PNAS
www.pnas.org

Supplementary Information for

Structure–function characterization of an insecticidal protein GNIP1Aa, a member of an MACPF and β-tripod families

Jelena Zaitseva, Daniel Vaknin, Christian Krebs, James Doroghazi, Sara L. Milam, Deepa Balasubramanian, Nicholas B. Duck, Joerg Freigang

Jelena Zaitseva
Email: jelena.zaitseva@agro.basf-se.com

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1815547116

Supplementary Methods

GNIP1Aa overexpression and purification

The construction of *the MBP-gnip1Aa* plasmid was described (1). Briefly, the open reading frame of *gnip1Aa* was cloned into an *E. coli* expression vector based on pMAL-c4X. The final protein fusion consists of three parts: the full-length wild type GNIP1Aa protein, the N-terminally attached maltose-binding protein (MBP), and a linker in between with Factor Xa recognition site.

For crystallization, the wild-type GNIP1Aa was over-expressed as the MBP-GNIP1Aa fusion and purified as described (1) with slight modifications. Briefly, BL21 Star™ (DE3) cells were freshly transformed with *MBP-gnip1Aa* and grown at 37ºC in Luria Broth supplemented with 0.2% glucose in the presence of carbenicillin (100 mg/L). Overexpression of protein was induced at OD_{600} 0.6 by addition of 1 mM IPTG. Cells were grown at 20ºC for the next 20 hours and then collected by centrifugation at 4°C. The following purification steps were performed at 4°C.

Cells were resuspended in Buffer A [20 mM Tris-Cl pH 8, 200 mM NaCl, 1 mM EDTA, 1mM DTT] and lysed by three passes through a French Press cell at 16,000 psi. A lysed cell suspension was spun at 150,000 x g for 1 hr. Clarified supernatant was applied onto MBPTrap HP column. The protein was eluted with 10 mM maltose in the loading Buffer A. Protein cleavage was performed overnight at 4°C at 1:100 of the fusion protein-to-Factor Xa (M/M) in the elution buffer supplemented with 2 mM $CaCl_2$.

The resulting protein solution was dialyzed against [20 mM Na-acetate pH 5.0, 50 mM NaCl, 20% glycerol] and loaded on a HiTrap SP Sepharose HP column. GNIP1Aa was eluted with 0 - 0.5 M linear NaCl gradient and concentrated to 50-100 mg/mL in the elution buffer in an Amicon Ultra-15 Centrifugal Filter Units with Ultracel-10 membrane. The final protein is stable for a few months when stored at +4ºC and doesn't show any sign of precipitation and/or degradation during storage.

For the liposome assays, the wild-type GNIP1Aa, its mutants, and WCR- inactive Plu-MACPF were expressed and purified essentially the same as described above, except that the last purification step, the cation-exchange chromatography on a HiTrap SP Sepharose HP column, was performed in [20 mM Na-acetate pH 5.0, 50 mM NaCl] without glycerol.

Crystallization

For crystallization, protein solution was diluted to 5-10 mg/mL with distilled deionized water. Initial screening was performed with Crystallization Screens from Hampton Research. Crystals were grown by sitting-drop vapor diffusion at 18 °C at 1:1 ratio of protein and reservoir solution in a drop. The reservoir solution contained [100 mM citrate pH 5.6, 2% Tacsimate pH 5.0, and 14% (w/v) PEG 3350]. Crystals appeared overnight and grew to the final size of 0.3-0.5 mm in 2 to 4 days.

To obtain heavy-metal derivatives, native crystals were soaked in reservoir solution containing 10 mM thiomersal for 14 days.

For cryoprotection the crystals were transferred to the reservoir crystallization solution with 25% (v/v) ethylene glycol for 5 min. before flash freezing in liquid nitrogen.

Structure determination

Crystallographic data were collected on an Xcalibur Nova system from Oxford Diffraction. XDS (2) was used for data integration and scaling. Phases were determined by SIRAS using Crank (Crunch2/BP3/Solomon) (3), which identified the positions of four mercury atoms in the asymmetric unit of a thiomersal treated crystal. The model was built using Coot (4), refined with Refmac (5) against native data. An example for the resulting electron density map is given in Fig. S5. Crystals belong to the space group $P2_12_12_1$ with four GNIP1Aa monomers in the asymmetric unit (Table S1).

Plasmid constructions to test protein activity in bio-assays

To analyze activity of GNIP1Aa and its variants, a construct expressing untagged full-length wild type protein was used, *gnip1Aa* plasmid. The plasmid was constructed on the basis of pRSF1b-vector, carrying kanamycin resistance (Novagen/Millipore). The synthetic gene *gnip1Aa* was assembled from synthetic oligonucleotides and/or PCR products by GeneArt. The 1617 bp fragment was cloned into an expression vector using PstI and AscI cloning sites, and placed under control of T7*lac* promoter, resulting in *gnip1Aa* plasmid. The *plu-macpf* plasmid, expressing the negative control, was created the same way, using synthetic *plu-macpf* gene.

To generate 101 protein variants for the Ala-scanning of the C-terminal domain of GNIP1Aa, the required DNA fragments were assembled from synthetic oligonucleotides and/or PCR products by GeneArt. The final 853 bp products were cloned into *gnip1Aa* plasmid, using StuI and AscI cloning sites and replacing the original part of GNIP1Aa gene.

Substitution of other selected positions with alanine and generation of 4 double and 1 triple Ala-mutants in the C-terminal domain of GNIP1Aa was performed using *gnip1Aa* plasmid and QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's instructions. Total of 24 variants were created in-house.

All resulting plasmids were verified by DNA sequencing.

Protein expression for bioassays

To express GNIP1Aa protein, its mutants, and the negative control Plu-MACPF, BL21 Star™ (DE3) were transformed with a corresponding plasmid. For cell culture growth, Overnight Express™ Instant TB Medium (Novagen/Millipore) with kanamycin (50 mg/L) was inoculated with a few freshly transformed colonies and grown at 37°C at 275 rpm for 22 hours. Typically, 64-128 cultures were grown at the same time.

The following steps were performed at 4°C. The cells were collected by centrifugation and resuspended in a desired volume of [20 mM Tris-Cl pH 8.0]. To reliably identify the negative effect of alanine substitutions, the cell suspensions were concentrated 3-fold (the final volume of the tested suspensions was reduced to 1/3 of the original volume of cell cultures).

Suspensions were frozen and kept at -20°C for at least 12 hours or stored at this temperature up to a month. Immediately before submission to bioassays, samples were completely thawed, thoroughly resuspended, and 200 µL of each sample was submitted to bio-assay. The level of a target protein expression in an individual culture was evaluated by SDS-PAGE. The average concentration of the wild type GNIP1Aa in WCR

bio-assays was estimated to be around 90-120 μg/cm^2 diet (75-100 μM in the submitted samples), which was intentionally higher than previously calculated LC_{50} of ~60 μg/cm^2 diet (1).

Liposome Preparation

Liposomes were prepared from L-α-phosphatidylcholine (Egg PC 99%; 5 mg; Avanti Polar Lipids, Inc., Alabaster, Alabama). Dissolved in chloroform lipid was dried down under a stream of nitrogen gas. The resulting film was hydrated in 250 μl of [50 mM HEPES pH 7.4, 100 mM NaCl, 0.5 mM EDTA] and 12.5 mg calcein dye (3,3-bis[N, N-bis(carboxymethyl)-aminomethyl] fluorescein; Invitrogen/Thermo Fisher Scientific, Waltham, MA). The lipid suspension was vortexed, sonicated, and extruded using an Avanti Mini-Extruder with a 100 nm pore size filter (Avanti Polar Lipids, Inc., Alabaster, Alabama) in order to generate large, unilamellar liposomes. To remove the untrapped dye, the lipid suspension was loaded onto a Superdex 75 10/300 GL column (GE Healthcare Biosciences, Piscataway, NJ, USA) equilibrated with [50 mM HEPES pH 7.4, 100 mM NaCl, 0.5 mM EDTA], and eluted with the same buffer.

Liposome leakage assays

For liposome assays a purified (unactivated) protein was proteolytically activated: a protein sample was incubated with trypsin (Sigma-Aldrich, St. Louis, MO) at 1:25 (w/w) ratio in [20 mM Na-acetate pH 5.0, 50 mM NaCl] at 4 °C overnight. The reaction was stopped with 1 mM phenylmethanesulfonyl fluoride (PMSF; Sigma-Aldrich, St. Louis, MO), resulting in a trypsin-activated protein.

Protein (10 μl of 0.125-1 μM) was added to 200 μl of mixture, containing 0.5μl calcein-encapsulated PC liposomes and 199.5 μl [50 mM CHES pH 9.0] (10-50 μM lipid). The release of calcein was monitored by change in fluorescence using a Synergy H1 Hybrid Multi-Mode Microplate Reader (BioTek, Winooski, VT) for up to 20 minutes. The excitation and emission wavelengths were 480 and 520 nm, respectively. At the end of the run, 5 μl of 5% Triton X-100 (MP Biomedicals, Santa Ana, CA) was added to determine the maximum fluorescence signal. All experiments were performed in duplicate at room temperature.

Protein analysis

Protein concentration for structural studies was determined according to the method of Bradford (6) or by BCA protein assay (7) with bovine serum albumin (BSA) as a standard.

The quality of protein samples was routinely estimated by SDS-PAGE analysis using NuPAGE® Novex 4–12% gradient Bis-Tris pre-cast gels (Invitrogen/Thermo Fisher Scientific, Waltham, MA) run in MOPS-SDS Running buffer (Invitrogen/Thermo Fisher Scientic, Waltham, MA, USA). Protein bands were visualized by Imperial staining (Pierce/Thermo Fisher Scientific, Waltham, MA) following manufacturer's protocol.

To estimate the level of protein expression in the bacterial cell suspensions submitted to bio-assays, the starting samples were diluted 30-fold with the SDS-PAGE loading buffer, heated at 95^{o}C for 10 minutes, and then 5 μL of each lysate was analyzed by SDS-PAGE.

To confirm protein identity after crystallization, a protein crystal was first washed in the reservoir solution, then in 40% glycerol, dissolved in distilled deionized water, and then

run on SDS-PAGE gel. After separation by SDS-PAGE the protein sample was transferred to a polyvinylidene difluoride membrane (Invitrogen/Thermo Fisher Scientific, Waltham, MA, USA) and visualized with Coomassie© brilliant blue R-250 (Sigma-Aldrich, St. Louis, MO) following manufacturer's protocol. The protein band was cut out and submitted for the N-terminal sequencing analysis (Protein facility of the Iowa State University, Ames, IA). The first 15 amino acid residues were re-confirmed to be the same as in the wild type GNIP1Aa protein, starting from the first methionine. Essentially the same N-terminal sequencing analysis was performed for the trypsin-activated wild type GNIP1Aa to confirm that the N-terminus stayed intact after proteolytic cleavage.
Protein concentration for the liposome assays was determined by densitometry: the SDS-PAGE gels were scanned on the Odyssey® CLx LI-COR® Imaging System (LI-COR Biosciences, Lincoln, NE) using bovine serum albumin (BSA) as a standard.

Western corn rootworm bioassays
WCR eggs were received from Crop Characteristics (Farmington, MN, USA). Upon receipt, eggs were maintained at a target temperature of 17°C to 25°C depending on their level of maturity. Eggs that hatched within 24-48 hours were washed, disinfected with sodium hypochlorite solution, washed again, and placed into a dish with breathable tape. Nearly hatching eggs were used in all bioassays.
Bioassays were performed using diet-overlay methodology in Corning™ Costar™ Flat Bottom Polystyrene 24-well plates (Thermo Fisher Scientific, Waltham, MA, USA). Each well contained 1.0 mL of Bayer CropScience proprietary western corn rootworm diet. 40 µL of a sample was overlaid in each well. Plates were air dried and 5-10 near-hatching eggs were added per well. Plates were sealed with a Breathe-Easy sealing membrane (Research Products International Corp./Thermo Fisher Scientific, Waltham, MA, USA) and placed in a dark incubator at 25ºC and 85% humidity. Each submitted sample (buffer-resuspended bacterial culture or buffer control) was run in 4 technical replicates.
Bioassay plates were scored 5 days after incubation. All 4 replicates were evaluated simultaneously at the end of bioassay and a single score was recorded for all 4, described by two values, mortality and developmental delay or stunt. Mortality was evaluated on a scale of 0-100%, where 0 corresponds to all live larvae, and 100% - to all dead larvae. Developmental delay was scored on a scale of 0-4, where 0 corresponds to all healthy larvae, 1 - to non-uniform stunt, 2 - to uniform 1-25% stunt, 3 - to uniform 26-75% stunt, and 4 - to severe uniform 76-100% stunt, where 100% is a size of newly hatched larvae. The surface area of a diet was estimated to be 1.9 cm^2. Each experiment was replicated at least 8 times.

Sequence discovery and domain search
In order to find the largest set of sequences containing homology to the β-tripod fold, a local copy of the NCBI non-redundant (NR) database was searched using a single subdomain from the β-tripod fold as the seed sequence with the jackhmmer program (8). Sequence set 1 consists of the full-length sequences retrieved from NR. The resulting 1743 sequences were clustered into 232 groups using cd-hit with a clustering threshold of 40% identity (9). The cd-hit selected representatives from these groups (sequence set 2

consists of these full-length sequences) were used as input for InterProScan (10). Further searches of all sequences for the β-tripod domain, MACPF and Toxin_10 (Fig. S2) Pfam domains were performed with hmmsearch (8).
Sequence set 3 is a subset of sequence set 1 and consists of full-length sequences that contain MACPF and Toxin_10 domains.

Phylogeny

The insecticidal-associated domain phylogeny shown in Fig. S2 is created from the alignment of the full sequences in sequence set 3 (all MACPF and Toxin_10 domain containing sequences). The alignment was performed with mafft using a gap open penalty of 10 (11). This alignment was used to generate a domain phylogeny using FastTree (12). The FastTree newick file was visualized using graphlan (13). Taxonomy was determined based on the taxonomy string in the sequence definition line and a custom Perl script using the NCBI taxonomy database. The pesticidal-like protein phylogeny shown in Fig. S2 was also generated with FastTree with the –gamma option and graphlan for visualization. The 'D-X-G-S/T-G-X_3-D' motif was searched using the regular expression 'D.G[ST]G…D' in Perl.

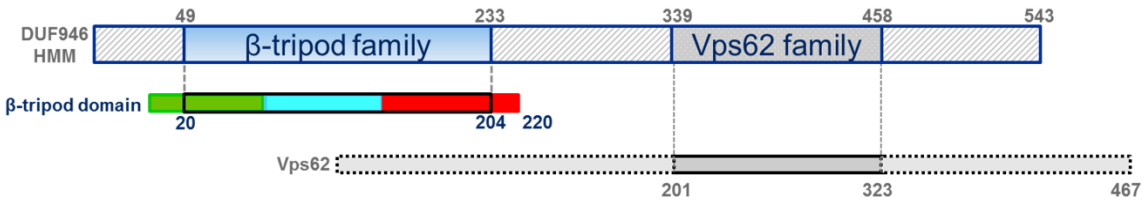

Fig. S1. Sequence alignment diagram of the DUF946 /Vps62 Hidden Markov Model (HMM), the β-tripod domain of GNIP1Aa, and Vps62 protein.
The β-tripod domain of GNIP1Aa and Vps62 protein are homologous to different parts of the DUF946 model. The β-tripod domain aligns to the N-terminal 185-residue region of the DUF946 HMM. Alignment region corresponds to segments 49-233 and 20-204 residues for HMM and β-tripod domain, respectively, with borders shown by dashed lines. Subdomains 1, 2, and 3 of the β-tripod domain are colored the same as in Fig. 2: green, cyan and red.
Vps62 protein aligns to the C-terminal part of the DUF946 model. Vps62 protein is shown in light grey with a darker grey region (201-323 residues) matching to the DUF946 HMM (339-458 residues) part. The DUF946 domain is suggested to be split into two protein domains/families, β-tripod and Vps62.

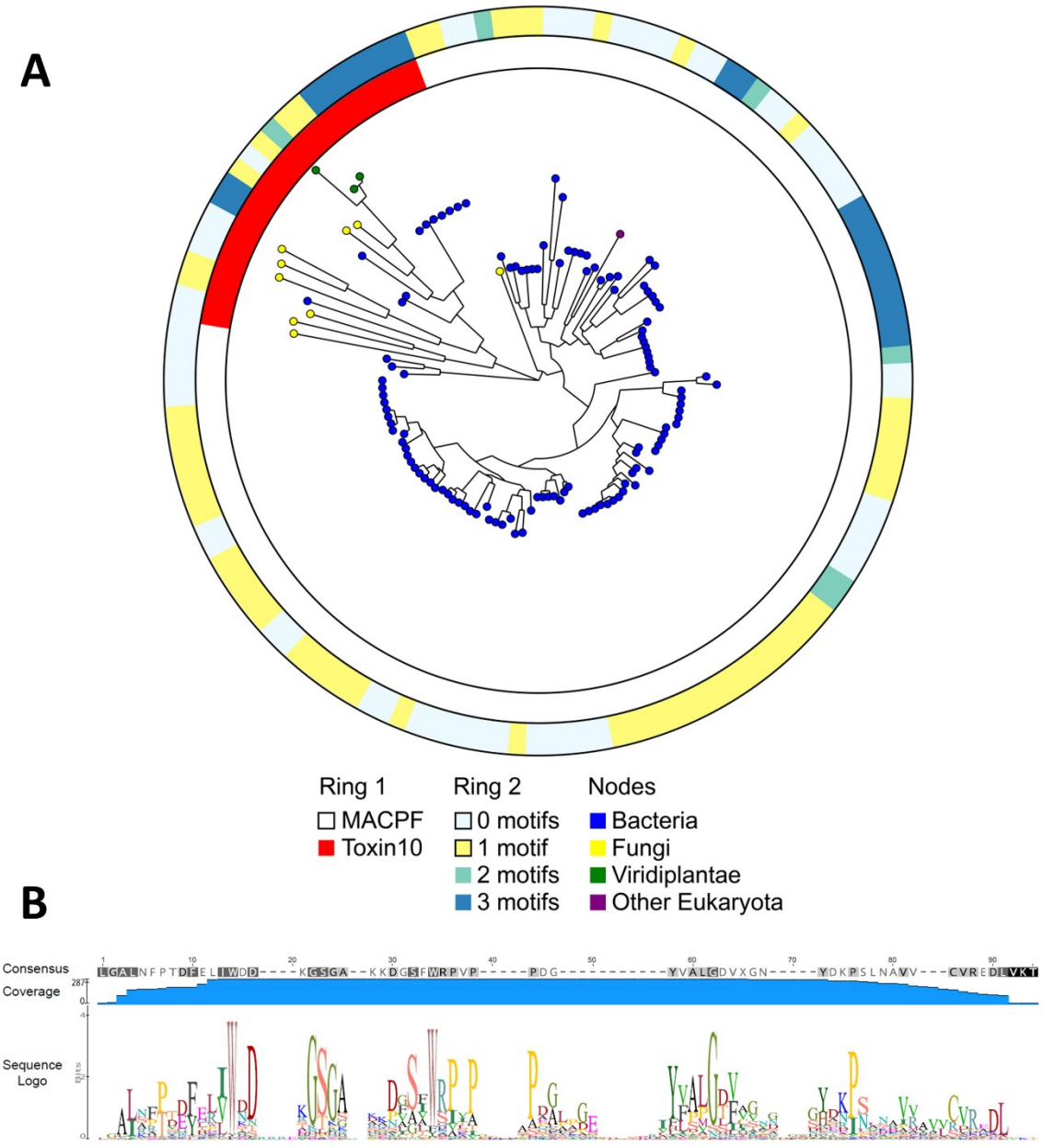

Fig. S2. Conservation of the 'D-X-G-S/T-G-X$_3$-D' motif in Insecticidal-Associated Proteins.

(A) This maximum likelihood tree shows three different types of information associated with each sequence. The node color is based on the taxonomic assignment of the sequences as described in the text. All sequences shown contain either MACPF or Toxin_10 domains. The Toxin_10 domain-containing sequences are highlighted in Ring 1 with red. Ring 2 shows the number of 'D-X-G-S/T-G-X$_3$-D' motifs contained within each full-length sequence, ranging from 0 to 3 occurrences.

(B) Sequence logo for the β-tripod domain from sequences shown in (A). The 'D-X-G-S/T-G-X$_3$-D'motif can be clearly seen as abundant in the consensus residues 16-30. Some

sequences have insertions in this region, adding length to the appearance of the full motif in the alignment. This logo has three different tracks. The top track indicates the consensus sequence; darker shading indicates higher levels of conservation. The middle track indicates the Coverage of the indicated positions across all sequences; coverage drops off at the ends of the sequence range, explaining the high conservation in the consensus sequence and low information content in the logo. The bottom track is the sequence logo itself.

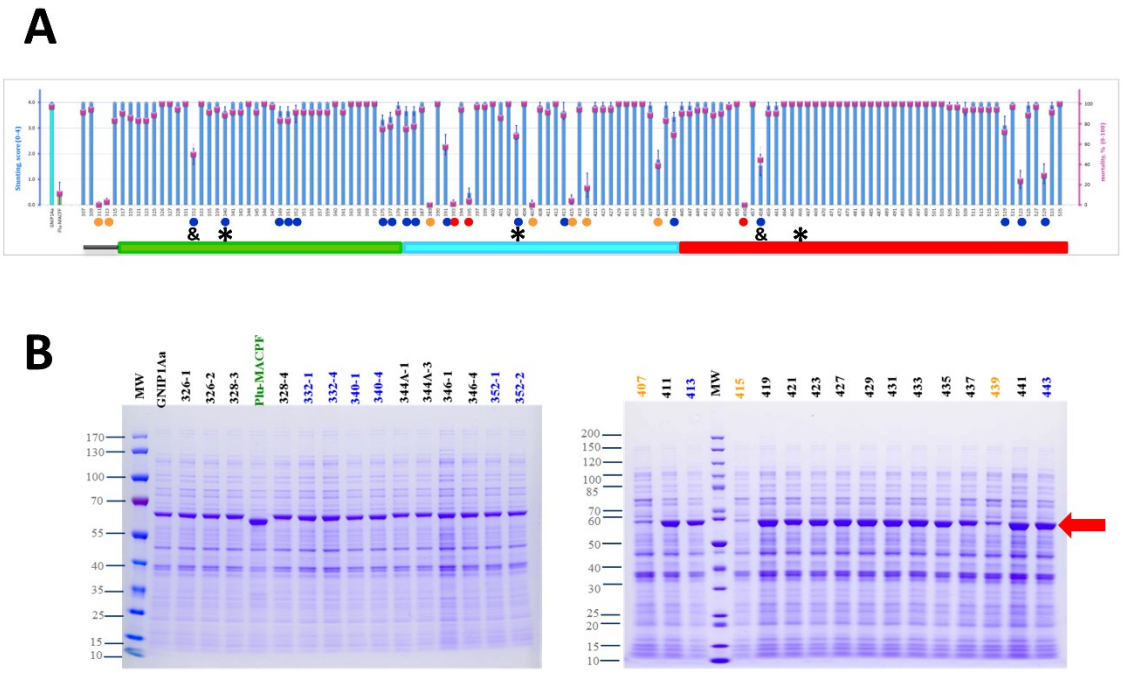

Fig. S3. WCR activity for all single site alanine mutants of GNIP1Aa.
A sample number indicates a position number for a GNIP1Aa1 residue replaced by an alanine; an extra number, if any, indicates a biological replicate.
(A) Activity of Ala-mutants in WCR bio-assays is represented by stunt (blue bars) and mortality (magenta squares) values. Data are represented as mean ± SEM.
GNIP1Aa: the wild type GNIP1Aa; positive control; its stunting activity is represented by a cyan-colored bar.
Plu-MACPF: the recombinant Plu-MACPF (PDB: 2QP2) protein, expressed in *E. coli* under the same conditions as GNIP1Aa; negative control; its background stunting activity is represented by a green-colored bar.
The colored dots indicate positions of the single-point alanine mutations that lead to the reduced or complete loss of GNIP1Aa toxicity to WCR and/or protein expression. Colors for dots are the same as highlighted colors for data in Table S2. Briefly, red dots – no activity (the negative control level) with a target protein expressed at the level of positive control; orange – lower or no protein expression resulting in lower or no activity; blue dots – reduced activity (equal or below stunt of 3.67) with a target protein expressed at the level of positive control; no dots – activity and expression similar to the wild type GNIP1Aa protein.
In addition, the positions negatively affected by double mutation are marked. Positions #1 in the conserved loops 1 and 3 (D332 and D458) are labeled with & symbol: the D332A/D458A is not toxic to WCR (see Fig. 6C). Aspartate residues at position #9 (D340, D403, and D466) for all three conserved loops are labeled with * symbol: the simultaneous replacement of any two aspartates with alanines at this position results in complete loss of protein activity (see Fig. 6C; D340A/D403A, D340A/D466A, and D403A/D466A mutants).

Each subdomain of the C-terminal domain of GNIP1Aa is shown as a colored rectangle using the same colors as in Fig. 2: green, cyan, red. The linker between the MACPF and β-tripod domains is shown as a grey bar in front of the green subdomain 1.
(B) SDS-PAGE analysis: Representative SDS-PAGE gels for the *E. coli* BL21 Star™ (DE3) lysates expressing the studied proteins.
MW: Molecular mass standards in kilodaltons.
The colors for numbers are the same as colors for dots in (A) or background colors in Table S2 and reflect the level of WCR activity and/or expression of the studied protein. The red arrow indicates the expected position of GNIP1Aa and/or its alanine mutant on the SDS-PAGE gel. The predicted molecular mass of GNIP1Aa is 58.9 kDa; Plu-MACPF – 57.0 kDa.

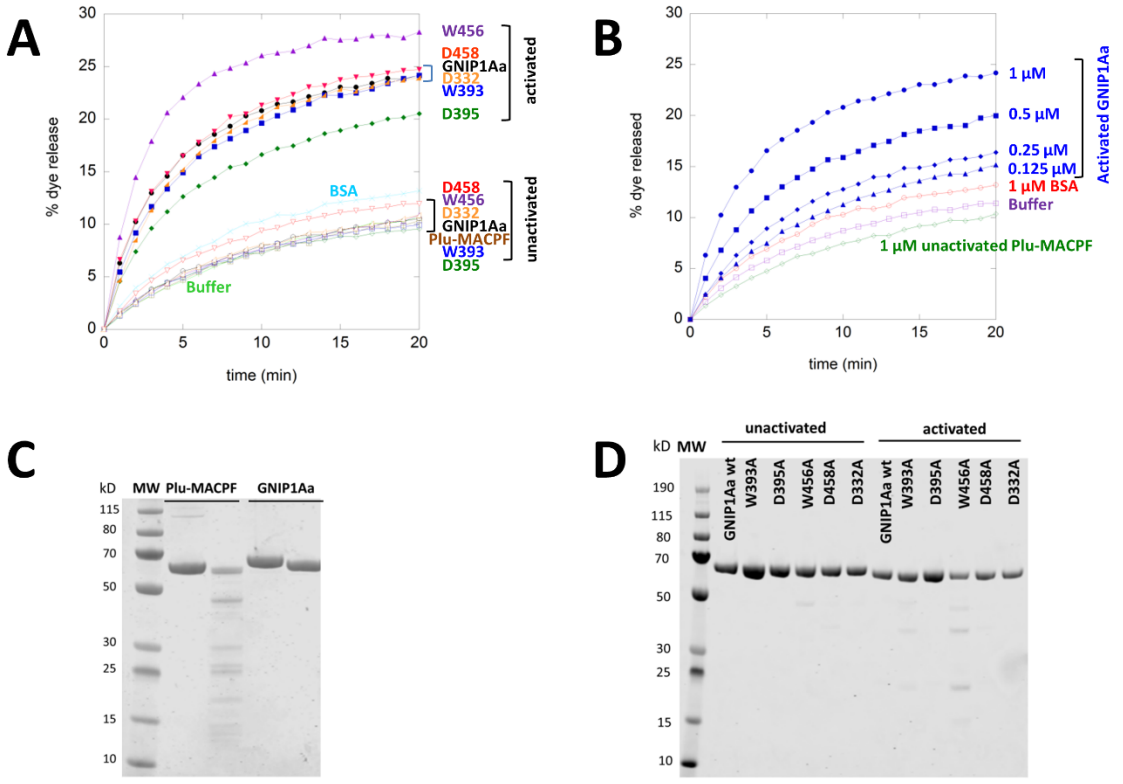

Fig. S4. Protein activation and liposome assays for GNIP1Aa and its mutants.
The GNIP1Aa is the wild type GNIP1Aa, the positive control; Plu-MACPF is the protein control with low background WCR activity. The letter and number indicate a one-letter code and position number for a residue of the wild type GNIP1Aa1 protein, replaced by an alanine.
(A) Representative traces of protein-induced fluorescent dye release from liposomes as a function of time for 5 mutant proteins, GNIP1Aa, Plu-MACPF, and negative controls. The dye release was normalized to 100% for total release. Open and closed symbols represent 1 µM unactivated and trypsin-activated samples, respectively. Protein samples shown: GNIP1Aa (black, circle), W393A (blue, square), D395A (green, diamond), W456A (purple, triangle), D458A (red, upside-down triangle), D332A (orange, sideways triangle), unactivated Plu-MACPF (brown, box with X), BSA (cyan, X), buffer* ([20 mM Na-acetate pH 5.0, 50 mM NaCl]; neon green, plus sign). * The buffer [20 mM Na-acetate pH 5.0, 50 mM NaCl, 1 mM PMSF, with/without trypsin] had a signal similar to BSA and/or [20 mM Na-acetate pH 5.0, 50 mM NaCl] buffer.
Plu-MACPF was completely degraded by trypsin treatment (C), so no activated version was available for this protein.
The higher signal for activated W456A could be explained by inherent instability of the mutant protein upon purification and its partial fragmentation observed on the Coomassie-stained gel before trypsin treatment (D). This unstable core may cause protein aggregation, leading to a false positive signal in the leakage assay. Another possible explanation is that the total protein concentration in liposome leakage assay is much

higher than the other proteins, as the concentration was determined by densitometry for the stable core, the single ~57 kDa band.
It is unknown why activated D395A has a lower leakage signal than the other activated proteins. Functional uniqueness of Asp395 was shown by complete loss of protein activity upon replacement by Ala; the only one among total of six Asp residues in three conserved loops. Perhaps D395 plays a role in protein oligomerization or pore formation, thus affecting efficiency of pore formation. Further studies are needed to define its role.
(B) Representative traces of protein-induced fluorescent dye release from liposomes as a function of time for a few concentrations of the activated GNIP1Aa. Protein concentrations shown: 1, 0.5, 0.25, and 0.125 μM (blue circle, square, diamond, and triangle, respectively). 1 μM BSA is represented by a red open circle. Buffer is represented by purple open square. 1 μM unactivated Plu-MacPF is represented by green open diamond.
(C) SDS-PAGE analysis of Plu-MACPF and GNIP1Aa trypsin activation. Lane 1: MW: Molecular mass standards in kilodaltons. Lane 2: Plu-MACPF. Lane 3: trypsin treated Plu-MACPF. Lane 4: GNIP1Aa. Lane 5: trypsin treated GNIP1Aa.
(D) SDS-PAGE analysis of GNIP1Aa mutants before and after trypsin activation. MW: Molecular mass standards in kilodaltons.

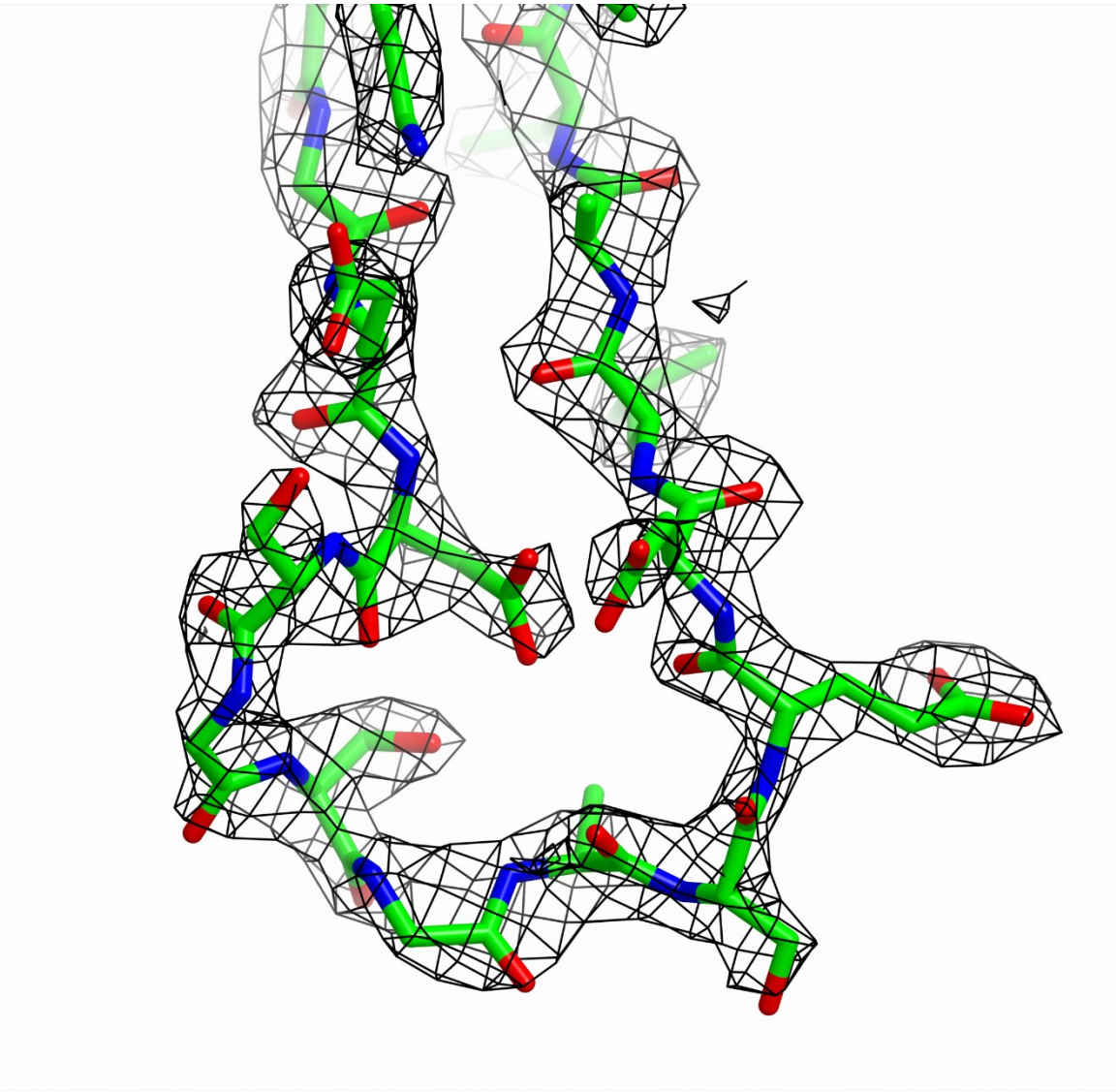

Fig. S5. Composite omit electron density map for residues surrounding the essential D332 from the first conserved loop of the C-terminal domain.
The view is the same as in Fig. 6A. The map is contoured at 1.5 s.

Table S1. X-ray data collection and refinement statistics

A. Data collection statistics

Dataset	Native	Thiomersal
Space group	P $2_12_12_1$	P $2_12_12_1$
Unit cell lengths a, b, c (Å)	78.68, 143.19, 209.62	78.91, 143.24, 209.80
Unit cell angles α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution range (Å)	15 .00 – 2.50 (2.65 – 2.50)	15 .00 – 2.60 (2.75 – 2.60)
R_{sym}*(%)	11.1 (36.4)	13.4 (39.2)
I / σ(I)	11.92 (4.11)	7.67 (2.49)
Completeness (%)	96.6 (97.7)	99.1 (98.3)
No. of unique observations	80029 (12826)	140052 (21471)
Redundancy	3.72	2.29

Values in parentheses are for the highest resolution shell.

*$R_{sym} = \sum_{hkl}\sum_{i}\left| I_i(hkl) - \overline{I(hkl)} \right| / \sum_{hkl}\sum_{i}\overline{I(hkl)}$

B. Refinement statistics

R_{cryst}*(%)	23.3 (29.8)
R_{free} (%)	27.1 (34.0)
No. of atoms	
Protein	16136
Solvent	452
r.m.s. deviations bonds (Å)	0.012
r.m.s. deviations angles (°)	1.550
Ramachandran statistics (%)	
Most favored	97.1
Additionally allowed	2.9
Disallowed	none

Values in parentheses are for the highest resolution shell.

* $R_{cryst} = \sum_{hkl}\left|F_{obs} - F_{calc}\right| / \sum_{hkl}\left|F_{obs}\right|$

Table S2. WCR activity of the single position alanine mutants of GNIP1Aa.

Data are represented as mean ± SEM.
GNIP1Aa: the original wild type GNIP1Aa protein, positive control
Plu-MACPF: the negative control protein with no WCR activity*
A sample number indicates a residue's position number in GNIP1Aa replaced by an alanine.
The highlighting colors are the same as in Fig. S3 and reflect the level of WCR activity and/or expression of the studied protein, compared to the positive control, GNIP1Aa:
red – no activity (similar or below the negative control) at the protein expression similar to the wild type GNIP1Aa;
orange – lower or no protein expression, resulting in lower or no protein activity;
blue – reduced activity at the protein expression similar to the wild type GNIP1Aa;
green – Plu-MACPF, a negative control protein with no WCR activity;
no highlighting color – activity and expression similar to the positive control, wild type GNIP1Aa protein.
* The low level of activity, stunt 0.44 and 11% mortality (with standard error of 0.44 stunt/11% mortality), reflects a typical background of live insect bio-assays. Higher protein concentrations of Plu-MACPF demonstrate similar background level of WCR activity.

GNIP1Aa position replaced by alanine	Average Stunt, score	Average Mortality, %	stunt, standard error	mortality, standard error, %
GNIP1Aa	4.00	97	0.00	3
Plu-MACPF	0.44	11	0.44	11
307	4.00	92	0.00	4
309	4.00	94	0.00	4
311	0.07	0	0.05	0
313	0.17	3	0.09	2
315	4.00	83	0.00	8
317	4.00	91	0.00	5
319	4.00	86	0.00	7
321	4.00	83	0.00	8
323	4.00	83	0.00	8
325	4.00	89	0.00	6
326	4.00	100	0.00	0
327	4.00	100	0.00	0
328	4.00	94	0.00	4
331	4.00	100	0.00	0
332	1.89	50	0.31	10
333	4.00	100	0.00	0
335	4.00	92	0.00	4

339	4.00	94	0.00	4
340	3.67	89	0.17	6
341	4.00	92	0.00	4
343	4.00	92	0.00	4
344	4.00	100	0.00	0
345	4.00	92	0.00	4
346	4.00	100	0.00	0
347	4.00	97	0.00	3
349	3.67	83	0.17	8
351	3.67	83	0.17	8
352	3.56	92	0.34	6
353	4.00	92	0.00	4
355	4.00	92	0.00	4
357	4.00	92	0.00	4
359	4.00	92	0.00	4
360	4.00	100	0.00	0
361	4.00	92	0.00	4
363	4.00	100	0.00	0
365	4.00	100	0.00	0
369	4.00	100	0.00	0
373	4.00	100	0.00	0
375	3.33	75	0.17	7
377	3.44	78	0.18	8
379	3.89	92	0.11	4
381	3.67	75	0.17	7
383	3.67	78	0.17	7
387	4.00	94	0.00	4
389	0.00	0	0.00	0
390	4.00	100	0.00	0
391	2.35	57	0.39	10
393	0.11	1	0.08	1
394	4.00	94	0.00	6
395	0.48	4	0.18	3
397	4.00	97	0.00	3
399	4.00	97	0.00	3
400	4.00	100	0.00	0
401	4.00	86	0.00	4
402	4.00	100	0.00	0
403	2.83	68	0.27	7
404	4.00	100	0.00	0
407	0.11	0	0.08	0
408	3.89	94	0.11	6
411	4.00	92	0.00	6
412	4.00	100	0.00	0

413	3.67	89	0.33	7
415	0.22	4	0.17	4
419	4.00	94	0.00	4
420	0.78	17	0.46	11
421	3.89	94	0.11	4
423	3.89	94	0.11	4
427	3.89	94	0.11	4
429	4.00	100	0.00	0
431	4.00	100	0.00	0
433	4.00	100	0.00	0
435	4.00	100	0.00	0
437	3.89	89	0.11	6
439	1.78	39	0.35	9
441	3.89	83	0.11	8
443	3.44	69	0.18	9
445	3.88	91	0.13	7
447	3.88	91	0.13	7
449	4.00	94	0.00	6
451	4.00	94	0.00	6
452	4.00	89	0.00	7
453	4.00	91	0.00	9
454	3.89	97	0.11	3
455	4.00	100	0.00	0
456	0.00	0	0.00	0
457	4.00	100	0.00	0
458	1.56	44	0.41	13
459	3.88	91	0.13	7
461	3.88	91	0.13	7
464	4.00	100	0.00	0
465	4.00	100	0.00	0
466	4.00	100	0.00	0
467	4.00	100	0.00	0
469	4.00	100	0.00	0
470	4.00	100	0.00	0
471	4.00	100	0.00	0
472	4.00	100	0.00	0
473	4.00	100	0.00	0
481	4.00	100	0.00	0
483	4.00	100	0.00	0
485	4.00	100	0.00	0
487	4.00	100	0.00	0
489	4.00	100	0.00	0
491	4.00	100	0.00	0
493	4.00	100	0.00	0

495	4.00	100	0.00	0
497	4.00	100	0.00	0
499	4.00	100	0.00	0
501	4.00	100	0.00	0
503	4.00	100	0.00	0
505	3.88	97	0.13	3
507	3.88	97	0.13	3
509	3.75	94	0.25	6
511	4.00	94	0.00	6
513	4.00	94	0.00	6
515	4.00	94	0.00	6
517	4.00	94	0.00	6
519	3.11	72	0.35	11
521	4.00	97	0.00	3
523	1.00	24	0.35	8
525	3.89	89	0.11	6
527	4.00	97	0.00	3
529	1.22	29	0.38	9
533	3.89	92	0.11	6
535	4.00	100	0.00	0

Table S3. List of Pfam domains associated with β-tripod domain.

Results of InterProScan against all of the representative sequences (sequence set 2, see Methods for details).

Domain	Pfam	Occurrences in Sequence Set 2
Eukaryota	PF12359	27
Eukaryota	PF12340	27
Eukaryota	PF06101	15
Eukaryota	PF00632	8
Eukaryota	PF00622	7
Eukaryota	PF13499	7
Eukaryota	PF00415	5
Eukaryota	PF00498	4
Eukaryota	PF00620	3
Eukaryota	PF13854	2
Eukaryota	PF13418	2
Eukaryota	PF14604	2
Eukaryota	PF00169	2
Eukaryota	PF13833	2
Eukaryota	PF13964	1
Eukaryota	PF07727	1
Eukaryota	PF00569	1
Eukaryota	PF06701	1
Eukaryota	PF07653	1
Eukaryota	PF13923	1
Eukaryota	PF05175	1
Eukaryota	PF13385	1
Eukaryota	PF02622	1
Eukaryota	PF09514	1
Eukaryota	PF00443	1
Eukaryota	PF00621	1
Eukaryota	PF00069	1
Eukaryota	PF00018	1
Eukaryota	PF01569	1
Eukaryota	PF14200	1
Eukaryota	PF14593	1
Eukaryota	PF00076	1
Bacteria	PF06101	42
Bacteria	PF05431	6
Bacteria	PF01823	4
Bacteria	PF02368	3
Bacteria	PF00652	2

Bacteria	PF00388	2
Bacteria	PF08448	1
Bacteria	PF08238	1
Bacteria	PF08685	1
Bacteria	PF01345	1
Bacteria	PF13472	1
Bacteria	PF12282	1
Bacteria	PF02518	1
Bacteria	PF01510	1
Bacteria	PF07568	1
Bacteria	PF01453	1
Fungi	PF06101	8
Fungi	PF01823	5
Fungi	PF11951	1
Fungi	PF01822	1
Fungi	PF04115	1
Fungi	PF00400	1
Viruses	PF12359	7
Viruses	PF12340	6
Viruses	PF12624	3
Viruses	PF06101	2
Viruses	PF14604	2
Viruses	PF03952	1
Viruses	PF06650	1
Viruses	PF00113	1
Viruses	PF01607	1
Viruses	PF06398	1
Viruses	PF00704	1
Viruses	PF04130	1
Viruses	PF02296	1
Viruses	PF00622	1
Viruses	PF00415	1
Viruses	PF00498	1
Viruses	PF00632	1
Viridiplantae	PF06101	48
Viridiplantae	PF12624	32
Viridiplantae	PF06650	23
Viridiplantae	PF06398	14
Viridiplantae	PF12359	3
Viridiplantae	PF00169	3
Viridiplantae	PF12340	3
Viridiplantae	PF01501	1
Viridiplantae	PF01266	1
Viridiplantae	PF00226	1

Viridiplantae	PF01134	1
Viridiplantae	PF13359	1
Viridiplantae	PF06136	1
Viridiplantae	PF13499	1
Viridiplantae	PF05832	1
Viridiplantae	PF10229	1
Eumetazoa	PF06101	3
Eumetazoa	PF01833	1
Eumetazoa	PF10162	1
Eumetazoa	PF00135	1
Eumetazoa	PF00007	1
Eumetazoa	PF01825	1
Eumetazoa	PF00002	1
Eumetazoa	PF01822	1

References

1. Sampson K*, et al.* (2017) Discovery of a novel insecticidal protein from Chromobacterium piscinae, with activity against Western Corn Rootworm, Diabrotica virgifera virgifera. *J Invertebr Pathol* 142:34-43.
2. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* 26(6):795-800.
3. Ness SR, de Graaff RA, Abrahams JP, & Pannu NS (2004) CRANK: new methods for automated macromolecular crystal structure solution. *Structure* 12(10):1753-1761.
4. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 60 (Pt 12 Pt 1):2126-2132.
5. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* 53(Pt 3):240-255.
6. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.
7. Smith PK*, et al.* (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 150(1):76-85.
8. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics (Oxford, England)* 14(9):755-763.
9. Li W & Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.
10. Jones P*, et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236-1240.
11. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30(4):772-780.
12. Price MN, Dehal PS, & Arkin AP (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one* 5(3):e9490.
13. Asnicar F, Weingart G, Tickle TL, Huttenhower C, & Segata N (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3:e1029.