Comparative analysis of amino acid usage of the Connexin super family in human and mouse

Anusua Paul and Debaprasad Mukherjee*

Bioinformatics group
Purabi Das School of Information Technology
Bengal Engineering and Science University
B. Garden, Shibpur, Howrah – 711103
West Bengal, India.

* Email id – dm@pdsit.becdu.ac.in

Index terms – Connexin, Amino acid usage, Evolutionary complexity.

Abstract

The connexins are a family of integral membrane proteins that oligomerise to form intercellular channels that are clustered at gap junctions. These channels are specialized sites of cell-cell contact that allow the passage of ions, intracellular metabolites and messenger molecules from the cytoplasm of one cell to its opposing neighbors. The amino acid usage of the domains of this family is analyzed utilizing the protein sequences from the Interpro and the UniProt databases and then using a PERL program. The Interpro domains of the "Connexin" family are found to occur in the proteins for a particular length of amino acid. These lengths vary throughout the protein. Nine domains are found to occur in a greater number of proteins in the mouse than in the human and two are found to occur in the same number in the proteins in both human and mouse. But the domain "Gap junction alpha-3 protein (Cx46)" does not occur in any protein of human or mouse. These domains are found to occur repeatedly in the proteins. Further we have seen that the domains are repeating themselves in the same number in each of the proteins they occur. The results also bring to our notice that though they are evolutionary related and they have the common ancestor, they show noticeable difference in their amino acid usage. When all the domains were analyzed together, the general trend of the average amino acid counts in both human and mouse was the non polar amino acid count was the highest followed by polar amino

acid, positively charged amino acid and negatively charged amino acid. But on analyzing these domains individually and for human and mouse separately this trend is not always followed. The domains – “Gap junction alpha-4 protein (Cx47)”, “Gap junction alpha-5 protein (Cx40)”, “Gap junction beta-1 protein (Cx32”, “Gap junction beta-2 protein (Cx26)”, “Gap junction beta-3 protein (Cx31)”, “Gap junction beta-4 protein (Cx31.1)” and “Gap junction beta-5 protein (Cx30.3)” are found to follow this trend but the domains – “Gap junction alpha-9 protein (Cx36)”, “Gap junction alpha-1 protein (Cx43)”, “Gap junction alpha-6 protein (Cx45)”and “Gap junction alpha-8 protein (Cx50)” donot follow this pattern. Further analysis also revealed that the domains “Gap junction alpha-4 protein (Cx47) ”, “Gap junction alpha-5 protein (Cx40) ”, “Gap junction beta-2 protein (Cx26)” and “Gap junction beta-3 protein (Cx31) ” have the maximum amino acid content in both human and mouse and the domains “Gap junction alpha-9 protein (Cx36) ”, “Gap junction alpha-1 protein (Cx43) ”, “Gap junction alpha-6 protein (Cx45) ”, “Gap junction alpha-8 protein (Cx50) ” and “Gap junction alpha-1 protein (Cx43)” have the least amino acid content. This is probably due to the length of the domains. In brief, this work gives a view of the amino acid usage of the Connexin family which leads to the complexity of the cell communication pathway. This analysis reveals the variation and thus the complexity of the connexin superfamily in the context of domain lengths, domain usage and amino acid usage in human and mouse.

Table 1 – To show the location, usage and the length of the domains.

IPR		NO. OF PROTEINS		NO. OF DOMAINS		LENGTH OF DOMAINS	
ID	DESCRIPTION	HUMAN	MOUSE	HUMAN	MOUSE	HUMAN	MOUSE
IPR002260	Gap junction alpha-9 protein (Cx36)	1	2	3	3 \3	[12\13\8]	[12\13\8] x 2
IPR002261	Gap junction alpha-1 protein (Cx43)	2	4	5\5	5\5\5\5	[11\10\11\11\10] x2	[11\10\11\11\10] x 4
IPR002262	Gap junction alpha-3 protein (Cx46)	0	0	0	0	0	0
IPR002263	Gap junction alpha-4 protein (Cx47)	3	4	1\1\1	1\1\1\1	[329] \ [291] \ [274]	[329] \ [329] \ [329] \ [22]
IPR002264	Gap junction alpha-5 protein (Cx40)	1	3	1	1\1\1	[357]	[357] \ [80] \ [357]
IPR002265	Gap junction alpha-6 protein (Cx45)	2	3	4\4	4\4\4	[7\12\11\13] \ [7\12\14\13]	[7\12\11\14] \ [7\12\11\13] x 2
IPR002266	Gap junction alpha-8 protein (Cx50)	1	1	5	5	[43\13\13\11\12]	[13\13\13\11\12]
IPR002267	Gap junction beta-1 protein (Cx32)	2	2	4\4	4\4	[8\10\6\9] x 2	[8\10\6\9] x 2
IPR002268	Gap junction beta-2 protein (Cx26)	2	4	1\1	1\1\1\1	[226] x 2	[226] x 4
IPR002269	Gap junction beta-3 protein (Cx31)	2	3	1\1	1\1\1	[250] \ [266]	[266] x2 \ [31]
IPR002270	Gap junction beta-4 protein (Cx31.1)	1	3	3	3\3\3	[224\12\11]	[233\12\11] x 3
IPR002271	Gap junction beta-5 protein (Cx30.3)	1	3	3	3\3\3	[241\10\13]	[241\10\13] x 3

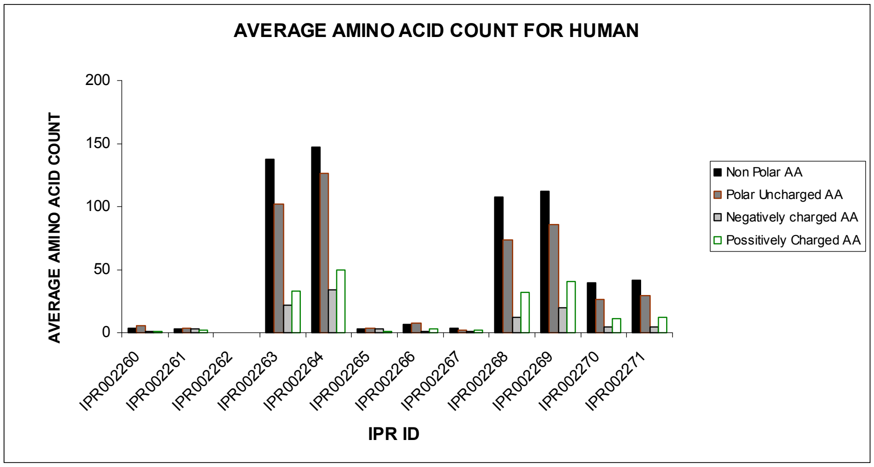

Figure 1: The Graph showing the average amino acid count for the human for each IPR.

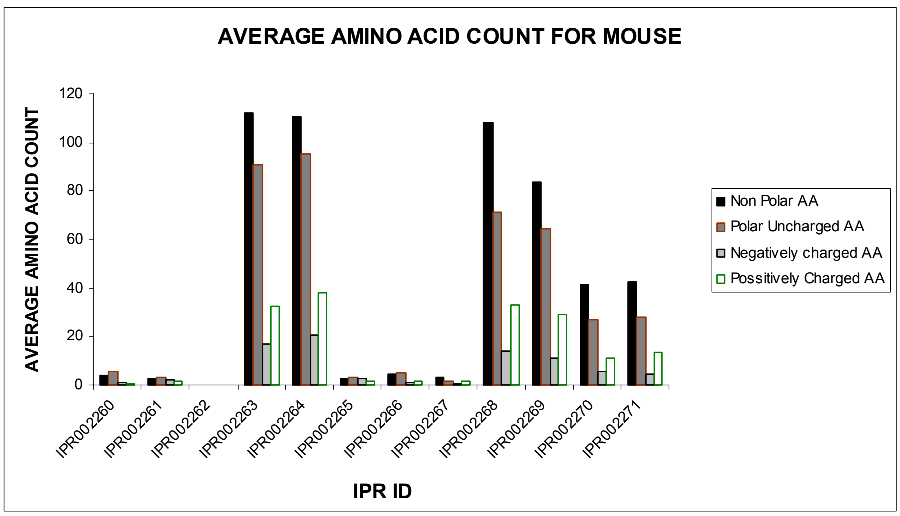

Figure 2: The Graph showing the average amino acid count for the mouse for each IPR.

References

1. Phelan P. , Bacon J.P. , Davies J.A. , Stebbings L.A. , Todman M.G. , Avery L. , Baines R.A. , Barnes T.M. , Ford C. , Hekimi S. , Lee R. , Shaw J.E. , Starich T.A. , Curtin K.D. , Sun Y.A. , Wyman R.J. Innexins: a family of invertebrate gap-junction proteins. Trends Genet. 14 348-349 1998 [PubMed: 9769729]

2. BERNARDI, G., and G. BERNARDI. 1986. Compositional constraints and genome evolution.

J. Mol. Evol. 24:1–11.