

Extending the Aerolysin Family: From Bacteria to Vertebrates

Pawel Szczesny^{1,2}, Ioan Iacovache³, Anna Muszewska¹, Krzysztof Ginalski⁴, F. Gisou van der Goot³, Marcin Grynberg^{1*}

1 Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland, **2** Institute of Experimental Plant Biology, University of Warsaw, Warsaw, Poland, **3** Faculty of Life Sciences, Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, **4** Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland

Abstract

A number of bacterial virulence factors have been observed to adopt structures similar to that of aerolysin, the principal toxin of *Aeromonas* species. However, a comprehensive description of architecture and structure of the aerolysin-like superfamily has not been determined. In this study, we define a more compact aerolysin-like domain – or aerolysin fold – and show that this domain is far more widely spread than anticipated since it can be found throughout kingdoms. The aerolysin-fold could be found in very diverse domain and functional contexts, although a toxic function could often be assigned. Due to this diversity, the borders of the superfamily could not be set on a sequence level. As a border-defining member, we therefore chose pXO2-60 – a protein from the pathogenic pXO2 plasmid of *Bacillus anthracis*. This fascinating protein, which harbors a unique ubiquitin-like fold domain at the C-terminus of the aerolysin-domain, nicely illustrates the diversity of the superfamily. Its putative role in the virulence of *B. anthracis* and its three dimensional model are discussed.

Citation: Szczesny P, Iacovache I, Muszewska A, Ginalski K, van der Goot FG, et al. (2011) Extending the Aerolysin Family: From Bacteria to Vertebrates. PLoS ONE 6(6): e20349. doi:10.1371/journal.pone.0020349

Editor: Sarah K. Highlander, Baylor College of Medicine, United States of America

Received: December 13, 2010; **Accepted:** April 29, 2011; **Published:** June 8, 2011

Copyright: © 2011 Szczesny et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: P.S. and K.G. acknowledge the support from Foundation for Polish Science (FNP). K.G. acknowledges the support from European Molecular Biology Organization (EMBO). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: greenb@ibb.waw.pl

Introduction

Among bacterial toxins, the most widely distributed and largest family is of pore-forming toxins (PFTs). Within the PFT family, the largest sub-family is currently composed of the Cholesterol Dependent Cytolysins (CDCs), which are produced by Gram-positive bacteria of *Bacillus*, *Clostridium*, *Streptococcus*, *Listeria* and *Arcanobacterium* genera. Interestingly, structural, but not sequence, similarity have recently been found with proteins produced by the vertebrate immune system, more specifically with the C8 and C9 components of the complement cascade and with perforin [1–3]. These latter proteins all contain a so called MACPF domain, and have a structure very similar to that of CDCs. Perforin like proteins are also found in Protozoans and in particular in the human parasites *Toxoplasma gondii* and *Plasmodium falciparum* [1].

A second family has started to emerge as new structures of PFTs are being discovered: the aerolysin family. Aerolysin is produced by *Aeromonas* species, but related proteins are present in both Gram-positive and Gram-negative bacteria, plants and eukaryotes. It shares high sequence identity with alpha-toxin from *Clostridium septicum* [4], as well as with enterolobin produced by the seeds of the Brazilian tree *Enterolobium contortiquum* [5,6]. The aerolysin family was subsequently extended, based on the analysis of conserved motifs [7], to hydralysins produced by *Cnidaria*, ε-toxin from *Clostridium perfringens* [8], a hemolytic lectin from the parasitic mushroom *Laetiporus sulphureus* [9] and structurally to parasporin-2 from *Bacillus thuringiensis* [10]. These similarities are reflected in protein structure classifications, such as SCOP or CATH, in which

aerolysins and epsilon toxins (in case of SCOP) belong to the same superfamily.

These various toxins are thought to share the same overall mode of action. The toxin is produced by the bacterium as a soluble protein that can either be a precursor, as is the case for aerolysin itself. The aerolysin precursor is called proaerolysin and activation consists in the proteolytic processing of a C-terminal peptide [11]. The soluble toxin diffuses towards its target cell where it binds via specific surface receptors, which are GPI-anchored proteins for aerolysin and *C. septicum* alpha-toxin, but likely to differ for other members of the family. Once receptor bound and proteolytically activated, the toxin undergoes circular polymerization, generating ring like structures that subsequently insert into the membrane and form a pore. While aerolysin and ε-toxin form heptamers, the stoichiometry might differ between members. The membrane-inserting portion represents only a small fraction of the entire protein. It is thought to cross the membrane in β-barrel conformation, as shown for the leukocidin-like fold alpha hemolysin of *Staphylococcus aureus*. The sequence of the transmembrane domain of aerolysin family pore-forming toxins is thus characterized by the alternating pattern of polar and hydrophobic residues, rather than any distinct sequence conservation [11].

Here we analyzed the similarities between various aerolysin like structures and defined the "aerolysin domain." Searching through multiple genomes for proteins containing this domain, we reveal that members of the superfamily can be found in all kingdoms. Sequence variability and fusion events with other domains suggest aerolysin core is widely used and may serve for diverse functions.

Results

Determination of the conserved core of the aerolysin domain

Following the procedure described in the method section, we identified 338 sequences in the NCBI non-redundant database that have a detectable similarity to the aerolysin domain. When performing the alignment of these sequences, a subset of which is shown in Fig. 1, we could identify the following conserved common core: two β -strands $\beta 1$ and $\beta 2$ followed by what in aerolysin corresponds to the membrane insertion β -hairpin, again followed by two additional β -strands $\beta 3$ and $\beta 4$. As observed in the structures (Fig. 2), the first β -strand usually does not adopt a β conformation along its whole length, which was denoted as separation into two individual β -strands numbered as $\beta 1a$ and $\beta 1b$. Interestingly, the best preserved pattern in the alignment was that of the insertion loop, where polar and non polar residues alternated in all depicted structures (Fig. 3) suggesting that this loop might, as for aerolysin, cross the membrane in a β -hairpin conformation upon oligomerization into a β -barrel structure.

In the aerolysin structure, these four β -strands and the insertion β -hairpin are part of the so-called domains 3 and 4, which respectively form a twisted anti-parallel beta-sheet with an amphipathic β -hairpin and a beta-sandwich (Fig. 2D). The polypeptide chain traverses the boundary between these 2 structural domains five times. Based on the analysis of the four

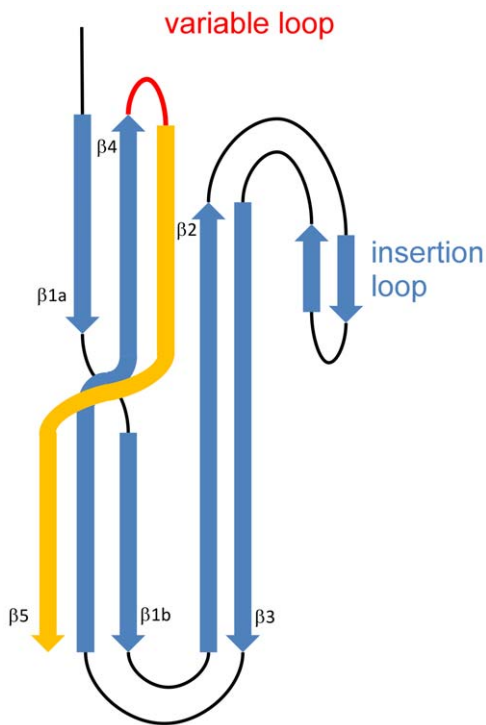


Figure 1. A simplified topology of a common core for aerolysin-like β -PFTs. Five β -strands that span the structure of common core of aerolysin-like toxins have been numbered from 1 to 5. The first β -strand in most cases does not maintain extended secondary structure through, therefore we divided it into two (denoted as $\beta 1a$ and $\beta 1b$). β -strands in so-called “insertion” loop are not strictly preserved, so we did not number them. The fifth strand (marked in orange) is not present in the alignment. Due to different lengths of the variable loop connecting $\beta 4$ and $\beta 5$, we were unable to precisely align the last strand (see text).

doi:10.1371/journal.pone.0020349.g001

known structures (Fig. 2), one would predict that the aerolysin domain consists of five β -strands, with an insertion loop between strands 2 and 3 and a variable loop between strands $\beta 4$ and $\beta 5$ which can range from a few residues, as in the hemolytic pore-forming lectin from *L. sulphureus*, to multiple secondary structure elements such as in aerolysin (Fig. 3).

Our sequence analysis however reveals that what we define as the “aerolysin domain” is shorter, by 100 residues, than the publicly available domain definition, explaining why establishing similarity between all aerolysin-like domains was not an easy task. Indeed even when we removed out the variable loop between $\beta 4$ and $\beta 5$, we were unable to construct a reliable alignment consisting of all five major β -strands. Reliable similarity (depicted by blue strands) ended at the $\beta 4$; further similarities along the sequence (between sequences of different $\beta 5$ s) were only due to biased, beta-strand-like, residue composition of that segment. This is probably due to different evolutionary pressure on that element: presence of a variable loop suggests that this strand is susceptible to phase shifts in the sequence, i.e. to shifts of amino acids along the β -strand as it needs to adapt to the presence of different insertions in the loop (although such shifts are unlikely to alter protein function). Additionally, conservation of this element might not be needed due to stable structural scaffold provided by hairpins $\beta 1:\beta 4$ and $\beta 2:\beta 3$. Thus, the middle strand of the fold is the necessary part of the topological unit, but it does not belong to highly conserved sequence core. We use the historical convention (domains 3 and 4) and our convention interchangeably. When talking about aerolysin-like structures as a whole, we use historical convention as it refers to two distinct structural units. However, when we refer to sequence-structure relationship in aerolysin superfamily, we prefer the $\beta 1$ – $\beta 5$ convention, as it better corresponds to relations between various structural elements.

Species distribution of the aerolysin domain

Proteins containing the aerolysin domain defined in here were found in all kingdoms of life (Table 1). Approximately 90% of the identified proteins were found in Proteobacteria, Firmicutes and Fungi. Examples of most interesting, not previously known family members are shown in Table 2, while the complete list is attached as Table S1.

Less than 30% of the proteins were annotated as hypothetical of putative proteins. The other 70% had a function confirmed by an experiment or assigned by similarity, and these were almost exclusively from the three groups mentioned before: Proteobacteria, Firmicutes and Fungi. A notable exception is a cytotoxin from *Pseudomonas* phage phiCTX, which has an experimentally confirmed toxic function [12]. The presence of a signal sequence was predicted in 139 proteins, around ~40% of the total, not surprisingly, since many sequences correspond to aerolysin fragments. Also almost 30 (~30%) sequences, with no assigned function, had a predicted signal peptide.

About half of the eukaryotic and one third of the prokaryotic species harboring aerolysin domain containing proteins are considered non-pathogenic. These bacterial species are aquatic and some pathogenic ones are of the same provenience, e.g. *Aeromonas salmonicida* and *Vibrio splendidus*. In Eukaryotes, except for fungi, cnidarians and the Brazilian fish *Thalassophryne nattereri*, there are no other predatory or pathogenic species. The archaeal Methanosarcinaceae, that possess aerolysin-like genes, are anaerobic methanogens.

Aerolysin in the context of other domains

In our analysis we identified several domain topologies, both in Prokaryotes and Eukaryotes (Fig. 4). In bacteria the majority of

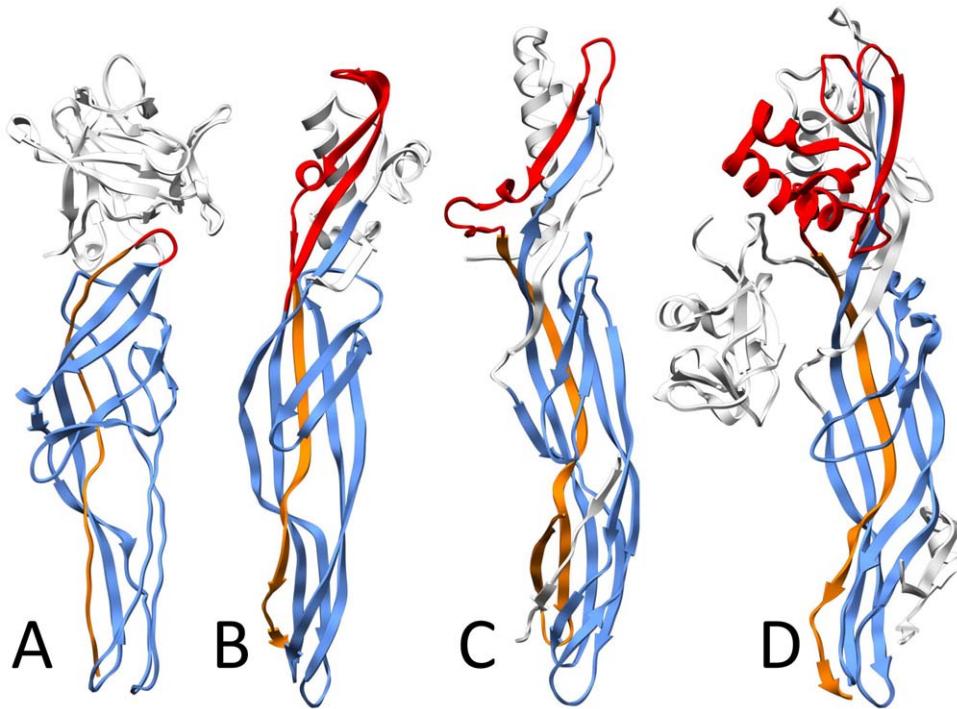


Figure 2. Topology of a common core for aerolysin-like β -PFTs mapped onto structures. **A:** hemolytic lectin, PDB code 1w3f; **B:** parasporin, PDB code 1ztb; **C:** epsilon-toxin, PDB code 1uyj; **D:** proaerolysin, PDB code 3c0n. Color scheme as on Fig. 1: blue - core with conserved sequence; red - variable loop; orange - the fifth, weakly conserved β -strand (see text). doi:10.1371/journal.pone.0020349.g002

aerolysins are either single pore-forming lobe or a pore-forming domain fused to an N-terminal C-lectin type structure, as found in aerolysin.

Aerolysin turns out to be amongst the most complex members. The aerolysin domain is extended N-terminally into a so-called domain 2, involved in binding to the glycan core of Glycosylphosphatidylinositol anchored proteins [13,14] (Fig. 2D). Fused to this GPI-binding domain is an N-terminal lectin domain [15] involved in binding to N-linked sugars present on the polypeptide moiety of GPI-anchored proteins [16].

We also identified two novel bacterial fusions. The first is the *Bacteroides uniformis* ATCC 8492 protein (gi: 160892167) from a

recently sequenced *B. uniformis* genome obtained from the human gut (data obtained from the Human Microbiome Project). The unique feature of this protein is the presence of a prokaryotic membrane lipoprotein lipid attachment site at the N terminus. In Prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached [17]. It seems then that *Bacteroides* protein is specifically bound to the plasma membrane, which is proven by the PSORT cell location program that predicts this aerolysin-like protein to be extracellular. The

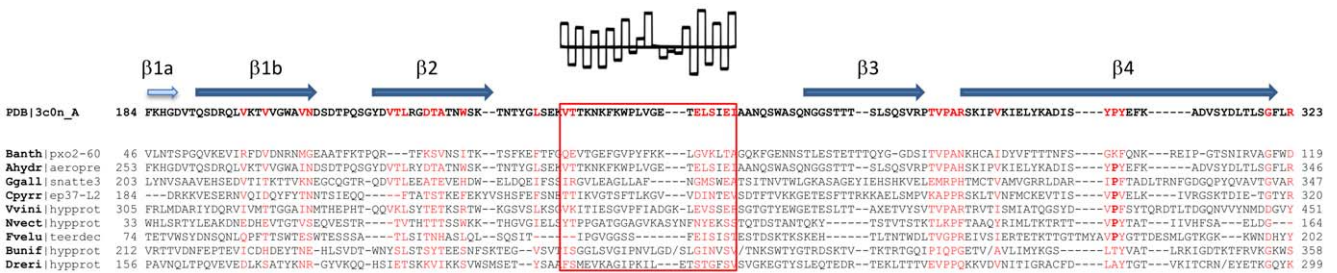


Figure 3. Alignment of the conserved common core of aerolysin like β -PFTs. Beta strands corresponding to the structure of proaerolysin (3c0n) and previously shown topology are marked above the alignment. The B1a is marked in light blue as only part of it seem to hit the alignment. A putative “insertion loop” is marked with red box. This fragment has calculated an average hydrophobicity according to Kyte-Doolittle scale in each column and shown above the alignment. Abbreviations: **Banth**|pXO2-60 - pXO2-60 protein, gi: 51704196 [*Bacillus anthracis*]; **PDB**|3c0n_A - proaerolysin, PDB structure 3c0n, chain A [*Aeromonas hydrophila*]; **Ahydr**|aeropre - aerolysin-3 precursor, gi: 2501301 [*Aeromonas hydrophila*]; **Ggall**|snatte3 - protein similar to nattering-3, gi: 118105776 [*Gallus gallus*]; **Cpyrr**|ep37-L2 - ep37-L2 protein, gi: 2339973, [*Cynops pyrrhogaster*]; **Vvini**|hypprot - hypothetical protein, gi: 147838248 [*Vitis vinifera*]; **Nvect**|hypprot - hypothetical protein, gi: 156349328 [*Nematostella vectenensis*]; **Fvelu**|teerdece - TEER-decreasing protein, gi: 3551186 [*Flammulina velutipes*]; **Bunif**|hypprot - hypothetical protein, gi: 160892167 [*Bacteroidetes uniformis*]; **Drieri**|hypprot - hypothetical protein, gi: 162139040 [*Danio rerio*]. doi:10.1371/journal.pone.0020349.g003

Table 1. Species distribution of proteins belonging to aerolysin-like superfamily with a number of proteins that have assigned a function.

Phylum	No. of proteins	No. of proteins with assigned function
Bacteria:	225	193
- Actinobacteria	7	0
- Bacteroidetes/Chlorobi	3	0
- Cyanobacteria	1	0
- Firmicutes	61	53
- Proteobacteria	151	140
- Tenericutes	2	0
Archaea:	3	0
- Euryarcheota	3	0
Eukaryota:	109	49
- Amoebozoa	1	0
- Fungi/Metazoa	100	48
- Viridiplantae	8	1
Viruses:	1	1
- Caudovirales	1	1
TOTAL:	338	243

The functional annotation could be inferred from sequence similarity. Text annotation is not based on experimental evidence ("predicted toxin" or "similar to aerolysin" count as functional annotation; "putative transmembrane protein", "hypothetical protein" or "predicted protein" are treated as no functional assignment).

doi:10.1371/journal.pone.0020349.t001

other case is the *Bacillus anthracis* protein pXO2-60 (gi: 10956450; alternative names: BXB0074/GBAA_pXO2_0074) located on the pXO2 plasmid, known to be involved in virulence [18,19]. The aerolysin domain of this protein is most similar to the *Clostridium perfringens* epsilon-toxin. While it harbors no N-terminal fusion, it carries a β -grasp fold at the C-terminus. Importantly, the *B. anthracis* gene has been shown to be highly regulated by the general virulence regulator, AtxA [20] - it is up-regulated an order of magnitude more than the anthrax toxin and capsule genes. Together with the finding that the pXO2 plasmid is indispensable for anthrax virulence [21], these data suggest that pXO2-60 protein may play an important role in *B. anthracis* cytotoxicity.

Eukaryotic proteins containing aerolysin domains can also be found in fungi. They are present in several species of pathogenic *Basidiomycota* and three *Ascomycota*. The common theme for all of them is arboreal/plant pathogenesis (*Laetiporus sulphureus*, *Agrocybe chaxingu*, *Pleurotus eryngii* var. *Ferule*, *Ganoderma lucidum*, *Flammulina velutipes*, *Hericium erinaceum*). The only exceptions from the rule are members of the *Coprinus* genus (*C. cinerea* and *C. comatus*) that are saprophytes. The presence of a toxin in this genus suggests that these fungi may in fact be pathogenic; another hypothesis is that the aerolysin domain is used for defense against other pathogens or when hunting. For instance, *C. comatus* is known to attack nematodes [22,23]. Our analysis suggests it is indeed a pore-forming protein belonging to the aerolysin family. Other aerolysins from *Basidiomycota* were identified in a 'fungal pathogens of plants' sequencing project (Fu, M., Wu, Z., Lin, Q. and Xie, L., unpublished).

The eukaryotic aerolysin family members that seem to be most similar in structure (and function) to canonical bacterial toxins are composed of an N-terminal lectin domain followed by the pore-forming domain. We detected three new combinations of this type.

In the grapevine *Vitis vinifera* and the wheat *Triticum aestivum* two N-terminal agglutinin and a C-terminal aerolysin-like domains constitute a protein that in the latter organism was shown to be involved in defense against insects and maybe other pathogens [24,25]. Agglutinin from *Amaranthus caudatus* (amaranthin) is a lectin from the ancient South American crop, amaranth grain. Although its biological function is unknown, it can agglutinate A, B and O red blood cells, and has a carbohydrate-binding site that is specific for the methyl-glycoside of the T-antigen found linked to serine or threonine residues of cell surface glycoproteins [26]. The structure of this lectin is a beta-trefoil fold and forms a homodimer with each monomer composed of two domains [27,28]. The authors suggest that HFR-2, the aerolysin-like protein from wheat, may normally function in defense against certain insects or pathogens. They also propose that as virulent insects (larvae) manipulate the physiology of the susceptible host; the HFR-2 protein may be forced to insert in plant cell membranes at the larvae feeding sites and by forming pores provides water, ions and other small nutritive molecules to the developing larvae [25]. We found another lectin-like fusion in two fish: *Danio rerio* and *Salmo salar* (e.g., gi: 162139040 and 209732252, respectively). Most probably the same topology would be found in other fish species when sequenced. Amazingly, *D. rerio* possesses as many as 16 copies of the gene encoding the same aerolysin-like protein. The last lectin-aerolysin topology is quite unusual and may be a virtual translation of a false CDS (gi: 148223884). We discovered it in the African clawed frog *Xenopus laevis*, but it is not present in its close relative, *Xenopus tropicalis*. This protein contains a tandem of immunoglobulin-like domains at the N terminus, followed by a transmembrane motif and a duplication of large fragments located at the C terminus. The latter is composed of a tachylectin-2-like domain and an aerolysin-like domain (gi: 165970884). In *X. tropicalis*, immunoglobulin domains are in a different ORF than the lectin-toxin tandem (gi: 171847007 and 73853870, respectively). The Japanese horseshoe crab *Tachypleus tridentatus* tachylectin-2 binds specifically to *N*-acetylglucosamine and *N*-acetylgalactosamine and is a part of the innate immunity host defense system of this crab [29]. Tachylectin-2 is a protein displaying a five-bladed β -propeller structure. Tachylectin-2 exhibits five virtually identical binding sites, one in each β -sheet. The high number of binding sites within a single polypeptide chain strongly suggests the recognition of carbohydrate surface structures, possibly on microbial pathogens with a fairly high ligand density [30], however, this notion has not been proven yet. Similarly, immunoglobulin domains that are present at the N terminus are also involved in innate immunity [31]. Using TMHMM2.0 program, we show that the N-terminal portion of the protein is located inside the cell. Activity of immunoglobulin-like domains in cytoplasm have already been shown in other systems [32].

In the eukaryotic family of aerolysins we identified two exceptional fusions. First, it is a crystallin domain tandem followed by an aerolysin domain. Such topology is found in the Japanese newt *Cynops pyrrhogaster* epidermis-specific protein EP37. This organism is armed with several homologues that are present in skin, gastric epithelium and fundic glands of an adult newt and in the swimming larva [33–36], however the authors did not identify the C-terminal toxin domain. N-terminal domains are non-lens β/γ -crystallin domains that have a Greek key structure. Crystallins were so named when they were recognized as the proteins that provide the crystallin lens of the vertebrate eye with its indispensable transparency and unique refractive properties [37]. Because lens cells live as long as their host, crystallins also have to live long. Crystallin ancestors can be tracked to bacteria and can be used for diverse purposes, e.g. as enzymes, toxins or antistress

Table 2. Representatives of previously unknown members of the aerolysin-like superfamily.

Annotation	GI number(s)	Species
hypothetical protein Bsph_1071	169826655	<i>Lysinibacillus sphaericus</i>
conserved hypothetical protein	169338027	<i>Clostridium botulinum</i>
hypothetical protein SvirD4_00927	256799387	<i>Streptomyces viridochromogenes</i>
hypothetical protein pxo2_60	10956450	<i>Bacillus anthracis</i>
hypothetical protein HborDRAFT_3048	227882342	<i>Halogeometricum borinquense</i>
hypothetical protein HCH_03563	83646295	<i>Hahella chejuensis</i>
hypothetical protein MmarC7_0032	150401960	<i>Methanococcus maripaludis</i>
predicted protein	154291554	<i>Botryotinia fuckeliana</i>
conserved hypothetical protein	255557038	<i>Ricinus communis</i>
predicted protein	242210318	<i>Postia placenta</i>
unnamed protein product	171678343	<i>Podospora anserina</i>
Predicted proteins	169848522, 169850135, 169862541,	<i>Coprinopsis cinerea</i>
hypothetical protein	256084754	<i>Schistosoma mansoni</i>
hypothetical proteins IscW	241709479, 241644210, 241838790, 241607979, 242002442	<i>Ixodes scapularis</i>
yolk proteins 1 and 2	20269940, 20269942	<i>Hyphantria cunea</i>
anarchy 1	67848428	<i>Apis mellifera</i>
Hypothetical proteins	165972443, 160333564, 156739307, 61651762, 189520990, 160333607, 125820446, 125820386, 189520988, 125820390, 189520992, 125820369, 160333636, 162139040, 125820384, 194579003	<i>Danio rerio</i>
Natterin	75571591, 75571593, 75571592, 75571590	<i>Thalassophryne nattereri</i>
Natterin-like protein	225714938	<i>Esox lucius</i>
Natterin-like protein	213515472	<i>Salmo salar</i>
hypothetical proteins	148223884, 168693479, 195972837	<i>Xenopus laevis</i>
hypothetical protein BRAFLDRAFT_68290	219416017	<i>Branchiostoma floridae</i>
Ep37, gep proteins	2339971, 1169472, 2339973, 6069419, 2339969	<i>Cynops pyrrhogaster</i>
PREDICTED: similar to ep37-L2	149419178	<i>Ornithorhynchus anatinus</i>
PREDICTED: similar to natterin 3	118105776	<i>Gallus gallus</i>

Names represent original NCBI annotations.
doi:10.1371/journal.pone.0020349.t002

proteins [38,39]. The topology found in EP37 is in accordance with the canonical crystallin proteins with a tandem of crystallin motifs. In solution β -crystallins are known to form dimers whereas γ -crystallins are monomeric. β -crystallins assembly into higher order complexes: tetramers and the likely assembly of this protein in the lens is that of higher heterooligomers [40]. This conformation fits the oligomeric nature of aerolysin.

The second exceptional topology can be found in Brazilian venomous fish *Thalassophryne nattereri* proteins named natterins [e.g. gi: 75571591]. There are 5 known paralogues of this toxin with 4 full-length proteins (natterins 1–4). These are known to cause nociception and oedema [41]. Fractions containing natterin were positively tested for the kininogenase activity, but no similarity to aerolysin toxins has been proposed or shown. The N-terminal region of natterins is occupied by two DM9 domains first discovered in *Drosophila* [42], but with no function ascribed. We have discovered the similarity of twin DM9 to a half of the MFP2 protein from pig roundworm *Ascaris suum*. MFP2 increases the rate of movement *in vitro* in *Ascaris* sperm cells and appears to function by increasing the rate of the major sperm protein (MSP) polymerization, possibly in a manner analogous to formins in actin-based systems [43,44]. Authors suggest MFP2 could increase the rate of MSP polymerization by increasing the nucleation rate, or by increasing the amount of polymerization competent MSP, or

by decreasing the termination rate. MSP provides sperm locomotion by the assembly and disassembly of filaments and replaces actin in the nematode filament structure [45,46]. Function of DM9 in natterins is unknown. We discovered a similar protein from the red jungle fowl (*Gallus gallus*) (gi: 118105776), with no function assigned to date. Interestingly, gene coding for the chicken protein overlaps head-to-tail with a chromatin modifying protein 2A (CHMP2A) (gi: 124249308). This DNA region is not assigned to any *G. gallus* chromosomes.

Architecture of the aerolysin superfamily

We made a cluster map of sequences of the aerolysin superfamily (see Fig. 5) that included all hits to our profile within E-value of 100 as assessed with HMMER3 (<http://hmmer.janelia.org/>) (see Methods). Major groups of aerolysin-like sequences form distinct clusters and the largest cluster is understandably composed of aerolysins and alpha toxins. Other proteins are spread across several small groups, which nevertheless form a distinct cluster within the core. Several clusters are formed by toxins with known structure and recognized similarity to aerolysins, such as ETX or Mtx2 toxins, hydralysins and parasporins or insect yolk-like proteins (Fig. 5). Other groups are mainly composed of proteins of similar phylogenetic origin, which is not surprising given that they are identified based on similarity (see Methods). Distinct

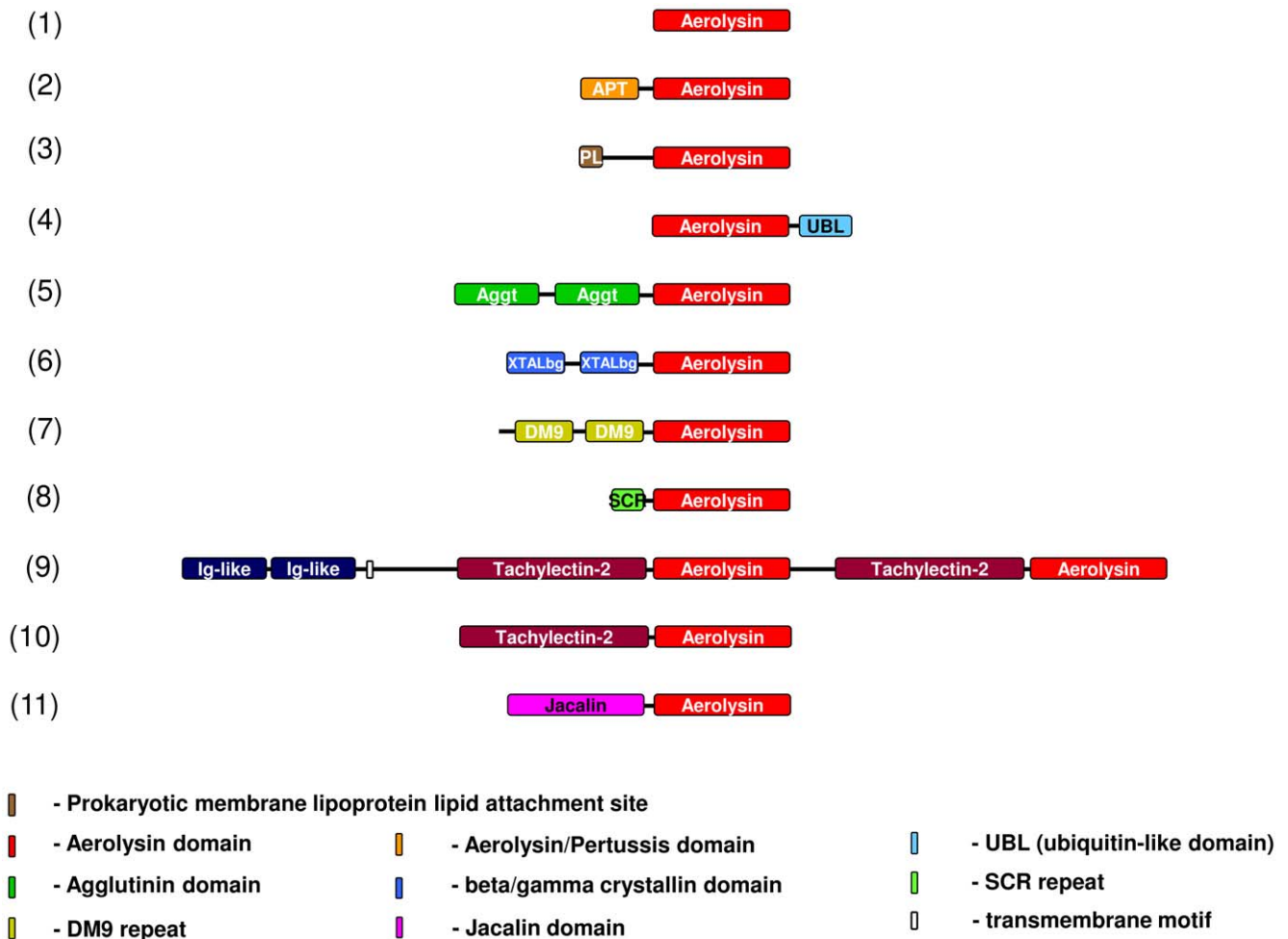


Figure 4. Domain organization of the Aerolysin/ETX pore-forming superfamily. Proteins 1–4 are present in bacteria, 5–11 in Eukaryotes. Domain organization representatives: (1) alpha-toxin, gi:452163 [*Clostridium septicum*]; (2) Hemolysin-3, gi:2501300 [*Aeromonas hydrophila*]; (3) hypothetical protein BACUNI_04630, gi:160892167 [*Bacteroides uniformis* ATCC 8492]; (4) pXO2-60 protein, gi: 10956450 [*Bacillus anthracis*]; (5) hypothetical protein CAN71829, gi:147838248 [*Vitis vinifera*]; (6) ep37-L2, gi:2339973 [*Cynops pyrrhogaster*]; (7) Natterin-3 precursor, gi:75571591 [*Thalassophryne nattereri*]; (8) hypothetical protein NEMVEDRAFT_v1g221281, gi:156349328 [*Nematostella vectensis*]; (9) hypothetical protein LOC494812, gi:148223884 [*Xenopus laevis*]; (10) hypothetical protein LOC613112, gi:73853870 [*Xenopus tropicalis*]; (11) hypothetical protein LOC568775, gi:162139040 [*Danio rerio*]. doi:10.1371/journal.pone.0020349.g004

clusters are formed by proteins from *Nematostella*, *Cynops*, *Ixodes* or *Danio rerio*. Also plant and fungi proteins tend to form separate clusters. All these clusters form an easily recognizable supergroup with many interconnections (Fig. 5).

Interesting small groups are formed at the boundaries of the core. These distant clusters are formed by hemolytic lectins, known to have structure similar to aerolysins and ETX toxins; lysenins, pXO2-60 and anthrax protective antigen. The last one has a clearly different structure from aerolysins despite sharing a very similar mode of pore-formation [47] and as such does not belong to the aerolysin superfamily. Both the distance and number of connections were not sufficient to distinguish between a false-positive case, such as protective antigen and a true positive such as hemolytic lectin. Therefore we decided to fall back on E-value scores and assume that pXO2-60 is a border-defining member of the aerolysin superfamily.

Structural model of pXO2-60 (aerub)

pXO2-60 is composed of three elements: signal peptide, aerolysin-domain and a β -grasp fold, also found in ubiquitin. The signal peptide is predicted to be cleaved at position 29 (AAA-

BBB). We have modeled the 3D structure of the exported part using two template structures: luyj (*Clostridium perfringens* epsilon toxin) and lubi (ubiquitin protein).

The aerolysin domain of pXO2-60 contains the complete core of the superfamily with small additions in front of β 1 and after β 5 (Fig. 6). These fragments adopt the β -strand conformation and they extend two β -sandwich structures originally named domain III and domain IV in first analyses of aerolysins structures. The small β -strand is positioned between the hairpin of β 4: β 5 and the insertion loop. Two small β -strands that follow β 5 extend β -sheets of β 4- β 5- β 1 and β 3- β 4, respectively. The variable loop present in other members of the superfamily is only a few residues long in pXO2-60.

Following aerolysin domain there is a domain that most likely adopts a β -grasp fold. We call it an ubiquitin fold domain because ubiquitin is the most prominent protein with such a scaffold. The domain was modeled on a template of ubiquitin protein (lubi, see Fig. 7), however β -grasp fold is found in proteins of many different functions, such as translation initiation factors, immunoglobulin-binding proteins, glutamine synthetases, ferredoxin-like proteins or GTP binding proteins. It is also found at the C-terminus of

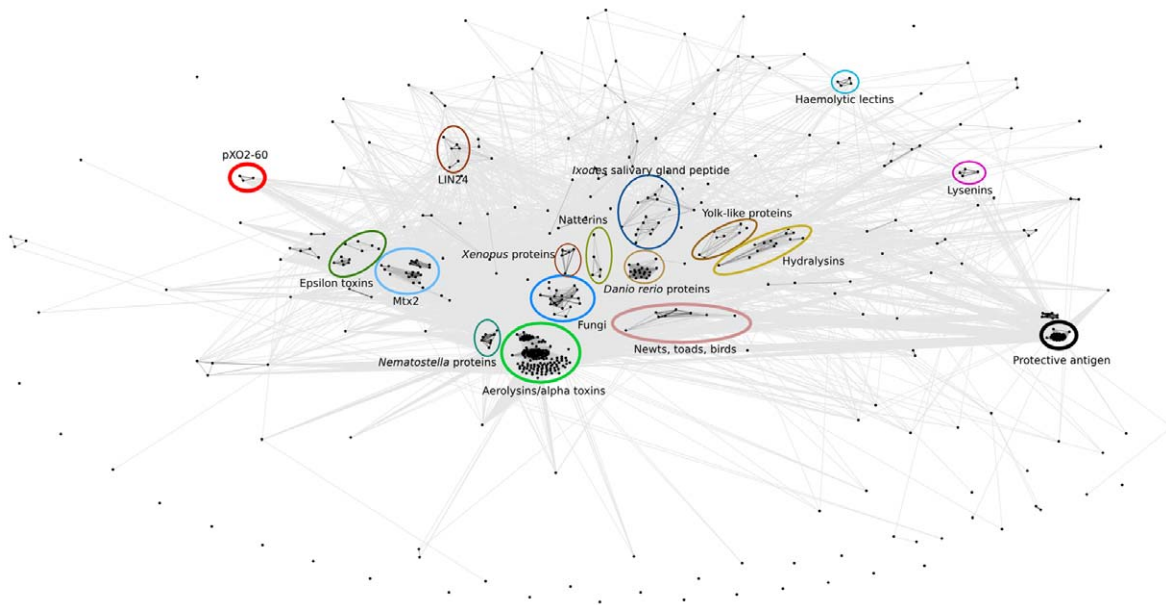


Figure 5. Cluster map of aerolysin superfamily. All sequenced from nr database similar to aerolysin core profile at E-value of 100 and better were used including false positives. Major groups of proteins are highlighted with ovals and described. Presence of a clear false positive (protective antigen) is highlighted to indicate difficulties in assessing the border of the superfamily. doi:10.1371/journal.pone.0020349.g005

staphylococcal/streptococcal toxins, where at least in some cases it mediates dimerization [48]. High structural similarity of proteins of diverse functions and sequences complicates the assignment of the C-terminal domain of pXO2-60 to any of these groups. Sequence identity at 20% and similarity at 40% over ca. 80 residues between this fragment and the closest template are not enough to clearly classify this domain within specific proteins with the β -grasp fold.

Model of pXO2-60 exemplifies the diversity of members of the aerolysin superfamily. Its compact structure, with no insertions in the variable loop, represents a new type on the structural map of the aerolysin-like fold. Presence of the ubiquitin-like domain at the C-terminus is also unique – pXO2-60 is the only protein from the superfamily with such a fusion.

Discussion

In this study we have shown that aerolysin-like pore forming toxins share not only structural features, but also their similarity is recognizable at the sequence level. Additionally, we were able to define the real common theme for the family and have shown that it is much smaller and compact than previously thought. That information helped us to significantly increase the coverage of the family of aerolysin-domain containing proteins and find a large number of previously unknown members.

Analysis of the common structural core has shown that the functionally important unit may be as small as the amphipathic loop and the five β -strands. The fifth C-terminal β -strand that connects two halves of the complex seems to be structurally required, however its sequence does not seem to be conserved. This may be due to a variable loop immediately preceding this segment that imposes constantly changing sequence onto this segment. Other explanation could be that this β -strand must have only general properties of an extended structure because structural constraints of the β -sheet will likely keep it in place.

It is important to note that an attempt to define a common core for aerolysin-like proteins together with identification of several new members of the superfamily has been published shortly before

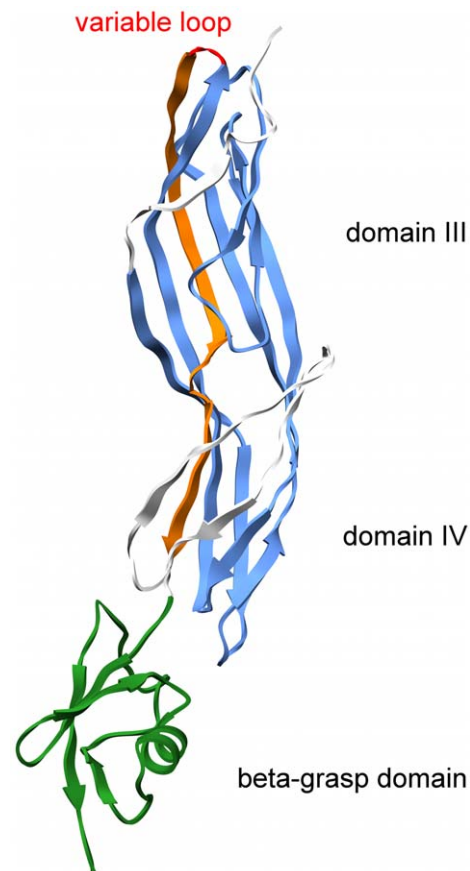


Figure 6. Structural model of pXO2-60. Color scheme: blue - core with conserved sequence; red – variable loop; orange – the fifth, weakly conserved β -strand; green – β -grasp domain modeled on ubiquitin protein. Two parts of the conserved core that form b-sandwich structures are marked as domain III and IV (historically). Variable loop that connects β 4 and weakly conserved β 5 is barely visible because of its short length. doi:10.1371/journal.pone.0020349.g006


```

pxO2-60 237 IKIHLGFTNGDDYYLKPE-----EVKEKILDRVYLDD 268
          :  ::  |  :  |  |  ||  ||  |  :
PDB:1UBI 1 MQIFVKTLTGKTITLVEVPSDTIENVKAKIQDKEGIPP 38

pxO2-60 269 DKKKVY----VLE---DAAEFNGQSGIDLSIKVIDRTN 299
          |  ::  ||  :  :  |  |  :  ::  |
PDB:1UBI 39 DQQRLLIFAGKQLEDGRTLSDYNIQK--ESTLHLVLR 74

```

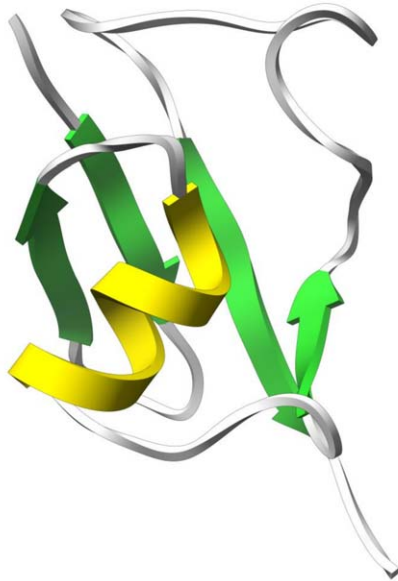


Figure 7. Alignment between ubiquitin fold domain of pXO2-60 and ubiquitin protein 1ubi. Green segments correspond to beta conformation, while orange to alpha conformation. Model with the same color scheme is shown below alignment. doi:10.1371/journal.pone.0020349.g007

submitting this manuscript in a review concerning *Laetiporus sulphureus* lectin and the aerolysin family [49]. While some results of that study significantly overlap with ours, its authors use traditional approach to describe the structural organization of aerolysin core (i.e. two domains).

Our analysis suggests that both prokaryotic and eukaryotic family members can be used either for attack or for defense. It is obvious in such species as bacterial pathogens *Clostridium* and *Aeromonas* or the eukaryotic Brazilian fish *Thalassophryne nattereri*. Cnidarians, newts, most probably non-pathogenic bacteria and other species may use it to defend against attackers. An uncertain point is the presence of aerolysins in some previously thought-to-be saprophytic or mycorrhizal fungi (*Coprinopsis cinerea*, *Laccaria bicolor*): are these toxins used for defense or are important in other functions (e.g. mycorrhiza)? Most puzzling case is of the honeybee anarchy 1 aerolysin-like protein, ‘a genetic locus for worker sterility in a social insect’ (cited after annotation in gi: 67848428). Its function seems to be unrelated to either attack or defense. Does it play a role in a physiological process? Is it an element in sterility regulation?

Our analysis also allowed the identification of previously unknown domain topologies. In Eukaryotes unexpected fusions include crystallin and DM9 domains. We hypothesize their functions to be similar to lectin domains, which is the facilitation of cell membrane binding. In bacteria, novel fusions were identified for proteins expressed in *Bacillus anthracis* and *Bacteroides uniformis*. The latter has an N-terminal membrane lipoprotein lipid attachment site that suggests a unique mechanism of cell attachment. *B. anthracis* pXO2-60 protein is the only one to possess a C-terminal ubiquitin fold domain. It will

be of interest to determine whether this ubiquitin fold has a role on the outside of the target cell, or whether it is at some point translocated, possibly through the channel formed by the aerolysin-domain in the cytoplasm of the host cell. The determination of a precise structure, or function, of this domain would help to qualify it to anything in between a classical ubiquitin, a mechanical cap to control the aerolysin pore, an immunoglobulin-binding (Ig-binding) domain or a ligand binding domain [50–52]. Also it would be interesting to verify if the N-terminal segment of pXO2-60 positioned between hairpin b4–b5 and the insertion loop has similar inactivation function to parasporins where the N-terminus forms a β -sheet with the insertion loop blocking its rearrangement and/or movement.

An interesting far-fetching hypothesis may be driven from experiments by Welkos et al. and Chand et al. that show that the lack of protective antigen does not significantly diminish the *B. anthracis* virulence in the mouse model [53–55]. The presence of protective capsule genes on the pXO2 plasmid may be insufficient to kill, therefore the presence of another toxin on this plasmid suggests its involvement in virulence. We hope future experiments will soon test our supposition.

Methods

To identify aerolysin-like sequences, we made queries at the National Center for Bioinformatics Information (NCBI) using BLAST and PSI-BLAST [56,57] on the non-redundant database (nr). Because the sensitivity of these tools is low, we have also used methods based on Hidden Markov Models. In the first step we

compared the profile built on the alignment of the conserved aerolysin domain, available at the PFAM database, to the nr database with HMMER (citation) implemented as FastHMMER tool at MPI Toolkit website [58]. Based on results from this step, we have built a manual profile containing all aerolysin-like sequences and used it again against the nr database, however this time using the newly released HMMER3 package (<http://hmmer.janelia.org/>). We have employed an E-value of 10 in order to identify as many hits as possible and, then, we have manually analyzed the distribution to assess which threshold to apply to the results. This was necessary, as the software was still in alpha development stage. The twilight-zone hits were analyzed with HHpred [59] to obtain an external confirmation of the presence of the aerolysin domain. As a result, we have applied an E-value of 0.001 as a final threshold to hits from HMMER3. Additionally, all low-complexity sequences were removed from the final list; sequences of these hits were extracted using the Entrez service (www.ncbi.nlm.nih.gov/Entrez/). A sequence alignment was generated on the basis of pairwise alignments of the sequences and PFAM aerolysin domain using HMM-HMM comparison tool HHpred [59]. These were later manually adjusted to satisfy hydrophobic pattern of beta-strands.

For cluster map we have extracted all sequences from NCBI non-redundant database at an E-value of 100 using the hmmsearch program from HMMER3 package; that set included false positives. All vs. all comparisons were done with the jackhammer application from the same package with default of 5 iterations.

Domain annotation for selected proteins was obtained using combined effort of Interpro [60], SMART [61], Pfam [62], HHpred [59], Bioinfo.pl Metaserver [63] and FFAS03 [64] tools followed by a manual verification. No single E-value threshold was

used in all cases. Presence of signal peptide was predicted using SignalP 3.0 server [65]. We have also searched for transmembrane helices with TMHMM2.0 server [66].

Both full Aerub sequence as well as single N- and C-terminal domain regions were analyzed with 3D-Jury server [63] to identify optimal templates and to derive reliable sequence-structure mappings using consensus alignment approach and 3D assessment [67]. The three dimensional model of Aerub was built with the MODELLER program [68] using as templates the structure of *Clostridium perfringens* epsilon-toxin (PDB code 1uyj) [8] for N-terminal Aerolysin/ETX pore-forming domain and the structure of human ubiquitin (PDB code 1ubi) [69] for C-terminal beta-grasp fold domain.

Figures of structures were prepared using Chimera [70].

Supporting Information

Table S1 List of proteins identified as members of the aerolysin superfamily containing Uniprot ID, protein name and species.

(XLS)

Acknowledgments

We are grateful to Dr Paulo Nuin for careful reading of the manuscript.

Author Contributions

Conceived and designed the experiments: PS MG. Performed the experiments: PS KG AM MG. Analyzed the data: PS FGvdG II MG. Contributed reagents/materials/analysis tools: KG. Wrote the paper: PS MG.

References

- Bischofberger M, Gonzalez MR, van der Goot FG (2009) Membrane injury by pore-forming proteins. *Curr Opin Cell Biol* 21: 589–595.
- Rosado CJ, Buckle AM, Law RHP, Butcher RE, Kan W, et al. (2007) A common fold mediates vertebrate defense and bacterial attack. *Science* 317: 1548–1551.
- Hadders MA, Beringer DX, Gros P (2007) Structure of C8alpha-MACPF reveals mechanism of membrane attack in complement immune defense. *Science* 317: 1552–1554.
- Ballard J, Sokolov Y, Yuan WL, Kagan BL, Tweten RK (1993) Activation and mechanism of *Clostridium septicum* alpha toxin. *Mol Microbiol* 10: 627–634.
- Bittencourt SET, Silva LP, Azevedo RB, Cunha RB, Lima CMR, et al. (2003) The plant cytolytic protein enterolobin assumes a dimeric structure in solution. *FEBS Lett* 549: 47–51.
- Sousa MV, Richardson M, Fontes W, Morhy L (1994) Homology between the seed cytolytic enterolobin and bacterial aerolysins. *J Protein Chem* 13: 659–667.
- Sher D, Fishman Y, Zhang M, Lebendiker M, Gaathon A, et al. (2005) Hydralysins, a new category of beta-pore-forming toxins in cnidaria. *J Biol Chem* 280: 22847–22855.
- Cole AR, Gibert M, Popoff M, Moss DS, Titball RW, et al. (2004) *Clostridium perfringens* epsilon-toxin shows structural similarity to the pore-forming toxin aerolysin. *Nat Struct Mol Biol* 11: 797–798.
- Mancheño JM, Tateno H, Goldstein IJ, Martínez-Ripoll M, Hermoso JA (2005) Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J Biol Chem* 280: 17251–17259.
- Akiba T, Abe Y, Kitada S, Kusaka Y, Ito A, et al. (2009) Crystal structure of the parasporin-2 *Bacillus thuringiensis* toxin that recognizes cancer cells. *J Mol Biol* 386: 121–133.
- Iacovache I, van der Goot FG, Pernot L (2008) Pore formation: an ancient yet complex form of attack. *Biochim Biophys Acta* 1778: 1611–1623.
- Nakayama K, Kanaya S, Ohnishi M, Terawaki Y, Hayashi T (1999) The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. *Mol Microbiol* 31: 399–419.
- Diep DB, Nelson KL, Raja SM, Pleshak EN, Buckley JT (1998) Glycosylphosphatidylinositol anchors of membrane glycoproteins are binding determinants for the channel-forming toxin aerolysin. *J Biol Chem* 273: 2355–2360.
- Gordon VM, Nelson KL, Buckley JT, Stevens VL, Tweten RK, et al. (1999) *Clostridium septicum* alpha toxin uses glycosylphosphatidylinositol-anchored protein receptors. *J Biol Chem* 274: 27274–27280.
- Rosjohn J, Buckley JT, Hazes B, Murzin AG, Read RJ, et al. (1997) Aerolysin and pertussis toxin share a common receptor-binding domain. *EMBO J* 16: 3426–3434.
- Hong Y, Ohishi K, Inoue N, Kang JY, Shime H, et al. (2002) Requirement of N-glycan on GPI-anchored proteins for efficient binding of aerolysin but not *Clostridium septicum* alpha-toxin. *EMBO J* 21: 5047–5056.
- Hayashi S, Wu HC (1990) Lipoproteins in bacteria. *J Bioenerg Biomembr* 22: 451–471.
- Green BD, Battisti L, Koehler TM, Thorne CB, Ivins BE (1985) Demonstration of a capsule plasmid in *Bacillus anthracis*. *Infect Immun* 49: 291–297.
- Uchida I, Sekizaki T, Hashimoto K, Terakado N (1985) Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. *J Gen Microbiol* 131: 363–367.
- Bourgogne A, Drysdale M, Hilsenbeck SG, Peterson SN, Koehler TM (2003) Global effects of virulence gene regulators in a *Bacillus anthracis* strain with both virulence plasmids. *Infect Immun* 71: 2736–2743.
- Heninger S, Drysdale M, Lovchik J, Hutt J, Lipscomb MF, et al. (2006) Toxin-deficient mutants of *Bacillus anthracis* are lethal in a murine model for pulmonary anthrax. *Infect Immun* 74: 6067–6074.
- Luo H, Mo M, Huang X, Li X, Zhang K (2004) *Coprinus comatus*: A basidiomycete fungus forms novel spiny structures and infects nematode. *Mycologia* 96: 1218–1224.
- Tomita T, Ishikawa D, Noguchi T, Katayama E, Hashimoto Y (1998) Assembly of flammutoxin, a cytolytic protein from the edible mushroom *Flammulina velutipes*, into a pore-forming ring-shaped oligomer on the target cell. *Biochem J* 333(Pt 1): 129–137.
- Velasco R, Zharkikh A, Troggo M, Cartwright DA, Cestaro A, et al. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2: e1326.
- Puthoff DP, Sardesai N, Subramanyam S, Nemacheck JA, Williams CE (2005) Hfi-2, a wheat cytolytic toxin-like gene, is up-regulated by virulent Hessian fly larval feeding double dagger. *Mol Plant Pathol* 6: 411–423.
- Rinderle SJ, Goldstein IJ, Remsen EE (1990) Physicochemical properties of amaranthin, the lectin from *Amaranthus caudatus* seeds. *Biochemistry* 29: 10555–10561.
- Transue TR, Smith AK, Mo H, Goldstein IJ, Saper MA (1997) Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat Struct Biol* 4: 779–783.
- Rinderle SJ, Goldstein IJ, Matta KL, Ratcliffe RM (1989) Isolation and characterization of amaranthin, a lectin present in the seeds of *Amaranthus*

- caudatus*, that recognizes the T- (or cryptic T)-antigen. *J Biol Chem* 264: 16123–16131.
29. Okino N, Kawabata S, Saito T, Hirata M, Takagi T, et al. (1995) Purification, characterization, and cDNA cloning of a 27-kDa lectin (L10) from horseshoe crab hemocytes. *J Biol Chem* 270: 31008–31015.
 30. Beisel H, Kawabata S, Iwanaga S, Huber R, Bode W (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J* 18: 2313–2322.
 31. Bork P, Holm L, Sander C (1994) The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol* 242(4): 309–20.
 32. Mues A, van der Ven PF, Young P, Fürst DO, Gautel M (1998) Two immunoglobulin-like domains of the Z-disc portion of titin interact in a conformation-dependent way with telethonin. *FEBS Lett* 428: 111–114.
 33. Takabatake T, Takahashi TC, Takeshima K, Takata K (1991) Protein Synthesis during Neural and Epidermal Differentiation in *Cynops* Embryo. *Development, Growth & Differentiation* 33: 277–282.
 34. Takabatake T, Takahashi TC, Takeshima K (1992) Cloning of an Epidermis-specific *Cynops* cDNA from Neurula Library. *Development, Growth & Differentiation* 34: 277–283.
 35. Ogawa M, Takahashi TC, Takabatake T, Takeshima K (1998) Isolation and characterization of a gene expressed mainly in the gastric epithelium, a novel member of the ep37 family that belongs to the $\beta\gamma$ -crystallin superfamily. *Development, Growth & Differentiation* 40: 465–473.
 36. Ogawa M, Takabatake T, Takahashi TC, Takeshima K (1997) Metamorphic change in EP37 expression: members of the $\beta\gamma$ -crystallin superfamily in newt. *Development Genes and Evolution* 206: 417–424.
 37. Augusteyn RC, Stevens A (1998) Macromolecular structure of the eye lens. *Progress in Polymer Science* 23: 375–413.
 38. Jaenicke R, Slingsby C (2001) Lens crystallins and their microbial homologs: structure, stability, and function. *Crit Rev Biochem Mol Biol* 36: 435–499.
 39. Piatigorsky J (2003) Crystallin genes: specialization by changes in gene regulation may precede gene duplication. *J Struct Funct Genomics* 3: 131–137.
 40. D'Alessio G (2002) The evolution of monomeric and oligomeric beta gamma-type crystallins. Facts and hypotheses. *Eur J Biochem* 269: 3122–3130.
 41. Magalhães GS, Lopes-Ferreira M, Junqueira-de-Azevedo ILM, Spencer PJ, Araújo MS, et al. (2005) Natterins, a new class of proteins with kininogenase activity characterized from *Thalassophryne nattereri* fish venom. *Biochimie* 87: 687–699.
 42. Ponting CP, Mott R, Bork P, Copley RR (2001) Novel protein domains and repeats in *Drosophila melanogaster*: insights into structure, function, and evolution. *Genome Res* 11: 1996–2008.
 43. Buttery SM, Ekman GC, Seavy M, Stewart M, Roberts TM (2003) Dissection of the *Ascaris* sperm motility machinery identifies key proteins involved in major sperm protein-based amoeboid locomotion. *Mol Biol Cell* 14: 5082–5088.
 44. Grant RP, Buttery SM, Ekman GC, Roberts TM, Stewart M (2005) Structure of MFP2 and its function in enhancing MSP polymerization in *Ascaris* sperm amoeboid motility. *J Mol Biol* 347: 583–595.
 45. Italiano JE, Roberts TM, Stewart M, Fontana CA (1996) Reconstitution in vitro of the motile apparatus from the amoeboid sperm of *Ascaris* shows that filament assembly and bundling move membranes. *Cell* 84: 105–114.
 46. Bottino D, Mogilner A, Roberts T, Stewart M, Oster G (2002) How nematode sperm crawl. *J Cell Sci* 115: 367–384.
 47. Abrami L, Reig N, van der Goot FG (2005) Anthrax toxin: the long and winding road that leads to the kill. *Trends Microbiol* 13: 72–78.
 48. Al-Shangiti AM, Naylor CE, Nair SP, Briggs DC, Henderson B, et al. (2004) Structural relationships and cellular tropism of staphylococcal superantigen-like proteins. *Infect Immun* 72: 4261–4270.
 49. Mancheño JM, Tateno H, Sher D, Goldstein IJ (2010) *Laetiporus sulphureus* lectin and aerolysin protein family. *Adv Exp Med Biol* 677: 67–80.
 50. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, et al. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* 253: 657–661.
 51. Murzin AG (1992) Familiar strangers. *Nature* 360: 635.
 52. Burroughs AM, Balaji S, Iyer LM, Aravind L (2007) A novel superfamily containing the beta-grasp fold involved in binding diverse soluble ligands. *Biol Direct* 2: 4.
 53. Welkos SL (1991) Plasmid-associated virulence factors of non-toxicogenic (pX01-) *Bacillus anthracis*. *Microb Pathog* 10: 183–198.
 54. Welkos SL, Vietri NJ, Gibbs PH (1993) Non-toxicogenic derivatives of the Ames strain of *Bacillus anthracis* are fully virulent for mice: role of plasmid pX02 and chromosome in strain-dependent virulence. *Microb Pathog* 14: 381–388.
 55. Chand HS, Drysdale M, Lovchik J, Koehler TM, Lipscomb MF, et al. (2009) Discriminating virulence mechanisms among *Bacillus anthracis* strains by using a murine subcutaneous infection model. *Infect Immun* 77: 429–435.
 56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
 57. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
 58. Biegert A, Mayer C, Remmert M, Söding J, Lupas AN (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res* 34: W335–339.
 59. Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21: 951–960.
 60. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, et al. (2009) InterPro: the integrative protein signature database. *Nucleic Acids Research* 37: D211–D215.
 61. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res* 37: D229–232.
 62. Finn RD, Mistry J, Tate J, Coggill J, Heger A, et al. (2009) The Pfam protein families database. *Nucleic Acids Research* 38: D211–D222.
 63. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19: 1015–1018.
 64. Jaroszewski L, Rychlewski L, Li Z, Li W, Godzik A (2005) FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Research* 33: W284–W288.
 65. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340: 783–795.
 66. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305: 567–580.
 67. Ginalski K, Rychlewski L (2003) Protein structure prediction of CASP5 comparative modeling and fold recognition targets using consensus alignment approach and 3D assessment. *Proteins* 53 Suppl 6: 410–417.
 68. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5: Unit 5.6.
 69. Ramage R, Green J, Muir TW, Ogunjobi OM, Love S, et al. (1994) Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* 299(Pt 1): 151–158.
 70. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.