Seminars in Cell & Developmental Biology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Cell & Developmental Biology

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

Structural analysis of key gap junction domains—Lessons from genome data and disease-linked mutants

Donglin Bai*

Department of Physiology and Pharmacology, The University of Western Ontario, London, ON, Canada N6A 5C1

ARTICLE INFO

Article history:
Received 26 November 2015
Accepted 26 November 2015
Available online xxx

Keywords:
Gap junction channel
Sequence logo
Structure function relationship

ABSTRACT

A gap junction (GJ) channel is formed by docking of two GJ hemichannels and each of these hemichannels is a hexamer of connexins. All connexin genes have been identified in human, mouse, and rat genomes and their homologous genes in many other vertebrates are available in public databases. The protein sequences of these connexins align well with high sequence identity in the same connexin across different species. Domains in closely related connexins and several residues in all known connexins are also well-conserved. These conserved residues form signatures (also known as sequence logos) in these domains and are likely to play important biological functions. In this review, the sequence logos of individual connexins, groups of connexins with common ancestors, and all connexins are analyzed to visualize natural evolutionary variations and the hot spots for human disease-linked mutations. Several gap junction domains are homologous, likely forming similar structures essential for their function. The availability of a high resolution Cx26 GJ structure and the subsequently-derived homology structure models for other connexin GJ channels elevated our understanding of sequence logos at the three-dimensional GJ structure level, thus facilitating the understanding of how disease-linked connexin mutants might impair GJ structure and function. This knowledge will enable the design of complementary variants to rescue disease-linked mutants.

Contents

1. Introduction

Gap junctions (GJs) are clusters of membrane channels between neighboring cells mediating direct intercellular communication in many physiological processes [1,2]. GJs are composed of connexin proteins. A total of twenty-one connexin genes are identified in the human genome (20 in the mouse genome), which can be

Abbreviations: Cx, connexin; E1, extracellular domain 1; E2, extracellular domain 2; GJ, gap junction; HB, hydrogen bond; M1-4, transmembrane domains 1-4; NT, amino terminal domain; CT, carboxyl terminal domain.

* Tel.: +1 519 850 2569.
E-mail address: donglin.bai@schulich.uwo.ca

http://dx.doi.org/10.1016/j.semcdb.2015.11.015
1084-9521/

Table 1
Nomenclature of connexins and their gene symbols.

Grouping	Human			Mouse		
	Protein name	Gene symbol	Previous gene symbol	Protein name	Gene symbol	Previous gene symbol
α	Cx43	*GJA1*		Cx43	*Gja1*	
	Cx46	*GJA3*		Cx46	*Gja3*	
	Cx37	*GJA4*		Cx37	*Gja4*	
	Cx40	*GJA5*		Cx40	*Gja5*	
				Cx33	*Gja6*	
	Cx50	*GJA8*		Cx50	*Gja8*	
	Cx59	*GJA9*	*GJA10*			
	Cx62	*GJA10*		Cx57	*Gja10*	
β	Cx32	*GJB1*		Cx32	*Gjb1*	
	Cx26	*GJB2*		Cx26	*Gjb2*	
	Cx31	*GJB3*		Cx31	*Gjb3*	
	Cx30.3	*GJB4*		Cx30.3	*Gjb4*	
	Cx31.1	*GJB5*		Cx31.1	*Gjb5*	
	Cx30	*GJB6*		Cx30	*Gjb6*	
	Cx25	*GJB7*				
γ	Cx45	*GJC1*	*GJA7*	Cx45	*Gjc1*	*Gja7*
	Cx47	*GJC2*	*AJA12*	Cx47	*Gjc2*	*Gja12*
	Cx30.2	*GJC3*	*GJE1*	Cx29	*Gjc3*	*Gje1*
δ	Cx36	*GJD2*	*GJA9*	Cx36	*Gjd2*	*Gja9*
	Cx31.9	*GJD3*	*GJC1*	Cx30.2	*Gjd3*	*Gjc1*
	Cx40.1	*GJD4*		Cx39	*Gjd4*	
ε	Cx23	*GJE1*		Cx23	*Gje1*	

This table was modified from a version provided by Dr. Gerald M. Kidder who led the nomenclature committee presentation at the 2005 International Gap Junction Conference at Whistler, British Columbia, Canada. The official nomenclature of these gene names can be obtained at: http://www.genenames.org/genefamilies/GJ.

classified into α, β, γ, δ, and ε groups based on sequence homology (Table 1) [3]. All connexins are believed to share the same topological structure with four transmembrane domains (M1-4) connected via the first and the second extracellular loop domains (E1 and E2, respectively) and the cytoplasmic loop domain (CL) with the amino terminus (NT) and carboxyl terminus (CT) located intracellularly (Fig. 1A) [3,4]. Six connexins oligomerize to form a gap junction hemichannel (also known as connexon) and two hemichannels dock together at their extracellular domains to form a gap junction channel [4]. Tissue cells commonly express more than one type of connexin, a situation which enables the formation of homomeric and heteromeric gap junctions [1,4]. Different tissues usually express a distinct subset of connexins, and gap junctional communication between different tissue cells often requires heterotypic GJs [5–9]. Numerous recombinant expression studies have revealed that both heteromeric interactions and heterotypic docking of different connexins are specific and only possible between compatible connexins [5,7,10–14]. It is not clear why there should be so many different connexins and what unique sequences of residues in different connexins are important to determine their resultant GJ properties and heteromeric/heterotypic compatibilities. Mutations in the coding regions of many connexin genes are found to associate with inherited human diseases including hearing loss, cataracts, skin diseases, peripheral and central neurodegenerative diseases, cardiac arrhythmias and developmental disorders [15–21]. The vast majority of these disease-linked connexin mutants are missense mutations with only one mutated amino acid residue, providing information concerning the importance of that particular residue. It is important to explore the structural and functional mechanisms of these disease-linked mutants. We begin to learn that some disease-linked mutants impaired intra- or inter-subunit interactions which could be responsible for their GJ channel impairment [6,22–26]. Restoration of lost/impaired interactions would be an interesting strategy to rescue the function of these disease-linked connexin mutants [6]. Similar strategies have been successful in rescuing other types of ion channels [27–29].

Advances in DNA sequencing technology allowed the completion of the human genome project and the genomic sequencing of several other model animal species [30]. Connexin genes from many different vertebrate species are being sequenced at an accelerated pace. More and more connexin sequences are available and categorized according to their homology to the human connexins [31]. In preparing this review, the patterns of available vertebrate connexin gene sequences were analyzed with respect to three key domains to reveal regions that are conserved or variable among different species either within the same evolutionary group of connexins (e.g. all β-group connexins) or between all available connexin sequences. Well-conserved residues, non-conserved residues and outliers can be identified in connexin domains, which can be visualized by sequence logos (or signatures) derived from multiple sequence alignments [32,33]. In contrast to advances in DNA sequencing technology, detailed structural information about gap junction channels is limited and often the resolution is not high enough to resolve the interactions between individual residues [34,35]. A high resolution crystal structure of the human Cx26 homomeric–homotypic gap junction channel was resolved in 2009 and is the only gap junction structure with detailed near atomic resolution [36,37]. Fortunately, sequences of other connexins align well with Cx26 displaying high sequence identity in the structurally resolved domains, arguing that their GJ channels may be structurally similar to that of Cx26. Homology structure modeling of connexins closely related to Cx26 has been developed and is a powerful approach to address the structure-function relationship of GJs. Here some recent advances along this line will be discussed, including investigation of conserved sequence logos and their possible roles in GJ structure and function as well as their non-conserved sequences and outliers. The focus is on NT, E1 and E2 domains as they have been shown to be important for GJ channel oligomerization, gating, permeation and docking [38–41]. Approaches are emphasized rather than providing complete information on every domain.

2. Human connexin alignment and sequence logos

Connexins are homologous proteins and it is well known that two sections of these polypeptides are highly conserved: NT-M1-E1-M2 and M3-E2-M4. In contrast, the CL and CT domains vary

Please cite this article in press as: Bai D. Structural analysis of key gap junction domains—Lessons from genome data and disease-linked mutants. Semin Cell Dev Biol (2015), http://dx.doi.org/10.1016/j.semcdb.2015.11.015