

Chapter 2

Distribution of MACPF/CDC Proteins

Gregor Anderluh, Matic Kisovec, Nada Kraševc
and Robert J. C. Gilbert

Abstract Membrane Attack Complex/Perforin (MACPF) and Cholesterol-Dependent Cytolysins (CDC) form the MACPF/CDC superfamily of important effector proteins widespread in nature. MACPFs and CDCs were discovered separately with no sequence similarity at that stage being apparent between the two protein families such that they were not, until recently, considered evolutionary related. The breakthrough showing they are came with recent structural work that also shed light on the molecular mechanism of action of various MACPF proteins. Similarity in structural properties and conserved functional features indicate that both protein families have the same evolutionary origin. We will describe the distribution of MACPF/CDC proteins in nature and discuss briefly their similarity and functional role in different biological processes.

Keywords Cholesterol-dependent cytolysins • MACPF/CDC protein superfamily • Membrane attack complex • Perforin • Protein evolution

G. Anderluh (✉) · M. Kisovec · N. Kraševc
Laboratory for Molecular Biology and Nanobiotechnology, National Institute of Chemistry,
Hajdrihova 19, 1000 Ljubljana, Slovenia
e-mail: gregor.anderluh@ki.si

G. Anderluh
Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111,
1000 Ljubljana, Slovenia

R. J. C. Gilbert (✉)
Division of Structural Biology, Wellcome Trust Centre for Human Genetics,
University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
e-mail: gilbert@strubi.ox.ac.uk

Abbreviations

APC- β	Apicomplexan perforin-like proteins C-terminal β -pleated domain
ASTN	Astrotactin
BRINP	Bone morphogenic protein/retinoic acid inducible neural-specific protein
CDC	Cholesterol-dependent cytolysins
MAC	Membrane attack complex
MACPF	Membrane attack complex/perforin
Mpg-1	Macrophage-expressed gene 1
PFN	Perforin

General Overview of Structural Similarity of MACPF/ CDC Protein Families

MACPF and CDC proteins are important effectors in the immune system and bacterial pathogenesis, where their observed effects are primarily mediated through membrane interactions and the formation of transmembrane pores. Their activity is, however, not restricted to membrane interactions and it will be important to determine the exact role and molecular mechanism of action for many representatives. MACPF proteins were initially defined by the sequence similarity among proteins of the Membrane Attack Complex (*MAC*) of the complement system and perforin (*PF*) [59, 83, 93]. These proteins have crucial roles in immune defence against viral and bacterial pathogens and in removal of cancer cells. For example, perforin is the only molecule of the human immune system that allows entry of granzyme into the target cells and initiation of apoptosis. The MAC is the terminal stage of the complement system of innate immunity. On the other hand, CDCs are predominately found in Gram-positive bacteria [29, 94], although a recent report indicates that this protein family may be more widespread as they are at least present in Gram-negative bacteria too [41] and diatoms and plants (see below). Because they were discovered separately, in quite separate niches, there was originally no sequence similarity apparent between the proteins of MACPF and CDC families. The realisation that these two families are evolutionary related was the result of intense structural work on MACPF proteins; in particular crystal structures of Plu-MACPF from bacteria *Photorhabdus luminescens* [77], Bth-MACPF from bacteria *Bacteroides thetaiotaomicron* [104], the MACPF domain of complement components C6 [5] and C8 [35, 62, 85] and mouse perforin (PFN) [57] revealed their similarity to the CDC protein family. Both protein families are now collectively referred to as the MACPF/CDC pore forming proteins [30, 78]. Structure-based alignment allows the identification of very limited sequence conservation of glycines present in all family members [30], which in fact

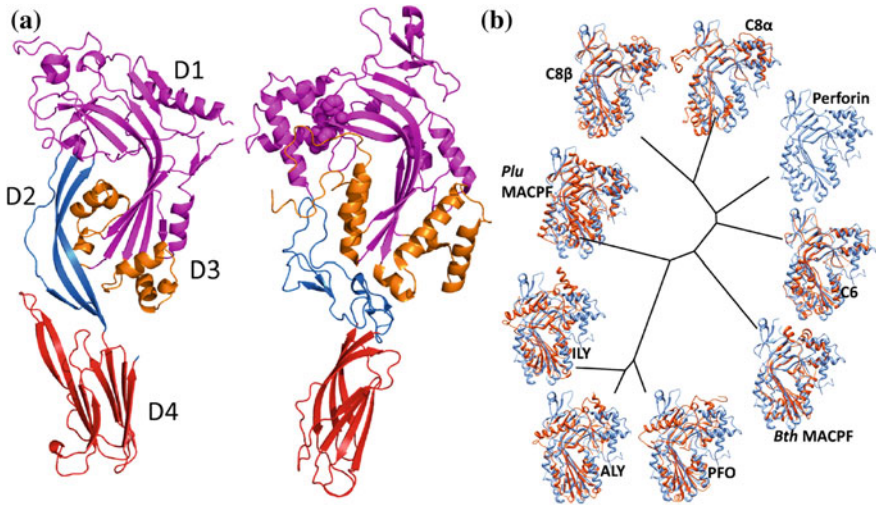


Fig. 2.1 Structure and similarities of CDC and MACPF proteins **a** Crystal structure of perfringolysin O (PFO), the first CDC to have its structure determined (*left*) [79] and mouse PFN (*right*) [57]. The original domains of the CDC are labelled D1–D4. The MACPF/CDC domain is shown in *magenta*, with helical regions that refold to form the transmembrane β -barrel coloured in *orange*. The membrane interacting domains in both proteins, immunoglobulin-like domain in PFO and C2 domain in PFN, are labelled *red*. The conserved MACPF residues in PFN are labelled with *spheres* (see below). **b** Structural phylogenetic tree of determined MACPF and CDC structures. The PFN structure is shown in *blue* and each MACPF/CDC structure in *orange*. The figure was adapted from [30]

represent a continuum in sequence space (see below) and several other shared features can be identified (see also [Chap. 4](#)).

The structure of CDCs is traditionally divided into four domains, each with particular functions during the pore-forming process (Fig. 2.1a) [29, 94]. The C-terminal domain 4 is used for the attachment of the protein to the lipid membrane, domain 1 is primarily used for contacts between the monomers, domain 3 is used for the formation of the final transmembrane β -barrel pore and domain 2 is a linking structure that provides flexibility needed during the process of pore formation (Fig. 2.1a) [29, 42]. See [Chaps. 4](#) and [5](#) in this volume for description of a detailed understanding of the CDC mechanism of action. Structural similarity between MACPFs and CDCs is centred on the central bent β -sheet composed of four strands and flanked on the sides by clusters of α -helices (the MACPF/CDC domain) [35, 77] (Fig. 2.1). From the structural comparisons it seems that domains 1 and 3 of CDCs could be seen as a single functional unit with a fold that is conserved throughout the tree of life (Fig. 2.1b). The universally-conserved glycines permit conformational rearrangements of this region of the protein to allow insertion of part of it into the lipid membrane, forming a transmembrane pore.

Distribution of MACPF Proteins

The MACPF protein family (Pfam PF01823) is the biggest family of eukaryotic pore-forming proteins and the current version of the Pfam database (version 27.0) [75] contains 1,329 unique sequences. The MACPF-containing proteins are widespread in nature and may be found in all domains of life [30]. Proteins with the MACPF domain are most abundant in metazoans and many representatives may be found in higher eukaryotes. Some organisms possess large numbers of genes containing MACPF domains—for example amphioxus (*Branchiostoma floridae*) possesses 29 such genes [43]. In this organism approximately 10 % of all genes are defence-related and since MACPF proteins represent an important part of the immune system such a high number is not surprising. MACPF proteins are also abundantly represented in Apicomplexans, where they have a crucial role in tissue and cell invasion and egress [47]. MACPF proteins were even found in anguillid herpesvirus 1 [96].

Prokaryotic representatives, primarily from the taxonomic groupings Chlamydia, Proteobacteria and Bacteroidia, may be found in various habitats and come from the following genera: bacteria of the human microbiome (*Alistipes*, *Bacteroides*), bacteria from the sea (*Dokdonia*, *Leeuwenhoekia*, *Plesiocystis*, *Zunongwangia*), from land waters (*Beggiatoa*), plant pathogens (*Calvibacter*, *Ralstonia*), bioluminescent pathogens of insects (*Photorhabdus*), human pathogens (*Chlamydia*, *Chlamydophila*, *Chryseobacterium*, *Filifactor*, *Porphyromonas*), cyanobacteria (*Trichodesmium*), etc.

MACPF proteins are abundantly represented in mammalian genomes. In the human genome, for example, there are at least 12 genes that code for a MACPF domain (Fig. 2.2). Complement proteins C6, C7, C8 α , C8 β and C9, perforin and macrophage-expressed gene 1 (Mpg-1) have important roles in the immune system, while astrotactins (ASTN) 1 and ASTN2 and bone morphogenic protein/retinoic acid inducible neural-specific protein (BRINP) 1, BRINP2, and BRINP3 have important roles in the neuronal system (see below). Some of these MACPF-containing proteins have relatively simple domain structures, for example Mpg-1 and the BRINPs have a MACPF domain at the N-terminus and a C-terminal region of unknown structure and function. Mpg-1 is an interesting example of a MACPF protein that is permanently tethered to the membrane in contrast to other well-known MACPF proteins such as complement proteins and perforin. Some of these have a complex structure containing many different associated domains (see Chaps. 6 and 10 for a detailed domain organisation of complement proteins and perforin, respectively), while ASTN1 and ASTN2 are the largest human MACPF proteins with multiple domains. They are composed of two N-terminal transmembrane helices, three epidermal growth factor-like domains in the central part followed by the MACPF domain and finally a fibronectin type-III domain on the C-terminus.

Although family membership is simply identified, there is in fact only limited sequence conservation in the MACPF family, as the sequences are around 20 %

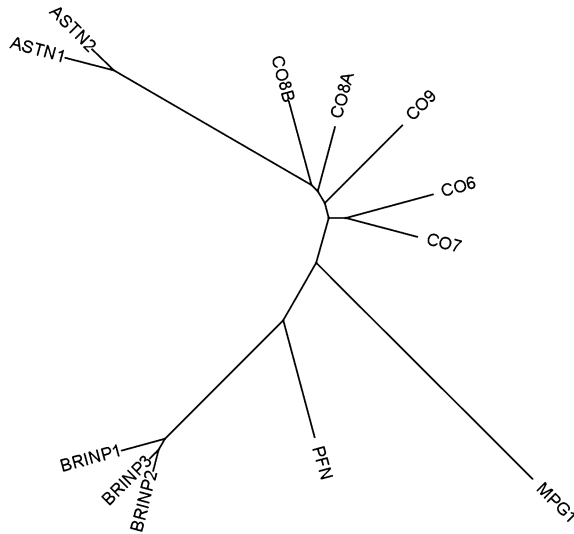
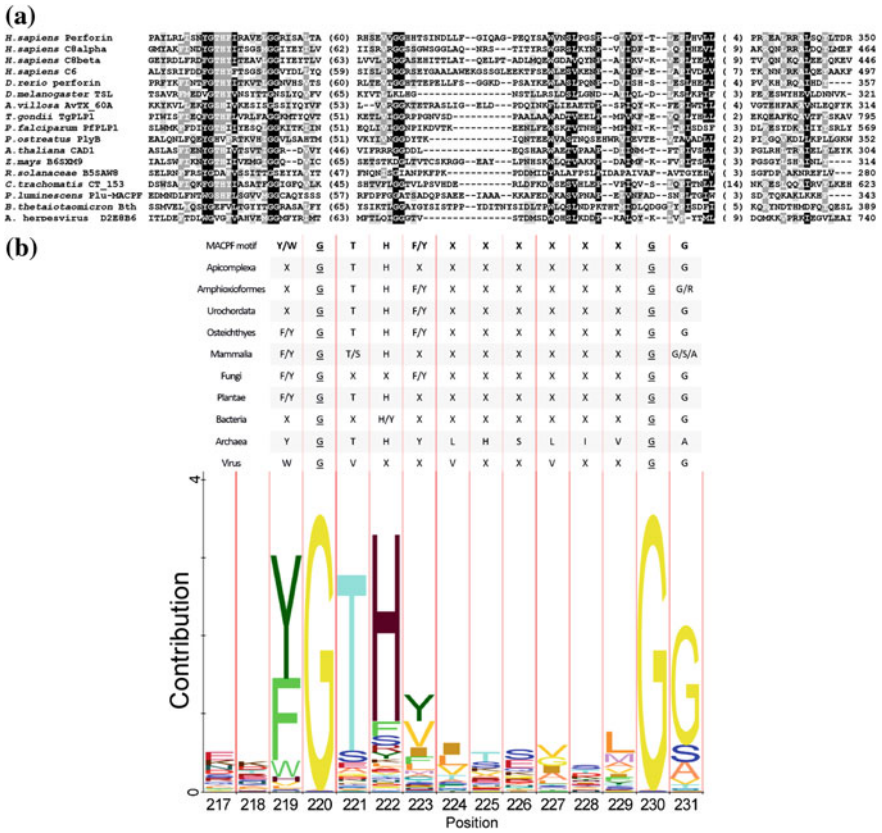


Fig. 2.2 Maximum-likelihood phylogenetic tree of human MACPF proteins. Protein sequences from UniProtKB (www.uniprot.org) were aligned using PSI-Coffee on T-Coffee server [51] and the phylogenetic tree was constructed with RAxML BlackBox [87]. The tree was depicted using *Archaeopteryx* [37]. *Sequence names* ASTN1 (Astrotactin 1, O14525), ASTN2 (Astrotactin 2, O75129), CO8B (complement component C8 β , P07358), CO8A (complement component C8 α , P07357), CO9 (complement component C9, P02748), CO6 (complement component C6, P13671), CO7 (complement component C7, P10643), MPG1 (macrophage-expressed gene 1 protein, Q2M385), PFN (P14222), BRINP1 (BMP/retinoic acid-inducible neural-specific protein 1, O60477), BRINP2 (Q9C0B6), BRINP3 (Q76B58)

identical (Fig. 2.3a). The most conserved amino acids are within the so-called MACPF motif [73] (Fig. 2.3b). This motif may be found in the structure at the top of the twisted β -sheet and particularly conserved are two glycine residues at the hinge between the upper and lower segment of the β -sheet (Fig. 2.1a). Other amino acid residues are not so well conserved (Fig. 2.3b). Thus the low sequence conservation presents a great challenge in discovering novel MACPF-containing proteins, as is highlighted by the fact that the original discovery of MACPFs in mammals and of CDCs in Gram-positive bacteria hid their deep evolutionary relationship until their structures were solved. This is demonstrated nicely by our recent analysis of MACPF proteins in fungi detailed in the next section.

MACPF Proteins in Fungi

Proteins with a MACPF/CDC domain are present in fungi as well. It is believed that more than 1.5×10^6 fungal species exist, with diverse lifestyles. They are important pathogens, saprophytes and symbionts. A huge expansion of available fungal



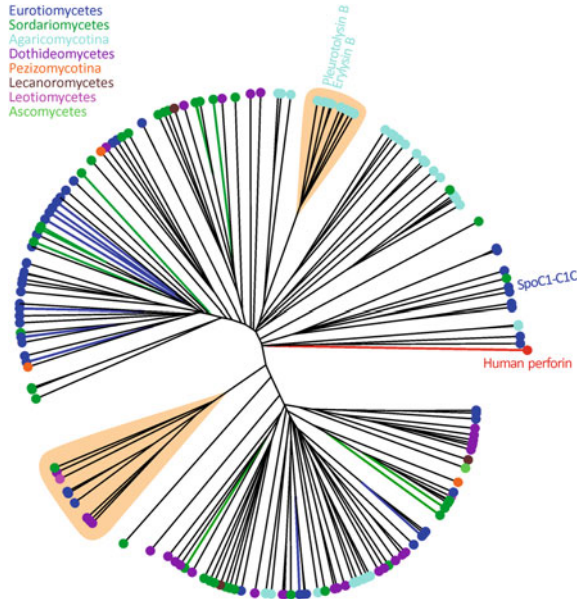


Fig. 2.4 Grouping of putative fungal MACPF proteins in relation to taxonomy, neighbour-joining tree without distance corrections was constructed after aligning sequences by using MAFFT algorithm. Taxon representation is according to JGI Mycosom [32]. The distribution of 9 putative MACPF proteins from *Aspergillus wentii* is marked with blue branches and 11 putative MACPF proteins from *Fusarium graminearum* in green. Orange background highlight branches of MACPF protein-containing genes that are in neighbourhood of aegerolysin protein. The position of functionally characterised proteins is denoted by names

fungi with a filamentous growth form (Fig. 2.4). A short signature sequence Y/F-G-X₂-F/Y-X₆-G-G was identified in 151 of the putative MACPF-containing proteins (Fig. 2.3b), verifying their identity. Some MACPF coding genes have a large number of introns, 8–13, in comparison to two introns in an average fungal gene and, consequently, annotation was difficult (interestingly, note that the MACPFs of Api-complexans are all mono-exonic except one, PPLP1/SPECT2, which has seven exons). The phylogenetic evolutionary tree was generated by using MAFFT [48] and a Neighbor-Joining approach. The sequences fall into two groups (Fig. 2.4). The first contains primarily sequences confirmed by Pfam and they are grouped together with PFN. In the second group are mostly sequences confirmed only by PHYRE2 but containing the typical MACPF/CDC signature as well.

We found the distribution of putative MACPF proteins in filamentous fungi uneven and with no clear correlation to taxonomy (Fig. 2.4). For example, in Agaricomycotina MACPF proteins were identified in 45 % of species, while in Eurotiomycetes they were present in 69 % of species (Table 2.1). The numbers of putative MACPF proteins present in a single fungal species are from one and up to nine. Some MACPF/CDC proteins within one species group together in the phylogenetic tree, while some do not, as is marked on Fig. 2.4 for 9 putative

Table 2.1 Taxonomic distribution of MACPF proteins in fungi

Taxon	Number of species in taxon	Number of species with MACPF sequence	Percent of species with MACPF sequence (%)	Number of MACPF sequences
Agaricomycotina	37	17	45	38
Pezizomycotina	4	3	75	3
Eurotiomycetes	42	29	69	60
Dothideomycetes	39	18	46	34
Lecanoromycetes	1	1	100	3
Leotiomycetes	6	2	33	2
Sordariomycetes	35	16	46	40
Ascomycetes	1	1	100	1

MACPF proteins from *Aspergillus wentii* (blue branches on Fig. 2.4) or for 11 putative MACPF proteins from *Fusarium graminearum* (green branches).

Among the strains with putative MACPF domains are those with declared pathogen lifestyles, like plant pathogens *Cochliobolus carbonum* (Northern Corn Leaf Spot), *C. heterostrophus* (Southern Corn Leaf Blight), *C. miyabeanus* (Brown Spot of rice), *C. victoriae* (Victoria Blight of oats), *Fusarium graminearum* (Wheat head blight), *F. oxysporum* (Panama disease) and *Moniliophthora perniciosa* (Witches'-broom disease), or animal pathogens like: *Arthrotrix oligospora* (Nematode trapping), *Cordyceps militaris* (Caterpillar fungus), *Trichophyton equinum* (Horse ringworm), *T. rubrum* (Athlete's foot) and *T. tonsurans* (Scalp ringworm). Some are also saprophytic, like soil fungus *Chaetomium globosum* and *Aspergillus nidulans*, *Postia placenta* (Brown rot) or edible Oyster mushroom *Pleurotus ostreatus* (White-rot), *Pleurotus eryngii* (Boletus of the steppes) and food producing *Aspergillus oryzae* (Yellow koji).

SpoC1-C1C from *Aspergillus nidulans* is among those rare fungal MACPF proteins with assigned function, as a conidium-specific member of the SpoC1 gene cluster [33]. It has been shown for some fungal MACPF proteins, like pleurotolysin B or erylysin B, that they act in concert with additional smaller proteins, the aegerolysins [71, 82, 92]. It is interesting to note that in some cases these two components are found as neighbouring genes in the genome. The subsets of MACPF proteins with an aegerolysin component as a neighbour are denoted by an orange background on Fig. 2.4. The interactions between the components and pore-forming capacity of fungal MACPF proteins are described in detail in Chap. 14.

In conclusion, only a little is known about the function of MACPF proteins in filamentous fungi. From the taxonomic distribution one can clearly conclude that they are definitely not part of the core proteome. Fungal MACPF proteins most probably contribute to different specific processes, while some of them were found in secretomes (without a typical signal sequence recognised) [18], others being intracellular, and some of them may be involved in pathogenesis, however, most probably not all of them. Large intron number, sporadic taxonomic distribution, and different numbers of diverse MACPF proteins per fungal species may imply involvement of horizontal transfer mechanisms in specific environmental niches.

Domain Organisation of MACPF Proteins

In most cases one MACPF/CDC domain exists per protein, although some examples with two or even three MACPF/CDC domains have been identified [30]. It is not however uncommon for other protein domains to be associated with MACPF proteins. The domain architecture is quite simple in bacteria, where MACPFs are located N-terminally to other domains and as a rule there is only one additional domain present. Metazoan MACPF/CDC proteins have much more complex architectures and several other domains may be present on either side of the MACPF domain. These domains are quite often involved in molecular interactions needed to exert the biological function of a particular MACPF protein. For example many MAC proteins contain additional domains that are important for regulation of MAC assembly at the surface of the target cell (see the details in Chap. 6). Many MACPF-containing proteins are involved in membrane interactions and pore formation and, therefore, there is quite a large number of small β -sheet rich domains associated with membrane interactions. For example, the C2 domain is present at the C-terminus of PFN and is used for calcium-specific membrane attachment [74, 99]. In other proteins such domains may facilitate cell attachment through interactions with sugars, such as the MIR domain, Jacalin domain (a mannose-binding lectin domain), B-lectin domain, MABP domain, etc. Apicomplexan MACPF proteins contain 1–3 so-called APC- β domains (Apicomplexan Perforin-like proteins C-terminal β -pleated domains) [47]. These are small modules of ~ 55 amino acids and, according to sequence comparisons at least, are unique to apicomplexan proteins (though see discussion in Chap. 4). For many of these cases the functional significance of particular domains for cell-surface interactions remains to be verified.

There are, however, many MACPF-containing proteins that do not possess additional domains (such as torso-like protein from *Drosophila melanogaster*). In such cases it needs to be determined whether these proteins are able to interact with lipid membranes and if such interactions are needed at all to exert their physiological activity. However, it is known that some MACPF proteins associate with other components in order to successfully complete formation of pores. For example, complement monomers of C9 associate with the C5b-8 complex [4, 5, 36, 62] (see Chap. 6). The MACPF protein pleurotolysin B from the fungus *Pleurotus ostreatus* needs a smaller component, pleurotolysin A for efficient membrane interactions and pore formation [71, 92] (see Chap. 14); also, an interesting combination of disulphide linked MACPF and lectin domains was reported recently [23].

Functional Properties of MACPF Proteins

Some of the described functions of MACPF-containing proteins are listed in Table 2.2. The most studied and understood role is that of MAC proteins and perforin in the immune system. These proteins help with the removal of unwanted

Table 2.2 Functions of MACPF proteins

Biological process	Function	Examples
Immunity	Membrane attack complex of vertebrates	C6, C7, C8 α , C8 β , C9
	Killing of unwanted cells	Perforin
	Killing of bacteria and innate immune defence in invertebrates	Mpg-1 and homologues
	Pathogen-induced immune response	CAD1
Development	Neuronal development	Astrotactin1, BRINP proteins
	Insect embryo development	Tsl
	Sea urchin embryo development	Apextrin
Pathogenesis	Tissue invasion and cell egress	Apicomplexan MACPF proteins
	Bacterial pathogenesis	Plu-MACPF
	Toxins	PsTX-60A

For references see text.

cells from the organism, either microorganisms [72], virally infected cells or cancer cells [98]. The removal is achieved through efficient formation of transmembrane pores at the cell membrane surface or within endosomes and is described in detail elsewhere in this book (Chaps. 6 and 10). In plants a MACPF protein CAD1 was shown to have a role in plant responses induced by pathogens [67].

The MACPF protein family is, however, characterised by a variety of functions that may not always be linked to pore formation or membrane association (Table 2.2). One prominent role of MACPF-containing proteins is in the development of insects [88], sea urchins [34] and mammalian neurons [1, 49, 106]. In apicomplexans MACPF proteins are implicated in parasite invasion of tissues and egress from cells [47] (see also Chap. 12 for a more detailed description of MACPF protein function in apicomplexan parasites). In bacteria it was proposed that they are important for pathogenic bacteria (e.g. *Chlamydia*), however their role in pathogenicity is not yet entirely understood [73, 77] (see also Chap. 13 for the role of MACPF protein in *Chlamydia*). MACPF-containing proteins were also suggested to have a role in protein secretion, nutrient uptake systems or to have a preventive role through molecular mimicry [104].

Involvement of MACPF proteins in the native immunity of lower eukaryotes and as a toxic arsenal for water-borne organisms is described in more detail in the following sections.

Macrophage-Expressed Gene 1 (Mpg-1)

Mpg-1 protein, termed also perforin-2, was originally identified by differential screening of a murine macrophage cDNA library. It showed macrophage and

differentiation stage-specific expression and its expression was increased upon transition of murine fetal liver hematopoietic cells into macrophages [86]. Mice contain another Mpg-1 homologue, termed EPCS50, which is primarily expressed in trophoblast giant cells in the ectoplacental cone and the parietal yolk sac [39]. It was also shown that Mpg-1 is increased in prion-inoculated brain tissue of mice, however the physiological function of Mpg-1 in mice brains was not clarified and is not known [55]. Thus, very little was known about the function of Mpg-1 until recently, where two studies demonstrated its importance for the killing of intracellular bacteria. In one, it was shown that Mpg-1 has an important role in elimination of intracellular bacteria [64] because the expression of its mRNA in mouse embryonic fibroblasts and a fibroblast NIH-3T3 cell line was shown after exposure of cells to several cytokines (IFN- α , IFN- β and IFN- γ). The same study showed that Mpg-1 mRNA was induced by exposure of fibroblasts to *Escherichia coli* and *Mycobacterium smegmatis*. Increased levels of Mpg-1 mRNA were associated with increased bactericidal activity and Mpg-1 also increased the susceptibility of intracellular bacteria to lysozyme. The second recent report showed that Mpg-1 is needed for elimination of intracellular *Chlamydia trachomatis* cells in macrophages [25]. *Chlamydia* infection caused an increase in Mpg-1 mRNA and in protein levels of Mpg-1 in murine macrophages. Production of Mpg-1 mRNA could also be induced by heat-killed *Chlamydia* in HeLa epithelial cells but not with viable bacteria and this led to the realisation that *Chlamydia* can actively suppress the production of Mpg-1 mRNA in such cells. This study also showed that Chlamydiae were susceptible to killing by ectopically expressed Mpg-1. In summary, these results demonstrated that expression of the Mpg-1 gene can be induced in other cells than macrophages, i.e. fibroblasts and HeLa cells, and that it is required for elimination of intracellular bacteria.

Mpg-1 represents an important constituent of innate immune defence in evolutionarily lower organisms. For example it was discovered in the sponge *Suberites domuncula* [103], cnidarians [65], the planarian *Schmidtea mediterranea* [6], molluscs, oysters [38] and abalones [52, 63, 100]. The Mpg-1 protein is evolutionarily very conserved, for example sponge and human Mpg-1 share 46 % similar and 28 % identical residues [103]. However, the evolutionary history of Mpg-1 is not clear, as it is not found in the genomes of *Drosophila* or *Caenorhabditis*. Its presence in gastropods and bivalves indicates that it may originate from an ancient ancestral gene that was lost in insects and nematodes [38]. Recent studies on perforin evolution are in agreement with this proposition and suggest that Mpg-1 is the precursor of perforin, which arose from duplication of an ancient Mpg-1 gene [21].

Marine invertebrates are in constant contact with bacteria and many genome-encoded effectors are an important constituent of their innate immunity. Haemocytes are specialised blood cells of the mollusc's cellular immune response. They circulate in the hemolymph and are able to recognise foreign antigens through specific pattern recognition motifs. Pathways for pathogen destruction include specific transduction pathways that are similar to vertebrate pathways [44]. It was observed that Mpg-1 is kept at very low levels in haemocytes and is upregulated

upon stress or infection with bacteria [52]. In different experimental settings the up-regulation of Mpg-1 was found upon challenge with bacteria [38, 52, 100, 101], ostreid herpesvirus 1 [76], parasite *Schistosoma mansoni* [45] or lipopolysaccharides in sponges and planarians [6, 103], further indicating its important and ancient role in innate immunity. Kemp et al. also showed that the protein levels of Mpg-1 were upregulated upon stress with bacterial exposure and that protein levels were upregulated in a similar manner to mRNA levels [52]. This was shown also for Mpg-1 protein from sponge [103]. Some Mpg-1 homologues have been expressed, purified to homogeneity and tested for antibacterial activity. Mpg-1 from Pacific oyster *Crassostrea gigas* was shown to possess antibacterial activity against Gram-positive and Gram-negative bacteria [38], while recombinant Mpg-1 protein from sponge exhibited antibacterial activity against Gram-negative bacteria [103].

Toxins with the MACPF Domain

MACPF proteins were described also as a part of the toxic arsenal of lower eukaryotes. Cnidarians evolved special organelles called cnidocysts, which are considered as one of the most complex cellular secretory products. It is widely believed that cnidocysts serve for storage and active delivery of toxic compounds in cnidarian venom [53, 61, 91]. Many different cytolytic protein families were identified in Cnidaria [9, 11, 28]; for example Nagai et al. have reported isolation of two toxins, PsTX-60A and PsTX-60B, from the nematocysts of the sea anemone *Phyllodiscus semoni*. These two proteins represent major venom toxins and were haemolytic (ED₅₀ values 300 and 600 ng/ml using 0.8 % suspension of sheep red blood cells) and lethal to shrimps (LD₅₀ value around 800 µg/kg for intraperitoneal injection) [68]. A similar protein was purified from the nematocyst venom of another sea anemone *Actinaria villosa*. This toxin, termed AcTX-60A, was lethal to mice (minimal lethal dose for intraperitoneal injection less than 250 µg/kg). Its primary structure was determined from mRNA and for the first time these proteins were recognised as members of the MACPF family [70]. A subsequent report of the full length sequence of PsTX-60B reconfirmed similarity to MACPF domains [80]. Proteins similar to PsTX-60A were found by proteomic analysis also in nematocyst venom from the jellyfish *Olindias sambaquiensis* [102] and similar proteins were found in the genomes of the hydrozoan *Hydra magnipapillata*, sea anemone *Nematostella vectensis* and in coral *Acropora millepora* [65]. Furthermore, EST clones, publicly available in GenBank, of sea anemones *Aiptasia pallida* [89] and *Metridium senile* also contain PsTX-60A homologues. It was shown previously that *M. senile* contains an 80 kDa polypeptide cytolytic metridiolysin, reminiscent of CDCs based on functional and structural properties [14] and it is tempting to suggest that metridiolysin belongs to the MACPF protein family. It seems that MACPF-containing protein toxins are not only restricted to sea anemones, but are present also in other cnidarians and they clearly represent another family of cnidarian pore-forming toxins [11].

The domain organisation of cnidarian MACPF domain-containing toxins resembles that of perforin. A MACPF domain is placed at the N-terminus and it is followed by an EGF-like domain. There is, however, a notable exception that the C2 domain is missing at the C-terminus and it will be interesting to determine the molecular details of membrane attachment and which parts of the protein allow initial interaction with the membrane. The PsTX-60A protein has also a typical signal sequence and a propeptide at the very N-terminus, a structural organisation that is shared by other nematocyst toxins [10]. However, it is not clear if all cnidarian MACPF-containing proteins are toxins expressed in cnidocytes. As noted by Miller et al. it is unclear if Hy-MAC, a homologue of PsTX-60A in *Hydra*, is an orthologue, since the overall sequence identity is low [65]. Furthermore, expression of Hy-MAC was restricted to gland cells in the endoderm and not nematocytes [65]. In another study, a differentially expressed PsTX-60A homologue gene, termed *anklet*, was also shown to be expressed in the newly differentiated gland cells of the basal disk and it was suggested that this gene is involved in the formation and maintenance of the basal disk in *Hydra* by exerting a cytotoxic role [7].

Another toxin that contains the MACPF domain was reported recently in an aquatic snail [23]. Protein PcPV2 from *Pomacea canaliculata* is a perivitellin involved in egg defence against predation. PcPV2 is a neurotoxin with strong lethal effects against rodents and is used for defence purposes [23]; it is glycosylated and is composed of four 98 kDa heterodimers yielding a 400 kDa complex. The smaller 31 kDa component is similar to tachylectin-1, while the larger 67 kDa domain is similar to the MACPF domain [23]. Both domains are linked with a disulphide bond [27] and since the tachylectin-1 component of PcPV2 allows binding to cells this domain combination resembles the A-B toxin combination found in bacterial and plant lectins [23].

Distribution of CDC Proteins

CDCs (Pfam PF01289) were originally identified in a number of Gram-positive bacteria from the genera *Bacillus*, *Clostridium*, *Streptococcus* and *Listeria* where they are associated with invasion of and replication within animal hosts [29, 94]. The most well-known such CDCs include perfringolysin O from *Clostridium perfringens*, streptolysin O from *Streptococcus pyogenes*, pneumolysin from *Streptococcus pneumoniae* and listeriolysin O from *Listeria monocytogenes* [29, 94]. In each case the CDC has a distinctive role. Perfringolysin O is responsible for *Clostridia* gaining access deep into tissue to protect anaerobic respiration, leading to gangrene [40], pneumolysin is required for pneumococcal virulence via the effects it has on cell membranes that permit host invasion [97], and listeriolysin O is used to enable invasion and growth within cells [17]. There are, however, many less well-known examples and to snapshot the current state of knowledge we performed a protein sequence database search using the whole

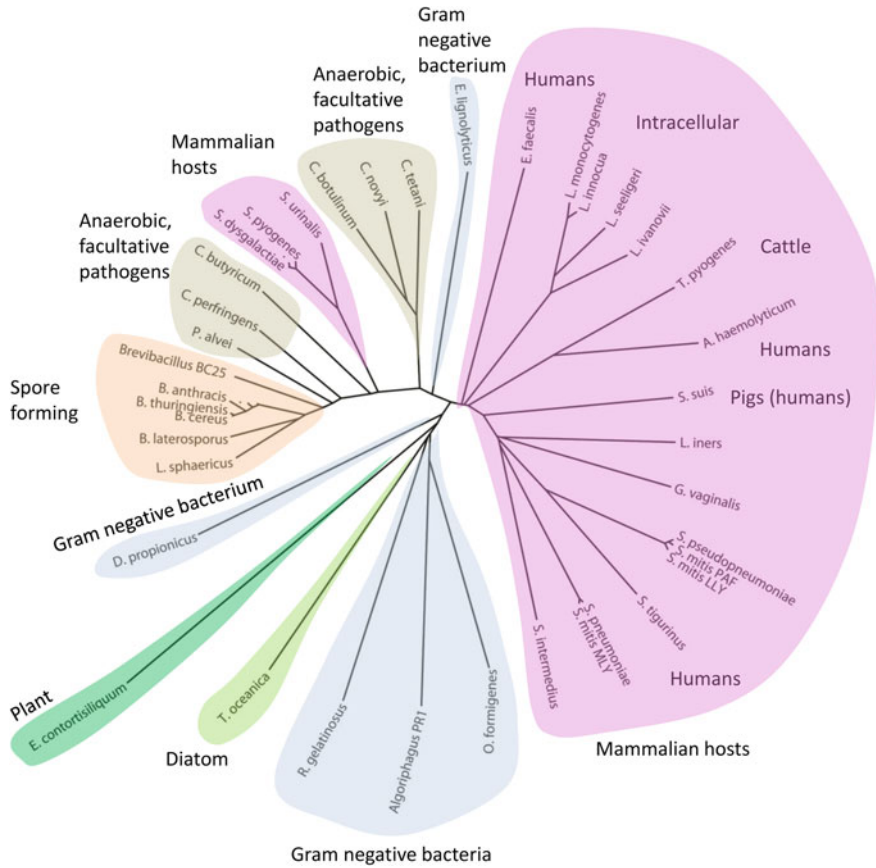


Fig. 2.5 Sequence-based phylogenetic tree of CDC proteins, *B. mycoides* and *B. weihenstephanensis* are not shown but share a branch with *B. anthracis*. *S. canis* shares a branch with *S. dysgalactiae*

perfringolysin O sequence, on the UniprotKB (<http://www.uniprot.org>). Of the 1,000 hits studied we found 40 unique sequences, and subsequently added the enterolysin from the Gram negative bacterium *Enterobacter lignolyticus* as identified alongside the CDC from *Desulfobulbous propionicus* by Tweten and colleagues [41], and the CDC identified in *Enterolobium contortisiliquum* (a member of the pea family, from South America) [26]. We pre-aligned the sequences in Uniprot and then used Clustal Omega [84] to perform alignments of the amino-terminal region of each protein (up to the end of the MACPF-homologous domain). We used the output alignments to construct phylogenetic trees using ClustalW2 Phylogeny [31, 56], which we then depicted using the program DRAWTREE (part of the PHYLIP package) [24] (Fig. 2.5). There are a number of interesting observations to be made on the basis of this analysis.

Firstly, in addition to the previously-identified CDCs from *E. lignolyticus* and *D. propionicus* we identified three more Gram-negative organisms which possess family members: one, *Oxalobacter formigenes* is a commensal human bacterium which relies on the degradation of oxaloacetate for most of its energetic and carbon needs and which (at 2,100 genes) has one of the smaller bacterial genomes [54]. By contrast, *Rubrivirax gelatinosus* and the *Algoriphagus* species PR1 are aquatic organisms; *R. gelatinosus* is a photosynthetic and nitrogen-fixing aquatic beta-proteobacteria [69], while *Algoriphagus* PR1 (*marchipongonensis*) is distinguished by the fact it is the prey of the choanoflagellate *Salpingoeca rosetta* [2, 3]. Choanoflagellates are the closest known relatives of animals, species of eukaryote of which some are unicellular and some multicellular colonies [95]. Even more surprisingly, we identified a form of CDC produced by the diatom *Thalassiosira oceanica* [60], which, after enterolobin from *E. contortisiliquum* is the second identified eukaryotic CDC. Genomic analysis of this species demonstrates 5 % gene acquisition from diverse taxonomic groups and so is consistent with the acquisition of a CDC by horizontal transfer from an aquatic bacterium [60]. The percent identity between the CDC of *T. oceanica* and that of *D. propionicus* is 23.9 %, that with the *Algoriphagus* PR1 26 % and that with *C. perfringens* perfringolysin 27.5 %. Its identity with the human-specific intermedilysin from *S. intermedius* is 25.2 % and to enterolobin, its neighbour in the phylogeny, is 13.9 % (see Fig. 2.6 for examples of alignment). The greater difference between the CDCs in *T. oceanica* and *E. contortisiliquum* reflects the much more developed evolutionary history of the plant versus the diatom, and the relative stasis in lifestyle and selective pressure experienced by bacterial lineages. The fact that sequences of CDCs in prokaryotic organisms which have not encountered each other for millions of years remain relatively stable at ~25 % identity suggests a strong selective pressure which is also, of course, reflected in the much greater conservation demonstrated by the MACPF/CDC fold [30]. The fact that the *T.oceanica* CDC is not more similar to those of bacteria nearby in ecological niche and within the phylogenetic tree than it is to those of bacteria much more distant phylogenetically and in ecology suggests that the diatom either acquired its CDC from the nuclear genome of its ancestor heterotrophic eukaryote, or due to an endosymbiotic event, or that any horizontal gene transfer that occurred was very ancient.

How ancient was the acquisition of a CDC by an ancestor of *T.oceanica*? A version of this phylogeny lacking just enterolobin did not present a single root, but instead a series of branches from a central common line. It is very interesting that this phylogenetic analysis based solely on sequence positions the diatom and plant CDCs adjacent to one another, and branching from a single point close to the origin of species. This leads to four of the five Gram negative CDCs alongside the diatom and plant CDCs forming a clade distinct from all other known family members, rooted at the origin of the tree. The absence of CDC proteins in many species argues against its presence in the genome of the last universal common ancestor and, therefore, the CDC gene transfer to diatoms and plants appears to have occurred from (endosymbiosis of) an ancestor of these Gram negative species

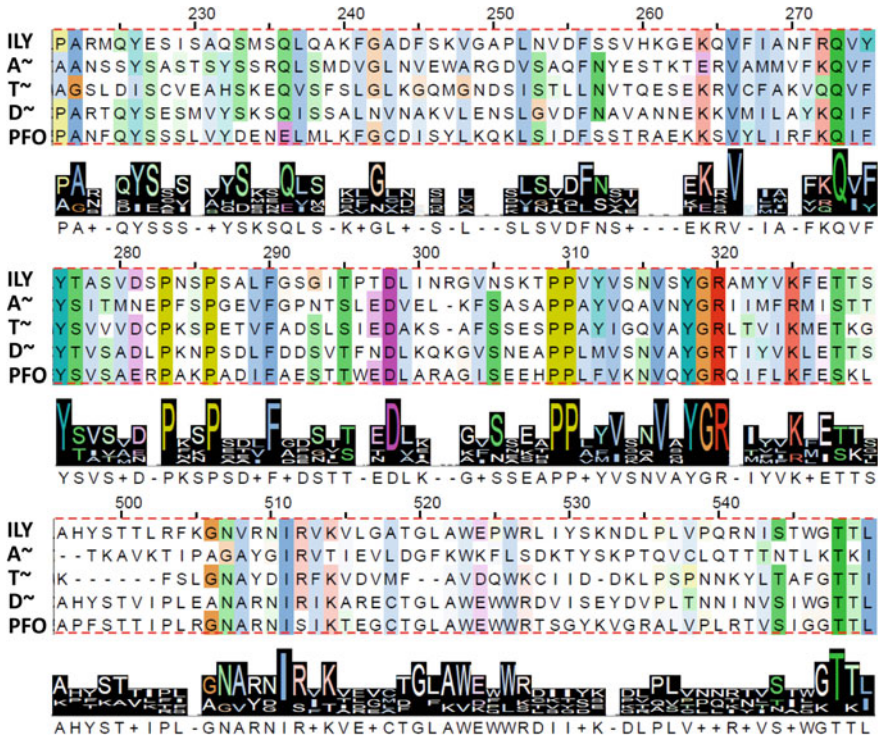


Fig. 2.6 Two regions of CDCs aligned. The first two sets of sequence comprise the core of the MACPF homology domain including the TMH helices; then final sequence comparison encompasses the carboxy-terminal domain (domain 4) undecapeptide. *ILY* intermedilysin from *S. intermedius*; *A~* the *Algoriphagus* species CDC; *T~* the *Thalassiosira oceanica* CDC; *D~* the *Desulfolobus* CDC and *PFO* perfringolysin O from *C. perfringens*

of which the evolutionarily closest example now known is *D. propionicus*. Another possibility is that a CDC was present in the last universal common ancestor as some sort of a housekeeping gene and was subsequently lost if not needed in particular species.

Diatoms evolved via a two-stage endosymbiotic process [12]. First, a cyanobacterium was captured by a eukaryotic heterotroph and this generated the origin of plants and red and green algae. Second, a red alga was captured by a different eukaryotic heterotroph and this gave rise to the stramenophile lineage that includes diatoms (emerging about 250 million years ago), macroalgae and plant parasites. Since our plant and diatom examples of CDCs derive from a single root this suggests that the CDC in diatoms (and in *E. contortisiliquum*) derives from the first endosymbiosis and thereby from an original cyanobacterium which shared a common (non-photosynthetic) ancestor with *D. propionicus*. In their paper on the *D. propionicus* and *E. lignolyticus* CDCs, Tweten and colleagues [41] discuss the possibility that a CDC may have helped to fend off endocytosis of primitive

bacteria. It seems plausible that diatoms and plants result from a successful endosymbiotic event involving a prokaryote possessing a CDC.

Overall, the pore-forming (MACPF) domain phylogeny for CDCs shows a structure strongly consistent with the lifestyles adopted by the producing organisms. Thus, one clade encompasses the majority of the *Streptococcal* species (those of the *mitis* grouping) [19, 66, 105] alongside others infecting humans such as *Gardnerella vaginalis* and *Lactobacillus iners*. Here, *S. mitis* stands out as a species of bacterium expressing three different CDC proteins: lectinolysin, mitilysin and the platelet aggregation factor. Its clade shares a common root to the centre of the phylogeny with other bacteria with similar or the same hosts (*Arcanobacterium haemolyticum*, first isolated from Pacific islanders, and *T. pyogenes*) [16, 90] and separate branches containing the intracellular *Listeria* bacteria and the originally commensal bacterium *Enterococcus faecalis* now found also as a drug resistant pathogen [58]. The position of the *mitis* grouping of *Streptococci* at a location distant from the pyogenic group (*S. pyogenes*, *S. canis*, *S. urinalis* and *S. dysgalactiae*) [20] is in line with their relative distance based on a published analysis of seven housekeeping genes that confirmed the phylogenetic positioning of the *dysgalactiae* species, among others [46].

Interposed between the two groups of *Streptococci* lie an isolated Gram negative example (from *E. lignolyticus*) and anaerobic bacteria (*Clostridium novyi*, *C. botulinum* and *C. tetani*) which can also cause human disease [40]. This common grouping matches the shared anaerobic ecological niche of *E. lignolyticus* and the Clostridial species. The *Clostridium* producers of CDC proteins also include *P. alvei*, *C. perfringens* and *C. butyricum* which are similarly anaerobic and associated with plants, humus and rotting material and again are potential human pathogens. These share an ecological niche with the spore-forming bacilli infecting livestock (*B. anthracis*) and insects (*B. thuringiensis*, *L. sphaericus*, *B. laterosporus* and *Brevibacillus*) [15, 22].

Evolutionary Implications

While CDCs are primarily cytolytic proteins, not all MACPF proteins have lytic activity and other biological functions are possible. This echoes the situation with eukaryotic proteins of the Bcl-2 family, which are similar to bacterial colicins. Although some Bcl-2 proteins are able to interact with membranes and form pores they do not all exhibit lytic activity and function at different stages of apoptosis—some to promote it, others to inhibit [8]. Various proteins that contain the MACPF domain are clearly evolutionary extremely conserved. For example Mpg-1 is present in sponge where it has a clear role in the immune response [103]. Similarly apextrins are widespread in cnidaria, echinoderms and apicomplexans [65]. Interestingly, apextrins seem to be specifically expressed in the ectoderm during metamorphosis in two distant evolutionary groups, cnidaria and echinoderms [65]. It is, however, extremely difficult to reconstruct the evolutionary history of the

MACPF/CDC superfamily with the current set of available genomic information and protein structures (see for example the distribution of MACPFs in fungi in section “[MACPF Proteins in Fungi](#)”). Until recently there were no examples of the two protein families that could be retrieved by sequence searches. However, now one can find MACPFs and CDCs with 20 % identity. For example the CDC of the diatom *Thalassiosira oceanica* is 22 % identical in its MACPF/CDC domain to the MACPF of *T. pseudonana*. The former was identified as a CDC by searching with the perforingolysin O sequence, the latter as a MACPF by searching with a *Plasmodium yoelii* perforin-like protein.

When reconstructing a CDC phylogeny we have identified not only known prokaryotic examples but also examples of eukaryotic proteins which, in our view, are likely to have arisen due to an endosymbiotic event. Mirroring the finding of eukaryotic CDCs, a preliminary phylogenetic study of 82 examples of MACPFs from unique genera highlights several bacterial examples of MACPFs as well as examples in coelacanths, a sea squirt, fungi (*Arthrotrichyza oligospora* and *Fusarium oxysporum*) and even the diatom *T. pseudonana*. Indeed, the bacterium *Enterococcus faecalis* appears to have a MACPF-lineage protein 19.9 % identical to its CDC (data not shown). This level of similarity alongside a pairwise alignment of the CDC from *T. oceanica* and the MACPF from *T. pseudonana* that suggests 22 % sequence identity indicates the possibility of constructing a single sequence-based phylogeny for the MACPF/CDC superfamily but this is outside the scope of this chapter. The analysis we have been able to perform does, however, show that there was indeed one original protein domain now found in at least two different sequence families, and that it was present at a point in extreme antiquity before the lineages giving rise to current bacteria, diatoms, plants, animals and fungi diverged.

Although MACPF/CDC proteins may belong to at least two different sequence families, they belong to one structural class, which has been conserved throughout evolution. Understanding why the structure has been so deeply conserved will require further work but to a first approximation it is reasonable to conclude that the capacity it has to refold from an aqueous monomeric conformation into an oligomeric membrane-inserted conformation lies at the heart of the matter. The CDCs and MACPFs may not be considered as “just” toxins or even exclusively toxins: in many cases of pathogenic bacteria, for example, they are not the distinctive protein at the heart of the mechanism of disease (e.g. not in anthrax, not in botulism). Perhaps we should, after all, think of the MACPF/CDCs as house-keeping genes, genes that conferred the capacity to remodel membranes in fundamentally helpful ways, for example in establishing colony formation and as seen today in developmental processes, and that only subsequently did they get co-opted as agents of attack and defence in bacteria, apicomplexans and humans, among other cases [30].

Elsewhere we have reported a structural phylogenetic analysis [30] and it is worth highlighting the fact that however useful such an analysis is in terms of tying together shared fold and function, it does not provide a measure of evolutionary distance in time. In that case, because there is no simple concept of a

“mutational clock” in play (as there is in the sequence data we have studied here, albeit one running faster or slower in different organisms with different nucleotide base biases, accuracies of genome replication and environments) it is very hard to know what the meaning of a tree branch length is. The sequence analysis we describe here gives a more informative measure of the evolutionary process; alongside each other the sequential and structural phylogenies provide a compelling case for understanding the unity of the CDC and MACPF families. Further structural work and novel genomic information will aid in understanding further the evolutionary history of this interesting protein superfamily.

Acknowledgments We would like to thank to Sabina Belc for genome mining and The Slovenian Research Agency for financial support.

References

1. Adams NC, Tomoda T, Cooper M, Dietz G, Hatten ME (2002) Mice that lack astrotactin have slowed neuronal migration. *Development* 129:965–972
2. Alegado RA, Ferriera S, Nusbaum C, Young SK, Zeng Q, Imamovic A, Fairclough SR, King N (2011) Complete genome sequence of *Algoriphagus* sp. PR1, bacterial prey of a colony-forming choanoflagellate. *J Bacteriol* 193:1485–1486
3. Alegado RA, Grabenstatter JD, Zuzow R, Morris A, Huang SY, Summons RE, King N (2013) *Algoriphagus machipongonensis* sp. nov., co-isolated with a colonial choanoflagellate. *Int J Syst Evol Microbiol* 63:163–168
4. Aleshin AE, Discipio RG, Stec B, Liddington RC (2012) Crystal structure of C5b-6 suggests a structural basis for priming the assembly of the membrane attack complex (MAC). *J Biol Chem* 287:19642–19652
5. Aleshin AE, Schraufstatter IU, Stec B, Bankston LA, Liddington RC, DiScipio RG (2012) Structure of complement C6 suggests a mechanism for initiation and unidirectional, sequential assembly of membrane attack complex (MAC). *J Biol Chem* 287:10210–10222
6. Altincicek B, Vilcinskis A (2008) Comparative analysis of septic injury-inducible genes in phylogenetically distant model organisms of regeneration and stem cell research, the planarian *Schmidtea mediterranea* and the cnidarian *Hydra vulgaris*. *Front Zool* 5:6
7. Amimoto Y, Kodama R, Kobayakawa Y (2006) Foot formation in *Hydra*: a novel gene, anklet, is involved in basal disk formation. *Mechanisms Develop* 123:352–361
8. Anderluh G, Lakey JH (2008) Disparate proteins use similar architectures to damage membranes. *Trends Biochem Sci* 33:482–490
9. Anderluh G, Maček P (2002) Cytolytic peptide and protein toxins from sea anemones (Anthozoa: Actiniaria). *Toxicon* 40:111–124
10. Anderluh G, Podlesek Z, Maček P (2000) A common motif in proapoptins of Cnidarian toxins and nematocyst collagens and its putative role. *Biochim Biophys Acta* 1476:372–376
11. Anderluh G, Sepčić K, Turk T, Maček P (2011) Cytolytic proteins from cnidarians—an overview. *Acta Chim Slovenica* 58:724–729
12. Armbrust EV (2009) The life of diatoms in the world’s oceans. *Nature* 459:185–192
13. Arnaud MB, Cerqueira GC, Inglis DO, Skrzypek MS, Binkley J, Chibucos MC, Crabtree J, Howarth C, Orvis J, Shah P, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G, Wortman JR (2012) The *Aspergillus* genome database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic Acids Res* 40:D653–D659

14. Bernheimer AW, Avigad LS, Kim K (1979) Comparison of metridiolysin from the sea anemone with thiol-activated cytolysins from bacteria. *Toxicon* 17:69–75
15. Berry C (2012) The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. *J Invertebr Pathol* 109:1–10
16. Billington SJ, Jost BH, Cuevas WA, Bright KR, Songer JG (1997) The *Arcanobacterium* (*Actinomyces*) *pyogenes* hemolysin, pyolysin, is a novel member of the thiol-activated cytolysin family. *J Bacteriol* 179:6100–6106
17. Birmingham CL, Canadien V, Kaniuk NA, Steinberg BE, Higgins DE, Brumell JH (2008) Listeriolysin O allows *Listeria monocytogenes* replication in macrophage vacuoles. *Nature* 451:350–354
18. Braaksma M, Martens-Uzunova ES, Punt PJ, Schaap PJ (2010) An inventory of the *Aspergillus niger* secretome by combining in silico predictions with shotgun proteomics data. *BMC Genomics* 11:584
19. Choi SM, Cho BH, Choi KH, Nam TS, Kim JT, Park MS, Kim BC, Kim MK, Cho KH (2012) Meningitis caused by *Streptococcus suis*: case report and review of the literature. *J Clin Neurol* 8:79–82
20. Collins MD, Hutson RA, Falsen E, Nikolaitchouk N, LaClaire L, Facklam RR (2000) An unusual *Streptococcus* from human urine, *Streptococcus urinalis* sp. nov. *Int J Syst Evol Microbiol* 50:1173–1178
21. D'Angelo ME, Dunstone MA, Whisstock JC, Trapani JA, Bird PI (2012) Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. *BMC Evol Biol* 12:59
22. Djukic M, Poehlein A, Thurmer A, Daniel R (2011) Genome sequence of *Brevibacillus laterosporus* LMG 15441, a pathogen of invertebrates. *J Bacteriol* 193:5535–5536
23. Dreon MS, Frassa MV, Ceolin M, Ituarte S, Qiu JW, Sun J, Fernandez PE, Heras H (2013) Novel animal defenses against predation: a snail egg neurotoxin combining lectin and pore-forming chains that resembles plant defense and bacteria attack toxins. *PLoS One* 8:e63782
24. Felsenstein J (1997) An alternating least squares approach to inferring phylogenies from pairwise distances. *Syst Biol* 46:101–111
25. Fields KA, McCormack R, de Armas LR, Podack ER (2013) Perforin-2 restricts growth of *Chlamydia trachomatis* in macrophages. *Infect Immun* 81:3045–3054
26. Fontes W, Sousa MV, Aragao JB, Morhy L (1997) Determination of the amino acid sequence of the plant cytolysin enterolobin. *Arch Biochem Biophys* 347:201–207
27. Frassa MV, Ceolin M, Dreon MS, Heras H (2010) Structure and stability of the neurotoxin PV2 from the eggs of the apple snail *Pomacea canaliculata*. *Biochim Biophys Acta* 1804:1492–1499
28. Frazao B, Vasconcelos V, Antunes A (2012) Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. *Marine Drugs* 10:1812–1851
29. Gilbert RJ (2005) Inactivation and activity of cholesterol-dependent cytolysins: what structural studies tell us. *Structure* 13:1097–1106
30. Gilbert RJ, Mikelj M, Dalla Serra M, Froelich CJ, Anderlüh G (2013) Effects of MACPF/CDC proteins on lipid membranes. *Cell Mol Life Sci* 70:2083–2098
31. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* 38:W695–W699
32. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, Otilar R, Poliakov A, Ratnere I, Riley R, Smirnova T, Rokhsar D, Dubchak I (2012) The genome portal of the department of energy joint genome institute. *Nucleic Acids Res* 40:D26–D32
33. Gwynne DI, Miller BL, Miller KY, Timberlake WE (1984) Structure and regulated expression of the SpoC1 gene cluster from *Aspergillus nidulans*. *J Mol Biol* 180:91–109
34. Haag ES, Sly BJ, Andrews ME, Raff RA (1999) Apextrin, a novel extracellular protein associated with larval ectoderm evolution in *Helicoidaris erythrogramma*. *Dev Biol* 211:77–87

35. Hadders MA, Beringer DX, Gros P (2007) Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. *Science* 317:1552–1554
36. Hadders MA, Bubeck D, Roversi P, Hakobyan S, Forneris F, Morgan BP, Pangburn MK, Llorca O, Lea SM, Gros P (2012) Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. *Cell Rep* 1:1–8
37. Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356
38. He X, Zhang Y, Yu Z (2011) An Mpeg (macrophage expressed gene) from the *Pacific oyster Crassostrea gigas*: molecular characterization and gene expression. *Fish Shellfish Immunol* 30:870–876
39. Hemberger M, Himmelbauer H, Ruschmann J, Zeitz C, Fundele R (2000) cDNA subtraction cloning reveals novel genes whose temporal and spatial expression indicates association with trophoblast invasion. *Develop Biol* 222:158–169
40. Hickey MJ, Kwan RY, Awad MM, Kennedy CL, Young LF, Hall P, Cordner LM, Lyras D, Emmins JJ, Rood JI (2008) Molecular and cellular basis of microvascular perfusion deficits induced by *Clostridium perfringens* and *Clostridium septicum*. *PLoS Pathog* 4:e1000045
41. Hotze EM, Le HM, Sieber JR, Bruxvoort C, McInerney MJ, Tweten RK (2013) Identification and characterization of the first cholesterol-dependent cytolysins from gram-negative bacteria. *Infect Immun* 81:216–225
42. Hotze EM, Tweten RK (2012) Membrane assembly of the cholesterol-dependent cytolysin pore complex. *Biochim Biophys Acta* 1818:1028–1038
43. Huang S, Yuan S, Guo L, Yu Y, Li J, Wu T, Liu T, Yang M, Wu K, Liu H, Ge J, Huang H, Dong M, Yu C, Chen S, Xu A (2008) Genomic analysis of the immune gene repertoire of *amphioxus* reveals extraordinary innate complexity and diversity. *Genome Res* 18:1112–1126
44. Humphries JE, Yoshino TP (2003) Cellular receptors and signal transduction in molluscan hemocytes: connections with the innate immune system of vertebrates. *Integrat Comp Biol* 43:305–312
45. Ittiprasert W, Miller A, Myers J, Nene V, El-Sayed NM, Knight M (2010) Identification of immediate response genes dominantly expressed in juvenile resistant and susceptible *Biomphalaria glabrata* snails upon exposure to *Schistosoma mansoni*. *Mol Biochem Parasitol* 169:27–39
46. Jensen A, Kilian M (2012) Delineation of *Streptococcus dysgalactiae*, its subspecies, and its clinical and phylogenetic relationship to *Streptococcus pyogenes*. *J Clin Microbiol* 50:113–126
47. Kafsack BF, Carruthers VB (2010) Apicomplexan perforin-like proteins. *Commun Integr Biol* 3:18–23
48. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
49. Kawano H, Nakatani T, Mori T, Ueno S, Fukaya M, Abe A, Kobayashi M, Toda F, Watanabe M, Matsuoka I (2004) Identification and characterization of novel developmentally regulated neural-specific proteins, BRINP family. *Mol Brain Res* 125:60–75
50. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protocols* 4:363–371
51. Kemena C, Notredame C (2009) Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics* 25:2455–2465
52. Kemp IK, Coyne VE (2011) Identification and characterisation of the Mpeg1 homologue in the South African abalone, *Haliotis midae*. *Fish Shellfish Immunol* 31:754–764
53. Klug M, Weber J, Tardent P (1989) Hemolytic and toxic properties of *Hydra attenuata* nematocysts. *Toxicon* 27:325–339
54. Knight J, Deora R, Assimos DG, Holmes RP (2013) The genetic composition of *Oxalobacter formigenes* and its relationship to colonization and calcium oxalate stone disease. *Urolithiasis* 41:187–196

55. Kopacek J, Sakaguchi S, Shigematsu K, Nishida N, Atarashi R, Nakaoko R, Moriuchi R, Niwa M, Katamine S (2000) Upregulation of the genes encoding lysosomal hydrolases, a perforin-like protein, and peroxidases in the brains of mice affected with an experimental prion disease. *J Virol* 74:411–417
56. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
57. Law RH, Lukoyanova N, Voskoboinik I, Caradoc-Davies TT, Baran K, Dunstone MA, D'Angelo ME, Orlova EV, Coulibaly F, Verschoor S, Browne KA, Ciccone A, Kuiper MJ, Bird PI, Trapani JA, Saibil HR, Whisstock JC (2010) The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* 468:447–451
58. Lawley TD, Walker AW (2013) Intestinal colonization resistance. *Immunology* 138:1–11
59. Lichtenheld MG, Olsen KJ, Lu P, Lowrey DM, Hameed A, Hengartner H, Podack ER (1988) Structure and function of human perforin. *Nature* 335:448–451
60. Lommer M, Specht M, Roy AS, Kraemer L, Andreson R, Gutowska MA, Wolf J, Bergner SV, Schilhabel MB, Klostermeier UC, Beiko RG, Rosenstiel P, Hippler M, Laroche J (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biol* 13:R66
61. Lotan A, Fishman L, Loya Y, Zlotkin E (1995) Delivery of a nematocyst toxin. *Nature* 375:456
62. Lovelace LL, Cooper CL, Sodetz JM, Lebioda L (2011) Structure of human C8 protein provides mechanistic insight into membrane pore formation by complement. *J Biol Chem* 286:17585–17592
63. Mah SA, Moy GW, Swanson WJ, Vacquier VD (2004) A perforin-like protein from a marine mollusk. *Biochem Biophys Res Commun* 316:468–475
64. McCormack R, de Armas LR, Shiratsuchi M, Ramos JE, Podack ER (2013) Inhibition of intracellular bacterial replication in fibroblasts is dependent on the perforin-like protein (perforin-2) encoded by macrophage-expressed gene 1. *J Innate Immun* 5:185–194
65. Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, Agata K, Bosch TC (2007) The innate immune repertoire in cnidaria—ancestral complexity and stochastic gene loss. *Genome Biol* 8:R59
66. Mitchell J (2011) *Streptococcus mitis*: walking the line between commensalism and pathogenesis. *Mol Oral Microbiol* 26:89–98
67. Morita-Yamamuro C, Tsutsui T, Sato M, Yoshioka H, Tamaoki M, Ogawa D, Matsuura H, Yoshihara T, Ikeda A, Uyeda I, Yamaguchi J (2005) The *Arabidopsis* gene CAD1 controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol* 46:902–912
68. Nagai H, Oshiro N, Takuwa-Kuroda K, Iwanaga S, Nozaki M, Nakajima T (2002) Novel proteinaceous toxins from the nematocyst venom of the Okinawan sea anemone *Phyllodiscus semoni* Kwietniewski. *Biochem Biophys Res Commun* 294:760–763
69. Nagashima S, Kamimura A, Shimizu T, Nakamura-Isaki S, Aono E, Sakamoto K, Ichikawa N, Nakazawa H, Sekine M, Yamazaki S, Fujita N, Shimada K, Hanada S, Nagashima KV (2012) Complete genome sequence of phototrophic betaproteobacterium *Rubrivivax gelatinosus* IL144. *J Bacteriol* 194:3541–3542
70. Oshiro N, Kobayashi C, Iwanaga S, Nozaki M, Namikoshi M, Spring J, Nagai H (2004) A new membrane-attack complex/perforin (MACPF) domain lethal toxin from the nematocyst venom of the Okinawan sea anemone *Actinaria villosa*. *Toxicon* 43:225–228
71. Ota K, Leonardi A, Mikelj M, Skočaj M, Wohlschlager T, Kunzler M, Aebi M, Narat M, Krizaj I, Anderlüh G, Sepčić K, Maček P (2013) Membrane cholesterol and sphingomyelin, and ostreolysin A are obligatory for pore-formation by a MACPF/CDC-like pore-forming protein, pleurotolysin B. *Biochimie* 95:1855–1864
72. Podack ER, Deyev V, Shiratsuchi M (2007) Pore formers of the immune system. *Adv Exp Med Biol* 598:325–341

73. Ponting CP (1999) Chlamydial homologues of the MACPF (MAC/perforin) domain. *Curr Biol* 9:R911–R913
74. Praper T, Beseničar MP, Istinič H, Podlesek Z, Metkar SS, Froelich CJ, Anderluh G (2010) Human perforin permeabilizing activity, but not binding to lipid membranes, is affected by pH. *Mol Immunol* 47:2492–2504
75. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301
76. Renault T, Faury N, Barbosa-Solomieu V, Moreau K (2011) Suppression subtractive hybridisation (SSH) and real time PCR reveal differential gene expression in the Pacific cupped oyster, *Crassostrea gigas*, challenged with Ostreid herpesvirus 1. *Develop Comparat Immunol* 35:725–735
77. Rosado CJ, Buckle AM, Law RH, Butcher RE, Kan WT, Bird CH, Ung K, Browne KA, Baran K, Bshattanyk-Puhlovich TA, Faux NG, Wong W, Porter CJ, Pike RN, Ellisdon AM, Pearce MC, Bottomley SP, Emsley J, Smith AI, Rossjohn J, Hartland EL, Voskoboinik I, Trapani JA, Bird PI, Dunstone MA, Whisstock JC (2007) A common fold mediates vertebrate defense and bacterial attack. *Science* 317:1548–1551
78. Rosado CJ, Kondos S, Bull TE, Kuiper MJ, Law RHP, Buckle AM, Voskoboinik I, Bird PI, Trapani JA, Whisstock JC, Dunstone MA (2008) The MACPF/CDC family of pore-forming toxins. *Cell Microbiol* 10:1765–1774
79. Rossjohn J, Feil SC, McKinstry WJ, Tweten RK, Parker MW (1997) Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. *Cell* 89:685–692
80. Satoh H, Oshiro N, Iwanaga S, Namikoshi M, Nagai H (2007) Characterization of PsTX-60B, a new membrane-attack complex/perforin (MACPF) family toxin, from the venomous sea anemone *Phyllo-discus semoni*. *Toxicon* 49:1208–1210
81. Schuster-Bockler B, Schultz J, Rahmann S (2004) HMM Logos for visualization of protein families. *BMC Bioinform* 5:7
82. Shibata T, Kudou M, Hoshi Y, Kudo A, Nanashima N, Miyairi K (2010) Isolation and characterization of a novel two-component hemolysin, erylysin A and B, from an edible mushroom, *Pleurotus eryngii*. *Toxicon* 56:1436–1442
83. Shinkai Y, Takio K, Okumura K (1988) Homology of perforin to the ninth component of complement (C9). *Nature* 334:525–527
84. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539
85. Slade DJ, Lovelace LL, Chruszcz M, Minor W, Lebioda L, Sodetz JM (2008) Crystal structure of the MACPF domain of human complement protein C8 alpha in complex with the C8 gamma subunit. *J Mol Biol* 379:331–342
86. Spilsbury K, O'Mara MA, Wu WM, Rowe PB, Symonds G, Takayama Y (1995) Isolation of a novel macrophage-specific gene by differential cDNA analysis. *Blood* 85:1620–1629
87. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 57:758–771
88. Stevens LM, Frohnhof HG, Klingler M, Nusslein-Volhard C (1990) Localized requirement for torso-like expression in follicle cells for development of terminal Anlagen of the *Drosophila embryo*. *Nature* 346:660–663
89. Sunagawa S, Wilson EC, Thaler M, Smith ML, Caruso C, Pringle JR, Weis VM, Medina M, Schwarz JA (2009) Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone *Aiptasia pallida* and its dinoflagellate endosymbiont. *BMC Genomics* 10:258
90. Tan TY, Ng SY, Thomas H, Chan BK (2006) *Arcanobacterium haemolyticum* bacteraemia and soft-tissue infections: case report and review of the literature. *J Infect* 53:e69–e74
91. Tardent P (1995) The cnidarian cnidocyte, a hightech cellular weaponry. *BioEssays* 17:351–362

92. Tomita T, Noguchi K, Mimuro H, Ukaji F, Ito K, Sugawara-Tomita N, Hashimoto Y (2004) Pleurotolysin, a novel sphingomyelin-specific two-component cytolysin from the edible mushroom *Pleurotus ostreatus*, assembles into a transmembrane pore complex. *J Biol Chem* 279:26975–26982
93. Tschopp J, Masson D, Stanley KK (1986) Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. *Nature* 322:831–834
94. Tweten RK (2005) Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. *Infect Immun* 73:6199–6209
95. Umen J, Heitman J (2013) Evolution of sex: mating rituals of a pre-metazoan. *Curr Biol* 23:R1006–R1008
96. van Beurden SJ, Bossers A, Voorbergen-Laarman MH, Haenen OL, Peters S, Abma-Henkens MH, Peeters BP, Rottier PJ, Engelsma MY (2010) Complete genome sequence and taxonomic position of anguillid herpesvirus 1. *J Gen Virol* 91:880–887
97. Vernatter J, Pirofski LA (2013) Current concepts in host-microbe interaction leading to pneumococcal pneumonia. *Curr Opin Infect Dis* 26:277–283
98. Voskoboinik I, Smyth MJ, Trapani JA (2006) Perforin-mediated target-cell death and immune homeostasis. *Nat Rev Immunol* 6:940–952
99. Voskoboinik I, Thia MC, Fletcher J, Ciccone A, Browne K, Smyth MJ, Trapani JA (2005) Calcium-dependent plasma membrane binding and cell lysis by perforin are mediated through its C2 domain: A critical role for aspartate residues 429, 435, 483, and 485 but not 491. *J Biol Chem* 280:8426–8434
100. Wang GD, Zhang KF, Zhang ZP, Zou ZH, Jia XW, Wang SH, Lin P, Wang YL (2008) Molecular cloning and responsive expression of macrophage expressed gene from small abalone *Haliotis diversicolor* supertexta. *Fish Shellfish Immunol* 24:346–359
101. Wang KJ, Ren HL, Xu DD, Cai L, Yang M (2008) Identification of the up-regulated expression genes in hemocytes of variously colored abalone (*Haliotis diversicolor* Reeve, 1846) challenged with bacteria. *Develop Comparat Immunol* 32:1326–1347
102. Weston AJ, Chung R, Dunlap WC, Morandini AC, Marques AC, Moura-da-Silva AM, Ward M, Padilla G, da Silva LF, Andreakis N, Long PF (2013) Proteomic characterisation of toxins isolated from nematocysts of the South Atlantic jellyfish *Olindias sambaquiensis*. *Toxicon* 71:11–17
103. Wiens M, Korzhnev M, Krasko A, Thakur NL, Perovic-Ottstadt S, Breter HJ, Ushijima H, Diehl-Seifert B, Müller IM, Müller WE (2005) Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway. Induction of a perforin-like molecule. *J Biol Chem* 280:27949–27959
104. Xu Q, Abdubek P, Astakhova T, Axelrod HL, Bakolitsa C, Cai X, Carlton D, Chen C, Chiu HJ, Clayton T, Das D, Deller MC, Duan L, Ellrott K, Farr CL, Feuerhelm J, Grant JC, Grzechnik A, Han GW, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Kozbial P, Krishna SS, Kumar A, Lam WW, Marciano D, Miller MD, Morse AT, Nigoghossian E, Nopakun A, Okach L, Puckett C, Reyes R, Tien HJ, Trame CB, van den Bedem H, Weekes D, Wooten T, Yeh A, Zhou J, Hodgson KO, Wooley J, Elsliger MA, Deacon AM, Godzik A, Lesley SA, Wilson IA (2010) Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont *Bacteroides thetaiotaomicron*. *Acta Crystallogr, Sect F: Struct Biol Cryst Commun* 66:1297–1305
105. Zbinden A, Mueller NJ, Tarr PE, Eich G, Schulthess B, Bahlmann AS, Keller PM, Bloemberg GV (2012) *Streptococcus tigurinus*, a novel member of the *Streptococcus mitis* group, causes invasive infections. *J Clin Microbiol* 50:2969–2973
106. Zheng C, Heintz N, Hatten ME (1996) CNS gene encoding astrotactin, which supports neuronal migration along glial fibers. *Science* 272:417–419



<http://www.springer.com/978-94-017-8880-9>

MACPF/CDC Proteins - Agents of Defence, Attack and
Invasion

Anderluh, G.; Gilbert, R. (Eds.)

2014, XI, 329 p. 57 illus., 52 illus. in color., Hardcover

ISBN: 978-94-017-8880-9