at least one, and in most cases 2 to 6 introns (7). The positions of these introns within the coding regions do, however, show some conservation, particularly among more closely related species but also to a lesser degree among highly divergent groups.

The ubiquity of actin in eukaryotes, and its strongly conserved amino acid sequence, have led to the suggestion that primitive actin genes may have contained numerous introns and that introns have been lost at different positions as species diverged. The loss of introns during evolution was first indicated by studies of insulin genes, where it was noted that in the rat an intron loss occurred in one of the two copies of this gene (8).

This paper reports an analysis of the actin gene from a multicellular fungus, the thermophile Thermomyces lanuginosus. The reasons for studying this organism, which can grow at temperatures up to 60°C, were several. First, it was of interest to know if yeast, which contains a single actin gene having one intron, is representative of other fungi, with respect to actin gene arrangement. Many plants and animals contain families of actin genes, and within one species each actin gene may have unique intron-exon arrangements (7). Second, the macromolecules (proteins and nucleic acids) of thermophilic organisms often exhibit greater thermal stability than those from mesophiles (9, 10). Because actin is a vital component of the cell, it might be anticipated that in the thermophile some changes will have occurred during evolution to adapt the molecule to the higher temperatures. Finally, the promoter sequences which regulate the transcription of eukaryotic protein coding genes often exhibit elements that have been conserved throughout evolution - i.e. a TATA-box, an upstream sequence element often containing -CCAAT- or -CCGCC- motifs, and an enhancer (11, 12). Furthermore, promoters from one species will often work in diverse organisms. Yeast promoters appear to be somewhat different from those of other eukaryotes in that the position of the TATA box is variable and promoter elements often contain AT-rich tracts (13, 14). Since actin is so highly conserved, it was of interest to compare the 5' flanking region of the Thermomyces actin gene, to that of yeast, to see if any putative promoter elements are also conserved between the thermophile and mesophile.

Examination of the Thermomyces actin gene reveals that its actin protein is highly homologous to that of yeast (90%), although the pattern of codon utilization is such that the wobble position is much more GC-rich. Furthermore, unlike the yeast actin gene, there are numerous introns (five). Three of the intron positions have not been found to occur simultaneously in any other eukaryote, suggesting that if intron loss has occurred during evolution the actin gene in this fungus may resemble an ancestral form of the gene. Other features of the gene and its flanking regions are discussed in relation to GC content and putative regulatory sequences.

MATERIALS AND METHODS

The actin gene from Thermomyces lanuginosus (ATCC 16455) was isolated from a library (gift of W. Wong, see ref. 15) of a partial EcoRI digest of genomic DNA cloned into a λ Charon 4A phage vector. The probe for plaque hybridizations was a nick-translated fragment of the yeast actin gene (gift of D. Gallwitz), extending from Xho I to Hind III (see ref. 16).

Hybridizations were carried out on nitrocellulose membranes at 30°C in 50% formamide, 5XSSC, 5X Denhardt's, 0.1% SDS, using standard procedures (17). Inserts contained within positive plaques were subcloned into pBR322 using E. coli HB101 (17), and sequenced using the method of Maxam and Gilbert (18).

Nuclease S_1 mapping (19) was used to identify the site of mRNA initiation. Total RNA was extracted using SDS and phenol from mycelia grown at 50°C in liquid medium containing soluble starch (12.5 g/l), yeast extract (2.5 g/l), $MgSO_4$ (0.5 g/l), and K_2HPO_4 (1.0 g/l). Poly A⁺ RNA was purified on oligo dT cellulose (17) and used for the S_1 analysis. The probe was 5'-labeled, using polynucleotide kinase, at a unique NcoI site found at the initiation codon by purifying on a 5% polyacrylamide gel the EcoRI to NcoI fragment containing the 5' flanking region, labeling and separating the strands on a 4% non-denaturing polyacrylamide gel using standard procedures (17, 18). The probe strand complementary to the mRNA was hybridized to the samples at 42°C in 50% formamide, 0.4 M NaCl, 10 mM Pipes, pH 6.4, for 16 hours. The hybrids were then digested with 50, 150 or 500 units of S_1 nuclease (Boehringer Mannheim) for 4 hours at 25°C as previously described (19). Following phenol/ $CHCl_3$ extraction, the protected fragments were analysed on denaturing 6% polyacrylamide gels.

RESULTS

Cloning and Sequencing the Thermomyces Actin Gene

Approximately 10^5 plaques containing the Thermomyces lanuginosus genomic library were screened with a nick-translated fragment of the yeast actin gene. A clone was identified, and subsequently confirmed by sequence analysis (see below), that carried the entire actin gene on two EcoRI fragments. These inserts were simultaneously subcloned into the EcoRI site of pBR 322. A clone, pThac 11, was identified that carried them in the same orientation as was present in the λ genomic clone. A restriction map of this subcloned DNA and the strategy for sequencing are shown in Figure 1. The nucleotide sequence of the gene, with adjacent 5' and 3' flanking regions, is shown in Figure 2.

Figure 1. Restriction map and sequencing strategy for the actin gene of *Thermomyces lanuginosus*. The genomic segment consisting of two EcoRI fragments was present in pThac11 and the restriction sites indicated were used to generate 5' termini for end-labeling and sequence determination (see Methods).

Gene Structure and Amino Acid Sequence

The derived amino acid sequences is shown in Figure 2. The protein contains 375 amino acids and shows 90% homology to yeast actin. Five introns are present in the gene, occurring within codons 3, 13, 32 and 42 and after codon 299. The presence of these introns was deduced in five ways. First, the conserved nature of actin protein sequences led to easy identification of the correct reading frame. Second, every intron identified follows the GT-AG rule (23) for 5' and 3' consensus sequences. Third, three of the introns occur in positions where introns are found in actin genes from other organisms. Fourth, the first intron is 149 nucleotides in length and if it were translated, the reading frame for the remainder of the gene, which is obviously homologous to other actins, would be destroyed. Finally, Northern blot analysis of *Thermomyces* Poly A⁺ RNA, using nick-translated pThac 11 as a probe, reveals an mRNA molecule of the predicted size, almost identical in length to that found in yeast (data not shown).

Although the amino acid sequence of the *Thermomyces* actin is very similar to that of yeast, in its gene there is a strong preference in codon utilization for codons that are more GC-rich (Figure 3). Similarly, the

```

1 GAATTGCTGTTTGTACTCGTAGCTGGCTGTTGGCCATTGAGCCGGCGGCAATGGATCTTGCCACACCGAGAAATGGCTATGCTGACATGCTCT
101 CTACTCCCTACAGTATGCCAAAGGGAAGCATTGTGACGGCAAGGAACACCGGATCATCAGGAGAGAGGGAGTGCCGAGGCCACCCAGCGGCCA
201 ATGCTCAGCTCGCTCACCTAGGCTCTGATGGTCCGGAGGACCGGGAAGAAATGGAATACAGGAGCAAAATATGGTATACGATGCTGGAATT
301 TAATCGGCTAGAGTTGAGTGGTCGCTTTAGTATTAAAGAGTTGGGAAAGCCGGATGTGATAAATTCAGCGCTTGTAGGGAAGAGGAGTGGCCAA
401 GTGGCTGTGTGCCAATCTGCCAACCGCCAGACTAAACCACTCTCGAATCGTTTCTCCGCTTCCCACTTAACCTTCTGATCATCATCTTCTCTGAG
501 TTCTCTCTCTCCCATCTCTCATCGATTCTCATCACCGTTGACCTTCTTCTCGCTGTTTTTTCCTAGTGTGTTCTTTTACTCTTTAATTCGACAC
601 AGTCACCATGGAAGTaaagcttgcctcctctccccctcctctggggtgcaggccccatttctctggacacctcagcgactctgctccccctgg
MetGlu
701 gacgcttcttggaatatacaaccacactcggacaagctaaatcgtatttcaactccgagAGGAAGTTGCTGCCCTGTTATTGACAATGggtatgt
GluGluValAlaAlaLeuValIleAspAsnGly
801 tttccagctcggtgcggccagccatcccaactctatctccccctattcctgctcccaatcaagcttgggagtagggcgagggtgctcgca
901 ttgctcttctcagcttggaatgaatcgaagctgagctggccaccagTTCGGGAATGTGCAAGGCCGCTTCCGCGTGACGATGCTCCCGAGC
SerGlyMetCysLeuAlaGlyPheAlaGlyAspAlaProArgAla
1001 TGTCTTCTgaagctccgacccctctgctccctcgcgacctccaagctcgagatagcagtagcgtactctggaaagcccttctcagctcactcag
ValPhe
1101 ggccttgggtcggtcgctcaattctctacgcgagcaatcactgagagagcagatCGTCCATCGTCGCTGCCCGCATCGGtaattccccctct
ProSerIleValGlyArgProArgHisHisGly
1201 cgtcggttaaacctctctaaatgagcttctaaactcgtatagTATCATGATTGGTATGGGTGAGAGGACTCGTACGTGGCGGATGAGGCCAGTGG
IleMetIleGlyMetGlyGlnLeuAspSerTyrValGlyAspGluAlaGlnSer
1301 AAGCGTGGTCTCGAGCTTGGGTACCGATCGAGCAGGTTGCTGTCACCACTGGGACGACATGGAGAGATCTGGCACCACACCTTCTACAATGAGT
LysArgGlyValLeuSerLeuArgTyrProIleGluHisGlyValValThrAsnTrpAspAspMetGluLeuIleTyrPheHisThrPheTyrAsnGlu
1401 TGGCTGTGCTCCCGAGGAGCACCCGCTCTTGAACGAGGCCCATCAACCCCAAGTCCAACGCTGAGAGAGATGACCCAGATCTGCTCGAGACCTT
LeuArgValAlaProGluGlnHisProValLeuLeuThrGluAlaProIleAsnProLysSerAsnArgGluLysMetThrGlnIleValPheGluThrPhe
1501 CAATGCGCGGCTCTGAGCTTCCATCCCAAGCGCTCTGCTCACTGTACGCTCGGGTCTGACCACTGGTATGCTCGCTGCTCCGTTGACGGTGTGACC
AsnAlaProAlaPheTyrValSerIleGlnAlaValLeuSerLeuTyrAlaSerGlyArgThrThrGlyIleValLeuAspSerGlyAspGlyValThr
1601 CACACGCTCCGATCTACGAGGGTTGCTCTGACCCAGCCATCTCGGCTATTGATATCGCTGGTGGCGACCTGACGGATACCTGATGAGAGATCCTGG
HisThrValProIleTyrGlyGlyPheAlaLeuThrHisAlaIleSerArgIleAspIleAlaGlyArgAspLeuThrAspTyrLeuMetHisIleLeu
1701 CCGAGCGCGCTACAGCTTCAACCACTCGCGAGCGTAAATGTGCGCGACATCAAGGAGAGCTCTGCTACGTTGCTTTCGACTTCGACGAGGAGCT
AlaGluArgGlyTyrSerPheThrThrAlaGluArgGluIleValArgAspIleLysGluLysLeuCysTyrValAlaPheAspPheGluGlnLeuLeu
1801 GCAGACCGCTGCTCAGAGTTCCAGCTCGGAGAGTCTGATGAGTCTCTGATGGTCAGGTATCACCATTGGCAATGAGGCGTCTCCGCTCCGAGGCT
GlnThrAlaAlaGlnSerSerThrLeuGluLysSerTyrGluLeuProAspGlyGlnValIleThrIleGlyAsnGluArgPheArgAlaProGluAla
1901 CTCTTCAACCTCTGCTGCTGGGTACGAGAGCGGTGTATCCACGAGACCACCTTCAACTCGATCATCAAGTGCATGTCGATGTCGATGAGGATCTGT
LeuPheAsnProAlaValLeuGlyHisGluSerGlyGlyIleHisGluThrThrPheAsnSerIleIleLysCysAspValAspValArgLysAspLeu
2001 ACGGCAATATCGTATGtaggtctctcccccttctgccccctgtgtgtgtgtgtggtcactaaacccctcgtagTCTGGTGGTACCACCATGTACCGGG
TyrGlyAsnIleValMet
2101 GTATTGCCACCGTATGCAAGAGGCTTACTGCTCTCGCCCGCTCGTGAAGATTAAAGATCATGCTCTCGGAGCGCAAGTACTCGCTCGGAT
GlyIleAlaAspArgMetGlnLysGluLeuThrAlaLeuAlaProSerSerMetLysIleLysIleAlaProGluArgLysTyrSerValTrpIle
2201 TGGTGGTTCGATTCTGGCTCGCTGCGACCTTCCAGCAGATGTGGGTCTCGAAGCAGGAGTACGATGAGAGCGGCTCTCGATGCTCCACCGCAAGTGC
GlyGlySerIleLeuAlaSerLeuSerThrPheGlnMetTrpValSerLysGlnGluTyrAspGluSerGlyProSerIleValHisArgLysCys
2301 TTCTAAGGTAGGACCGACAAATTATCCAGCTGGTATAGCAAAAGGAATCTGACGTGATTGGGTGTATATAGCAATCAAGAGCTT
Phe---

```

Figure 2. Nucleotide sequence of the actin gene and its 5' and 3' flanking regions from *Thermomyces lanuginosus*, and predicted amino acid sequence. Nucleotide position numbers are indicated at the left of each line, and amino acid positions are numbered below the protein sequence. Amino acids that differ from those present in yeast are underlined. The arrow indicates the site of transcription initiation determined by nuclease S₁ mapping (see Figure 4). Introns are shown in small case letters, and the boxed residues indicate the homology to the lariet branch point (PyXPyTPuAPy) involved in splicing of eukaryotic precursor mRNAs (see text).

First position	Second position								Third position				
	U		C		A		G						
		S.c.	T.l.		S.c.	T.l.		S.c.	T.l.				
U	Phe	2	0	Ser	14	1	Tyr	0	1	Cys	4	0	U
	Phe	12	15	Ser	12	6	Tyr	14	13	Cys	0	4	C
	Leu	2	0	Ser	3	1	Term	1	1	Term	0	0	A
	Leu	19	3	Ser	0	14	Term	0	0	Trp	4	4	G
C	Leu	2	2	Pro	6	4	His	3	2	Arg	5	10	U
	Leu	0	6	Pro	0	6	His	7	9	Arg	0	7	C
	Leu	2	0	Pro	13	0	Gln	14	1	Arg	0	1	A
	Leu	0	13	Pro	0	8	Gln	0	11	Arg	0	1	G
A	Ile	14	9	Thr	13	3	Asn	0	5	Ser	2	1	U
	Ile	16	21	Thr	8	18	Asn	9	5	Ser	0	4	C
	Ile	0	0	Thr	0	0	Lys	6	0	Arg	13	0	A
	Met	16	12	Thr	0	2	Lys	12	18	Arg	0	0	G
G	Val	14	4	Ala	15	13	Asp	11	10	Gly	29	24	U
	Val	12	21	Ala	11	15	Asp	9	10	Gly	0	5	C
	Val	0	0	Ala	0	0	Glu	26	3	Gly	0	1	A
	Val	0	1	Ala	0	1	Glu	1	26	Gly	0	0	G

Figure 3. Comparison of codon utilization patterns in the actin genes from the mesophile yeast and the thermophile *Thermomyces*. Indicated are the number of times that each codon is used in the 375 amino acid sequence of the protein. S.c. refers to the yeast *Saccharomyces cerevisiae*, and T.l. refers to *Thermomyces lanuginosus*. The data for yeast is taken from reference 16.

Thermomyces introns have an elevated GC content as compared to the yeast intron (54% versus 33%).

The 5' Flanking Regions

Nuclease S₁ mapping was used to locate the transcription initiation site on the *Thermomyces* actin gene (Figure 4). The probe was labeled at an NcoI site containing this codon (see Methods), and used to analyse both whole cell and Poly A⁺ RNA fractions. After S₁ digestion, a protected fragment 114 nucleotides in length was found. Since S₁-protected DNA fragments normally migrate on denaturing polyacrylamide gels as if they are one nucleotide longer than those fragments derived from chemical cleavage during sequencing (19), the cap site was concluded to be at a thymidylic acid residue 113 nucleotides from the labeled end of the probe, or 111 from the AUG codon.

A search for putative regulatory sequences upstream of the initiation site revealed two -CCAAT- motifs, centered 83 and 297 base pairs upstream of the start site (Figure 2). There is no consensus "TATA" box, but a -TAAA- sequence is found approximately 60 base pairs upstream, between the proximal

Figure 4. Determination of the site of initiation of transcription on the actin gene from *Thermomyces lanuginosus*. Total cellular RNA (10 µg) or poly A⁺ RNA (2 µg) samples were analysed by nuclease S₁ mapping using a probe labeled at an NcoI site found at the beginning of the translated portion of the actin gene (see Methods). Aliquots were digested with either 50, 150 or 500 units of S₁ enzyme and analysed on a denaturing 6% polyacrylamide gel. The marker lane (M) contains an A+G chemical sequencing reaction (17) of the probe, and the small dots indicate the band which remains after complete digestion. The sequence ladder numbers on the right correspond to those in Figure 2.

-CCAAT- motif and the transcription initiation site. A comparison of the 5' flanking regions of the *Thermomyces* and yeast actin genes shows that both the presence of these TATA-like and CCAAT sequences, and their positions relative to the messenger RNA initiation sites in these two genes, are conserved (Figure 5). Between the TATA box and the cap site these two species share 39% homology, made up largely by two conserved motifs, -TCTC- and -TTTCCT-, centered 20 and 40 base pairs upstream, respectively. Upstream of the TATA box region there is little obvious homology except for the two CCAAT boxes, underlining their putative role as promoter elements.

DISCUSSION

The actin gene from the thermophilic fungus *Thermomyces lanuginosus* has been isolated and sequenced. It encodes a protein of 375 amino acids, highly homologous with actins from other species (e.g. 90% and 78% homology to yeast (16) and corn (20), respectively). A comparison with the yeast actin (Figure 2) reveals that much of the difference is clustered in two regions of the protein, from amino acids 162 to 180 and 263 to 283. The amino terminal one-third of the proteins from these two species are highly conserved, as are those from other actins examined (16). The first six amino acids of actins have been noted to be highly variable among species (21), and the *Thermomyces* protein is no different in this regard. The presence of

Figure 5. Comparison of the 5' flanking region of the actin genes from yeast and *Thermomyces lanuginosus*. The boxed residues indicate TATA box and CCAAT motifs, and asterisks are placed wherever the two sequences are homologous. The arrows indicate the transcription initiation sites. Data for the yeast sequence is taken from reference 30.

three acidic residues following the initiation codon is unlike yeast, and is characteristic of multicellular eukaryotes (22).

The gene sequence encoding actin in *Thermomyces* exhibits several interesting features. First, it is interrupted by five introns, compared to only one in yeast. While most actin genes from multicellular eukaryotes have several introns, the positions of those in *Thermomyces* suggest that this fungus may have retained certain features of a putative ancestral gene, that have not been retained in other organisms examined to date. The ubiquity of actin throughout the eukaryotes predicts that the actin gene arose before eukaryotes diverged in evolution. It has been predicted that the primordial gene contained introns at all positions where they are found now, in diverse organisms, and that intron loss occurred in different patterns (7). This

		Codons at which introns are located																
	5' UT	4	13	18	32	41	63	84	105	121	150	204	267	299	307	327	353	
<i>Saccharomyces cerevisiae</i>	-	●	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Acanthamoeba</i>	-	-	-	-	-	-	-	●	-	-	-	-	-	-	-	-	-	
<i>Caenorhabditis elegans</i>	-	-	-	●	-	-	●	-	-	-	-	-	-	-	-	-	-	
<i>Drosophila melanogaster</i> Dm A-6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	-	-	
<i>Drosophila melanogaster</i> Dm A-4	-	-	●	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Drosophila melanogaster</i> Dm A-2	(-8)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>Drosophila melanogaster</i> Dm A-1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	●	-	-	
Sea urchin gene C	-	-	-	-	-	-	-	-	●	-	●	-	-	-	-	-	-	
Sea urchin gene J	-	-	-	-	-	●	-	-	●	-	●	●	-	-	-	-	-	
Chicken skeletal muscle	(-12)	-	-	-	-	●	-	-	-	●	●	●	-	-	-	●	-	
Rat skeletal muscle	(-12)	-	-	-	-	●	-	-	-	●	●	●	-	-	-	●	-	
Rat B-cytoplasmic	(-6)	-	-	-	-	●	-	-	●	-	-	●	-	-	-	●	-	
Human smooth muscle (aortic)	-	-	-	-	-	●	-	●	●	●	●	●	-	-	-	●	-	
Human cardiac muscle	-	-	-	-	-	●	-	-	-	-	●	●	●	-	-	●	-	
Soybean	-	-	-	●	-	-	-	-	-	-	●	-	-	-	-	-	●	
		↑	↑		↑	↑								↑				
<i>Thermomyces lanuginosus</i>		3	13		32	42								299				
		*	*			*												

Figure 6. Comparison of intron positions in the actin gene of *Thermomyces lanuginosus* with intron positions in actin genes from diverse eukaryotes. The data for the various organisms is obtained from the following references: *Saccharomyces cerevisiae* (30), *Acanthamoeba* (31), *Caenorhabditis elegans* (32), *Drosophila melanogaster* (33, 34) sea urchin (35), chicken skeletal muscle (36), rat skeletal muscle (7), rat β -cytoplasmic (37), human smooth muscle actin gene (aortic type) (38), human cardiac muscle (39), and soybean (40). Asterisks indicate the intron positions which are conserved between *Thermomyces* and other species.

hypothesis is strengthened considerably by the finding that there is an organism, *Thermomyces lanuginosus*, that has an actin gene in which introns have been retained at several positions which are characteristic of more restricted evolutionary groups (Figure 6). The intron at codon 3 is present in yeast, the one at codon 13 is present in one of the *Drosophila* actin genes (DmA-4), and the one at codon 42 is present in a sea urchin gene and all vertebrate actin genes. In vertebrates there are six actin isoforms known (skeletal muscle, cardiac muscle, and two smooth muscle types, and non-muscle β and γ cytoplasmic types), and genes encoding several of them have now been analysed (Figure 6). In addition to the conserved position at codon 41/42, the genes encoding these isoforms all have introns at codons 267 and 327/328, where *Thermomyces* does not. Interestingly, none of the intron positions in

the Thermomyces gene correspond to those observed to date in plants, although one of the plant intron positions, at codon 150, is identical to that seen in vertebrate muscle-type actin genes. In addition to the three Thermomyces intron positions already discussed, the presence of introns at codons 32 and 299 are to date unique to this organism. Consequently, it might be expected that they too represent vestiges of an ancestral gene. As shown in Figure 6, there are now at least sixteen intron positions documented for actin genes. It is unlikely that an ancestral gene having all of these sites exists today, but as more genes are examined there may be examples found like Thermomyces, that bridge some of the evolutionary groups. It is also possible that some of the intron sites adjacent to one another in fact represent only one site. For example, the Thermomyces position at codon 299 should maybe be thought of as being identical to the position at codon 307 found in two of the Drosophila actin genes. Only as more examples are studied will it become clear if such isolated intron sites are conserved between species. Finally, the presence of a somewhat novel actin gene in a thermophilic fungus suggests that species such as this may be useful in studies on the evolution of other eukaryotic macromolecules.

A second interesting feature of the actin gene structure in Thermomyces is the use of GC-rich codons. Most striking is the high GC content (74% versus 43% in yeast) in the third position of codons (Figure 3). Some genes from thermophilic bacteria have also been noted to evolve in this way (24), and recently it has been suggested that at least in bacteria there may be a bias toward AT or GC substitutions during mutation (25). That is, the mutagenic mechanisms may have evolved to favor GC formation. In the Thermomyces actin gene a higher overall GC content, as compared to the mesophile yeast, has also been achieved by events which leave the coding function of the gene essentially unchanged. Preliminary analysis of the ribosomal genes and spacer regions in the ribosomal DNA of Thermomyces support the notion that, as in bacteria, this fungus has a bias to GC mutagenesis mechanisms that serve to elevate the GC content without changing the ability of the gene to function (W. Wong, A. Wildeman, and R. Nazar, manuscript submitted).

Splicing in eukaryotic genes is now known to involve the formation of a lariat structure, where the 5' side of the intron is first cleaved and then joined to an acceptor site within the intron (27). This branch point, which has the consensus sequence PyXPuTPuAPy, is normally 20 to 40 base pairs from the 3' side of the intron (28, 29). All of the Thermomyces introns contain

this consensus sequence (Figure 2), but it lies nearer to the 3' site than in many other genes examined. The last two introns are very short (59 and 58 nucleotides, respectively), and the branch points are only eleven and twelve nucleotides, respectively, from the 3' sides of the introns. The first three introns have the putative branch points 20, 15 and 12 nucleotides from the 3' sides.

The expression of the Thermomyces actin gene is under the control of a 5' region that shares little overall homology with the yeast regulatory region, but retains elements characteristic of eukaryotic promoters. Both the yeast and Thermomyces genes contain TATA-like sequences in identical positions 60 bp upstream from the site of transcription initiation (Figure 5). They also share two CCAAT motifs, in analogous positions near -80 and -290. It is not known what contribution each of these elements makes to control of transcription, and that problem will most likely have to be solved using the yeast system because a gene transfer technology does not exist for Thermomyces. Although some typical promoter elements are conserved between the two genes, the Thermomyces region extending from -496 to +115 does not function as a promoter in yeast (unpublished observation). Thus, not surprisingly, additional regulatory sequences must be involved. Although some limited homologies in addition to TATA and CCAAT exist between the yeast and Thermomyces genes, the promoters appear to be unable to substitute for one another. Studies are currently underway to more precisely locate the yeast actin promoter sequences, and that knowledge may provide some insight into the Thermomyces control region. The most obvious difference between the two promoters is the lack of long AT-rich segments in Thermomyces. These segments are found in a number of yeast genes, and in some cases may themselves act to promote transcription in yeast (14). At present, very little is known about the transcription termination control sequences in fungal actin genes. Within the 3' flanking region analysed there is no -AATAAA- tract, a motif has been shown to be a signal for polyadenylation in eukaryotes (26). Additional sequence further downstream of the gene will have to be determined to see if this putative element is present.

References

1. Korn, E. (1978) Proc. Natl. Acad. Sci., U.S.A. **75**, 588-599.
2. Kaine, B.P. and Spear, B.B. (1982) Nature (London) **295**, 430-432.
3. Cupples, C. and Pearlman, R.E. (1986) Proc. Natl. Acad. Sci., U.S.A. **83**, 5160-5164.
4. Goldman, R., Pollard, T. and Rosenbaum, J., eds. (1976) Cell Motility, Cold Spring Harbor Conferences on Cell Proliferation (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), Vol. 3.

5. Novick, P. and Botstein, D. (1985) *Cell* 40, 405-416.
6. Shortle, D., Haber, J. and Botstein, D. (1982) *Science* 217, 371-373.
7. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) *Nature (London)* 298, 857-859.
8. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) *Cell* 20, 555-566.
9. Brock, T. (1985) *Science* 230, 132-138.
10. Muto, A. and Osawa, S. (1987) *Proc. Natl. Acad. Sci., U.S.A.* 84, 166-169.
11. Manley, J. (1983) *Prog. Nucl. Acids Res. Mol. Biol.* 30, 195-234.
12. McKnight, S. and Tjian, R. (1986) *Cell* 46, 795-805.
13. Chen, W. and Struhl, K. (1985) *EMBO J.* 4, 3273-3280.
14. Struhl, K. (1985) *Proc. Natl. Acad. Sci., U.S.A.* 82, 8419-8423.
15. Nazar, R.N. and Wong, W. (1985) *Proc. Natl. Acad. Sci., U.S.A.* 82, 5608-5611.
16. Gallwitz, D. and Sures, I. (1980) *Proc. Natl. Acad. Sci., U.S.A.* 77, 2546-2550.
17. Maniatis, F., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
19. Berk, A. and Sharp, P. (1977) *Cell* 12, 721-732.
20. Shah, D., Hightower, R. and Meagher, R. (1983) *J. Mol. Applied Genet.* 2, 111-126.
21. Vandekerckhove, J. and Weber, K. (1978) *Nature (London)* 276, 720-721.
22. Vandekerckhove, J. and Weber, K. (1979) *Differentiation* 14, 123-133.
23. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
24. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. and Oshima, T. (1984) *J. Biol. Chem.* 259, 2956-2960.
25. Muto, A. and Osawa, S. (1987) *Proc. Natl. Acad. Sci., U.S.A.* 84, 166-169.
26. Birnstiel, M., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
27. Ruskin, B., Krainer, A., Maniatis, T. and Green, M. (1984) *Cell* 38, 317-331.
28. Keller, E. and Noon, W. (1984) *Proc. Natl. Acad. Sci., U.S.A.*, 81, 7417-7420.
29. Krainer, A., Maniatis, T., Ruskin, B. and Green, M. (1984) *Cell* 36, 993-1005.
30. Gallwitz, D., Perrin, F. and Seidel, R. (1981) *Nucleic Acids Res.* 9, 6339-6350.
31. Nellen, W. and Gallwitz, D. (1982) *J. Mol. Biol.* 159, 1-18.
32. Files, J.G., Carr, S. and Hirsh, D. (1983) *J. Mol. Biol.* 164, 355-375.
33. Davidson, E.H. (1982) in *Genome Evolution*. Eds. G.A. Dover and R.B. Flavell, Academic, New York.
34. Fryberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981) *Cell* 24, 107-115.
35. Cooper, A.D. and Crain, Jr., W.R. (1982) *Nucleic Acids Res.* 10, 4081-4092.
36. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) *Nucleic Acids Res.* 10, 3861-3876.
37. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) *Nucleic Acids Res.* 11, 1759-1771.
38. Ueyama, H., Hamada, H., Battula, N. and Kakunaga, T. (1984) *Mol. Cell. Biol.* 4, 1073-1078.
39. Hamada, H., Petrino, M. and Kakunaga, T. (1982) *Proc. Natl. Acad. Sci., U.S.A.* 79, 5901-5905.
40. Shah, D.M., Hightower, R.C. and Meagher, R.B. (1983) *J. Mol. Appl. Genetics* 2, 111-126.