Biochimica et Biophysica Acta 1828 (2013) 4–14

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbamem

Review

Evolutionary analyses of gap junction protein families ☆

Federico Abascal, Rafael Zardoya *

Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales CSIC, José Gutiérrez Abascal, 2, 28006 Madrid, Spain

ARTICLE INFO

Article history:
Received 4 November 2011
Received in revised form 28 January 2012
Accepted 7 February 2012
Available online 14 February 2012

Keywords:
Connexin
Innexin
Pannexin
Molecular phylogeny
Maximum likelihood
Bayesian inference

ABSTRACT

Gap junctions are intercellular channels that link the cytoplasm of neighboring cells in animals, enabling straight passage of ions and small molecules. Two different protein families, pannexins and connexins, form these channels. Pannexins are present in all eumetazoans but echinoderms (and are termed innexins in non-chordates) whereas connexins are exclusive of chordates. Despite little sequence similarity, both types of proteins assemble into a common secondary structure with four hydrophobic transmembrane domains linked by one cytoplasmic and two extracellular loops. Although all pannexins and connexins are packed into hexamers forming single channels, only non-chordate pannexins (innexins) and connexins form gap junctions. Here, we revisit and review evolutionary features of pannexin and connexin protein families. For that, we retrieved members of both families from several complete genome projects, and searched for conserved positions in the independent alignments of pannexin and connexin protein families. In addition, the degree of evolutionary conservation was mapped onto the 3D structure of a connexon (i.e. the assembly of six connexins). Finally, we reconstructed independent phylogenies of pannexins and connexins using probabilistic methods of inference. Non-chordate (*Drosophila* and *Caenorhabditis*) pannexins (i.e. innexins) were recovered as sister group of chordate pannexins, which included *Ciona* paralogs and vertebrate pannexins (pannexin-1 and pannexin-3 were recovered as sister groups to the exclusion of pannexin-2). In the reconstructed phylogeny of connexins, subfamilies α and β were recovered as sister groups to the exclusion of subfamily γ, whereas δ and (the newly identified) ζ subfamilies were recovered at the base of the tree. A sixth highly divergent subfamily (ε) was not included in the phylogenetic analyses. Several groups of paralogy were identified within each subfamily. This article is part of a Special Issue entitled: The Communicating junctions, roles and dysfunctions.

Contents

1. Introduction

The acquisition of multicellularity was a major step in animal evolution that allowed continuous increase in morphological complexity, and opened numerous adaptive opportunities [1]. Comparative analyses of recently sequenced complete genomes of several organisms representing lineages linked to the origin and early evolution of animals, including the choanoflagellate *Monosiga* [2], the poriferan *Amphimedon* [3], the placozoan *Trichoplax* [4], and the cnidarians *Nematostella* [5] and *Hydra* [6] reveal that metazoan organismal complexity was achieved mainly through protein family expansions via gene duplication [3] as well as by protein domain shuffling [2].

Among metazoan multicellularity attributes, cell–cell and cell–matrix adhesion is fundamental [3,6]. Cell contacts include adheren, septate, tight, and gap junctions, among others [6]. Here, we will focus on and review evolutionary aspects of gap junctions, which are structures between appositional membranes separated by 2–4 nm

☆ This article is part of a Special Issue entitled: The Communicating junctions, roles and dysfunctions.

* Corresponding author. Tel.: +34 91 411 13 28x1129; fax: +34 91 564 50 78.
E-mail address: rafaz@mncn.csic.es (R. Zardoya).

0005-2736/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamem.2012.02.007

gap that are composed of intercellular channels connecting the cytoplasm of adjoining cells. Gap junctions, allow direct passage of ions and small molecules (<2000 Da) enabling cell-cell communication through electrical and biochemical coupling [7–10]. Gap junctions are constituted by innexins and connexins in non-chordates and chordates, respectively [7,9–13]. Innexins are the non-chordate homologs of the pannexin protein family (in this paper, we will use indistinctly the terms innexin or non-chordate pannexin), which seem to be present in all eumetazoans [7,12,14,15] except echinoderms (that apparently also lack gap junctions) [14,16]. Thus far, the chordate homologs of the pannexin protein family have been shown to form single membrane channels in vivo (at least Pannexin 1 and Pannexin 3) and limited ability to form gap junctions in some in vitro expression assays. The connexin protein family is restricted exclusively to chordates [6,7,12]. Despite their little sequence similarity, pannexins and connexins arrange into a common secondary structure with four hydrophobic transmembrane domains linked by one cytoplasmic and two extracellular loops, the latter containing conserved cysteyl residues that form intramolecular disulfide bonds and are essential for intercellular docking [12,14,17]. The N terminus of connexins is involved in channel gating whereas the rather long C terminus undergoes various post-translational modifications [9,10]. Remarkably, a similar structure is also found in tight junction (occludins and claudins) proteins [11,13,18]. Pannexins and connexins are packed around a central pore in hexamers forming single channels. The docking of innexin or connexin channels from two adjacent cells forms a gap junction, which directly connects the cellular cytoplasms and permits intercellular hydrophilic communication in non-chordates and chordates, respectively [17]. Gap junctions can be homo- or heteromeric (and homo- or heterotypic), and their permeability properties and physiological role may be unique depending on the specific protein subtype composition [9,10,14,17]. On the other hand, some chordate pannexins (1 and 3) act as single membrane channels (sometimes equivocally termed hemichannels in the literature), allowing communication with the extracellular space [8,14,19–21].

Pannexins and connexins are co-expressed in many chordate tissues, and their expression seems to be regulated by N-glycosylation and rapid turnover, respectively [17]. Both types of chordate proteins have distinct and complementary functions that remain to be fully understood [14,17,19,21,22]. Pannexins have been involved in triggering of the inflammosome, cell death, paracrine signaling via ATP release and regulation of intracellular Ca^{2+} leaking in the endoplasmic reticulum [14,17,21], whereas connexins are implicated in cell growth, differentiation, and development regulation [23,24]. On the other hand, many studies have investigated the functional roles of innexins, which include tissue regeneration, development, and electrical synapse formation [7,25,26].

There is controversy on whether pannexin and connexin protein families share a common ancestor (i.e., they belong to the same superfamily) or have an evolutionary independent origin, and thus the structural similarity is due to functional convergence [7,9,11,15,27]. We aligned divergent chordate pannexins and connexins and searched for statistically significant alignments that could provide hints on a putative common evolutionary origin of both protein families. Transmembrane helices of pannexins and connexins were properly aligned each other, several positions were conserved (e.g. extracellular cysteines, although connexins have three pairs of these residues and pannexins only two [21]); and some physic-chemical properties were shared (e.g. the charged nature of extracellular loops or the relative distribution of polar residues within the transmembrane helices). Regarding statistical significance of the observed similarities, the profile–profile comparison tool COMPASS [28] for remote similarity detection yielded a significant *e*-value (1.94e-06). However, HHALIGN [29], which uses hidden Markov models to perform profile–profile alignment with or without predicted secondary structures, fail to render a significant *e*-value. Given that the above results were not compelling, we were unable to discern among competing hypotheses on the evolutionary origin of pannexins and connexins. Homologous proteins can share little sequence similarity and still maintain similar structures and functions [30,31]. If pannexins and connexins are homologous, evolution superseded most positions and eroded footprints of a shared ancestry in the primary sequence [11,15], what seems not unreasonable considering the little similarity of chordate and non-chordate pannexins. Determining the 3D structure of a pannexin channel, and comparing it with the one of a connexon [32] would help clarifying this interesting question.

Diversification of many protein families is normally achieved through gene duplication followed by paralog structural and functional divergence [33]. Therefore, in order to understand the evolution of protein families, it is mandatory the reconstruction of robust phylogenies that allow identifying gene duplication events, and thus distinguishing between groups of orthology and paralogy. Reconstructed phylogenies set frameworks onto which the structural and functional properties of the diverse members of a protein family can be mapped, allowing the inference of underlying evolutionary processes involved in the generation of molecular diversity. Although several phylogenies of pannexins and connexins were reconstructed [11,13,15,34,35], the high rate of new genomes sequenced and the corresponding increasing availability of sequence data from previously underrepresented metazoan lineages prompted us to reconstruct new phylogenies for pannexins and connexins using state-of-the-art probabilistic methods of phylogenetic inference. Several previous attempts to reconstruct the phylogeny of these proteins also included tight junction proteins (i.e., occludins and claudins) in the alignments despite limited sequence similarity among the four types of proteins [11,13,15]. Here, we opted for reconstructing independent phylogenies of pannexins and connexins since we cannot fully ascertain that they have a common evolutionary origin. Moreover, this approach maximizes positional homology in the alignments. The new reconstructed phylogenies are aimed to set a robust evolutionary framework for comparative studies on gap junction proteins.

2. Pannexins

Despite intercellular communication was originally described in crayfish [36], isolation of gap junction proteins of non-chordates, i.e. the innexins, not occurred until recently, and much later than the discovery of chordate connexins [7]. Indeed, it came as a surprise the little sequence similarity of innexins with respect to connexins. From an evolutionary perspective, we can think of innexins as the original set of animal gap junction proteins [9,14,27]. They are absent from poriferans and placozoans, but present in all eumetazoans [6] except echinoderms [14,16]. For etymological reasons, the discovery of homologs of innexins in chordates prompted the renaming of the whole family to pannexins [37] (nomenclature that we are following in this study).

Pannexins are about 300–600 amino acids long [15], and their alignment shows few conserved sites (Fig. 1) due to the extreme sequence divergence between non-chordate (i.e. innexins) and chordate pannexins. Largest conservation within non-chordate pannexins takes place at the extracellular loops (two cysteines and a tryptophane), at the YYQWV motif before the second transmembrane domain, at an invariant proline at the end of the second transmembrane domain, a tryptophan nearby, and a rather conserved tyrosine residue close to the beginning of the third transmembrane domain (Fig. 1; [7,9,11,15]). The rest of the molecule is very variable. Comparison of nematode and fruit fly innexins identified only one site specifically conserved in each group, near the beginning of the third transmembrane helix (Fig. 1). This result suggests that divergence of innexins in invertebrates is likely due to genetic drift rather than triggered by functional diversification. Chordate pannexins not only present conserved extracellular loops, but also show remarkable conservation at the intracellular loop and the N- and C-terminal regions (Fig. 1). Many residues are found to be specific of each of the three