

Review

# Structure–Function Relationships of the Repeat Domains of RTX Toxins

Ulrich Baumann 

Institute of Biochemistry, University of Cologne, Zùlpicherstrasse 47, D-50674 Cologne, Germany; ubaumann@uni-koeln.de; Tel.: +49-221-470-3208

Received: 21 October 2019; Accepted: 9 November 2019; Published: 12 November 2019



**Abstract:** RTX proteins are a large family of polypeptides of mainly Gram-negative origin that are secreted into the extracellular medium by a type I secretion system featuring a non-cleavable C-terminal secretion signal, which is preceded by a variable number of nine-residue tandem repeats. The three-dimensional structure forms a parallel  $\beta$ -roll, where  $\beta$ -strands of two parallel sheets are connected by calcium-binding linkers in such a way that a right-handed spiral is built. The  $\text{Ca}^{2+}$  ions are an integral part of the structure, which cannot form without them. The structural determinants of this unique architecture will be reviewed with its conservations and variations together with the implication for secretion and folding of these proteins. The general purpose of the RTX domains appears to act as an internal chaperone that keeps the polypeptide unfolded in the calcium-deprived cytosol and triggers folding in the calcium-rich extracellular medium. A rather recent addition to the structural biology of the RTX toxin is a variant occurring in a large RTX adhesin, where this non-canonical  $\beta$ -roll binds to ice and diatoms.

**Keywords:** RTX toxin; type I secretion; calcium; internal chaperone; protein folding; tertiary structure

**Key Contribution:** Folding of a polypeptide chain in the extracellular medium encounters a major obstacle due to the lack of chaperones there. The RTX proteins carry a repeat domain which is a calcium-triggered internal chaperone that aids in secretion and extracellular folding.

## 1. Introduction

The repeats-in-toxins (RTX) family represents a collection of proteins from mainly Gram-negative bacterial origin [1–4]. These polypeptides vastly vary in size including molecular weights from about 10 kDa to larger than 1 MDa. Functional diversity is present as well: While many members of this family are cytolytic toxins, e.g., *Escherichia coli* hemolysin HlyA or *Bordetella pertussis* adenylate cyclase CyaA, other members are the nodulation-signaling protein NodO protein from *Rhizobacterium leguminosarum*, or possess hydrolytic activity, e.g., proteases such as the serralysins from *Pseudomonas aeruginosa* or lipases like LipA from *Serratia marcescens* (reviewed in [5]). A common denominator of probably all of these proteins is the secretion by a type I secretion system (T1SS), which requires a C-terminal non-cleavable signal sequence that is generally preceded by a variable number of nonapeptide tandem repeats [6–10]. These repeats coined the denotation of the RTX toxins [3].

The T1SS machinery itself consists of an inner-membrane ABC transporter that recognizes the C-terminal secretion signal of the passenger protein, a membrane fusion protein spanning the periplasmic space, and an outer membrane pore protein. The three components are thought to assemble into a continuous exit passage upon engagement of the ABC transporter with the secretion signal. Therefore, the traditional view is that secretion occurs without periplasmic intermediates, but there are at least some exceptions to this [11]. Owing to the narrow structure of the outer membrane protein tunnel, as seen paradigmatically in the crystal structure of TolC, secretion most likely requires an

unfolded or only partially folded passenger protein. The necessity for passage in the unfolded state was for the first time experimentally proven for HasA, maybe the only example for a protein being secreted by a type I system not possessing the characteristic RTX nonapeptide motifs (reviewed in [8,12]). On the other hand, HasA secretion is dependent on the chaperone SecB, contrary to the RTX proteins of which the secretion is chaperone-independent. SecB's engagement with HasA delays its folding by an order of magnitude, which is necessary considering that the C-terminal location of the translocation signal requires complete synthesis of the polypeptide chain and the very fast folding of HasA.

The nonapeptide repeats, which coin the denotation of the RTX proteins, possess the consensus sequence  $^1\text{XUXGGXGD}^9$ , where U stands for a non-polar residue, e.g., leucine. These repeats were early shown to bind  $\text{Ca}^{2+}$  ions [13]. There is a positive correlation of the number of repeats with the molecular weight of the protein, thus varying the number of nonapeptides from about 4 to more than 50. Larger number of repeats usually come in blocks of about six to eight, where they are arranged in a tandem fashion. These blocks are separated by linkers of variable length and amino acid sequence. Deviations from the consensus sequence mentioned above occur mainly at the edges of the blocks where the structural constraints are less strict.

The functional significance of the repeat domain (RD or RTX domain) has been less clear for some time. For example, the repeats of *E. coli* hemolysin HlyA have early on been implied in binding to receptors of erythrocytes [13], but it was also clear that many RTX proteins would not bind to erythrocytes. The calcium-binding property has also been implicated in the formation of cation pore channels, which again is not a feature of all RTX-containing proteins.

Beside the RTX repeats, another conspicuous structural feature of all these proteins is the rare occurrence of cysteine residues, with most proteins containing either none cysteines or just one cysteine. In other words, there is a lack of disulfide bonds, which are otherwise quite frequent in extracellular proteins.

In analogy to the role of SecB in the secretion of HasA, one could assume that the RTX domains also delay folding of the respective proteins and that folding is triggered by the extracellular binding of calcium ions, of which the concentration is sub-micromolar in the cytosol, but millimolar in the extracellular medium.

The requirements of the actual translocation process for the RTX repeats and concomitantly for calcium ions has been less clear for some time. While a number of studies agreed that the RTX repeats are not the secretion signal, Létoffé and Wandersman [14] showed that the RTX repeats are not involved in recognition by the T1SS but are needed for the efficient translocation of larger passenger proteins.

In this review, the focus will be on the structural features of the RTX domain and its implications for folding and secretion of the polypeptide chain.

## **2. RTX Repeats Form a Unique Parallel $\beta$ -Roll Domain Which Contains $\text{Ca}^{2+}$ Ions as an Integral Part**

The first three-dimensional structure of a protein containing RTX repeats was that of the alkaline protease AprA from *Ps. aeruginosa* (PDB entry 1KAP) [15], which features five RTX repeats in its C-terminal region. Other experimental structures with RTX domains that are available in the PDB are cited in Table 1.

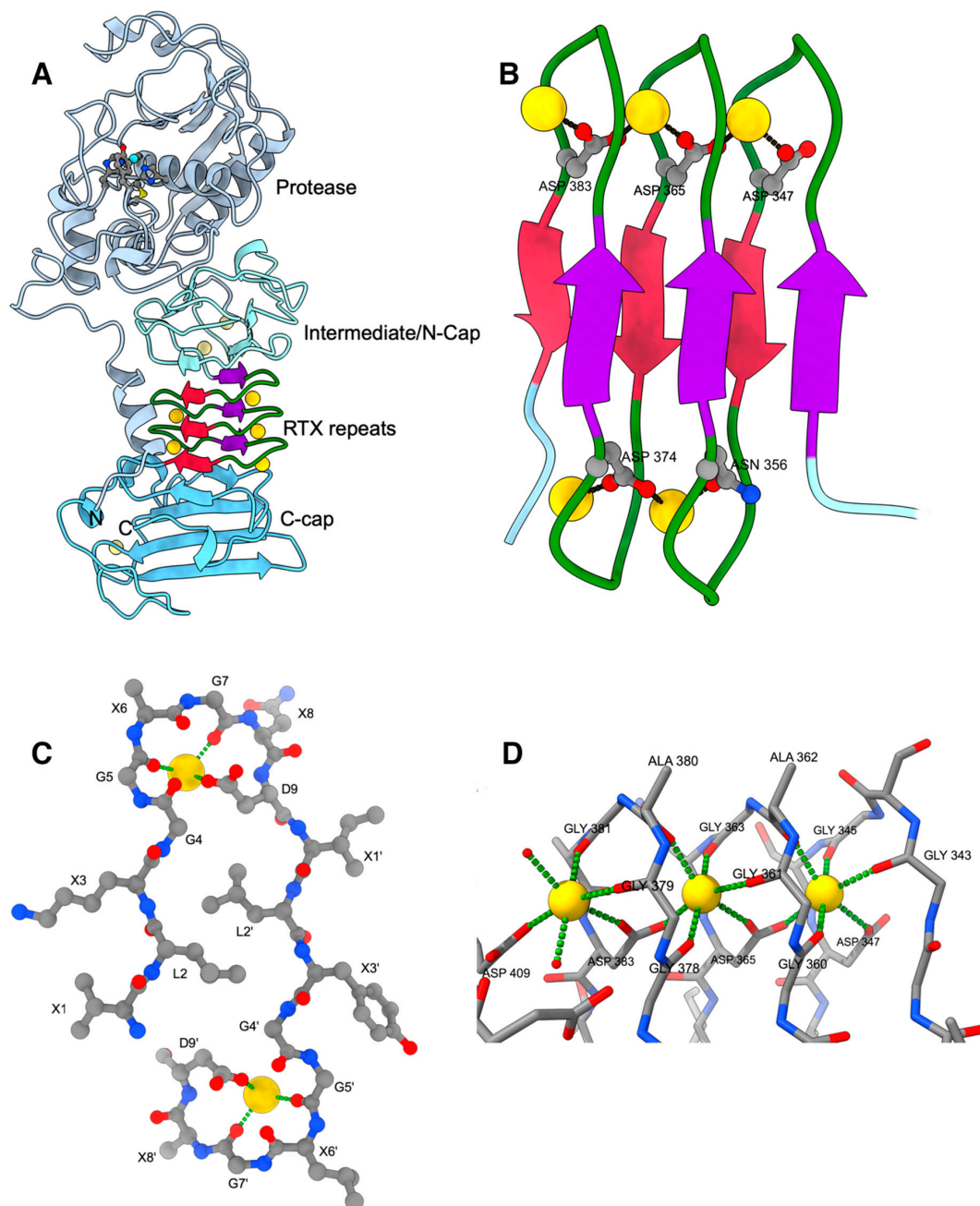
**Table 1.** RTX protein structure in the PDB ([www.rcsb.org](http://www.rcsb.org))

Protein	PDB Entry	No of Repeats XUXGGXGXD
<i>Ps. aeruginosa</i> alkaline protease APRA_PSEAE	1KAP [15], 1JIW [16], 1AKL [17], 3VI1	5
<i>S. marcescens</i> metalloprotease PRZN_SERMA	1SAT [18], 1SMP [19], 1AF0, 4I35, 5D7W [20]	5
<i>S. marcescens</i> sp. Metalloprotease PRTZN_SERME	1SRP [21]	5
<i>Erwinia chrysanthemi</i> metalloprotease PrtC PRTC_DICCH	1GO8 [22], 1K7G [23], 1K7I, 1K7Q, 3HB2 [24], 3HBU, 3HBV, 3HDA	5
<i>Caulobacter vibroides</i> S-Layer protein A0A0H3C8J1_CAUVN	5N97, 5N8P [25]	4
<i>Azotobacter vinelandii</i> mannrinan-C5-Epimerase ALGE4_AZOVI	2AGM, 2ML1, 2ML2, 2ML3 [26]	6
<i>Bordetella pertussis</i> adenylate cyclase CYAA_BORP1	5CVW, 5CXL [27]	8
<i>Serratia marcescens</i> lipase LipA Q5933_SERMA	2QUA, 2QUB [28]	5 + 8
<i>Ps. Fluorescence</i> extracellular lipase Q9RBY1_9PSED	2Z8X, 2Z8Z, 2ZJ6, 2ZJ7, 2ZVD, 3A6Z, 3A70 [29]	5 + 8
<i>Pseudomonas</i> sp. psychrophilic metalloprotease O69771_9PSED	1H71 [30], 1G9K, 1O0Q, 1O0T, 1OM6, 1OM7, 1OM8, 1OMJ	5
<i>Flaviobacterium</i> sp. alkaline metalloprotease DOVMS8_9FLAO	3U1R [31]	5
<i>Marinomonas primoryensis</i> ice-binding protein region IV A1YY3_9GAMM	3P4G [32]	12 19-residue repeats
<i>Marinomonas primoryensis</i> ice-binding protein region V A1YY3_9GAMM	5 JUH [33]	4

AprA is a 50 kDa metalloprotease that is secreted by a type I secretion system into the extracellular medium. It is homologous to the 50 kDa protease from *Serratia marcescens* (serralyisin) and the metalloproteases PrtA, PrtB, PrtC, and PrtG from *Erwinia chrysanthemi*, where the latter have been also extensively studied with respect to their secretion mechanism [34–36]. All these proteases belong to the serralyisin family [37], which is annotated as family M10B in the MEROPS protease data base [38]. Many or all of them are encoded on an operon together with the necessary components for secretion, i.e., for alkaline protease the ABC transporter aprD, the membrane fusion protein aprE, and the outer membrane protein aprF, followed by the protease-encoding aprA gene [39]. Some of the serralyisins producing organisms also carry a gene for an inhibitor of the protease, which is located at the end of the operon. The inhibitors are small 10 kDa proteins that are located in the periplasm of the bacteria, although type I secretion usually occurs in one step without periplasmic intermediates [16,19,40]. As with the other proteins secreted by the type I system, serralyisins contain a C-terminal non-cleavable secretion signal, which is preceded by five copies of the characteristic 9-residue sequence motif <sup>1</sup>XUXGGXGXD<sup>9</sup>.

Serralyisins possess an N-terminal metzincin-type protease domain (Figure 1A) [41], which harbors the characteristic HEXXHXXGXXHP metzincin metalloprotease sequence motif with the ultimate proline being characteristic for the serralyisins. The three histidines in this motif are zinc ligands and the glutamic acid serves as catalytic base. About 30 residues downstream of this motif resides the S-[T/V/L/I]-MSYW met-turn finger print, where the name-coining methionine of the metzincins is

located, together with a tyrosine residue that functions as electrophilic agent stabilizing the tetrahedral transition state of the peptide bond hydrolysis reaction.



**Figure 1.** (A) Cartoon of AprA (PDB entry 1KAP).  $\text{Ca}^{2+}$  ions are shown as golden spheres. The N-terminal helix and the protease domain are depicted in grey with the active-site  $\text{Zn}^{2+}$  as cyan sphere together with the residues mentioned in the text. The  $\beta$ -roll is depicted in crimson and purple with green calcium-binding loops. The intermediate/N-cap and C-cap subdomains are depicted in different shades of light blue. (B) Close-up of the parallel  $\beta$ -roll. (C) Side-view of one turn of the parallel  $\beta$ -roll depicting 2 tandem nonapeptides of sequence XLXGGXGXGD with X being variable amino acids. (D) Details of the octahedral  $\text{Ca}^{2+}$  coordination by the GGXGXGD turn. The left  $\text{Ca}^{2+}$  ion is the most accessible one and the only that can be exchanged by  $\text{Pb}^{2+}$ , for example. Its coordination sphere is completed by water molecules depicted as small red spheres. All figures were prepared with ChimeraX [42].

After the protease domain there is an extended  $\beta$ -sheet architecture, which can be formally subdivided into three parts: an intermediate/N-cap domain, the repeat domain containing the RTX motif, and a C-terminal domain, which caps the RTX domain and where the secretion signal is located within the last ~20 amino acids. Most important in the context of RTX toxins is the middle part, which was termed parallel  $\beta$ -roll, sometimes also called a parallel  $\beta$ -helix, although the latter refers originally to a different architecture. The parallel  $\beta$ -roll is built up by the succession of the RTX nonapeptide repeats together with bound calcium ions. Here, the first three residues of a nonapeptide  $^1\text{XUXGGXGXD}^9$  form a short beta strand, while the following six amino acids build a turn that binds to calcium ions and connects this nonapeptide to the following one downstream in the sequence. Further continuation of this leads to a right-handed spiral of  $\beta$ -strands with calcium ions bridging the aspartic acid residues at positions  $n$  and  $n + 18$  (Figure 1B,C). This results in a packing of two parallel  $\beta$ -sheets against each other with opposite strand directions. The characteristic  $\beta$ -sheet main-chain hydrogen bonds are formed only within the two sheets and not between them, i.e., the arrangement is rather a sandwich of two  $\beta$ -sheets than a  $\beta$ -barrel. The loops on top and bottom of the  $\beta$ -sandwich wrap around ‘pearls-on-a-string’ aligned calcium ions, where neighbouring  $\text{Ca}^{2+}$  ions are bridged by the Asp-9 carboxylate group of a particular repeat motif.

This architecture differs drastically from the parallel  $\beta$ -helix discovered at about the same time in the structure of pectate lyase C from *Erwinia chrysanthemi*, which does not contain repeating sequence motifs and which is built up by three  $\beta$ -sheets and which does not require metal ions for stability [43,44].

The requirements for the individual amino acids in the nine-residue repeats became well understood from the three-dimensional structure. Residues 1 to 3 XUX form a short, non-twisted  $\beta$ -strand with the sidechain of the second residue ‘U’ pointing into the interior of  $\beta$ -roll. Leucine residues are preferred because they nicely interdigitate forming a tight hydrophobic core of the sandwich, but other hydrophobic residues—e.g., valine or phenylalanine—are found as well. The sidechains of the X-residues point to the exterior of the  $\beta$ -roll, and residues with a high propensity for  $\beta$ -strands are most frequent at this position; for example, threonine residues are often located at the first position. The following six residues GGXGXD build two calcium-binding half sites, where the carbonyl oxygens of the Gly-5 and Gly-7 together with the *anti* lone pair of one oxygen of the carboxylate group of Asp-9 form half of the octahedral coordination sphere of a calcium ion (Figure 1D). The remaining three ligands completing calcium ion coordination originate from a nonapeptide motif located 18 residues further downstream in the sequence (i.e., the residues involved are located in the over the next motif). These ligands are the *syn* lone pair of an aspartic acid carboxylate group oxygen atom Asp-9’ and the carbonyl oxygens of residues Gly4’ and X6’.

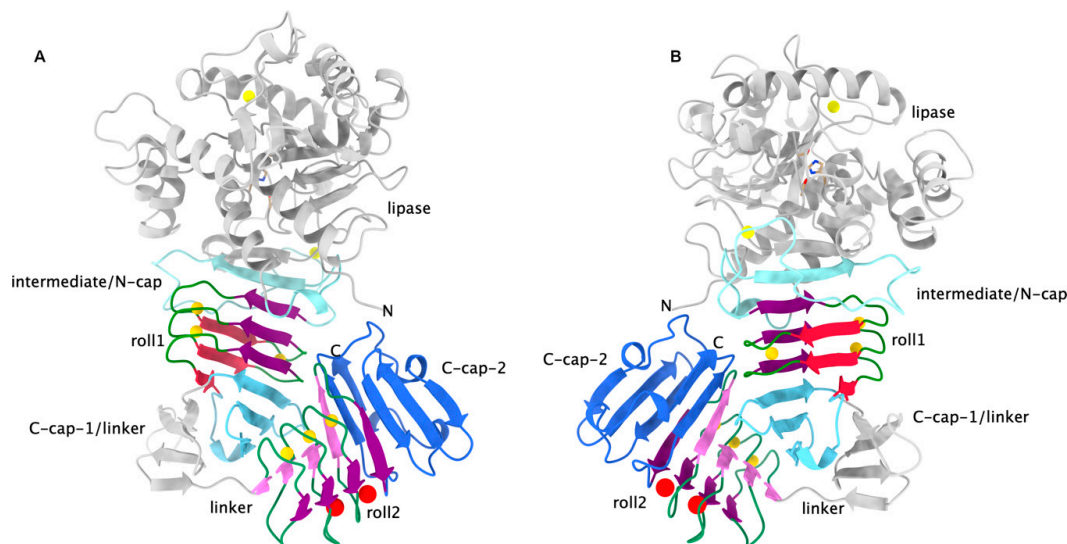
An analysis of the main chain dihedral angles revealed that the conformation of the amino acids at positions 5 and 7 are energetically unfavorable for non-glycine residues due to their positive phi-angles (‘Ramachandran disallowed’). Residue 4, which is also preferably a glycine, adopts a conformation with main-chain dihedral angles of  $\Phi/\Psi \approx -70/-30$ , a combination which is accessible for all amino acids. However, the presence of a sidechain as small as a methyl group would lead to clashes with the  $\text{Ca}^{2+}$ -coordinating carboxylate group of the aspartate residue at position 6.

It is very noteworthy that the parallel  $\beta$ -roll domain mediates contacts to the intermediate domain, on which the protease module resides, to the N-terminal helix, which connects in the tertiary structure the C-terminal end part of the structure with the protease domain, and to the C-cap structure located at the C-terminal end of the primary structure. Thus, structuring of the parallel  $\beta$ -roll during the folding process would definitely impact the folding of the entire molecule.

Beside the calcium ions bound to the RD, there are three additional  $\text{Ca}^{2+}$  sites, of which two are in the intermediate/N-Cap domain. Their purpose might be to confer additional stability to the tertiary structure, which as most RTX proteins lacks disulphide bonds, although these two sites in the intermediate domain share some structural similarity to those bound to the RTX repeats.

### 3. RTX Domains Containing more than One Block of Repeats Build a Compact Assembly that Probably Serves as a Folding Nucleus for the Functional Passenger Domains

Initially, we thought that polypeptides with a greater number of repeat units simply had a larger  $\beta$ -roll, i.e., that the additional repeat units were just added directly to the existing structure like adding additional turns to an  $\alpha$ -helix. However, closer inspection revealed that the nonapeptide repeats come in blocks of about six to eight and that these blocks are separated by linker with sequences very different from the consensus sequence described above. So far, there are only two structures available in the PDB data bank that feature multiple blocks of repeats (i.e., two of them), namely the extracellular lipase LipA from *Serratia marcescens* [28] (Figure 2) and of a related extracellular lipase from *Pseudomonas* sp. [29]. Besides the significant deviations of the RTX consensus sequence of the individual repeats, a major finding here was that the two RDs are not flexibly connected but interact strongly with each other. Furthermore, especially in the second block, the deviations from the consensus sequence are significant. For example, there are insertions of one residue and the crucial aspartic acid residue is substituted in every other repeat, leading to only one side of that  $\beta$ -roll being occupied by  $\text{Ca}^{2+}$  ions, while at the corresponding positions at the other edge the density has been interpreted as water molecules. The very similar *Pseudomonas* sp. MIS38 lipase also exhibits variations from the consensus sequence but has two calcium ions bound at that side of the  $\beta$ -roll as well [29].



**Figure 2.** Two views on LipA rotated by about 180 deg. around the vertical (PDB entry 2QUA). (A): The color coding is as in Figure 1. This time, there are two C-caps, one for each RTX  $\beta$ -roll. Water molecules occupying one edge of the second roll are shown as red spheres. Roll 2 has only at one side calcium ions bound. (B): Same as in A but rotated by about 180 degrees around the vertical. The interaction between the two  $\beta$ -roll domains is clearly visible.

The tight association of the two  $\beta$ -roll domains and both their interactions with the lipase domain led to the reanimation of the proposal for an internal chaperone function of the RTX repeats, an idea which had emerged already with the first crystal structure of AprA. The purpose of it would be to keep the polypeptide chain in an unfolded state in the cytosol, where the  $\text{Ca}^{2+}$  concentration is low, and to induce folding of the RDs and concomitantly of the whole protein extracellularly by the higher amounts of  $\text{Ca}^{2+}$  present there [28].

### 4. $\text{Ca}^{2+}$ -Induced Folding of RTX Domains

The  $\text{Ca}^{2+}$  ions bound by the RDs were recognized as integral part of the 3D structure of the folded RTX domains from the beginning on when the crystal structure of alkaline protease had been determined. The metal ions are buried in the structure. Only the calcium ions located at the edge

of the parallel  $\beta$ -roll structure are partially solvent-accessible and are coordinated in part by water molecules; only these ions can be removed by EDTA under non-denaturing conditions or be substituted for example by  $\text{Pb}^{2+}$  ions (in Figure 1D this is the upper left  $\text{Ca}^{2+}$  ion for example). Thus, complete removal of calcium ions from folded  $\beta$ -rolls requires—in addition to chelators—a denaturing agent, e.g., urea. The calcium ions bridge spatially adjacent aspartate residues that are 18 amino acids apart in the primary structure. The distance between the centers of adjacent carboxylate groups is only about 4.8 Å, like the distance between the calcium ions. Consequently, in the absence of calcium ions, the electrostatic repulsion of the carboxylate groups will make the folded structure unstable.

This was first demonstrated by folding experiments employing an artificially designed  $\beta$ -roll consisting of the chemically synthesized peptide  $\text{H}_2\text{N-WLS-[GGSGNDNLS]}_8\text{-COOH}$ , which contains eight exact repeats, and revealed in the absence of  $\text{Ca}^{2+}$  ions a random coil structure [45]. Addition of  $\text{Ca}^{2+}$  together with PEG8000 led to conformational changes and a CD spectrum consistent with the formation of a mainly  $\beta$ -sheet structure. In accordance with the crucial bridging of spatially adjacent carboxylate groups from the aspartic acid residues at position 9 of the consensus motif, it turned out that—for this artificial peptide—the presence of  $\text{Ca}^{2+}$  in concentrations of about 1 mM was essential for folding. Other cations such as  $\text{Mg}^{2+}$  or  $\text{Tb}^{3+}$  could not substitute for calcium, revealing an interesting specificity of the cation binding sites, especially considering the rather high intracellular  $\text{Mg}^{2+}$  concentrations (about 100 mM although most of it is bound), which are then incompetent to induce folding.

However, this artificially designed  $\beta$ -roll was only folding in the presence of rather high concentrations of PEG8000, an unspecific stabilizer of tertiary structure by molecular crowding. Similar observations were made using ‘repeat-domain-only’ fragments of *Bordetella pertussis* CyaA, where a polypeptide consisting solely of nine repeats (residues 1528-1612 or 1529-1612) did not bind to  $\text{Ca}^{2+}$  [46,47], while in another study an increase in  $\text{Ca}^{2+}$ -affinity by molecular crowding of the CyaA RD(1530-1680) by Ficoll was observed [48]. For the synthetic  $\beta$ -roll described above, the lability of the folded state may have at least in part arisen from a sub-optimally designed sequence with an asparagine residue at position 7, which should be replaced by other residues such as threonine.

However, there is another reason why peptides only comprising the RTX nonapeptides are less prone to fold spontaneously in the presence of  $\text{Ca}^{2+}$ . Interestingly, the folding kinetics of the synthetic  $\beta$ -roll revealed a lag phase, which depended on the peptide concentration, as did the folding of the  $\beta$ -roll. Finally, the appearance of polymers and insoluble fibrous aggregates was observed. This led to a model where cooperative folding is initiated by a few already structured monomers, of which the formation is overall rate-limiting. In accordance with this assumption, the lag phase and consequently overall folding could be accelerated by adding a small amount of pre-formed beta roll (‘seeding’). It was furthermore assumed that structure-induction and polymerization leading to fibrous insoluble aggregates occurred via edge-to-edge aggregation of the  $\beta$ -roll units, where an edge of the beta sheets of a folded monomer or oligomer interacts with unfolded molecules, thus providing a  $\beta$ -sheet nucleation face. Since the edges of a folded monomer are unprotected, edge-to-edge aggregation will occur, leading to large insoluble polymers resembling amyloid fibrils.

Indeed, other studies showed that a protection especially of the C-terminal edge of the  $\beta$ -roll is necessary for calcium-induced folding, as it is generally needed for avoiding aggregation of  $\beta$ -sheet structures. So it was early observed that the  $\beta$ -helix structure of pectate lyase employs a capping of the edges by loops and helices in order to avoid edge-to-edge aggregation [49]. This is exactly what apparently happens in the RTX domains, too; in Figures 1 and 2 these capping elements are denoted as N-cap/intermediate/linker and C-cap. The need for the C-terminally flanking regions of an RTX block was recognized during studies of the repeat domain of *Bordetella pertussis* CyaA, where it was found that a fragment including residues 1529–1681 corresponding to the most C-terminal RTX block could be stably expressed and refolded when the residues of the C-terminal segment were included. The crystal structure of this fragment revealed a typical beta roll structure where the C-terminal side is covered by additional beta strands and an alpha helix [27,46]. Another study showed that even replacement of the

C-terminal cap by YFP was sufficient for folding, although much higher  $\text{Ca}^{2+}$  concentrations were required [47].

The  $\text{Ca}^{2+}$ -induced folding of RTX domains which were flanked by additional segments has been in-depth characterized on various constructs of the RD of adenylate cyclase CyaA from *Bordetella pertussis* [47,48,50–55] as well as for example on the alkaline protease AprA from *Ps. aeruginosa* [56]. All these studies confirmed that folding requires calcium concentrations in the range of about 0.1 to 0.2 mM, which is about 1000 times higher than that present in the cytosol. In the absence of  $\text{Ca}^{2+}$ , all constructs exhibited random-coil characteristics while at the same time staying soluble—i.e., no aggregation was observed—owing to the highly hydrophilic nature of the RTX repeats.

All these results strongly hint at an unusual internal chaperone function of the RTX domains. They keep the polypeptide chains in the calcium-depleted cytosolic medium in a (mainly) unfolded state, which is required for secretion. Translocation to the extracellular medium then occurs in the unfolded state from the C-terminus, which bears the secretion signal, to the N-terminus. The exposure of the C-terminal region containing RTX repeats and adjacent segments to the high  $\text{Ca}^{2+}$  concentrations in the extracellular medium result in folding of the  $\beta$ -roll, which aids in structuring the whole polypeptide chain. This disorder-to-order transition was shown to be essential for secretion of the CyaA toxin from *B. pertussis* [52]. The vectorial C-to-N folding has been demonstrated in a study by Bumba et al. [27], who also could thereby show that this mechanism leads to a more efficient secretion by a ‘Brownian ratchet’, since the folded domains of the polypeptide chains are unable to slide back through the secretion tunnel. This consolidates the long-standing issue whether the RTX repeats are needed for secretion by type I secretion systems: small and/or unfolded domains can be fused to the C-terminal secretion signal without the intervening RTX repeats, while larger passengers will remain unfolded if fused to a large enough RTX domain, or, like HasA, will require chaperones that keep them in the unfolded state in the cytosol. The requirement for  $\text{Ca}^{2+}$  ions for secretion (in contrast to folding) has been recently examined in some detail as well with conflicting results. While one publication reported a HlyA secretion rate independent of the calcium concentration [57], another study showed a strong increase [27].

## 5. Structure of RTX Adhesins Reveals the Presence of Canonical and Non-Canonical RTX Repeats

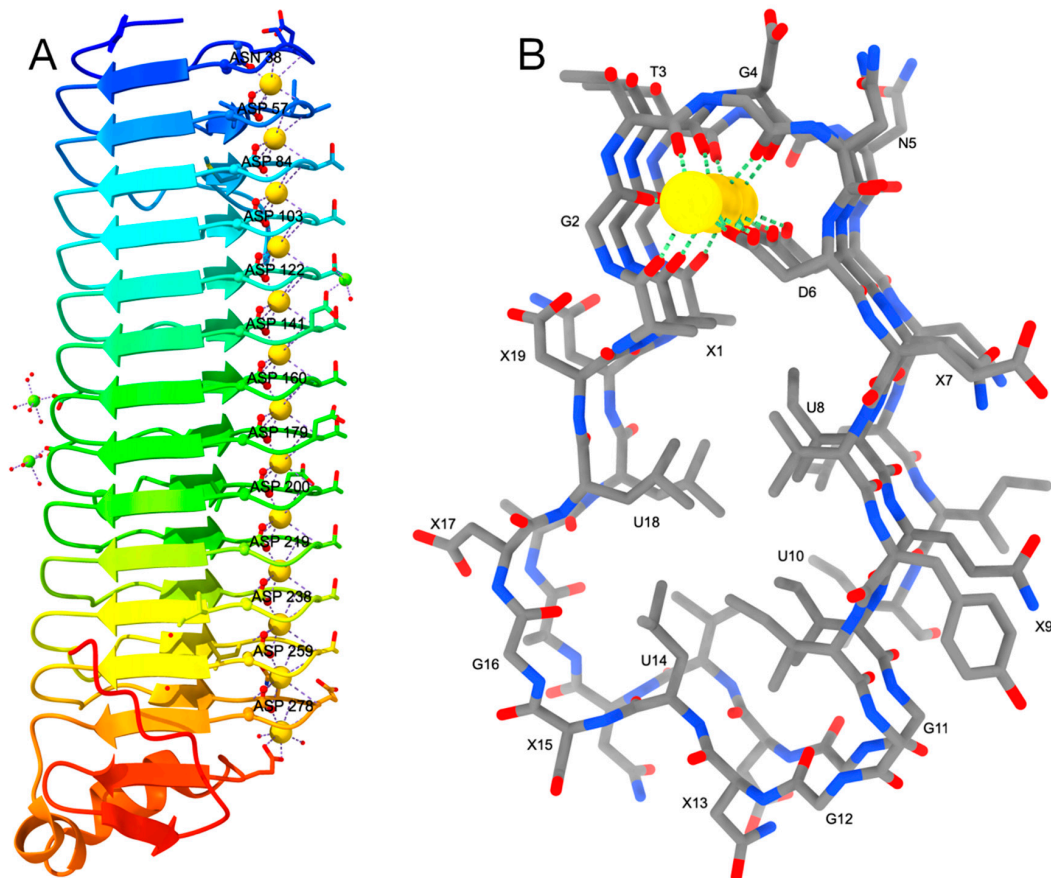
A very interesting addition to the structural repertoire of the RTX domains is represented by a hyperactive antifreeze protein MpIBP (IBP for ice binding protein) from the Gram-negative Antarctic bacterium *Marinomonas primoryensis*. The protein belongs to the RTX adhesins and has a molecular weight of about 1.5 MDa. It shows a very strong depression of the freezing point of water by interaction with lattice planes of hexagonal ice, but also anchors the bacterium to diatoms. The protein is retained in the outer membrane and possesses a very repetitive structure which is shared by other RTX adhesins and multifunctional autoprocessing RTX toxins (MARTX, [58]). In the context of the RTX repeat domains most interesting are here the so-called region-IV and region-V (RIV and RV), which contain the RTX repeats and the T1SS sequence. The structure of these two domains as well as of the other pieces have been determined [32,33,59].

RV is at the very C-terminal end and comprises four copies of the RTX-nonapeptide plus the C-cap including the secretion signal. Four  $\text{Ca}^{2+}$  ions are bound to the loops.

RIV, which is very closely located in the sequence upstream of RV, is the ice-binding domain of MpIBP and possesses an atypical  $\beta$ -roll where  $\text{Ca}^{2+}$  ions are only bound to one side of the protein (Figure 3). One turn of this new  $\beta$ -roll comprises 19 amino acids instead of 18, i.e.  $2 \times 9$ , in the canonical structure. The repeat sequence deviates from the canonical  $(\text{XUXGGXGXD})_2$  motif and reads now XGTGND-XUXU-GGXUXG-XUX, where the first six residues build the  $\text{Ca}^{2+}$ -binding turn like in the canonical structure, followed by a four-residue  $\beta$ -strand. The second six-residue loop has a very different sequence and consequently does not bind to  $\text{Ca}^{2+}$  ions, which therefore are all located at the other edge of the roll. Further differences are that the two  $\beta$ -sheets of the sandwich differ in length: one is four residues long, the other three, although small variations occur. Similar deviations were observed



in some parts of the second roll domain of *S. marcescens* lipase (see above). However, in the *MpIBP* these are rather the rule than the exception. The bound  $\text{Ca}^{2+}$  ions are described as not hexa-coordinated in an octahedral geometry but approximately hepta-coordinated. The aspartic acid from the second half site is providing both *syn* lone pairs of the two carboxylate oxygen atoms, although one of the oxygen atoms is more distant ( $\sim 3.0$  Å instead of 2.3 Å). Furthermore, all 12 19-residue repeats form one continuous  $\beta$ -roll and are not divided into smaller blocks.



**Figure 3.** (A) the parallel  $\beta$ -roll of the *MpIBP* is colored as rainbow from N-terminus (blue) to C-terminus (red). Oxygen atoms are painted in red,  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  ions are shown as golden and green spheres, respectively. Waters are shown as small red spheres. (B) Side-view of the roll depicting the motif XGTGND-XUXU-GGXUXG-XUX where U stands for a non-polar and X for a polar amino acid. Carbon atoms are shown in grey, nitrogen and oxygen in blue and red, respectively, and  $\text{Ca}^{2+}$  ions as golden spheres.

Since the aspartate residues bridge the calcium ions and they are deeply buried in the interior of the loops, it is very likely that also this segment adopts a random coil structure in the cytosol and folds in the extracellular medium. Thus, this domain may have a twofold role: aiding secretion, and binding to ice and diatoms.

## 6. Conclusions and Outlook

The RTX-domains is a calcium-triggered switch that keeps the polypeptides in a secretion-competent unfolded state in the cytosol and induces folding of the translocated proteins in the external, calcium-rich medium. Keeping the polypeptide unfolded and therefore dysfunctional in the cytosol may also mitigate possible toxic effects. It must be stressed that the tight binding of calcium ions is not reflected by a dissociation constant since the binding is accompanied by folding and concomitant burial of the ions and is therefore irreversible. Folding and formation of  $\text{Ca}^{2+}$  sites is

highly cooperative, and it requires  $\text{Ca}^{2+}$  in a concentration of at least 100  $\mu\text{M}$ . The calcium binding sites can be degenerated with large deviations from the canonical  $(\text{XUXGGXGXD})_2$  motif leading to an asymmetric distribution of the ions. The RTX domain also aids secretion, at least for larger passengers. While this appears to be common for all RTX proteins, the example of the antifreeze protein shows that these domains could also possess further functions, e.g., binding to ice surfaces or receptors; for CyaA the integrin receptor binding site has been mapped to the RTX domain [60]. The structures of complete toxins are not available yet, however owing to the advances in single particle cryo-EM they may be well under way. They would provide more insight into the arrangement of the RTX blocks and may help to discover additional functions. The specific calcium-induced folding of RTX domains has also raised interest in biotechnological applications [61,62].

**Funding:** This research was funded by the Swiss National Science Foundation, grant numbers 31-52398.97, 31-67253.01, and 3100A0-108262, as well as by a BASF-fellowship of the Studienstiftung des Deutschen Volkes.

**Acknowledgments:** This work had its origin in the laboratory of David B. McKay, to whom I am very grateful for the time being there. I like to thank my collaborators Cecile Wandersman, Philippe Delepelaire, Sylvie Létoffé, André Lazdunski, Allain Filloux, Robert D. Gray, Hauke Lilie, and Karl-Erich-Jäger, as well as all former members of my lab who worked with me on the RTX proteins: Thomas Hege, Anselm Oberholzer, Santina Russo, Mario Bumann, and Reto Meier. I apologize for not mentioning many studies, especially those on the functional aspects of the RTX proteins.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Benz, R. Channel formation by RTX-toxins of pathogenic bacteria: Basis of their biological activity. *Biochim. Biophys. Acta Biomembr.* **2016**, *1858*, 526–537. [[CrossRef](#)]
2. Linhartova, I.; Osicka, R.; Bumba, L.; Masin, J.; Sebo, P. *Microbial Toxins*; Gopalakrishnakone, P., Stiles, B., Alape-Girón, A., Dubreuil, J.D., Mandal, M., Eds.; Springer: Dordrecht, The Netherlands, 2015; pp. 1–29.
3. Welch, R.A. Pore-forming cytolysins of gram-negative bacteria. *Mol. Microbiol.* **1991**, *5*, 521–528. [[CrossRef](#)]
4. Welch, R.A. RTX Toxin Structure and Function: A Story of Numerous Anomalies and Few Analogies in Toxin Biology. *Curr. Top. Microbiol. Immunol.* **2001**, *257*, 85–111.
5. Linhartová, I.; Bumba, L.; Mašín, J.; Basler, M.; Osička, R.; Kamanová, J.; Procházková, K.; Adkins, I.; Hejnová-Holubová, J.; Sadílková, L.; et al. RTX proteins: A highly diverse family secreted by a common mechanism. *FEMS Microbiol. Rev.* **2010**, *34*, 1076–1112. [[CrossRef](#)]
6. Wandersman, C.; Delepelaire, P.; Letoffe, S.; Ghigo, J.M. A signal peptide-independent protein secretion pathway. *Antonie Van Leeuwenhoek* **1992**, *61*, 111–113. [[CrossRef](#)]
7. Thomas, S.; Holland, I.B.; Schmitt, L. The Type 1 secretion pathway—The hemolysin system and beyond. *Biochim. Biophys. Acta Mol. Cell Res.* **2014**, *1843*, 1629–1641. [[CrossRef](#)]
8. Delepelaire, P. Type I secretion in gram-negative bacteria. *Biochim. Biophys. Acta Mol. Cell Res.* **2004**, *1694*, 149–161. [[CrossRef](#)]
9. Kanonenberg, K.; Spitz, O.; Erenburg, I.N.; Beer, T.; Schmitt, L. Type I secretion system-it takes three and a substrate. *FEMS Microbiol. Lett.* **2018**, *365*. [[CrossRef](#)]
10. Spitz, O.; Erenburg, I.N.; Beer, T.; Kanonenberg, K.; Holland, I.B.; Schmitt, L. Type I Secretion Systems—One Mechanism for All. *Microbiol. Spectr.* **2019**, *7*. [[CrossRef](#)]
11. Smith, T.J.; Sondermann, H.; O’Toole, G.A. Type 1 Does the Two-Step: Type 1 Secretion Substrates with a Functional Periplasmic Intermediate. *J. Bacteriol.* **2018**, *200*. [[CrossRef](#)]
12. Debarbieux, L.; Wandersman, C. Folded HasA inhibits its own secretion through its ABC exporter. *EMBO J.* **2001**, *20*, 4657–4663. [[CrossRef](#)]
13. Ludwig, A.; Jarchau, T.; Benz, R.; Goebel, W. The repeat domain of Escherichia coli haemolysin (HlyA) is responsible for its  $\text{Ca}^{2+}$ -dependent binding to erythrocytes. *Mol. Gen. Genet.* **1988**, *214*, 553–561. [[CrossRef](#)]
14. Létoffé, S.; Wandersman, C. Secretion of CyaA-PrtB and HlyA-PrtB fusion proteins in Escherichia coli: Involvement of the glycine-rich repeat domain of Erwinia chrysanthemi protease B. *J. Bacteriol.* **1992**, *174*, 4920–4927. [[CrossRef](#)]

15. Baumann, U.; Wu, S.; Flaherty, K.M.; McKay, D.B. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: A two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* **1993**, *12*, 3357–3364. [[CrossRef](#)]
16. Hege, T.; Feltzer, R.E.; Gray, R.D.; Baumann, U. Crystal structure of a complex between *Pseudomonas aeruginosa* alkaline protease and its cognate inhibitor: Inhibition by a zinc-NH<sub>2</sub> coordinative bond. *J. Biol. Chem.* **2001**, *276*, 35087–35092. [[CrossRef](#)]
17. Miyatake, H.; Hata, Y.; Fujii, T.; Hamada, K.; Morihara, K.; Katsube, Y. Crystal structure of the unliganded alkaline protease from *Pseudomonas aeruginosa* IFO3080 and its conformational changes on ligand binding. *J. Biochem.* **1995**, *118*, 474–479. [[CrossRef](#)]
18. Baumann, U. Crystal structure of the 50 kDa metallo protease from *Serratia marcescens*. *J. Mol. Biol.* **1994**, *242*, 244–251. [[CrossRef](#)]
19. Baumann, U.; Bauer, M.; Létoffé, S.; Delepelaire, P.; Wandersman, C. Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi*. *J. Mol. Biol.* **1995**, *248*, 653–661. [[CrossRef](#)]
20. Wu, D.; Ran, T.; Wang, W.; Xu, D. Structure of a thermostable serralysin from *Serratia* sp. FS14 at 1.1 Å resolution. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **2016**, *72*, 10–15. [[CrossRef](#)]
21. Hamada, K.; Hata, Y.; Katsuya, Y.; Hiramatsu, H.; Fujiwara, T.; Katsube, Y. Crystal structure of *Serratia* protease, a zinc-dependent proteinase from *Serratia* sp. E-15, containing a beta-sheet coil motif at 2.0 Å resolution. *J. Biochem.* **1996**, *119*, 844–851. [[CrossRef](#)]
22. Hege, T.; Baumann, U. The conserved methionine residue of the metzincins: A site-directed mutagenesis study. *J. Mol. Biol.* **2001**, *314*, 181–186. [[CrossRef](#)]
23. Hege, T.; Baumann, U. Protease C of *Erwinia chrysanthemi*: The crystal structure and role of amino acids Y228 and E189. *J. Mol. Biol.* **2001**, *314*, 187–193. [[CrossRef](#)]
24. Oberholzer, A.E.; Bumann, M.; Hege, T.; Russo, S.; Baumann, U. Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. *Biol. Chem.* **2009**, *390*, 875–881. [[CrossRef](#)]
25. Bharat, T.A.M.; Kureisai-Ciziene, D.; Hardy, G.G.; Yu, E.W.; Devant, J.M.; Hagen, W.J.H.; Brun, Y.V.; Briggs, J.A.G.; Löwe, J. Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nat. Microbiol.* **2017**, *2*, 17059. [[CrossRef](#)]
26. Buchinger, E.; Knudsen, D.H.; Behrens, M.A.; Pedersen, J.S.; Aarstad, O.A.; Tøndervik, A.; Valla, S.; Skjåk-Bræk, G.; Wimmer, R.; Achmann, F.L. Structural and functional characterization of the R-modules in alginate C-5 epimerases AlgE4 and AlgE6 from *Azotobacter vinelandii*. *J. Biol. Chem.* **2014**, *289*, 31382–31396. [[CrossRef](#)]
27. Bumba, L.; Masin, J.; Macek, P.; Wald, T.; Motlova, L.; Bibova, I.; Klimova, N.; Bednarova, L.; Veverka, V.; Kachala, M.; et al. Calcium-Driven Folding of RTX Domain  $\beta$ -Rolls Ratchets Translocation of RTX Proteins through Type I Secretion Ducts. *Mol. Cell* **2016**, *62*, 47–62. [[CrossRef](#)]
28. Meier, R.; Drepper, T.; Svensson, V.; Jaeger, K.E.; Baumann, U. A calcium-gated lid and a large beta-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. *J. Biol. Chem.* **2007**, *282*, 31477–31483. [[CrossRef](#)]
29. Angkawidjaja, C.; You, D.J.; Matsumura, H.; Kuwahara, K.; Koga, Y.; Takano, K.; Kanaya, S. Crystal structure of a family I.3 lipase from *Pseudomonas* sp. MIS38 in a closed conformation. *FEBS Lett.* **2007**, *581*, 5060–5064. [[CrossRef](#)]
30. Aghajari, N.; Van Petegem, F.; Villeret, V.; Chessa, J.P.; Gerday, C.; Haser, R.; Van Beeumen, J. Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. *Proteins* **2003**, *50*, 636–647. [[CrossRef](#)]
31. Zhang, S.C.; Sun, M.; Li, T.; Wang, Q.H.; Hao, J.H.; Han, Y.; Hu, X.J.; Zhou, M.; Lin, S.X. Structure analysis of a new psychrophilic marine protease. *PLoS ONE* **2011**, *6*, e26939. [[CrossRef](#)]
32. Garnham, C.P.; Campbell, R.L.; Davies, P.L. Anchored clathrate waters bind antifreeze proteins to ice. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7363–7367. [[CrossRef](#)]
33. Guo, S.; Stevens, C.A.; Vance, T.D.R.; Olijve, L.L.C.; Graham, L.A.; Campbell, R.L.; Yazdi, S.R.; Escobedo, C.; Bar-Dolev, M.; Yashunsky, V.; et al. Structure of a 1.5-MDa adhesin that binds its Antarctic bacterium to diatoms and ice. *Sci. Adv.* **2017**, *3*, e1701440. [[CrossRef](#)]

34. Delepelaire, P.; Wandersman, C. Protein secretion in gram-negative bacteria. The extracellular metalloprotease B from *Erwinia chrysanthemi* contains a C-terminal secretion signal analogous to that of *Escherichia coli* alpha-hemolysin. *J. Biol. Chem.* **1990**, *265*, 17118–17125.
35. Delepelaire, P.; Wandersman, C. Characterization, localization and transmembrane organization of the three proteins PrtD, PrtE and PrtF necessary for protease secretion by the gram-negative bacterium *Erwinia chrysanthemi*. *Mol. Microbiol.* **1991**, *5*, 2427–2434. [[CrossRef](#)]
36. Létoffé, S.; Delepelaire, P.; Wandersman, C. Protease secretion by *Erwinia chrysanthemi*: The specific secretion functions are analogous to those of *Escherichia coli* alpha-haemolysin. *EMBO J.* **1990**, *9*, 1375–1382. [[CrossRef](#)]
37. Baumann, U. *Handbook of Proteolytic Enzymes*, 2nd ed.; Barrett, A.J., Rawlings, N.D., Woessner, F.J., Eds.; Elsevier: Amsterdam, The Netherlands, 2004; pp. 579–581.
38. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, 624–632. [[CrossRef](#)]
39. Duong, F.; Lazdunski, A.; Cami, B.; Murgier, M. Sequence of a cluster of genes controlling synthesis and secretion of alkaline protease in *Pseudomonas aeruginosa*: Relationships to other secretory pathways. *Gene* **1992**, *121*, 47–54. [[CrossRef](#)]
40. Létoffé, S.; Delepelaire, P.; Wandersman, C. Characterization of a protein inhibitor of extracellular proteases produced by *Erwinia chrysanthemi*. *Mol. Microbiol.* **1989**, *3*, 79–86. [[CrossRef](#)]
41. Stöcker, W.; Grams, F.; Baumann, U.; Reinemer, P.; Gomis-Rüth, F.X.; McKay, D.B.; Bode, W. The metzincins—Topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* **1995**, *4*, 823–840. [[CrossRef](#)]
42. Goddard, T.D.; Brilliant, A.A.; Skillman, T.L.; Vergenz, S.; Tyrwhitt-Drake, J.; Meng, E.C.; Ferrin, T.E. Molecular Visualization on the Holodeck. *J. Mol. Biol.* **2018**, *430*, 3982–3996. [[CrossRef](#)]
43. Yoder, M.D.; Keen, N.T.; Jurnak, F. New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science* **1993**, *260*, 1503–1507. [[CrossRef](#)]
44. Yoder, M.D.; Lietzke, S.E.; Jurnak, F. Unusual structural features in the parallel beta-helix in pectate lyases. *Structure* **1993**, *1*, 241–251. [[CrossRef](#)]
45. Lilie, H.; Haehnel, W.; Rudolph, R.; Baumann, U. Folding of a synthetic parallel beta-roll protein. *FEBS Lett.* **2000**, *470*, 173–177. [[CrossRef](#)]
46. Bauche, C.; Chenal, A.; Knapp, O.; Bodenreider, C.; Benz, R.; Chaffotte, A.; Ladant, D. Structural and functional characterization of an essential RTX subdomain of *Bordetella pertussis* adenylate cyclase toxin. *J. Biol. Chem.* **2006**, *281*, 16914–16926. [[CrossRef](#)] [[PubMed](#)]
47. Blenner, M.A.; Shur, O.; Szilvay, G.R.; Cropek, D.M.; Banta, S. Calcium-induced folding of a beta roll motif requires C-terminal entropic stabilization. *J. Mol. Biol.* **2010**, *400*, 244–256. [[CrossRef](#)]
48. Sotomayor-Pérez, A.C.; Subrini, O.; Hessel, A.; Ladant, D.; Chenal, A. Molecular crowding stabilizes both the intrinsically disordered calcium-free state and the folded calcium-bound state of a repeat in toxin (RTX) protein. *J. Am. Chem. Soc.* **2013**, *135*, 11929–11934. [[CrossRef](#)]
49. Richardson, J.S.; Richardson, D.C. Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754–2759. [[CrossRef](#)] [[PubMed](#)]
50. Chenal, A.; Guijarro, J.I.; Raynal, B.; Delepierre, M.; Ladant, D. RTX calcium binding motifs are intrinsically disordered in the absence of calcium: Implication for protein secretion. *J. Biol. Chem.* **2009**, *284*, 1781–1789. [[CrossRef](#)]
51. Chenal, A.; Karst, J.C.; Sotomayor Pérez, A.C.; Wozniak, A.K.; Baron, B.; England, P.; Ladant, D. Calcium-induced folding and stabilization of the intrinsically disordered RTX domain of the CyaA toxin. *Biophys. J.* **2010**, *99*, 3744–3753. [[CrossRef](#)]
52. O'Brien, D.P.; Perez, A.C.S.; Karst, J.; Cannella, S.E.; Enguéné, V.Y.N.; Hessel, A.; Raoux-Barbot, D.; Voegelé, A.; Subrini, O.; Davi, M.; et al. Calcium-dependent disorder-to-order transitions are central to the secretion and folding of the CyaA toxin of *Bordetella pertussis*, the causative agent of whooping cough. *Toxicon* **2018**, *149*, 37–44. [[CrossRef](#)]
53. Sotomayor-Pérez, A.C.; Ladant, D.; Chenal, A. Calcium-induced folding of intrinsically disordered repeat-in-toxin (RTX) motifs via changes of protein charges and oligomerization states. *J. Biol. Chem.* **2011**, *286*, 16997–17004. [[CrossRef](#)] [[PubMed](#)]

54. Sotomayor-Pérez, A.C.; Ladant, D.; Chenal, A. Disorder-to-order transition in the CyaA toxin RTX domain: Implications for toxin secretion. *Toxins* **2014**, *7*, 1–20. [[CrossRef](#)] [[PubMed](#)]
55. Szilvay, G.R.; Blenner, M.A.; Shur, O.; Cropek, D.M.; Banta, S. A FRET-based method for probing the conformational behavior of an intrinsically disordered repeat domain from Bordetella pertussis adenylate cyclase. *Biochemistry* **2009**, *48*, 11273–11282. [[CrossRef](#)]
56. Zhang, L.; Conway, J.F.; Thibodeau, P.H. Calcium-induced folding and stabilization of the Pseudomonas aeruginosa alkaline protease. *J. Biol. Chem.* **2012**, *287*, 4311–4322. [[CrossRef](#)]
57. Lenders, M.H.; Beer, T.; Smits, S.H.; Schmitt, L. In vivo quantification of the secretion rates of the hemolysin A Type I secretion system. *Sci. Rep.* **2016**, *6*, 33275. [[CrossRef](#)] [[PubMed](#)]
58. Satchell, K.J. Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu. Rev. Microbiol.* **2011**, *65*, 71–90. [[CrossRef](#)]
59. Guo, S.; Vance, T.D.R.; Stevens, C.A.; Voets, I.; Davies, P.L. RTX Adhesins are Key Bacterial Surface Megaproteins in the Formation of Biofilms. *Trends Microbiol.* **2019**, *27*, 453–467. [[CrossRef](#)] [[PubMed](#)]
60. El-Azami-El-Idrissi, M.; Bauche, C.; Loucka, J.; Osicka, R.; Sebo, P.; Ladant, D.; Leclerc, C. Interaction of Bordetella pertussis adenylate cyclase with CD11b/CD18: Role of toxin acylation and identification of the main integrin interaction domain. *J. Biol. Chem.* **2003**, *278*, 38514–38521. [[CrossRef](#)] [[PubMed](#)]
61. Bulutoglu, B.; Dooley, K.; Szilvay, G.; Blenner, M.; Banta, S. Catch and Release: Engineered Allosterically Regulated  $\beta$ -Roll Peptides Enable On/Off Biomolecular Recognition. *ACS Synth. Biol.* **2017**, *6*, 1732–1741. [[CrossRef](#)] [[PubMed](#)]
62. Bulutoglu, B.; Banta, S. Calcium-Dependent RTX Domains in the Development of Protein Hydrogels. *Gels* **2019**, *5*, 10. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).