Introns and protein revolution – An analysis of the exon/intron organisation of actin genes

S. Bagavathi, R. Malathi*

Department of Genetics, Dr. ALM Post Graduate Institute of Basic Medical Sciences, University of Madras, Taramani, Madras - 600113, India

Received 29 May 1996

Abstract A catalogue of intron positions obtained from a large number of actin genes has been compiled with a view to understanding the possible origin of intervening sequences. Actins are ubiquitous proteins conserved in evolution and an analysis of their gene structures from various organisms has revealed that there may be at least 25 intron positions distributed at different positions in the coding regions. A comparison of intron positions from a wide range of organisms from that of yeast to human actins shows that introns could be more ancestral in origin. The conservation in the observed intron patterns within the different tissue types hints at a possible functional significance of introns in present day actin genes.

Key words: Actin gene; Intron; Molecular evolution; Gene structure analysis

1. Introduction

Since the existence of split genes in eukaryotes much speculation has centred on the origin and function of intervening sequences. The debate on the origin of introns has led to 'introns-early' and 'introns-late' theories [1–4]. Evidence favouring both theories has been observed for different gene families [3,5–7]. In the present investigation the possible origin of intervening sequences has been traced for the actin gene family, based on comparative analysis of the gene structures.

Actins are highly ubiquitous [8] and are present as multigene families in protists, plants and animals, but as a single copy gene in yeast and *Tetrahymena* [9]. The multigene family members are differentially expressed and earlier reports have shown a remarkably high sequence conservation for the regulatory regions upstream for higher organisms [10]. There is an extensive conservation of coding sequences among the actin genes in contrast to the diversity observed in their exon/intron organisation. Therefore, a molecular structural analysis of the actin gene family could be very useful not only in understanding the evolutionary process of actin genes but it may also shed light on the evolution of introns themselves, i.e. whether introns are ancestral or have been acquired during the course of evolution.

Actins are involved in carrying out diversified functions in cells, including cell motility in animals and cytoplasmic streaming in plants, and are essential components of contractile apparatus in muscle cells [11,12]. In vertebrates, six actin isoforms have been identified: four muscle types (skeletal, cardiac, aorta-type smooth muscle, and enteric-type smooth muscle actins) and two non-muscle types (cytoplasmic β and γ actins) [13]. The muscle cytoplasmic variants of actin gene types are also observed for other higher eukaryotes like chicken [14–16] and invertebrates like sea star, sea urchin, etc. [17,18]. The different isoforms of human [19–21] and chicken among vertebrates, sea urchin, sea star, *Drosophila* [22], etc. from invertebrates and also gene sequences of actins from plants [23–27] and lower eukaryotes [19,28–30] have been considered and a hypothetical model demonstrating the lineage of introns is proposed. The results support the idea that a presumed ancestral actin gene contained a number of intron positions and that the present-day actins may have evolved through loss of some of these introns, while they could have been retained and conserved in certain actin genes, since they may be advantageous in the regulatory aspects of gene expression.

2. Materials and methods

Actin gene sequence data were stored and analysed on a PC AT 386. The majority of DNA sequence data was obtained through Bioinformatics Centre, IISc., Bangalore, apart from data collected from recent publications. As many as 20 sequences from organisms representing the fungal, plant and animal kingdoms were included in the analysis. A schematic representation showing the distribution of introns was generated using Microsoft Word and Aldus Pagemaker. A phenogram illustrating the evolutionary lineage of intron positions of actin genes from fungal, plant, and animal sources was generated.

3. Results and discussion

To analyse precisely the intron relationships among the various actin genes, we prepared a schematic representation of intron positions for actin genes coding from lower eukaryotes to higher eukaryotes (Fig. 1). The human actin genes studied included all the six isoforms – α and γ smooth muscle, α cardiac and α skeletal muscle and β and γ cytoplasmic actins [19–21]. Interestingly the comparison reveals a strict conservation of introns at positions 43, 270, and 330 amino acids (aa). However, there are other intron locations that are unique to particular muscle types. For instance, the intron at position 86 aa was restricted to the smooth muscle-type actin. An intron at position 123 has been found in all the muscle types except for cardiac and skeletal actins. Similarly, the

*Corresponding author. Fax: (91) (44) 4926709.

Abbreviations: aa, amino acids; m, muscle; cyt, cytoplasmic; Hum.α sm.mus., human α smooth muscle; Hum.γ sm.mus., human γ smooth muscle; Hum.α card., human α cardiac; Hum.α sk., human α skeletal; Hum.β cyt., human β cytoplasmic; Hum.γ cyt., human γ cytoplasmic; Rat sk.mus., rat skeletal muscle; Chick α sm.mus., chick α smooth muscle; Chick α card.mus., chick α cardiac muscle; Chick β cyt., chick cytoplasmic; D.mel., *Drosophila melanogaster*; C.eleg., *Caenorhabditis elegans*; T.sol., *Taenia solium*; A.thaliana., *Arabidopsis thaliana*; V.carteri., *Volvox carteri*; Therm.lan., *Thermomyces lanuginosis*; Asp.nid., *Aspergillus nidulans*; S.cer., *Saccharomyces cerevisae*

PII S0014-5793(96)00769-7

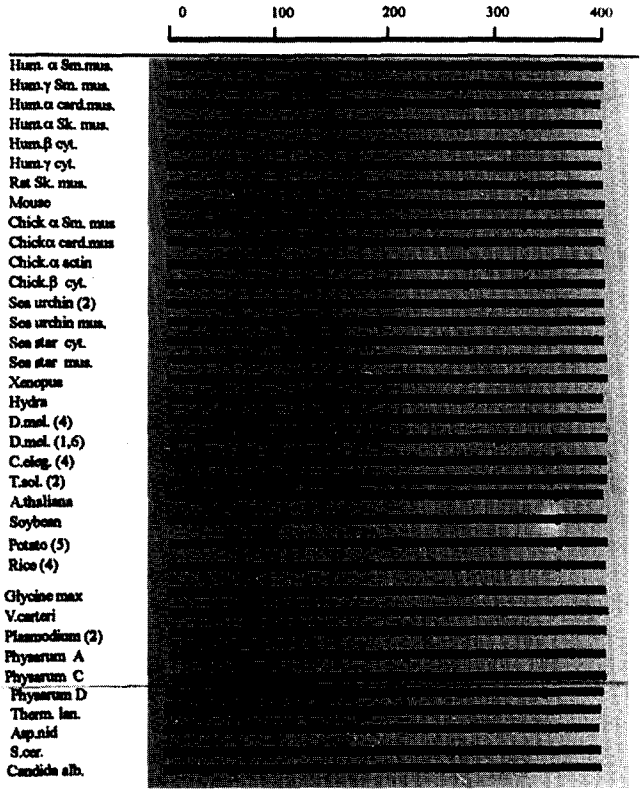

Fig. 1. The distribution of introns within the actin gene family. The presence of an intron is shown by a black bar. The intron locations are described by codon positions.

introns at position 152 and 206 aa are found in all the muscle types other than β and γ cytoplasmic actins. The intron patterns observed for chicken [14–16] and rat [31] actin genes are identical to those of human genes.

The invertebrates included organisms as diverse as hydra [32], *C. elegans* [33], *Drosophila* [22], sea star (m) [34] and sea urchin (m and cyt) actins [18,35]. It was observed that the introns at positions 43, 121, 152, and 327 could also be correlated to those positions already known for higher eukaryotes. For instance, the single intron of hydra at 86 aa coincides with the intron location found only in smooth muscle-type actins. It could be speculated that this feature may contribute to the function that may be common to hydra and the smooth muscle actins. Further investigation showed that the introns at 152, 206, and 267 aa are shared with higher organisms. Such a conservation of intron locations between vertebrates and invertebrates may suggest that they could have evolved from a common ancestral gene, with the intron positions maintained for the above mentioned sites.

All plant actin genes sequenced to date contain introns at three positions [23–27]. Of these, one is shared with vertebrates, one with *C. elegans* and one is plant specific, i.e. at 356 aa. The remarkable coincidence in intron positions for all the plant actins (described in Fig. 1) supports strongly that the plant actin genes are derived from a common ancestor. The coincidence of an intron at 20 aa with the nematodes, the presence of an intron within the 5′ RNA leader in soybean and also an intron at 152 with the higher organisms suggest that these introns should have been present in the precursor gene prior to the divergence of plants and lower eukaryotes. An interesting situation is found in the case of *V. carteri* (alga) actin [36]. This gene is interrupted at nine positions, a

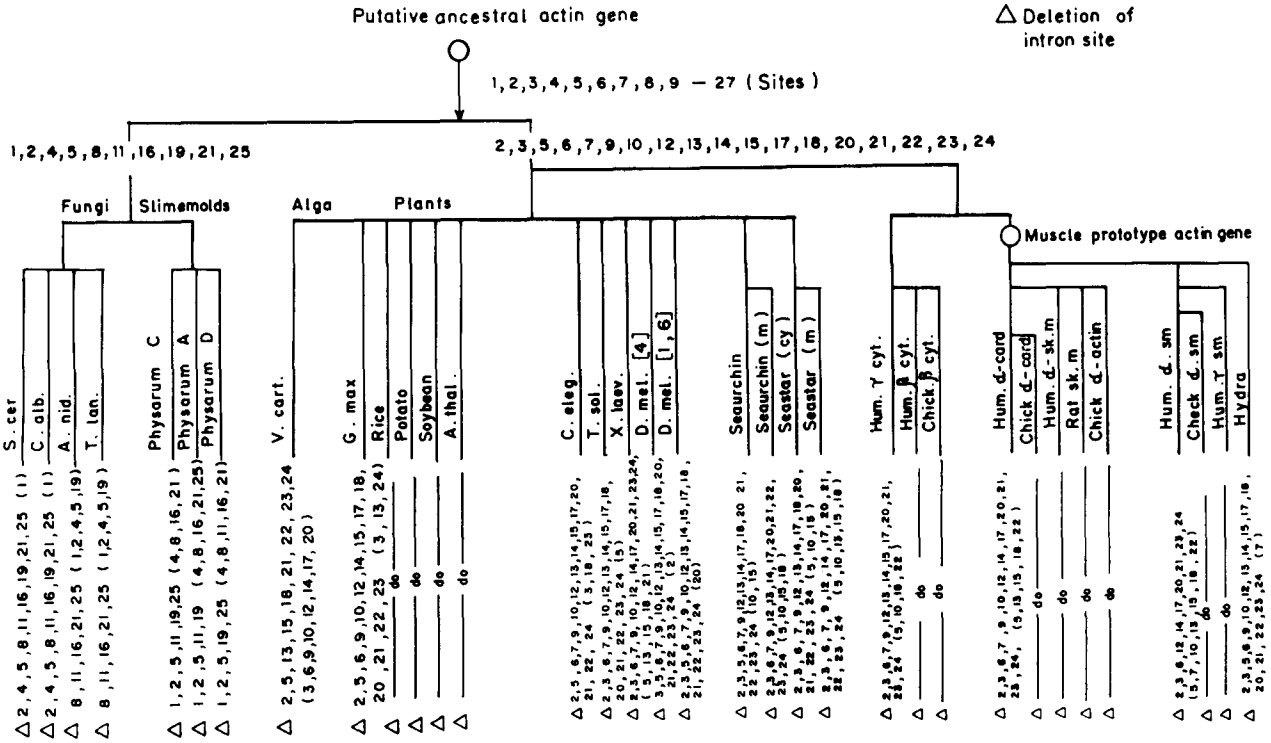

Fig. 2. A phenogram for the hypothetical evolutionary lineage of intron sites of actins. Intron positions are given as 25 different sites for the varying species. Triangles indicate deletions of intron sites.

number higher even than the muscle prototypes, of which three are shared with one or the other species of actin genes. These are the introns at 20 with higher plants and nematodes, 123 with sea star and other higher eukaryotes, and at 309 aa with *Drosophila*. That the intron positions are conserved for varying species such as algae, plants, invertebrates and higher eukaryotes suggests the possibility of an ancestral split gene organisation that could have been retained by the different species.

The lower eukaryotes considered for the analysis were slime molds [30] and fungi [9,28,29]. The two fungal genes (*Thermomyces lanuginoses* and *Aspergillus nidulans*) both shared one intron position at 41 aa with higher eukaryotes, while maintaining an intron at codon 2/3 among themselves and yeast. Similarly, the slime molds share two of their intron positions with fungal and animal kingdoms for interruptions at codons 27 (near identical to 31) and 322 (near identical to 327) respectively. Curiously, the number of introns in lower eukaryotes are high compared to higher organism and if one presumes these organisms to represent more of an ancestral form of the gene, then introns present in them could be considered more ancient in origin.

Based on these ideas, a model can be proposed to demonstrate the lineage of introns in evolution (Fig. 2). Accordingly, a putative ancestral actin gene, for example, had at least 25 introns that were then differentially lost to produce the observed distribution of introns found in modern actin genes. This becomes clear when one analyses Fig. 2, judged by the intron distribution within gene families. The fact that most actin intron positions can be clearly attributed to descent from intron loss explains that introns are very ancient. Thus, the findings support the 'intron-early' hypothesis, where introns have always been an integral part of the gene structure.

References

[1] Darnell, J.E. and Doolittle, W.F. (1986) Proc. Natl. Acad. Sci. USA 83, 1271–1275.
[2] Gilbert, W., Marchionni, M. and McKnight, G. (1986) Cell 46, 151–154.
[3] Cavalier-Smith, T. (1985) Nature 315, 283–284.
[4] Hickey, D.A., Benkel, B.F. and Abukashawa, S.M. (1989) J. Theor. Biol. 137, 41–53.
[5] Stone, E.M., Rothblum, K.N. and Schwartz, R.J. (1985) Nature 313, 498–500.
[6] Marchionni, M. and Gilbert, W. (1986) Cell 46, 133–141.
[7] Lonberg, N. and Gilbert, W. (1985) Cell 40, 81–90.
[8] Baird, W.V. and Meagher, R.B. (1987) EMBO J., 6, 3223–3231.
[9] Ng, R. and Abelson, J. (1980) Proc. Natl. Acad. Sci. USA 77, 3912–3916.
[10] Naka, Y., Nishihara, T., Sasayama, S., Miwa, T., Kamada, S. and Kakunaga, T. (1991) Gene 99, 285–289.
[11] Clarke, M. and Spudich, J.A. (1977) Annu. Rev. Biochem. 46, 797–822.
[12] Lazarides, E. and Revel, J.P. (1979) Sci. Am. 240, 100–113.
[13] Vandekerckhove, J. and Weber, K. (1978) J. Mol. Biol. 126, 783–802.
[14] Carroll, S.L., Bergsma, D.J. and Schwartz, R.J. (1986) J. Biol. Chem. 261, 8965–8976.
[15] Chang, K.S., Rothblum, K.N. and Schwartz, R.J. (1985) Nucleic Acids Res. 13, 1223–1237.
[16] Kost, T.A., Theodorakis, N. and Hughes, S.H. (1983) Nucleic Acids Res. 11, 8287–8301.
[17] Lee, J.J., Calzone, F.J., Britten, R.J., Angerer, R.C. and Davidson, E.H. (1986) J. Mol. Biol. 188, 173–183.
[18] Durica, D.S., Garza, D., Restrepo, M.A. and Hryniewicz, M.M. (1988) J. Mol. Evol. 28, 72–86.
[19] Miwa, T., Manabe, Y., Kurokawa, K., Kamada, S., Kanada, N., Bruns, G., Ueyama, H. and Kakunaga, T. (1991) Mol. Cell. Biol. 11, 3296–3306.
[20] Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 5901–5905.
[21] Nakajima-Iijima, S., Hamada, H., Reddy, P. and Kakunaga, T. (1985) Proc. Natl. Acad. Sci. USA 79, 6133–6137.
[22] Fyrberg, E.A., Bond, B.J., Hershey, N.D., Mixter, K.S. and Davidson, N. (1981) Cell 24, 107–116.
[23] McElroy, D., Rothenberg, M. and Wu, R. (1990) Plant Mol. Biol. 14, 163–171.
[24] Shah, D.M., Hightower, R.C. and Meagher, R.B. (1982) Proc. Natl. Acad. Sci. USA 79, 1022–1026.
[25] Pearson, L. and Meagher, R.B. (1990) Plant Mol. Biol. 14, 513–526.
[26] Drouin, G. and Dover, G.A. (1990) J. Mol. Evol. 31, 132–150.
[27] Nairn, C.J., Winesett, L. and Feri, R.J. (1988) Gene 65, 247–257.
[28] Fidel, S., Doonan, J.H. and Morris, N.R. (1988) Gene 70, 283–293.
[29] Wildeman, A.G. (1988) Nucleic Acids Res. 16, 2553–2564.
[30] Nader, W.F., Isenberg, G. and Sauer, H.W. (1986) Gene 48, 133–144.
[31] Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) Nucleic Acids Res. 11, 1759–1771.
[32] Fisher, D.A. and Bode, H.R. (1989) Gene 84, 55–64.
[33] Files, J.G., Carr, S. and Hirsh, D. (1983) J. Mol. Biol. 164, 355–375.
[34] Kowbel, D.J. and Smith, M.J. (1989) Gene 77, 297–308.
[35] Cooper, A.D. and Crain Jr., W.R. (1982) Nucleic Acids Res. 10, 4081–4092.
[36] Cresnar, B., Mages, W., Muller, K., Salbaum, J.M. and Schmitt, R. (1990) Curr. Genet. 18, 337–346.