

# Structure and activation mechanism of the Makes caterpillars floppy 1 toxin

Alexander Belyy<sup>1,\*</sup>, Philipp Heilen<sup>1,\*</sup>, Philine Hagel<sup>1</sup>, Oliver Hofnagel<sup>1</sup>, and Stefan Raunser<sup>1,#</sup>

<sup>1</sup>Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Otto-Hahn-Str. 11, 44227 Dortmund, Germany

\*The authors contributed equally.

#Correspondence should be addressed to: [stefan.raunser@mpi-dortmund.mpg.de](mailto:stefan.raunser@mpi-dortmund.mpg.de).

## Abstract

The bacterial Makes caterpillars floppy 1 (Mcf1) toxin promotes apoptosis in insects, leading to loss of body turgor and death. The molecular mechanism underlying Mcf1 intoxication is poorly understood. Here, we present the cryo-EM structure of Mcf1 from *Photorhabdus luminescens*, revealing a seahorse-like shape with a head and tail. While the three head domains contain two effectors, an activator-binding domain (ABD) and an autoprotease, the tail consists of two translocation as well as three receptor-binding domains. Rearrangement of the tail moves the C-terminus away from the ABD and allows binding of the host cell ADP-ribosylation factor 3, inducing conformational changes that position the cleavage site closer to the protease. This unique activation mechanism that is based on a hook-loop interaction results in three autocleavage reactions and the release of two toxic effectors. Unexpectedly, the BH3-like domain containing ABD is not an active effector. Our findings allow us to understand key steps of Mcf1 intoxication at the molecular level.

# Introduction

Many bacteria produce toxins as the major virulence factors during the infection of their hosts. The toxins of insecticidal bacteria have attracted remarkable attention in recent years since they often target specifically insect pests. Therefore, the development and application of insecticidal bacterial toxins have the potential to revolutionize pest management strategies and reduce the harmful effects of chemical insecticides on the environment<sup>1</sup>.

To date, the insecticidal toxins from *Bacillus thuringiensis* are the most commercially successful organic pest control agents<sup>2</sup>. However, the emergence of resistance stimulates the search for novel biopesticides<sup>3</sup>. Promising alternatives include highly efficient insecticidal toxins from entomopathogenic bacteria of the *Photorhabdus* genus: Tc toxins, large bacterial protein complexes that translocate cytotoxic enzymes into target cells using a syringe-like mechanism<sup>4,5</sup>; *Photorhabdus* Insect Related (Pir) Toxins, binary toxins similar to *Bacillus thuringiensis*  $\delta$ -endotoxins<sup>6</sup>; *Photorhabdus* Virulence Cassettes (PVC), phage-like protein-translocation structures that inject effector proteins into target cells<sup>7</sup>; and “Makes caterpillars floppy 1” toxin (Mcf1), which induce cell apoptosis by a mechanism that is not yet understood<sup>8</sup>. In addition to potential agricultural applications, the modular nature of these toxins allows them to be converted into medical delivery systems in gene and cancer therapy to target molecules of interest into cells<sup>9,10</sup>.

Mcf1 is a 324 kDa protein, which was discovered when *Manduca sexta* larvae were infected with *Escherichia coli* carrying fragments of the *P. luminescens* genome<sup>8</sup>. At the cellular level, activity of the toxin leads to apoptotic blebbing of membranes and fragmentation of nuclei<sup>11</sup>. Such apoptosis in the midgut cells leads to a rapid loss of body turgor and the characteristic “floppy” phenotype of the intoxicated insect within 12 hours and death of the organism at around 24 hours after intoxication<sup>8</sup>.

Mcf1 shares limited similarity to known proteins in four regions, as revealed by a whole-genome search. First, the N-terminal region contains a sequence of 15 amino acids (residues 911-925) resembling the BH3 domain of pro-apoptotic proteins. Therefore, it was logical to propose that this region induces apoptosis during cell intoxication with Mcf1<sup>12</sup>. Second, the central region of Mcf1 shares similarities with the 50 kDa Mcf-like effector domain of the multifunctional-autoprocessing repeats-in-toxin (MARTX) toxins of *Vibrio vulnificus*<sup>13,14</sup>. This domain is a cysteine protease that, upon autoproteolytic activation by host cell ADP-ribosylation factor (Arf) proteins, degrades Rab GTPases and induces cell apoptosis<sup>15</sup>. Third, Mcf1 possesses a highly hydrophobic sequence similar to that in the

translocation region of the *Clostridioides difficile* toxin B (TcdB), suggesting that Mcf1 can undergo a conformational change and penetrate the endosomal membrane, similar to TcdB<sup>16</sup>. Finally, the C-terminal region shows minimal similarity to the putative receptor-binding regions of RTX-like toxins<sup>11</sup>.

Interestingly, some *P. luminescens* strains also express a Mcf1 homolog, called Mcf2. These two toxins are similar in the central and C-terminal regions but differ strikingly at their N-terminal regions. The N-terminal region of Mcf2 does not contain a BH3-like domain, but probably exerts its activity through the HrmA-like domain of the plant pathogen *Pseudomonas syringae*<sup>17</sup>. However, the structure, mechanism of cellular internalization and molecular mechanism of action of Mcf1 and Mcf2 remain unknown.

In this study, we focus on Mcf1, determine its cryo-EM structure and functionally analyze the toxin. The structure reveals that Mcf1 is divided into an N-terminal head region and a C-terminal tail region. The head contains an N-terminal effector domain, an activator-binding domain, and a protease effector domain. The tail consists of two translocation and three receptor-binding domains. After translocation into target cells, Mcf1 interacts with the host cell ADP-ribosylation factor 3 (Arf3) that allosterically activates the protease effector domain using a unique pulling mechanism. The latter promotes three cleavage reactions and the release of two toxic effectors into the cytoplasm. Taken together, our results provide for the first time molecular insights into the architecture of Mcf1 and allow us to understand key steps of its mechanism of action.

## Results

### The architecture of Mcf1

To understand how Mcf1 is built in molecular detail, we heterologously expressed and purified the full-length Mcf1 complex from *Photorhabdus luminescens* and determined its structure using single particle cryo-EM to 3.6 Å resolution ([Supplementary Table 1](#), [Supplementary Fig. 1](#), [Supplementary Movie 1](#), [Methods](#)). Since the toxin proved to be inherently flexible, we performed a series of signal subtraction and local refinements of subregions of the complex. The resulting composite reconstruction allowed us to build 92% of the Mcf1 atomic model *de novo*. An additional dataset of C-terminally truncated Mcf1 ([Methods](#)) enabled us to build the remaining N-terminal region, completing the atomic model of Mcf1 encompassing 2,911 of 2,929 amino acids.

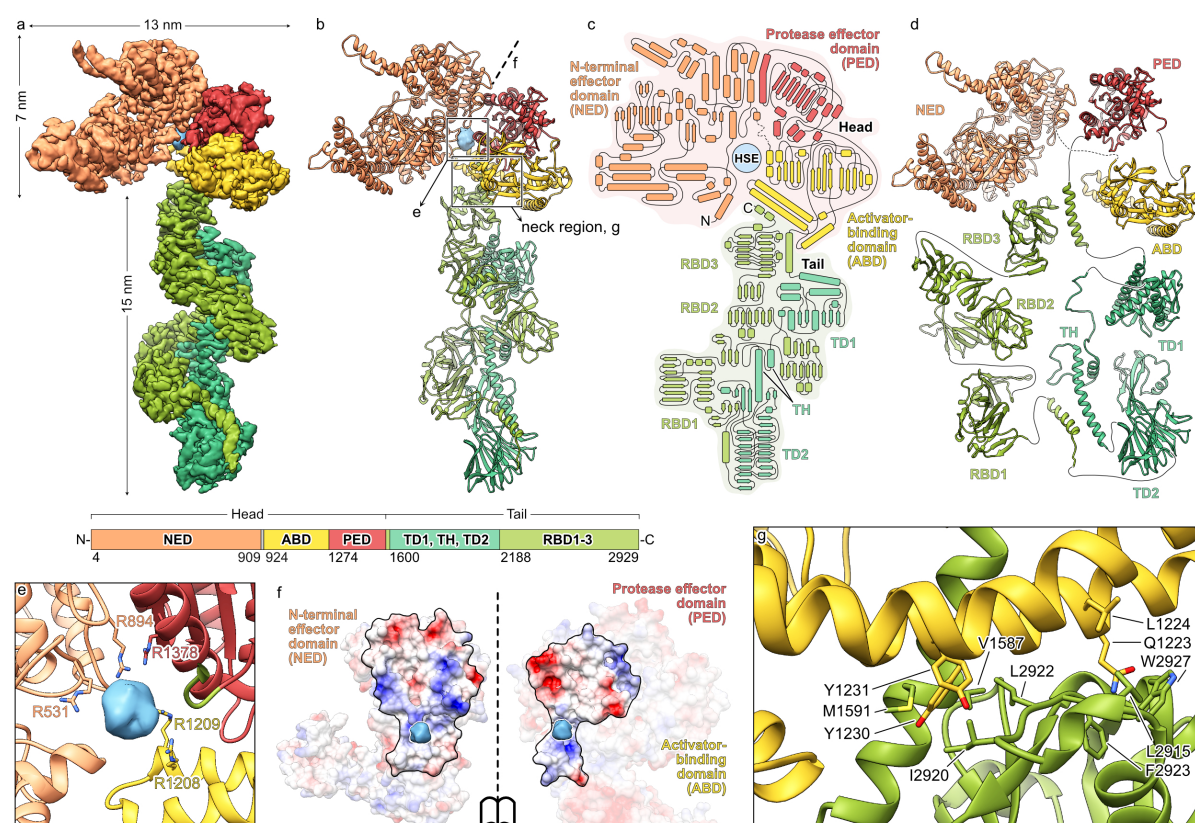
The structure reveals that Mcf1 is a monomeric, asymmetric protein. Based on its overall shape, reminiscent of a seahorse, it can be divided into two parts - an N-terminal head and a C-terminal tail region (Fig. 1). Both are composed of several domains. In order to name them and identify their potential function, we compared the sequence, the general architecture and the structure of the domains with those of other proteins (Supplementary Fig. 2, 3). This revealed that Mcf1 is built in a similar fashion to TcdA and TcdB toxins from *C. difficile*, although Mcf1 does not contain a combined repetitive oligopeptides (CROPs) domain<sup>18</sup> (Supplementary Fig. 3a).

The N-terminal domain of TcdA and TcdB comprises an effector domain, namely a glycosyltransferase<sup>19,20</sup>. Although the equivalent domain of Mcf1 has no structural or sequence homology to this domain, we named it N-terminal effector domain (NED). Indeed, a subregion of this domain has a structural homology to bacterial ADP-ribosyltransferases (RhsP2 from *Pseudomonas aeruginosa*<sup>21</sup> and Rhs1 from *Salmonella enterica*<sup>22</sup>) (Supplementary Fig. 2).

The second domain of the Mcf1 head is not found in TcdA or TcdB. It has structural homology to the *Burkholderia* lethal factor 1<sup>23</sup>, but one critical catalytic residue is missing (Fig. 3a). Therefore and due to studies that will be described below, we dubbed this domain activator-binding domain (ABD). The third domain of the Mcf1 head has a high homology to the cysteine protease domain (CPD) of TcdA and TcdB from *C. difficile* (Supplementary Fig. 2). In addition, it is homologous to the Mcf-like domain of the multifunctional-autoprocessing repeats-in-toxin (MARTX) from *Vibrio vulnificus*, which induces cell apoptosis by cleaving C-terminal regions of Rab proteins<sup>15</sup>. We therefore named this domain protease effector domain (PED), reflecting its potential dual function.

The N-terminal effector domain and the protease effector domain have a large interaction interface formed mostly by electrostatic interactions (Fig. 1f). In addition, we noticed a large spherical density of 10 Å diameter in the middle of the head that is coordinated by five arginines from all three head domains (Fig. 1e). Its location at the interface of the three domains and coordination suggests that this is a negatively charged molecule, which stabilizes the head region. We therefore call it head-stabilizing element (HSE).





**Fig. 1. The structure of Mcf1.** **a,b** Composite cryo-EM density (**a**) and atomic model (**b**) of Mcf1. **c,d** Schematic representation of the secondary structure (**c**) and domain architecture (**d**). **e** Density and coordination of the head-stabilizing element. **f** Surface at the interface of the three domains of the Mcf1 head (N-terminal effector domain, protease effector domain and activator-binding domain) coloured by electrostatic Coulomb potential from  $-10 \text{ kcal mol}^{-1}$  (red) to  $+10 \text{ kcal mol}^{-1}$  (blue). **g** Close-up view of the neck region, connecting the head and tail of Mcf1. NED – N-terminal effector domain, ABD – activator-binding domain, PED – protease effector domain, RBD – receptor-binding domain, TD – translocation domain, TH – transmembrane helices, HSE – head-stabilizing element.

The elongated tail starts with an  $\alpha$ -helical domain and a  $\beta$ -strand-rich domain (Fig. 1c and d), separated by a segment of 99 amino acids, 87 of which are hydrophobic, with high probability to form transmembrane helices (Supplementary Fig. 3b). Indeed, these domains have a striking structural similarity to the translocation apparatus of TcdA and TcdB toxins from *C. difficile* (Supplementary Fig. 2). We therefore termed them translocation domain 1 (TD1), transmembrane helix (TH) and translocation domain 2 (TD2). Interestingly, the translocation domains of Mcf1 match the recently derived consensus sequence of the evolutionary conserved translocase (ECT) of large clostridial toxins even better than the original *C. difficile* toxins<sup>16</sup> (Supplementary Fig. 3c). Since the common translocation apparatus is the hallmark of the large clostridial toxins (LCTs), our structural data clearly identify Mcf1 as a member of this bacterial toxin superfamily.

The tail of Mcf1 ends with three  $\beta$ -strand-rich domains. Only the latter shares structural similarity to a region of *C. difficile* toxins that has been connected to receptor binding (Supplementary Fig. 2). Since the other domains are also structurally reminiscent of the receptor-binding domains of Tc toxins<sup>4</sup> and diphtheria toxins<sup>24</sup>, we called these domains receptor-binding domains 1-3 (RBD1-3) (Fig. 1c and d). Interestingly, the receptor-binding domains are arranged such that they fold back onto the tail towards the head, resulting in the C-terminus of Mcf1 to be located within 2 nm distance from its N-terminus at the neck region (Fig. 1c and d). Moreover, the C-terminal region forms multiple hydrophobic interactions with a helix bundle of the activator-binding domain and supports the helix that directly connects the head and the tail region of the toxin, thus organizing the neck region (Fig. 1g). This unique fold is likely an important element of the toxin.

### Autoproteolysis of Mcf1

The structure revealed that Mcf1 possesses a complete protease domain. It has a typical  $\alpha/\beta$  hydrolase fold with a structurally correctly positioned catalytic triad comprising a cysteine, histidine and aspartic acid (Fig. 3a) indicating that it indeed functions as a protease. To test if this domain is active and responsible for Mcf1 processing as in the case of other toxins<sup>25,26</sup>, we intoxicated Sf9 insect cells with N-terminally or C-terminally Myc-tagged Mcf1 variants and monitored the toxin during intoxication using western blots of cell lysates. Wild-type Mcf1 was cleaved into a ~105 kDa N-terminal and ~150 kDa C-terminal fragment in a time-dependent manner (Fig. 2a), demonstrating that Mcf1 is proteolyzed inside the target cell. Interestingly, the C1379A variant of Mcf1, which impairs the catalytic center of the domain, results in Mcf1 to remain uncleaved, indicating that the protease effector domain is indeed responsible for the cleavage of Mcf1 by autoproteolysis.

When mapping the fragment sizes on our structure, we found that the N-terminal 105 kDa fragment lacks the activator-binding domain and protease effector domain and only corresponds to the N-terminal effector domain. The C-terminal 150 kDa fragment matches the tail region (Fig. 2b). Thus, the experiment demonstrates that Mcf1 is cleaved at least at two positions inside the cytoplasm of the target cell. This is different from other toxins, such as TcdA and TcdB from *C. difficile* that have a single cleavage site<sup>25,26</sup>.

### Arf3 triggers Mcf1 autoproteolysis

Although we could demonstrate that Mcf1 is autoproteolytically cleaved in cells, the toxin was stable in solution for long periods of time. This indicates that the protease apparently needs to

be activated by factor(s) of the target cell. In the case of TcdA and TcdB, the protease domain directly interacts with the eukaryote-specific molecule inositol hexakisphosphate (IP6) to activate the catalytic center by orienting the catalytic residues<sup>25,27</sup>. However, Mcf1 does not have an IP6-binding site and its catalytic center does not require reorganization since all three residues of the catalytic triad are in 3 Å proximity. Therefore, the mechanism of Mcf1 activation must be completely different from the one of LCT toxins.

To reveal this mechanism, we aimed to identify the host cell factor(s) that leads to the activation of the protease and subsequent processing of the toxin. The 50 kDa Mcf-like domain of MARTX from *V. vulnificus* binds to eukaryote-specific ADP-ribosylation factors (Arfs) in their GTP-bound state with nanomolar affinity<sup>13,14</sup>. Due to the structural homology, we chose one of these factors, Arf3, which is primarily located on the surface of the host Golgi apparatus and endosomes<sup>28</sup>, and tested whether it could stimulate proteolytic cleavage of Mcf1 *in vitro*. Incubation of N-terminal His-tagged and C-terminal FLAG-tagged Mcf1 with His-tagged Arf3 resulted indeed in the cleavage of Mcf1, indicating that Arf3 acts as an activator of Mcf1 autoproteolysis (Fig. 2c).

In the case of the Mcf-like domain of MARTX, it was shown that Arf binding mechanically moves the cleavage site into the catalytic center and thereby triggers cleavage<sup>14</sup>. Such a mechanism is not possible for Mcf1 since binding of Arf3 to the protease effector domain is sterically hindered by the presence of the N-terminal effector domain (Supplementary Fig. 4a). This suggests, that Arf3 binding to Mcf1 activates autoproteolysis in a different way.





its processing. We observed rapid and efficient Arf3-dependent toxin processing, confirming that Arf3 is the only essential eukaryotic factor in this *E. coli* system (Supplementary Fig. 4b). This rules out that another host cell factor is required for efficient Mcf1 autoproteolysis.

Membrane insertion and translocation of Mcf1 during intoxication of insect cells would inevitably result in conformational changes at the neck region, where the C-terminus of the protein makes strong hydrophobic interactions with the head of Mcf1 (Fig. 1g). To test the role of the C-terminal region of Mcf1 in the autoprocessing reaction, we purified a Mcf1 mutant devoid of 15 C-terminal amino acids ( $\Delta 15C$ ) and incubated it with Arf3 *in vitro*. Strikingly, the autocleavage reaction proceeded several times faster than with the wild-type toxin and yielded the same fragments (Fig. 2d). This result demonstrates that the C-terminus of Mcf1 is a regulatory element that prevents autoprocessing of Mcf1 prior to its translocation into the target cell.

### **Mcf1 undergoes three autocleavage reactions**

In our cell intoxication and *in vitro* cleavage experiments, we detected either two or multiple fragments of Mcf1, which is surprising given the fact that most known toxins are only cleaved at one position<sup>29</sup> (Fig. 2a). By combining the lengths of the different fragments and mapping them on our structure, we found that cleavage must occur at three different positions resulting in the separation of the N-terminal effector domain, activator-binding domain, protease effector domain, and the Mcf1 tail (Fig. 2b).

We used mass spectrometry to specify the first cleavage site (Methods) and found that cleavage occurs between lysine 912 and alanine 913 and results in the release of the 105 kDa N-terminal effector domain (Supplementary Fig. 4c). To confirm this, we intoxicated cells with the cleavage site-impaired Mcf1 L911A K912A variant (Supplementary Fig. 4d and e) and noticed a shift in the size of the N-terminal fragment from 105 kDa to 145 kDa, corresponding to the N-terminal effector domain and activator-binding domain. Although mass spectrometry could not clearly determine cleavage site 2 and 3, we could narrow down cleavage 2 by site-directed mutagenesis to the motif 1271-IQGG-1274 (Supplementary Fig. 4d and e).

Interestingly, in our *in vitro* assay we noticed one more 125 kDa C-terminal fragment, which did not appear in the cell intoxication assays (Fig. 2c and d). Mapping of the fragment on the structure suggested that it represents the tail region without the translocation domain 1. In contrast to the *in vitro* conditions, where the protease effector domain can access this region and excise the translocation domain 1, during cell intoxication this domain is membrane-bound and therefore inaccessible to the protease. This result further corroborates to the idea that the

conformational change during membrane penetration precedes toxin activation inside the host cells.

### **N-terminal effector and protease effector domains are toxic in eukaryotic cells**

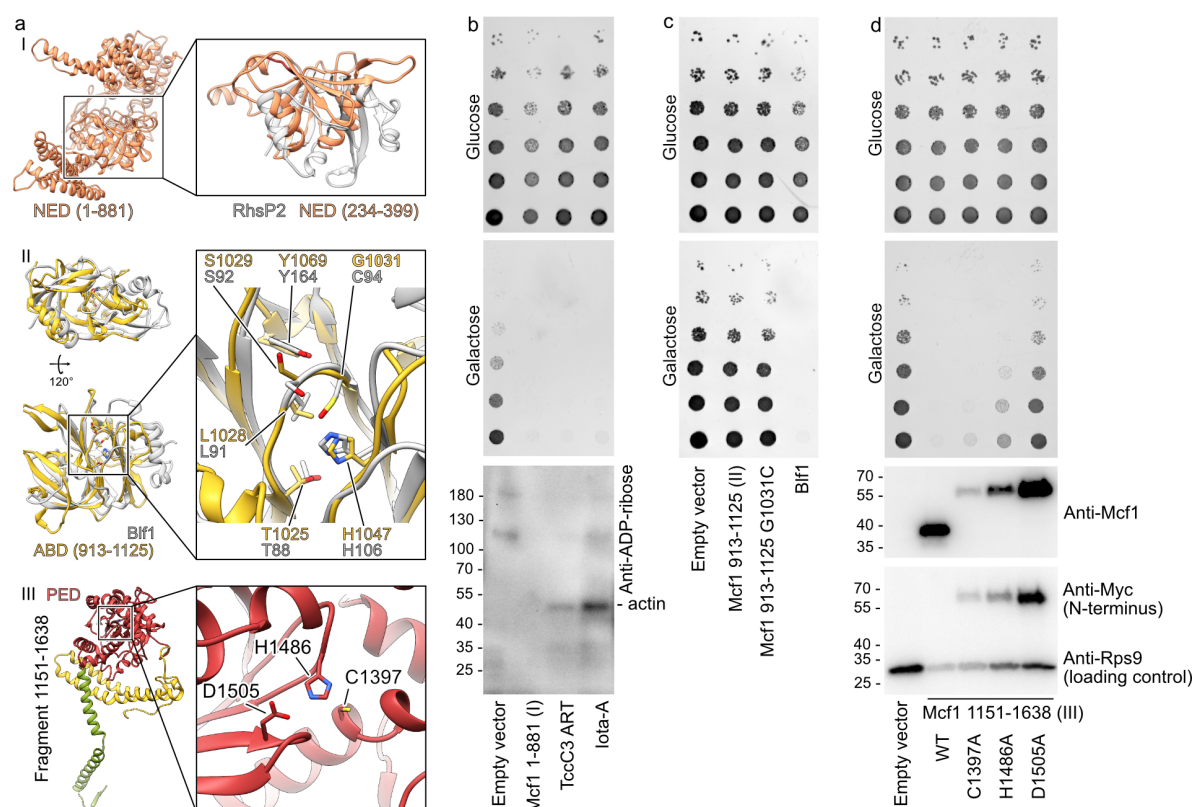
Intoxication of insect and mammalian cells with Mcf1 leads to apoptotic blebbing of membranes and fragmentation of nuclei<sup>11</sup> ([Supplementary Fig. 3d](#)). To find out which of the three fragments that Mcf1 excises from the membrane-embedded tail domain are actually cytotoxic in eukaryotic cells, we expressed these fragments in the *Saccharomyces cerevisiae* model that we previously established to study various bacterial effectors (i.e. TccC3<sup>30</sup>, ExoY<sup>31,32</sup> and Iota-A<sup>33</sup>).

Our analysis of yeast viability demonstrated that expression of the 105 kDa N-terminal effector domain was highly toxic ([Fig. 3b](#)). Interestingly, the N-terminal effector domain could not be split or truncated for more than a dozen amino acids without loss of function, revealing that, despite its relatively large size, this domain acts as a single toxic effector ([Supplementary Fig. 5a](#)). Based on the structural homology with ADP-ribosyltransferases ([Supplementary Fig. 2](#)), we tested whether expression of the Mcf1 effector leads to ADP-ribosylation of any proteins in yeast and whether the purified effector can ADP-ribosylate any insect cell molecules *in vitro*. While the ADP-ribosyltransferase TccC3 from *P. luminescens* was very active and rapidly modified actin *in vitro* and *in vivo*, we did not detect any ADP-ribosyltransferase activity of the Mcf1 N-terminal effector ([Supplementary Fig. 5b](#), [Fig. 3b](#)), indicating that it must have a different mechanism of action.

The second fragment released by Mcf1 corresponds to the 40 kDa activator-binding domain ([Fig. 2b and d](#)). Interestingly, this domain also contains a putative BH3-like domain (amino acids 911-925)<sup>12</sup>. Due to the function of BH3-like proteins in apoptosis, it has been speculated that this part of Mcf1 mediates toxicity in cells in a similar fashion. In addition, the activator-binding domain is structurally homologous to the *Burkholderia* lethal factor 1 (Blf1), which is a highly potent deamidase<sup>23</sup>. In contrast to the highly toxic Blf1, production of the activator-binding domain did not affect yeast growth ([Fig. 3c](#)). Yeast growth was also not affected by reintroduction of a catalytic cysteine, which is missing in the activator-binding domain but required for the deamidation reaction of Blf1 ([Fig. 3a](#)). Therefore, we conclude that this part of Mcf1 does not induce apoptosis and is in general not an active effector. This domain likely lost its enzymatic activity but remained an important structural element of the toxin.

Finally, we expressed a 56 kDa fragment (amino acids 1133-1619), which includes the complete protease effector domain, in yeast. Surprisingly, in addition to the fact that the domain

dramatically inhibited cell growth (Fig. 3d), the toxic protease effector domain was truncated from the N-terminus to 40 kDa (Fig. 3d). We then tested whether the reduced yeast growth was related to the proteolytic activity of the protease effector domain by mutating its catalytic residues. Our viability assays demonstrated that mutations in the catalytic center of the protease effector domain decreased cell toxicity (Fig. 3d). Therefore, our experiments suggest that the protease effector domain, matching its name, plays a dual function: it activates the Mcf1 toxin by proteolytic cleavage, and also directly mediates toxicity in cells. Interestingly, a similar protease domain from the *P. asymbiotica* PaTox toxin was shown to enhance the cytotoxic effects of the toxin, but did not show any toxic effects alone<sup>34</sup>. Therefore, a dual role of protease domains might be a key feature of *Photorhabdus* toxins that allows to increase their overall potency.



**Fig. 3. Toxicity of Mcf1 effectors.** a Structure of the three head domains and comparison with structural homologues of two of them<sup>21,23</sup>. b, c, d Growth phenotype assay with *S. cerevisiae* expressing Mcf1 fragments under galactose promoter in the experimental conditions with low (glucose) or high (galactose) Mcf1 fragment expression. The protein expression and the level of ADP-ribosylation was analyzed by western blot of cells grown on galactose-containing media with anti-Myc, anti-Mcf1, ADP-ribose binding reagent (anti-ADP-ribose) and anti-ribosomal protein S9 (RPS9) antibodies. Bif1 – *Burkholderia* lethal factor 1; NED – N-terminal effector domain; ABD – activator-binding domain; PED – protease effector domain; ART – ADP-ribosyltransferase. Western blot analysis was performed twice.



## The mechanism of Mcf1 activation

We have shown that the interaction of Mcf1 with Arf3 stimulates the autoproteolytic activity of the toxin and the release of three fragments, at least two of which induce cell death. However, the molecular mechanism of this activation remains unclear, since the first cleavage site is more than 20 Å away from the catalytic center of the protease effector domain. Thus, for the proteolytic reaction to occur, either the linker with the cleavage site or the protease effector domain must be moved ([Supplementary Fig. 6a](#)). To decipher the molecular mechanism of the activation process, we analyzed the Mcf1-Arf3 interaction by cryo-EM and single particle analysis. To obtain a stable complex, we used the C1397A Δ15C variant of Mcf1 to make sure that the protease is inactive and the C-terminus does not interfere with the interaction (see above) and determined structures of the Mcf1 variant in the absence and presence of Arf3 ([Supplementary Fig. 7, 8](#); [Supplementary Table 1](#)).

During particle selection and 2D classification, we noticed in both datasets that the Mcf1 head domain was clearly visible, while the tail domain was absent in most 2D classes or blurry in others ([Supplementary Fig. 7b and 8b](#)). This is in contrast to the single particle analysis of the wild type ([Supplementary Fig. 1b](#)) and suggests, that the removal of the last 15 amino acids destabilizes the neck regions leading to a high flexibility between the head and tail. Therefore, we only concentrated on the heads for the following analysis. During 3D refinement and classification, we noticed that the heads were also very heterogenous and the activator-binding domain was missing in most classes of Mcf1<sub>C1397A Δ15C</sub>. Nevertheless, we managed to determine a cryo-EM structure of the complete head of Mcf1<sub>C1397A Δ15C</sub> in absence of Arf3 at 3.5 Å resolution ([Supplementary Fig. 7b](#)).

In contrast to the head in absence of Arf3, the activator-binding domain was present in the Mcf1<sub>C1397A Δ15C</sub>-Arf3 map, and surprisingly, we observed an additional density close to the activator-binding domain and not to the protease effector domain. This suggests that Arf3 binds to the activator-binding domain and thereby stabilizes it. We then determined a cryo-EM structure of the complete head of Mcf1<sub>C1397A Δ15C</sub>-Arf3 at 4 Å resolution that let us unambiguously fit Arf3 in the additional density. The atomic model of the Mcf1-Arf3 interface revealed that Arf3 binds to all three helices of the activator-binding domain ([Supplementary Fig. 6c](#)), which is the same region where the C-terminus is located in wild-type Mcf1 ([Fig. 1g](#)). This observation provides a clear explanation why deletion of the extreme C-terminus enhanced autoproteolysis rate of Mcf1 ([Fig. 2d](#)) and therefore defines the C-terminus as important regulator of Mcf1 autoprocessing.

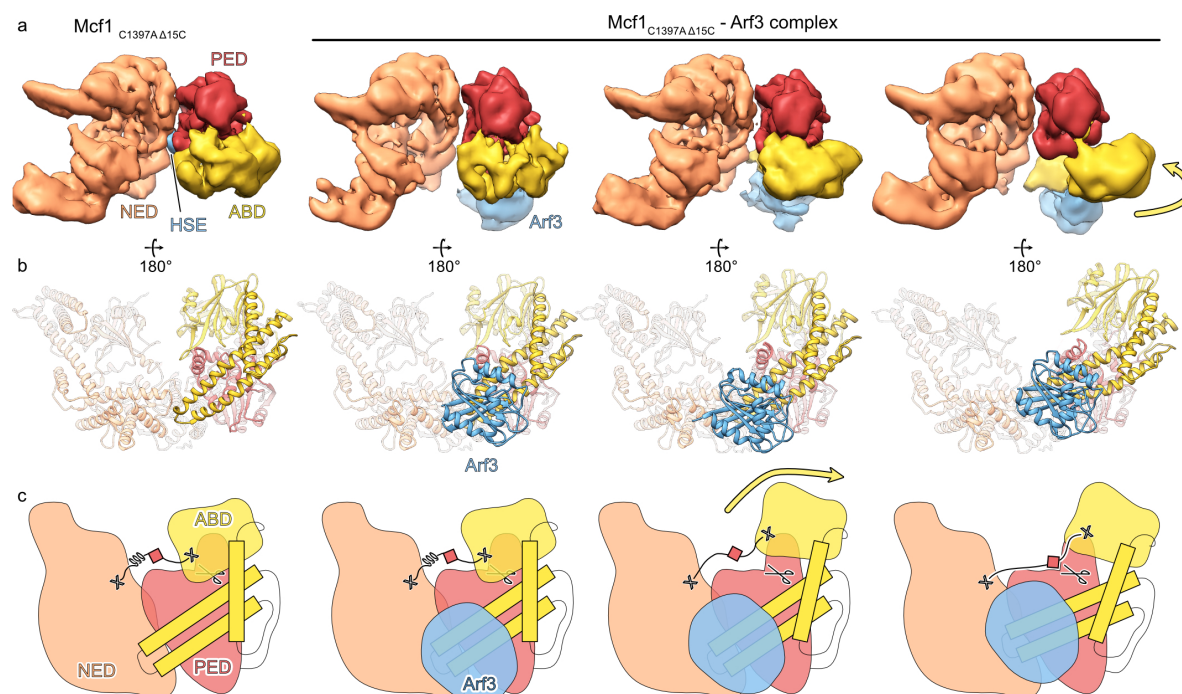
The comparison of the Mcf1<sub>C1397A Δ15C</sub>-Arf3 structure with Mcf1<sub>C1397A Δ15C</sub>, however, showed that the position of the protease domain is not altered and the cleavable linker is more or less in the same position, indicating that Arf3 binding does not induce direct conformational changes. We therefore went back to further analyze our 3D classes and found that Arf binding induces a swinging movement of the activator-binding domain away from the N-terminal effector domain (Fig. 4a, Supplementary Fig. 8b, Supplementary Movie 2). Taking into account that the N-terminal effector domain and the protease effector domain are structurally rigid, this movement pulls the cleavable linker towards the protease domain where it is proteolytically cleaved.

On a molecular level, we suggest that Arf3 binding strains the helices of the activator-binding domain and drives the movement of its N-terminal part (Fig. 4b and c). This movement is possible due to a hook-loop interaction between the activator-binding domain and the protease effector domain, organized by arginine 957 of the activator-binding domain and the loop 1291-1298 of the protease effector domain (Supplementary Fig. 6b). This structural feature allows the activator-binding domain to swing away from the N-terminal effector, and to pull the cleavable linker toward the catalytic center of the protease effector domain for proteolysis (Fig. 4c).

To test this mechanism, we first studied the role of the hook-loop interaction between the activator-binding and protease effector domains (Supplementary Fig. 6b). To this end, we engineered two Mcf1 mutants: in the first one we aimed to abolish the interaction between the activator-binding and protease effector domains by mutating arginine 957 to alanine, and in the second one, we introduced a mutation in tyrosine 1286 to make the loop more flexible and increase the freedom of movement of R957 during activation. When we then incubated these Mcf1 variants with Arf3, we observed nearly no autoprocessing of the R957A mutant and instant cleavage of the Y1286S variant (Supplementary Fig. 6d). These results clearly indicate the importance of the hook-loop interaction in coordinating the activator-binding domain movement and linker stretch.

Finally, we aimed to test the requirement of pulling the cleavable linker for Mcf1 activation. To this end, we engineered two Mcf1 mutants with one or two amino acid deletions (ΔK907 and ΔD908) upstream of the cleavage site to shorten the linker and another two Mcf1 mutants L1359G and R1360G to destabilize the rigid base that keeps the N-terminal effector and the protease effector domain rigid and immobile (Supplementary Fig. 6e). We then incubated these four mutants with Arf3 and found drastically reduced autoproteolytic activity (Supplementary Fig. 6f, g), thus demonstrating that the pulling of the linker is an essential step

in toxin autoproteolysis. Altogether, our structural and biochemical data reveal a unique activation mechanism of a bacterial toxin through allosteric displacement of the cleavage site toward the catalytic center of the protease induced by Arf (Fig. 4c).



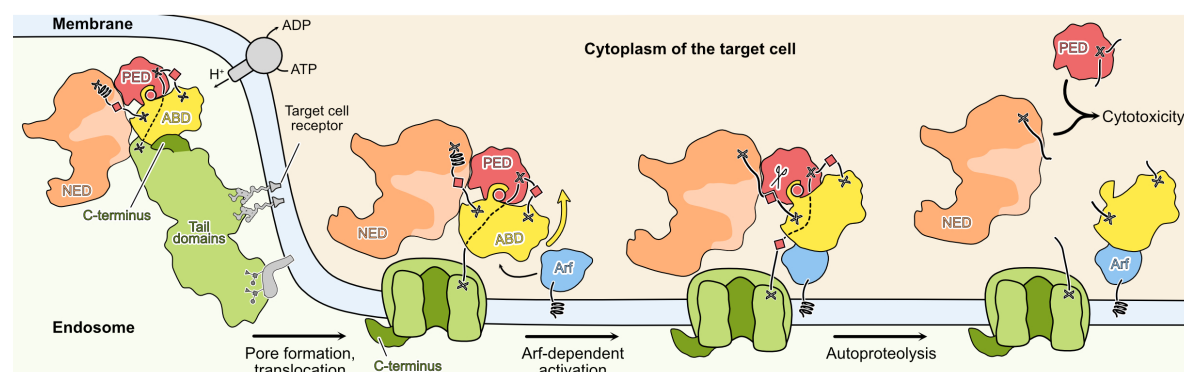
**Fig. 4.** Mechanism of activation of Mcf1. **a,b** 3D class averages (**a**) and molecular models (**b**) of the Mcf1<sub>C1397A Δ15C</sub> and Mcf1<sub>C1397A Δ15C</sub>-Arf3 complex datasets. **c** Schematic illustration of the mechanism of Mcf1 activation. Abbreviations: see Fig. 1.

## Discussion

In this study, we present a detailed structural and functional analysis of the Mcf1 toxin from *P. luminescens*, which allows us to confidently assign Mcf1 to the ABCD subfamily of bacterial toxins<sup>29</sup> next to TcdA and TcdB toxins from *C. difficile*, a major cause of health care-associated diarrhea and colitis<sup>35</sup>.

Mcf1 and large clostridial toxins share similarities in their domain architecture (Supplementary Fig. 2, 3a). However, Mcf possesses three compact receptor-binding domains instead of an extended and mobile CROPs domain, an additional activator-binding domain and a novel N-terminal effector with no homology to the glycosyltransferase of TcdA or TcdB<sup>36</sup>. Finally, we demonstrate that the protease domain, which is normally required only for autoprocessing, is a toxic effector of Mcf1 *per se*. Its mechanism of toxicity might be similar to the Mcf-like domain of *V. vulnificus* multifunctional-autoprocessing repeats-in-toxin

(MARTX), which induces cell apoptosis by cleaving C-terminal regions of Rab proteins<sup>15</sup>. Thus, Mcf1 provides an example of an evolutionary scramble between different toxins of different families and highlights the potential of its modular architecture for the design of customizable delivery systems with possible applications in cancer and gene therapy.



**Fig. 5.** Mechanism of action of Mcf1 toxin. Abbreviations: see Fig. 1.

Our structural and functional data together with previous analyses by other groups allow us to propose a molecular mechanism of action of Mcf1 (Fig. 5). After binding of the toxin to a yet-unknown protein and/or glycan receptors on the surface of the host cell, the toxin is endocytosed<sup>11</sup>. We propose that two positively charged regions on the side and the bottom of the toxin (Supplementary Fig. 9) form initial contacts with the negatively charged surface of the membrane. Acidification of the endosome triggers a conformational change inside the 600 amino acid-long delivery region, which forms a pore and translocates the head region of the toxin inside the target cell. After translocation, the activator-binding domain of Mcf1 has lost its interaction with the C-terminal region, which probably remains on the other side of the membrane. Therefore, the binding site for membrane-bound Arf proteins on the activator-binding domain becomes available for interaction. Formation of the Mcf1-Arf3 complex stimulates a swinging movement of the activator-binding domain that pulls the cleavable linker towards the catalytic site of the protease effector domain. After the first cleavage of the N-terminal effector domain, the protease effector domain catalyzes two more proteolysis reactions resulting in the release of the protease effector and activator-binding domain. Enzymatic activity of the Mcf1 effectors in the cytoplasm of the target cells impairs essential physiological pathways and ultimately leads to cell death. We believe that our new structural and mechanistic insights will help to engineer Mcf1 variants for ecological pest control.

## References

1. Xiao, Y. & Wu, K. Recent progress on the interaction between insects and *Bacillus thuringiensis* crops. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 374 Preprint at <https://doi.org/10.1098/rstb.2018.0316> (2019).
2. Palma, L., Muñoz, D., Berry, C., Murillo, J. & Caballero, P. *Bacillus thuringiensis* toxins: An overview of their biocidal activity. *Toxins* vol. 6 Preprint at <https://doi.org/10.3390/toxins6123296> (2014).
3. Bravo, A. & Soberón, M. How to cope with insect resistance to Bt toxins? *Trends in Biotechnology* vol. 26 Preprint at <https://doi.org/10.1016/j.tibtech.2008.06.005> (2008).
4. Meusch, D. *et al.* Mechanism of Tc toxin action revealed in molecular detail. *Nature* **508**, (2014).
5. Gatsogiannis, C. *et al.* Tc toxin activation requires unfolding and refolding of a  $\beta$ -propeller. *Nature* **563**, (2018).
6. Waterfield, N., Kamita, S. G., Hammock, B. D. & Ffrench-Constant, R. The *Photorhabdus* Pir toxins are similar to a developmentally regulated insect protein but show no juvenile hormone esterase activity. *FEMS Microbiol Lett* **245**, 47–52 (2005).
7. Vlisidou, I. *et al.* The *photorhabdus asymbiotica* virulence cassettes deliver protein effectors directly into target eukaryotic cells. *Elife* **8**, (2019).
8. Daborn, P. J. *et al.* A single *Photorhabdus* gene, makes caterpillars floppy (mcf), allows *Escherichia coli* to persist within and kill insects. *Proc Natl Acad Sci U S A* **99**, (2002).
9. Roderer, D., Schubert, E., Sitsel, O. & Raunser, S. Towards the application of Tc toxins as a universal protein translocation system. *Nat Commun* **10**, (2019).
10. Kreitz, J. *et al.* Programmable protein delivery with a bacterial contractile injection system. *Nature* (2023) doi:10.1038/s41586-023-05870-7.
11. Dowling, A. J. *et al.* The insecticidal toxin Makes caterpillars floppy (Mcf) promotes apoptosis in mammalian cells. *Cell Microbiol* **6**, (2004).
12. Dowling, A. J. *et al.* The Mcf1 toxin induces apoptosis via the mitochondrial pathway and apoptosis is attenuated by mutation of the BH3-like domain. *Cell Microbiol* **9**, (2007).



13. Herrera, A. *et al.* N-terminal autoprocessing and acetylation of multifunctional-autoprocessing repeats-in-toxins (MARTX) Makes Caterpillars Floppy-like effector is stimulated by adenosine diphosphate (ADP)-Ribosylation Factor 1 in advance of Golgi fragmentation. *Cell Microbiol* **22**, (2020).
14. Lee, Y. *et al.* Makes caterpillars floppy-like effector-containing MARTX toxins require host ADP-ribosylation factor (ARF) proteins for systemic pathogenicity. *Proc Natl Acad Sci U S A* **116**, (2019).
15. Herrera, A. *et al.* Vibrio MARTX toxin processing and degradation of cellular Rab GTPases by the cytotoxic effector Makes Caterpillars Floppy. doi:10.1101/2023.04.19.537381.
16. Orrell, K. E., Mansfield, M. J., Doxey, A. C. & Melnyk, R. A. The *C. difficile* toxin B membrane translocation machinery is an evolutionarily conserved protein delivery apparatus. *Nat Commun* **11**, (2020).
17. Waterfield, N. R. *et al.* The insecticidal toxin Makes caterpillars floppy 2 (Mcf2) shows similarity to HrmA, an avirulence protein from a plant pathogen. *FEMS Microbiol Lett* **229**, (2003).
18. Aktories, K., Schwan, C. & Jank, T. Clostridium difficile Toxin Biology. *Annu Rev Microbiol* **71**, (2017).
19. Just, I. *et al.* Glucosylation of Rho proteins by Clostridium difficile toxin B. *Nature* **375**, (1995).
20. Just, I. *et al.* The enterotoxin from Clostridium difficile (ToxA) monoglucosylates the Rho proteins. *Journal of Biological Chemistry* **270**, (1995).
21. Bullen, N. P. *et al.* An ADP-ribosyltransferase toxin kills bacterial cells by modifying structured non-coding RNAs. *Mol Cell* **82**, 3484-3498.e11 (2022).
22. Jurėnas, D. *et al.* Salmonella antibacterial Rhs polymorphic toxin inhibits translation through ADP-ribosylation of EF-Tu P-loop. *Nucleic Acids Res* **50**, 13114–13127 (2022).
23. Cruz-Migoni, A. *et al.* A Burkholderia pseudomallei toxin inhibits helicase activity of translation factor eIF4A. *Science* **334**, 821–824 (2011).
24. Choe, S. *et al.* The crystal structure of diphtheria toxin. *Nature* **357**, (1992).
25. Reineke, J. *et al.* Supplementary information: Autocatalytic cleavage of Clostridium difficile toxin B. *Nature* **446**, (2007).

26. Egerer, M., Giesemann, T., Jank, T., Fullner Satchell, K. J. & Aktories, K. Auto-catalytic cleavage of *Clostridium difficile* toxins A and B depends on cysteine protease activity. *Journal of Biological Chemistry* **282**, (2007).
27. Pruitt, R. N. *et al.* Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in *Clostridium difficile* toxin A. *Journal of Biological Chemistry* **284**, (2009).
28. Casalou, C., Ferreira, A. & Barral, D. C. The Role of ARF Family Proteins and Their Regulators and Effectors in Cancer Progression: A Therapeutic Perspective. *Frontiers in Cell and Developmental Biology* vol. 8 Preprint at <https://doi.org/10.3389/fcell.2020.00217> (2020).
29. Jank, T. & Aktories, K. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. *Trends in Microbiology* vol. 16 Preprint at <https://doi.org/10.1016/j.tim.2008.01.011> (2008).
30. Belyy, A. *et al.* Mechanism of threonine ADP-ribosylation of F-actin by a Tc toxin. *Nat Commun* **13**, (2022).
31. Belyy, A., Merino, F., Mechold, U. & Raunser, S. Mechanism of actin-dependent activation of nucleotidyl cyclase toxins from bacterial human pathogens. *Nature Communications* 2021 12:1 **12**, 1–9 (2021).
32. Belyy, A. *et al.* Actin activates *Pseudomonas aeruginosa* ExoY nucleotidyl cyclase toxin and ExoY-like effector domains from MARTX toxins. *Nature Communications* 2016 7:1 **7**, 1–14 (2016).
33. Belyy, A. *et al.* Roles of Asp179 and Glu270 in ADP-ribosylation of actin by *Clostridium perfringens* Iota Toxin. *PLoS One* **10**, (2015).
34. Bogdanovic, X. *et al.* A cysteine protease-like domain enhances the cytotoxic effects of the *Photobacterium asymbiotica* toxin PaTox. *Journal of Biological Chemistry* **294**, 1035–1044 (2019).
35. Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. & Kuijper, E. J. *Clostridium difficile* infection. *Nat Rev Dis Primers* **2**, 1–20 (2016).
36. Jank, T., Belyi, Y. & Aktories, K. Bacterial glycosyltransferase toxins. *Cell Microbiol* **17**, 1752–1765 (2015).
37. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**, (2002).



38. Kushnirov, V. V. Rapid and reliable protein extraction from yeast. *Yeast* **16**, (2000).
39. Stabrin, M. *et al.* TranSPHIRE: automated and feedback-optimized on-the-fly processing for cryo-EM. *Nat Commun* **11**, (2020).
40. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* vol. 14 Preprint at <https://doi.org/10.1038/nmeth.4193> (2017).
41. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, (2015).
42. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, (2019).
43. Moriya, T. *et al.* High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *Journal of Visualized Experiments* **2017**, (2017).
44. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
45. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, (2010).
46. Frenz, B., Walls, A. C., Egelman, E. H., Veessler, D. & Di Maio, F. RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat Methods* **14**, (2017).
47. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, (2018).
48. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, (2019).
49. Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol* **4**, (2021).
50. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, (2017).
51. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**, (2004).

52. López-Blanco, J. R. & Chacón, P. IMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J Struct Biol* **184**, (2013).

## Methods

### Plasmids, bacteria and yeast strains

The complete list of oligonucleotides, plasmids and strains used in this manuscript can be found in [Supplementary Table 2](#). *E. coli* were cultured in LB medium supplemented with kanamycin or ampicillin. *S. cerevisiae*, were cultivated on synthetic defined medium (Yeast nitrogen base, Difco) containing galactose or glucose and supplemented if required with histidine, uracil, tryptophan, adenine, or leucine. Yeast transformation was performed using the lithium-acetate method<sup>37</sup>. Yeast viability upon expression of recombinant genes was analyzed by serial 5-fold dilution of OD<sub>600</sub>-normalized yeast culture<sup>33</sup>. To analyze protein production in yeast, the cell lysate was first prepared by treatment of cells with 0.2 M NaOH for 5 minutes and boiling in 4-fold Laemmli buffer<sup>38</sup>. Then, the extract was separated by SDS-PAGE, followed by blotting on PVDF membrane, and incubation with anti-Myc (dilution 1 to 10000, clone 9B11 #2276, Cell Signaling Technology), anti-Mcf1 (dilution 1 to 10000, custom-made rabbit polyclonal antibodies), ADP-ribose binding reagent (Anti-pan-ADP-ribose binding reagent, dilution 1 to 3000, MABE1016, Merck) or anti-RPS9 serum (dilution 1 to 10000, polyclonal rabbit antibodies were a generous gift of Prof. S. Rospert), and visualized using secondary anti-mouse HRP (dilution 1 to 3000, #1706516, Bio-Rad) or anti-rabbit HRP-antibodies (dilution 1 to 3000, #1706515, Bio-Rad).

### Recombinant protein production and purification

Mcf1 and its variants were expressed overnight at 22 °C in the presence of 0.02 mM IPTG in *E. coli* BL21-CodonPlus (DE3)-RIPL cells transformed with plasmids listed in [Supplementary Table 2](#). Following the lysis by ultrasonication in the buffer containing 20 mM Tris pH 8 and 100 mM NaCl, the soluble fraction was loaded onto Ni-IDA resin, washed with the lysis buffer, and eluted with the same buffer supplemented with 250 mM imidazole. The eluates were loaded on a monoQ 5/50 column and eluted with a linear gradient from 100 to 500 mM NaCl in 20 mM Tris-HCl pH 8, and stored in aliquots at −20 °C.

Arf3 with the N-terminal deletion of 17 amino acids and a mutation Q71L was expressed overnight at 21 °C in the presence of 0.02 mM IPTG in *E. coli* BL21-

CodonPlus(DE3)-RIPL cells possessing the corresponding plasmid. After lysis by sonication, and centrifugation, the lysate supernatant was purified on Ni-IDA resin and a monoQ 5/50 column as described above for Mcf1. Finally, the sample was supplemented with GTP and MgCl<sub>2</sub> to concentrations of 1 and 10 mM, respectively, and stored in aliquots at −20 °C.

TccC3 ART was purified as described previously<sup>30</sup>.

### **Mcf1 cleavage assays**

In the total volume of 6 µl, 4 µg of Mcf1 in the 20 mM Tris pH 8 and 100 mM NaCl buffer were mixed with 4 µg of Arf3 Q71L in the buffer containing 20 mM Tris pH 8, 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 1 mM GTP, and incubated for the indicated amount of time at 22 °C. The reaction was stopped by addition of 4-fold Laemmli buffer. The samples were analyzed by SDS-PAGE and western blot with anti-FLAG (dilution 1 to 5000, clone M2, #F3165, Sigma Aldrich) or anti-His primary antibodies (dilution 1:3000, #H1029 Sigma Aldrich).

### **Intoxication assays**

Sf9 cells at a concentration of 1 million/ml were intoxicated with the wild-type or C1397A Mcf1 variant. If specified, Sf9 cells were treated with bafilomycin A1. Cells were imaged after 24 h incubation at 28 °C.

For the analysis of Mcf1 cleavage during intoxication, 5 ml of Sf9 cells at the concentration of 4-5 million/ml were intoxicated with 5 nM of Mcf1 and incubated at 28 °C gently shaking for the indicated amount of time. Cells were centrifuged at 1000 g, washed three times with 20 mM Tris pH 8, 100 mM NaCl buffer, and lysed by passing the suspension for 10 times through a needle with 26 gauge in diameter. The lysate was clarified by 2 min centrifugation at 1000 g, and mixed with Laemmli buffer. The samples were analyzed by SDS-PAGE, blotted onto PVDF membranes and stained with anti-Myc antibodies (dilution 1 to 2000, clone 9B11 #2276, Cell Signaling Technology).

### **Mass spectrometry**

To identify the cleavage site that separates the 105 and 225 kDa Mcf1 fragments, we performed mass-spectrometry analysis of the N-terminal effector polypeptide. To this end, we first co-expressed the full-length His-tagged Mcf1 and Arf3 Q71L in *E. coli* (Supplementary Fig. 4c), lysed the cells in 20 mM HEPES pH 7.5, 100 mM NaCl, and applied the supernatant to the Ni-IDA beads. After washing with lysis buffer and elution with 250 mM imidazole, we diluted the sample twice with 20 mM HEPES pH 7.5, loaded the eluate on the Mono S 5/50 column and

eluted with NaCl gradient. The fractions were loaded on SDS-PAGE, and the bands corresponding to the N-terminal effector were cut out and analyzed in quadruplicates by LC-MS/MS at Proteome Factory AG. The mass-spectrometry data is available in [Supplementary Dataset 1](#).

### **ADP-ribosylation assays**

30  $\mu$ l of Sf9 insect cell lysate with total protein concentration 3 mg/ml was mixed with 3  $\mu$ g of Mcf1 N-terminal effector or 2  $\mu$ g of TccC3 ART. The reaction was initiated by addition of biotinylated NAD to 16.7  $\mu$ M and then incubated for 30 min at 30 °C. After separation on SDS-PAGE and blotting on PVDF membrane, ADP-ribosylation was visualized by incubating the membrane with HRP-streptavidin (dilution 1 to 5000, Pierce #21130).

### **Common cryo-EM techniques**

All cryo-EM samples were prepared using glow-discharged copper C-flat 2-1 400 mesh grids. After application on the grid, samples were blotted at 100% humidity for 3 s (13°C, 0 s drain time, blot force -3 to +2) with a Vitrobot Mark IV (Thermo Fisher Scientific) and plunged into liquid ethane. All datasets were collected on a Titan Krios at 300 kV using a K3 detector (Gatan) with nominal magnification of 105,000 x in super-resolution mode (physical pixel size 0.9 Å), using the EPU software version 2.7 or 2.8 (Thermo Fisher Scientific). The defocus ranged from -1 to -2.5  $\mu$ m. The preprocessing of data was carried out on-the-fly with the TransPHIRE software (v1.4,<sup>39</sup>), including drift correction and dose-weighting with MotionCor2 (v1.3.0,<sup>40</sup>), as well as CTF estimation with CTFFIND4 (v4.1.13,<sup>41</sup>).

### **Cryo-EM of full-length Mcf1**

3  $\mu$ l of the freshly purified Mcf1 sample at a concentration of 1.1 mg/ml, supplemented with 0.002% of Tween-20, were applied onto a grid, blotted and plunged into liquid ethane.

We collected two datasets on a Titan Krios, equipped with an energy filter (GIF BioQuantum, Gatan) set to a slit width of 20 eV. Exposure time of 3 s over 60 frames resulted in an accumulated dose of 71 e<sup>-</sup>/Å<sup>2</sup> for the first and 69 e<sup>-</sup>/Å<sup>2</sup> for the second dataset. 7,907 and 15,000 movies were collected for the first and the second datasets respectively. The last 5,955 micrographs of the second dataset were collected with 30° stage tilt to overcome the preferred orientation of Mcf1 inside the thin vitreous ice layer.

The acquired micrographs were manually inspected, and images with large ice contaminations, excessive drift or micrographs with a resolution limit above 6 Å were

discarded. Particles were picked using SPHIRE-crYOLO (version 1.6<sup>42</sup>), with a model trained on 10 hand-picked micrographs. The selected particles (787,549 from 7,653 micrographs of the first dataset, and 1,983,207 from 14,088 micrographs of the second dataset) were extracted with a box size of 300 pixels and used in a 3D refinement in MERIDIEN (SPHIRE package version 1.4<sup>43</sup>) with coarse shifts (search range 25, step range 5 pixels). The calculated projection parameters were used to recenter and re-extract properly centered particles with a box size of 320 pixels. The centered particles were 2D classified using the iterative stable alignment and clustering approach ISAC (SPHIRE version 1.4; pre-alignment was switched off and translation search range was set to 0) to remove erroneous picks.

The following processing steps were performed differently for the first and second datasets (Supplementary Fig. 1). For the first dataset, the 787,549 particles from 2D classification were used for a 3D refinement with MERIDIEN and resulted in a 4.3 Å reconstruction. Alternating particle polishing in RELION version 3.1<sup>44</sup> with 3D refinement in SPHIRE improved the average resolution of the reconstruction to 3.4 Å. This density, as well as the density obtained by signal subtraction of the head was used to build the C-terminal half of the protein. We noticed that the density corresponding to the N-terminal part of the toxin was of lower resolution and appeared at lower visualization thresholds. Therefore, the particles and their projection parameters were imported into RELION 3.1 and were subjected to 3D classification without image alignment (T=4, two classes, 25 iterations), with a soft mask covering the N-terminal half of the protein and a reference reconstruction low-pass filtered to 20 Å. Out of 363,044 particles, only 47,522 possessed the density corresponding to the N-terminal region. The low number of such particles and their preferred orientation in the thin vitreous ice layer hindered us from obtaining a high-resolution reconstruction for robust model building of the N-terminal part of the toxin. Therefore, we acquired a second, larger dataset with 14,088 micrographs, including 5,955 collected with a stage tilt of 30°.

Applying the same pre-processing strategy as for the first dataset, 711,705 particles were selected for the first 3D refinement, which resulted in a reconstruction with an average resolution of 4.6 Å with more isotropy. To obtain a high-resolution reconstruction of the head region, we similarly performed a 3D classification in RELION with a mask covering the head region and obtained 219,318 particles. These particles were combined with the particles from the first dataset after 3D classification. The following rounds of signal subtraction, 3D classification, 3D refinement and particle polishing, resulted in the final reconstruction at the resolution of 3.6 Å, which was used to build the N-terminal part of the toxin.

The complete composite model of Mcf1 was built *de novo* in three steps. First, using one original and one signal subtracted reconstruction obtained from the first dataset, we built the tail domain. Secondly, we used the reconstruction from particles of the first and the second dataset to build the head domain except for residues 216-436. Finally, we took the structure of the missing residues from the Mcf1<sub>C1397A Δ15C</sub> structure and merged them together to generate the complete Mcf1 structure. The model building was performed manually in Coot 0.8.9.2<sup>45</sup> or automatized by trRosetta version 3<sup>46</sup> and further refined in Isolde 1.4<sup>47</sup> and PHENIX version 1.16<sup>48</sup>.

The composite cryo-EM density for Fig. 1a was postprocessed using DeepEMhancer<sup>49</sup>.

### Cryo-EM of Mcf1<sub>C1397A Δ15C</sub>

Freshly purified protein was diluted to 0.7 mg/ml (2.2 μM) and mixed with Tween-20 to the final concentration of 0.008% to improve ice quality. 3 μl of the sample were applied onto a grid, blotted and plunge-frozen in liquid ethane.

The dataset was collected using a Titan Krios, equipped with a postcolumn energy filter (slit width of 15 eV). Image stacks with 60 frames were collected with a total exposure time of 3.5 s and a total dose of 60.85 e<sup>-</sup>/Å<sup>2</sup>. 9,032 images were acquired and 6,987 of them were used for processing. After on-the-fly preprocessing, the particle picking was performed using SPHIRE-crYOLO version 1.7. 2,202,784 particles were then extracted in RELION 3.1.0 with pixel size 1.8 Å and box size 150 pixels, and transferred into CryoSPARC version 4.1.0<sup>50</sup>. After 3 independent 2D classifications (in 150, 175 or 200 classes) and *ab initio* 3D refinement (in 3 classes), we ended up with 544,954 clean particles and a new *ab initio* 3D model, which we later refined to 3.72 Å reconstruction by non-uniform refinement. Then, we reextracted and unbinned particles to a pixel size 0.9 Å, performed two series of particle polishing in RELION 3.1.0, and finally computed the 3.08 Å resolution reconstruction using local refinement with a mask covering the N-terminal effector and the protease effector domain.

In the next step, we took coordinates of the particles used in the reconstruction and used them to train a new crYOLO picking model. Picking resulted in 2,181,255 particles, which were extracted in RELION 3.1.0 with a pixel size of 1.8 Å and a box size of 150 pixels and transferred into CryoSPARC version 4.1.0. After a 2D classification and a removal of junk particles, we performed a non-uniform 3D refinement with 904,685 clean particles, followed by a round of alignment-free 3D classification in 10 classes. Particles from class 7, which showed a complete activator-binding domain, were unbinned, and subjected to two rounds of non-uniform refinement in CryoSPARC and polishing in RELION. Finally, after the last clean-



up by alignment-free 3D classification, we ended up with 53952 particles, refining to 3.46 Å resolution reconstruction covering the N-terminal effector domain, protease effector domain and activator-binding domain of Mcf1.

To build the atomic models, we rigid-body fit the previously built structure of the Mcf1 head into the density in UCSF Chimera version 1.14<sup>51</sup>, build the missing amino acids and refined the structure in Isolde version 1.4<sup>47</sup>, trRosetta software version 3<sup>46</sup> and Phenix<sup>48</sup> version 1.17.

### **Cryo-EM of the Mcf1<sub>C1397A Δ15C</sub>-Arf3 complex**

Freshly purified Mcf1<sub>C1397A Δ15C</sub> at 2 mg/ml (6.2 μM) was mixed with 5-fold molar excess of Arf3 Q71L, diluted to 0.5 mg/ml, and mixed with Tween-20 to the final concentration of 0.006% to improve ice quality. The sample was applied onto a grid, blotted and plunge-frozen in liquid ethane.

The dataset was collected using a Titan Krios, equipped with a C<sub>s</sub>-corrector and a postcolumn energy filter (slit width of 15 eV). Image stacks with 60 frames were collected with the total exposure time of 2 s and a total dose of 62.6 e<sup>-</sup>/Å<sup>2</sup>. 18,671 images were acquired and 15,693 of them were used for processing. After on-the-fly preprocessing, the particle picking was performed using crYOLO version 1.7. 1,491,851 particles were then extracted in RELION 3.1.0 with pixel size 1.8 Å and box size 150 pixels, and transferred into CryoSPARC version 4.1.0. After 3 independent 2D classifications (in 100, 150 or 200 classes) and *ab initio* 3D refinement (in 3 classes), we ended up with 770,049 clean particles and a new *ab initio* 3D model, which we later refined to 4.3 Å reconstruction by non-uniform refinement, followed by a round of alignment-free 3D classification in 10 classes. Then, we took class 5 with the largest number of particles and proceeded with a non-uniform refinement and unbinning to the pixel size of 0.9 Å. The particles were refined once again in the original pixel size in CryoSPARC and polished in RELION. The final 3D refinement in CryoSPARC resulted in the reconstruction of 4 Å resolution, which we used for model building.

To build the atomic model of the Mcf1<sub>C1397A Δ15C</sub>-Arf3 complex, we rigid-body fitted the Mcf1<sub>C1397A Δ15C</sub> and Arf3 models from PDB 6II6<sup>14</sup> into the density, and then performed a flexible fitting using iMODFIT Chimera plugin<sup>52</sup>, version 1.2. The resulting model was refined in Isolde version 1.4<sup>47</sup> and Phenix version 1.17<sup>48</sup>.



## Acknowledgements

We thank S. Bergbrede and M. Hülseweh for the excellent technical assistance, S. Rospert for providing us with anti-RPS9 serum, A. Rai for advice on biochemistry of Arf proteins and providing us with Arf6 for initial tests, W. Oosterheert, C. Gatsogiannis, O. Sitsel and Y. Belyi for fruitful discussions. This work has been funded by the Max Planck Society (S.R.). A.B. was supported by an EMBO long-term fellowship and a stipend of the Humboldt foundation.

## Author contribution

S.R. and A.B. designed the project. A.B. and P. Heilen performed wet lab experiments. P. Hagel performed initial cryo-EM experiments. A.B. prepared cryo-EM specimens. O.H. collected cryo-EM data. A.B. and P. Heilen analyzed cryo-EM data and built atomic models. A.B., P. Heilen and S.R. wrote the manuscript with input from all co-authors.

## Data availability

The coordinates for the cryo-EM structures of the full-length Mcf1, Mcf1<sub>C1397A Δ15C</sub>, and the Mcf1<sub>C1397A Δ15C</sub>-Arf3 complex have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-17440 (composite map; original and low-resolution consensus maps – EMD-17437, 17438, 17437, 17450), EMD-17436 and EMD-17435. The corresponding molecular models have been deposited at the wwPDB with accession codes PDB 8P52, 8P51 and 8P50. Uncropped gels, western blots and corresponding loading controls can be found in [Supplementary Fig. 10](#).