# Chapter 5
# Genomic Organization of Connexins

The general genomic structure of connexin genes is rather simple. The 5′-untranslated region (5′-UTR) of connexin mRNA is part of exon 1 that is separated from exon 2 by a long intron. Exon 2 harbours the entire open reading frame and the subsequent 3′-untranslated region (3′-UTR). However, with the accumulation of new experimental data, it became evident that the genomic organization of several connexin genes refutes this simplicity. The identification of different splice isoforms of several connexin genes demonstrated that exon 1 harbouring the 5′-UTR could be spliced in an alternate manner possibly due to alternate promoter usage. However, it should be emphasized that these transcript isoforms vary only in their 5′-untranslated region, leaving the coding region unaltered. Moreover, there are well-known exceptions to the above-proposed genomic structure. For example, Cx36, Cx39, and Cx57 coding regions are interrupted by another intron. The variation in the genomic organization of various connexin genes can be ascertained by the following paragraphs that describe the structure of several individual connexin genes.

## 5.1 Cx32

The genomic organization of Cx32 gene consists of two exons, separated by a 6.1 kb intron. Exon1 constitutes the 5′-UTR, while the entire coding region and 3′-UTR are present in exon 2. However, multiple alternatively spliced transcripts of Cx32 gene have been identified in the mammalian species. For example, two different transcripts were identified in the rat and human, while in cow and mice three different transcripts have been identified. Additionally, two promoter elements have been identified in the rat Cx32 gene, and the usage of these promoters is known to be cell type specific. In hepatocytes Cx32 mRNA is transcribed from the promoter 1 (P1) that is located upstream of exon 1, while in Schwann cells it is transcribed from an alternative promoter 2 (P2) that is located upstream of exon 1B. In humans, Cx32 gene consists of three exons, designated as 1, 1B, and 2. Exons 1 and 1B



M.U. Hussain, *Connexins: The Gap Junction Proteins*, SpringerBriefs
in Biochemistry and Molecular Biology, DOI 10.1007/978-81-322-1919-4_5

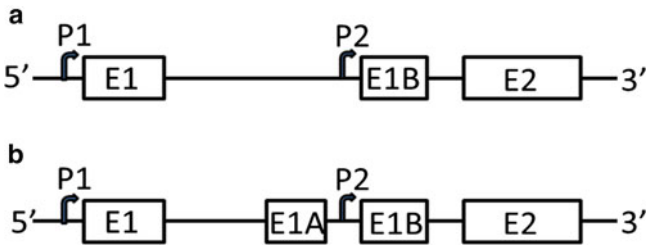


**Fig. 5.1** Genomic organization of Cx32: (**a**) human and rat, (**b**) mouse and bovine. "E" represents exons, "P" represents promoter

are alternatively spliced with exon 2 to produce mRNAs with different 5′-UTRs. Transcription initiation occurs from two different sites, and each initiation site is driven by cell type-specific promoter. In liver, the promoter P1 that is located more than 8 kb upstream of the translation start codon is used, while in nerve cells promoter P2, located 497 bp upstream from the translation start codon, is used for the transcription. Mouse Cx32 gene contains at least four exons designated as 1, 1A, 1B, and 2. Three main transcripts have been identified, which arise by the alternative splicing of exon 1, 1A, and 1B with exon 2 (E1/E2, E1A/E2, E1B/E2). The three transcripts are transcribed from two promoters specifically active in different tissues. Transcript E1/E2 (exon 1/exon 2) is transcribed in the hepatocyte from the promoter P1 that is upstream of exon 1, and transcript E1A/E2 is transcribed in embryonic cells and liver from the same promoter P1. However, transcript E1B/E2 is transcribed in the Schwann cell using promoter P2 upstream of exon 2 (Fig. 5.1). The specificity of P1 promoter is determined by the presence of cell type-specific binding site for the hepatocyte nuclear factor-1 (HNF-1). In addition, the promoter P1 contains putative binding site for various ubiquitous transcription factors, such as Sp1/Sp3, nuclear factor 1 (NF-1), and Yin Yang 1 (YY1). Similarly, the specificity of promoter P2 is driven by the presence of binding sites for nerve-specific transcription factors, like Sox10 and early growth response gene-2 (Egr2/Knox20).

## 5.2 Cx40

Similar to Cx32, various transcripts of Cx40 have been identified in different cell types and in different organisms. Two transcripts are found in the human, three in mouse, and one in the rat. Human Cx40 gene contains at least three exons designated as 1A, 1B, and 2, which are spliced in cell type-specific manner. Transcript E1A/E2 (exon 1A/exon 2) is transcribed in human umbilical cord vein endothelial cells (HUVEC), whereas transcript E1B/E2 (exon 1B/exon 2) is transcribed in human placenta. However, in the human heart, both transcripts have been identified in various regions, like the left atrium, right atrium, left ventricle, and right ventricle.

In mice, the three transcripts arise due to the differential splicing of exon I as exons 1A, 1B, and AS. This results in the formation of three transcript of Cx40 (E1A/E2, E1B/E2, and EAS/E2), having same coding region, however differing in the 5′-UTR. The different transcripts of mouse Cx40 show differential tissue expression. For example, E1A/E2 is ubiquitously expressed, while EAS/E2 is expressed abundantly in the oesophagus. Rat Cx40 is only expressed as single transcript that arises from exon 1 and exon 2 of the Cx40 gene.

## 5.3 Cx43

Cx43 gene was known to possess two exons, that is, exon 1 containing most of the 5′-UTR and exon 2 containing little part of 5′-UTR and the entire coding region along with the 3′-UTR. However, this general genomic organization of Cx43 has been challenged recently by the identification of many more noncoding exons. It has been demonstrated that, at least in mice, there exist six exons (E1A, E1B, E1C, E1D, E1E, and E2), out of which E1A, E1B, E1C, E1D, and E1E form the 5′-UTR, while E2 contains the coding region and the 3′-UTR. The alternative splicing of these exons results in the production of various transcripts, and these different transcripts are transcribed from three different promoters (P1–P3) (Fig. 5.2). In mice, differential promoter usage and alternative splicing result in the formation of different Cx43 transcripts, which include E1A/E2, E1As/E2, E1A/E1E/E2, E1B/E2, E1Bs/E2, E1Bs/E1D/E2, E1C/E2, and E1C/E1D/E2. In the mouse Cx43 gene, the P1 promoter is located upstream of exon 1A, which was earlier regarded as the main promoter of Cx43 gene. Promoter P2 is located within exon 1A, and promoter P3 is located upstream of exon 1C. Similarly in rat, four exons have been identified (E1A, E1B, E1C, and E2), the alternate splicing of which results in the production of six different Cx43 mRNA transcripts, which includes E1A/E2, E1As/E2, $E1A_L$/E2, E1B/E2, E1Cs/E2, and E1C/E2. In humans, the differential promoter usage and alternate splicing of Cx43 are yet to be identified.

Moreover, Cx43 is the only known connexin, which possesses a pseudogene (ψCx43). The pseudogene is located on human chromosome 5, while the main Cx43 gene is located on the human chromosome 6. The ψCx43 lacks an intron between exon 1 and exon 2, and it has been found that the ψCx43 possesses an open reading frame which is almost similar to that of main Cx43, albeit with some amino acid changes. Interestingly, it has been shown that the Cx43 pseudogene is transcribed specifically in several breast cancer cell lines and not in the normal mammary

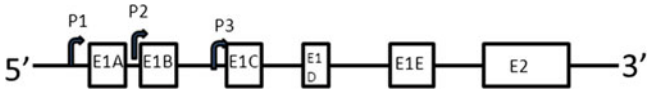


**Fig. 5.2** Genomic organization of Cx43: "E" represents exons, "P" represents promoter. E1A–E1E are noncoding exons, while E2 is the coding exon

epithelial cells. There seems to be inverse relation between the expression of Cx43 transcript and ψCx43 transcript. Moreover, it has been demonstrated that the Cx43 pseudogene can be translated in an in vitro translation system and that the Cx43 pseudogene product inhibits cell growth similar to that of Cx43.

## 5.4 Cx45

Cx45 gene varies a little from the general connexin gene structure. Cx45 gene contains three exons, with 5′-UTR represented by exon 1, exon 2, and part of exon 3. However, the entire coding region and 3′-UTR are present in exon 3. Differential splicing of exon 1 gives rise to various transcripts, differing in 5′-UTR. Accordingly, various alternative spliced transcripts of Cx45 have been identified; each of them shares exon 2 and exon 3, while exon 1A, 1B, and 1C are alternatively spliced with exons 2 and 3. For example, mice Cx45 gene is composed of five exons, designated as exons 1A, 1B, 1C, 2, and 3. Exons 1A, 1B, 1C, and 2 are part of 5′-UTR, while exon 3 contains the remaining 5′-UTR, entire coding sequence, and the 3′-UTR. Additionally, some transcripts have been shown to possess only exon 2 and exon 3. The different transcripts of Cx45 have been shown to arise either due to differential splicing or due to multiple promoter usage. For example, the promoter element, which transcribes the Cx45 mRNA containing exon 1A/2/3, is nearly ubiquitous; the promoter element which transcribes the Cx45 transcript containing exon 1B/2/3 is found in colon, while the transcript containing E1C/2/3 is found in the bladder, lung, skeletal muscle, ovary, and heart.

## 5.5 Cx26

Cx26 gene possesses the basic genomic organizations, having only two exons, with exon 1 as noncoding and exon 2 as the coding and having 3′-UTR. However, the length of exon 1 and intron (between exon 1 and exon 2) varies among different species. For example, in the mice, the length of exon 1 is 234 bp and that of intron is 3.8 kb. Similarly, in the human Cx26 gene, the length of exon 1 is 160 bp, while that of intron is 3.148 kb. The promoter element of Cx26 gene is known to possess binding sites for various transcription factors, like Sp1/Sp3, with a prominent TATA box.

## 5.6 Cx31

The Cx31 gene comprises exon 1A, exon 1B, and exon 2. Exons 1A and 1B constitute the 5′-UTR, and exon 2 possesses the coding region and 3′-UTR. However, there are variations from the general structure of Cx31 gene in different organisms.

For example, mice possess the same genomic structure as described above, while in rat, the transcripts with exon 1A and exon 2 have been identified. Mouse 1B is transcribed from a proximal promoter region 561 bp upstream and serves as a basal promoter in both mouse embryonic stem (ES) cells and a mouse keratinocyte-derived cell line. The promoter element of Cx31 does not possess TATA box and contains many GATA-2/GATA-3 binding sites. Moreover, putative binding sites for NFκB, CCAAT-box, cEBPα/CEBβ, cAMP response element, and multiple E-box/E-box are also known to be present in the promoter region.

## 5.7 Cx30

Gene organization of Cx30 consists of at least six exons, named as exons 1, 2, 3, 4, 5, and 6. Among these, exon 1 to exon 5 constitute the 5′ UTR, while exon 6 contains the coding information and the 3′-UTR. Differential splicing of exon 1 to exon 5 results in the formation of various transcripts of Cx30. In humans, all six exons have been identified with tissue-specific differential expression. For example, in hair follicle keratinocytes, four different transcripts are known to express, and these transcripts share only exon 5 and exon 6. The transcripts of Cx30 are transcribed from a TATA box-containing promoter element, flanking upstream of exon 1. The promoter element contains putative binding sites for Sp1 and early growth gene (Egr) transcription factor.

## 5.8 Cx36, Cx39, and Cx57 Genes

Exceptions to the general genomic organization of connexins are the Cx36, Cx39, and Cx57 genes. The coding region of these connexin genes is interrupted by introns. The coding regions of Cx36, Cx39, and Cx57 genes are located on two (or more) different exons. In rat Cx36, the coding region is interrupted by a 1.14 kb intron at the N-terminal region, separating the first 71 bp from the rest of the coding region. Similarly, the coding region of the mouse Cx39 is also interrupted by a 1.5 kb intron. In mouse retinal Cx57 gene, a part of C-terminal domain is spliced with another exon 3. As a result, 97.6 % (480 amino acids) of the coding region of mouse Cx57 gene is located on exon 2, whereas the residual 2.4 % (12 amino acids) is encoded by the third exon, which is separated by an intron of about 4 kb. The connexin genes whose coding region is interrupted by introns, the exons are required to be spliced properly for the proper translation. Otherwise, alternative splicing would lead to a dramatic modification of the connexin coding region.