Molecular Immunology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Molecular Immunology

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

C1q: A fresh look upon an old molecule

Nicole M. Thielens[a,*], Francesco Tedesco[b], Suzanne S. Bohlson[c], Christine Gaboriaud[a], Andrea J. Tenner[d,*]

[a] Univ. Grenoble Alpes, CEA, CNRS, IBS, F-38000 Grenoble, France
[b] Istituto di Ricovero e Cura a Carattere Scientifico, Istituto Auxologico Italiano, Milan, Italy
[c] DesMoines University, DesMoines, IA, USA
[d] University of California, Irvine, CA, USA

ARTICLE INFO

Keywords:
Complement C1q
Autoimmunity
Pregnancy
Wound healing
Tumor
Neurobiology

ABSTRACT

Originally discovered as part of C1, the initiation component of the classical complement pathway, it is now appreciated that C1q regulates a variety of cellular processes independent of complement activation. C1q is a complex glycoprotein assembled from 18 polypeptide chains, with a C-terminal globular head region that mediates recognition of diverse molecular structures, and an N-terminal collagen-like tail that mediates immune effector mechanisms. C1q mediates a variety of immunoregulatory functions considered important in the prevention of autoimmunity such as the enhancement of phagocytosis, regulation of cytokine production by antigen presenting cells, and subsequent alteration in T-lymphocyte maturation. Furthermore, recent advances indicate additional roles for C1q in diverse physiologic and pathologic processes including pregnancy, tissue repair, and cancer. Finally, C1q is emerging as a critical component of neuronal network refinement and homeostatic regulation within the central nervous system. This review summarizes the classical functions of C1q and reviews novel discoveries within the field.

1. The complement C1q molecule: from early appearance in evolutionary time to recent recombinant expression

At the time of its discovery at the end of the 19th century complement was described as a heat-labile factor able to cooperate with antibodies to eliminate bacteria. By the mid-1920s complement was recognized as consisting of four components (C1 to C4, with names assigned in order of their discovery), and their isolation from serum was achieved in the early 1940s (Sim et al., 2016). The macromolecular nature of C1 as a complex of three protein entities, namely C1q, the recognition subunit, and its associated proteases C1r and C1s, was defined in the early 1960s, i.e. more than half a century ago (Lepow et al., 1963; Naff et al., 1964). The fact that C1q and the initiating proteins of the classical complement pathway were first discovered is likely related to the high concentrations of these proteins in serum, by comparison with the corresponding recognition proteins (mannan-binding lectin (MBL), ficolins) and associated proteases (MBL-associated serine proteases) of the lectin complement pathway. From this point of view C1q can be considered undoubtedly as an old molecule.

In addition, C1q is also “old” in being present early in evolutionary time. An orthologue of vertebrate C1q proteins has been identified in amphioxus, the modern survivors of the ancient basal chordate lineage (Gao et al., 2014), well before the appearance of jaw vertebrates and adaptive immunity. C1q from *Branchiostoma japonicum* (BjC1q) possibly activates a primitive C1q-mediated complement system, through the binding of Ig-domain-containing amphioxus proteins as well as C4-activating proteases. BjC1q can also bind mammalian C1r and C1s proteases, as well as human IgG, and thus can replace human C1q to activate the human classical pathway. The C1q molecule further evolved with three different A, B and C subunits, possibly during the diversification of cartilaginous fishes, with clear sequence signatures characterizing their functional evolutionary diversification (Gao et al.,

Abbreviations: Aß, amyloid peptide; AD, Alzheimer’s disease; AGE, advanced glycation end-product; C, complement; cC1qR, receptor for the collagen tail of C1q; CCP, complement control protein; CNS, central nervous system; CR, complement receptor; CRT, calreticulin; DC, dendritic cell; DC-SIGN, DC-specific intercellular adhesion molecule (ICAM)-3 grabbing non integrin); DC-SIGNR, DC-SIGN-related protein; ECM, extracellular matrix; EGF, epidermal growth factor; EMI, domain present in proteins of the EMILIN family; ER, endoplasmic reticulum; gC1qR, receptor for the globular “heads” of C1q; HMGB1, high mobility group protein B1; HSP, heat-shock protein; IC, immune complexes; ICAM, intercellular adhesion molecule; LA, LDL receptor-class A; LAIR-1, leukocyte associated immunoglobulin like receptor 1; LHR, long homologous repeat; LPS, lipopolysaccharide; LRP1, low density lipoprotein receptor related protein 1; LTA, lipoteichoic acid; MBL, mannose-binding lectin; PS, phosphatidylserine; RAGE, AGE receptor; SLE, systemic lupus erythematosus; SP, surfactant protein; SR, scavenger receptor

* Corresponding authors.
E-mail addresses: Nicole.thielens@ibs.fr (N.M. Thielens), atenner@uci.edu (A.J. Tenner).

http://dx.doi.org/10.1016/j.molimm.2017.05.025
Received 3 May 2017; Received in revised form 27 May 2017; Accepted 29 May 2017

2014; Tariq et al., 2015; Goshima et al., 2016). Although only one C1q, C1r, C1s protein sequence has been obtained from the liver transcriptome of a hammerhead shark, the presence of three C1q chains in the nurse shark has been reported at the protein level, suggesting that C1q could be expressed in extra-hepatic sites in sharks as in mammals (Goshima et al., 2016). A recent study details the C1q sequences in the Chinese goose and the corresponding evolutionary relationships of the three C1q subunits in duck, chicken, bird and alligator (Tariq et al., 2015).

1.1. C1q ultrastructure and polypeptide chain composition

Electron microscopy studies in the 1970s revealed the typical shape of C1q as resembling a bunch of flowers, with six peripheral globular regions each connected by fibrillar strands to a central bundle of fibers (Shelton et al., 1972; Svehag et al., 1972). In parallel, detailed biochemical studies revealed that C1q is a complex glycoprotein assembled from 18 polypeptide chains of three different types named A, B, and C of 29, 27, and 23 kDa, respectively. Each chain comprises an N-terminal collagen-like sequence and a C-terminal globular gC1q module, with disulfide bridges linking the N-terminal ends of the A and B chains and of two C chains (Reid and Porter, 1976). Each A-B dimer associates with a C chain, resulting in a basic subunit comprised of two disulfide-linked heterotrimeric collagen-like triple helices prolonged by globular domains (Fig. 1). Biochemical characterization of the C1q protein culminated in the determination of the amino acid sequences of the entire A and B chains and that of the collagen-like region of the C chain by N-terminal Edman sequencing (Reid, 1979; Reid et al., 1982). Both cDNA and genomic clones were isolated shortly afterwards for the A and B chains (Reid, 1985), and the isolation of the cDNA sequence of the C chain allowed completion of the entire derived amino acid sequence of the C1q molecule (Sellar et al., 1991). The three genes were found to be aligned 5′-3′ in the same orientation and in the order A-C-B on human chromosome 1p.

1.2. C1q functional domains

The presence of two distinct types of structures in the C1q molecule, *i.e.* the globular heads and the collagen-like fragments, allowed their isolation by limited proteolysis of C1q with collagenase and pepsin, respectively (Brodsky-Doyle et al., 1976; Hughes-Jones and Gardner, 1979). Functional studies showed that the globular regions are mainly responsible for target recognition, including binding to the Fc region of immunoglobulins (Hughes-Jones and Gardner, 1979), but also to bacterial and viral surface proteins, as well as altered self elements (apoptotic cells, amyloid and prion proteins) (Gaboriaud et al., 2011). The collagen-like regions mediate immune effector mechanisms, including complement activation through interaction with the C1r and C1s proteases (Siegel and Schumaker, 1983) and enhancement of phagocytosis through interaction with cell receptors (Bobak et al., 1987).

Isolation of the globular region of C1q allowed determination of its X-ray crystal structure, revealing a compact heterotrimeric assembly with a calcium ion bound at the top (Gaboriaud et al., 2003). Each of the three gC1q modules exhibits distinct surface patterns of charged and hydrophobic residues, which likely endow them with a diversity of specific recognition properties. In addition, the compact structure of the trimer allows ligand binding through residues contributed by two or three chains, thereby broadening the recognition spectrum of C1q (Gaboriaud et al., 2003). This structure confirmed the structural homology between tumor necrosis factor (TNF) and gC1q-containing proteins, previously revealed in the crystal structures of the homotrimeric gC1q domain of mouse adiponectin (Shapiro and Scherer, 1998) and collagen X (Bogin et al., 2002).

1.3. Recombinant C1q and site-directed mutagenesis studies

Detailed investigation of the structure-function relationships of C1q required its production in a recombinant form. The first achievements consisted in bacterial expression of soluble fusion proteins of the individual gC1q domains (A, B and C) with maltose binding protein (MBP), which were purified by affinity chromatography on an amylose resin. Characterization of their interaction properties revealed that these domains have likely evolved as functionally autonomous entities with differential ligand-binding properties (Kishore et al., 2003). Site-directed mutagenesis revealed that the binding sites for a number of C1q target molecules, including IgG, IgM, pentraxins and bacterial components such as *Salmonella* lipopolysaccharides and *Klebsiella* outer membrane protein K36, are closely overlapping (Kojouharova et al., 2003; Kojouharova et al., 2004; Roumenina et al., 2006; Zlatarova et al., 2006; Roumenina et al., 2008), but differ from the binding sites for the viral glycoproteins gp41 of HIV-1 and gp21 of HTLV-1 (Kojouharova et al., 1998; Kishore et al., 2003).

Production of full-length recombinant C1q has been a major technical bottleneck for a long time, owing to the three chain structure, particular inter-chain disulfide pattern and numerous post-translational modifications of the molecule. For example, the collagen-like sequences of C1q contain the repeating Gly-X-Y triplet where X is often a proline residue and Y a hydroxyproline or hydroxylysine residue, the latter being frequently modified with glucosylgalactosyl disaccharide units (Reid, 1979; Pflieger et al., 2010). However, this challenge was achieved in 2013 with the production of the full-length recombinant protein in stably transfected mammalian 293-F cells (Bally et al., 2013). The recombinant C1q molecule is similar to serum-derived C1q, as judged from biochemical analysis and electron microscopy imaging, and retains the ability to associate with the C1r and C1s proteases, to recognize physiological C1q ligands including IgG and pentraxin 3, and to trigger complement activation. The production of recombinant C1q opened the way for deciphering the interaction properties of this protein by site-directed mutagenesis, as illustrated by the identification of the C1s-C1r-C1r-C1s–binding residues in the C1q collagen-like stems (Bally et al., 2013). The availability of recombinant C1q should allow engineering of C1q molecules lacking specific functions (*i.e.* complement activation or binding to a protein partner) which will represent unique tools to study the role of C1q in various cellular contexts.

The latest progress in recombinant C1q production was the generation of a single-polypeptide form of the human C1q globular region (C1q-scGR) (Moreau et al., 2016), based on a strategy previously used

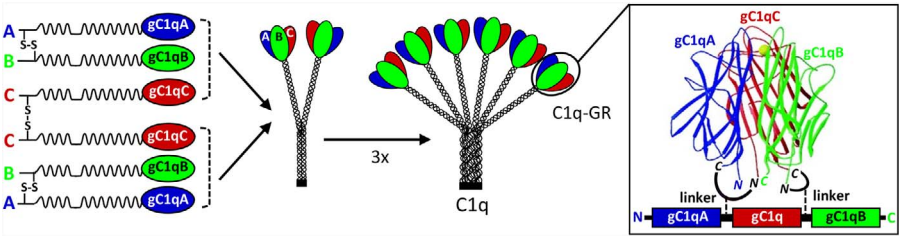

Fig. 1. Schematic representation of the assembly of the C1q molecule and of the recombinant single chain globular region. C1q is assembled from three polypeptide chains (A, B and C) each containing an N-terminal collagen-like sequence and a C-terminal globular gC1q module. A particular inter-chain disulfide pattern results in a basic subunit comprised of two heterotrimeric collagen-like stalks prolonged by globular domains (C1q-GR). Three subunits associate to yield the full-length protein with a typical shape of a bouquet of six flowers. The crystal structure of the trimeric C1q-GR shows that the N- and C- terminal ends of the 3 chains are in close proximity which allowed insertion of short linkers between the gC1qA-gC1qC and gC1qC-gC1qB modules to yield a recombinant single chain construct.