Expansion and neofunctionalization of actinoporin-like genes in Mediterranean mussel 
(*Mytilus galloprovincialis*)

Neža Koritnik, Marco Gerdol, Gašper Šolinc, Tomaž Švigelj, Simon Caserman, Franci Merzel, Ellie Holden, Justin LP Benesch, Francesco Trenti, Graziano Guella, Alberto Pallavicini, Maria Vittoria Modica, Marjetka Podobnik, Gregor Anderluh

**Supplementary materials**
Supplementary Materials and Methods

Identification of Mytiporin-Related Sequences in Other Molluscan Genomes

A similar HMMER-based approach was used to screen all available annotated molluscan genomes for mytiporin-related sequences. Particular attention was placed on the detection of mytiporins in *Mytilus edulis* (v. PEIMed, credits to Genome Atlantic, DFO, and PEIMSO) and *Mytilus coruscus*, combining the data from the two available versions of the genome assembly, i.e., MCOR1.1 (Li et al. 2020) and Mcoruscus_HiC (Yang et al. 2021). A complete list of the species screened (24 bivalves, 18 gastropods, 5 cephalopods, and 1 chiton) is provided in table 1. Only complete or nearly complete gene models, i.e., those including a full porin domain, were selected for further analyses.

Cloning, Expression, and Purification of Monomeric Mytiporin-1

The cDNA sequence of MYTP1 was amplified by polymerase chain reaction (PCR) using sense (5'-TCACATATGCTGATTGACTGGGCTGC-3') and antisense (5'-TATAAGCTTACCCACGATCGGGACTTTTTC-3') oligonucleotides containing Ndel and HindIII restriction sites, respectively, and using a synthetic gene (GenScript, USA) for MYTP1 as a template. Amplified cDNA was inserted into a modified pET24a vector containing an encoded recognition site for the TEV protease, allowing expression of the protein with a TEV-cleavage site followed by a histidine tag at its C-terminus. The DNA construct was verified by nucleotide sequencing. PCR, restriction and ligation reactions, cloning, and agarose gel electrophoresis were performed according to standard protocols. The vector containing the MYTP1 sequence was used to transform the *Escherichia coli* strain BL21(DE3).

A colony grown overnight at 37 ºC on Luria-Bertani plates with kanamycin (30 μg/ml) was inoculated into 100 ml of Luria-Bertani medium with kanamycin (30 μg/ml) and incubated overnight at 37 ºC. A sample (60 ml) of the overnight culture was inoculated in 3 l of Terrific broth medium with kanamycin (30 μg/ml) and incubated at 37 ºC until an absorbance at 600 nm of 0.7 was reached. Expression of the recombinant protein was induced by 0.2 mM isopropyl β-D-1-thiogalactopyranoside. The protein was expressed overnight at 18 ºC. Cells were centrifuged at 6,721 × g for 6 min and resuspended in approximately 100 ml of cell lysis buffer (50 mM Tris/HCl, 300 mM NaCl, 5% (v/v) glycerol, pH 7.0). The cell suspension was sonicated twice for 7 min with the Cole-Parmer Ultrasonic Processor on ice and centrifuged for 20 min at 50,000 × g and 4 ºC. The supernatant was passed through a filter with a pore size of 0.22 μm and loaded onto a Tricorn 16/20 nickel-chelate affinity column (GE Healthcare Life Sciences) filled with 30 ml of nickel-nitrilotriacetic acid Superflow resin (Qiagen). The column was connected to the ÄKTA Purifier chromatography system (GE Healthcare Life Sciences) and equilibrated with phosphate-buffered saline (1.8 mM KH₂PO₄, 10.1 mM Na₂HPO₄, 140 mM NaCl, 2.7 mM KCl; PBS) containing an additional 150 mM NaCl, 5% (v/v) glycerol, and 10 mM imidazole (pH 7.0), which was used to wash the unbound proteins. MYTP1 was eluted from the column with 300
mM imidazole in the same buffer and then dialyzed overnight at 4 °C in PBS containing an additional 150 mM NaCl and 5% (v/v) glycerol (pH 7.0).

TEV protease (with a final concentration that was 25-fold less than the estimated molar concentration of MYTP1) and dithiothreitol (with a final concentration of 1 mM) were added to the protein solution. The reaction mixture was incubated overnight at 4 °C and then separated by nickel-chelate chromatography as described above, after which MYTP1 without histidine tag was collected in the unbound fraction and TEV protease in the bound fraction. The unbound fraction was purified by gel filtration using a HiLoad 16/60 Superdex 75 prep grade column connected to an ÄKTA Purifier chromatography system (GE Healthcare Life Sciences). MYTP1 with high purity was collected in storage buffer (50 mM sodium acetate, 300 mM NaCl, and 5% (v/v) glycerol, pH 5.0).

**Cloning, Expression, and Purification of Porins from Conus andremenezi and Cumia reticulata**

DNA sequences coding for ALPs from *C. andremenezi* (Cand) and *C. reticulata* (Col16 and Col26) were amplified by PCR using the oligonucleotides 5'-TATATCATATGTTGTTACTCTCTT-3' and 5'-ATATAAAGCTTGTGCCGTGTGTTGTAAGGATGA-3' for Cand, 5'-TATATCATATGCAAGTGTGGATGA-3' and 5'-ATATAAAGCTTGTACAGCCAGCCA-3' for Col16, and 5'-TATATCATATGCGTGGCGCTCCAC-3' and 5'-ATATAAAGCTTATTGCGAAAGGCCCC-3' for Col26. The templates were synthetic genes ordered from Invitrogen (Thermo Fisher Scientific). All sense oligonucleotides contained *Nde*I, and all antisense oligonucleotides contained *Hind*III restriction sites. The amplified cDNA was inserted into a pET24a vector containing a histidine tag at its C-terminus. The DNA construct was verified by nucleotide sequencing. Cloning and expression procedures were as described for MYTP1.

Cand purification was the same as described for MYTP1, except that the supernatant of the cell lysate was loaded onto a Tricorn 16/20 nickel-chelate affinity column (GE Healthcare Life Sciences) filled with 10 ml of nickel-nitrilotriacetic acid Superflow resin (Qiagen). The fraction eluted with 300 mM imidazole was concentrated to 15 mL and loaded onto a HiLoad 26/60 Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated in the same buffer as the gel filtration column for MYTP1.

For Col16 and Col26 purification, cells were centrifuged at 6,721 × g for 6 min and resuspended in approximately 100 ml of cell lysis buffer (50 mM Na-phosphate buffer, 200 mM NaCl, 5% (v/v) glycerol, pH 7.2). The cell suspension was sonicated twice for 7 min with the Cole-Parmer Ultrasonic Processor on ice and centrifuged at 50,000 × g for 20 min at 4 °C. The supernatant was passed through a filter with a pore size of 0.22 μm and loaded onto a Tricorn 16/20 nickel-chelate affinity column (GE Healthcare Life Sciences) filled with 30 ml of nickel-nitrilotriacetic acid Superflow resin (Qiagen). The column was connected to the ÄKTA Purifier chromatography system (GE Healthcare Life Sciences) and equilibrated with the binding buffer (50 mM Na-phosphate buffer, 200 mM NaCl, 10% (v/v) glycerol, pH 7.2). Col16 and Col26 were eluted from the column with a 60% gradient of elution buffer.
(50 mM Na-acetate buffer, 200 mM NaCl, pH 4.0). Fractions containing target proteins were loaded onto a HiLoad 26/60 Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated in 50 mM Na-acetate, 200 mM NaCl, 10% (v/v) glycerol, pH 5.0. Fractions containing Col16 or Col26 were collected, concentrated with ultrafiltration, and stored at −20 °C.

**Preparation of Mytiporin-1 Pores**

Monomeric MYTP1 (300 μg) was mixed with POPG-containing LUVs at a molar ratio of 1:100 in PBS (pH 7.4). LUVs were prepared in PBS as described above. The mixture was incubated for 30 min at 37 ºC while shaking at 250 rpm. Lipids were solubilized by adding Tween-80 (at a final percentage of 2% (v/v)) for 15 min at 37 ºC while shaking at 250 rpm. The mixture was centrifuged at 16,100 × g for 3 min, diluted 40-fold in pore buffer (50 mM Tris/HCl, 0.25 mM Brij-35 or Brij-58, pH 7.4), and loaded onto the Resource S (GE Healthcare Life Sciences) cation-exchange chromatography column. MYTP1 pores were eluted from the column by a salt gradient at approximately 100 mM NaCl.

**Circular Dichroism Spectroscopy**

Circular dichroism spectra were measured using a Chirascan™ CD-spectrometer (Applied Photophysics). Spectra were scanned in the far-UV region from 195 to 250 nm at 25 ºC. The bandwidth was set to 0.5 nm, and the time-per-point to 1 s. Samples contained proteins in 5 mM Na-phosphate buffers with pH values of 5.0 (Col16, Col26) or 7.0 (MYTP1, Cand). The concentration of Cand, Col16, and Col26 in the samples was 0.1 mg/ml, and the concentration of MYTP1 was 0.2 mg/ml. A 1-mm-pathlength cuvette was used for the measurements. All spectra were scanned 10 times.

**Calcein Release Assay**

Large unilamellar vesicles (LUVs) in PBS (pH 7.4) containing 50 mM calcein/NaOH were prepared by extrusion of MLVs using LipoFast™ lipid extruder (Avestin) and polycarbonate membranes with a pore size of 100 nm. The LUVs consisted of the following lipids: (1) POPG, (2) POPC and SM at a 1:1 molar ratio, and (3) POPC. A 10 ml Sephadex G-50 Superfine column (GE Healthcare) was used to remove the external calcein pool from the LUV suspension. Lipid concentration was determined using the LabAssayTM Phospholipid (Fujifilm Wako Diagnostics) colorimetric assay. LUVs were also analyzed by fluorescence measurements before and after the addition of the detergent Triton-X100. All fluorescence measurements were performed using the SynergyTM MX microplate reader (BioTek) at excitation and emission wavelengths of 485 nm and 520 nm, respectively. Calcein efflux from LUVs in the presence of MYTP1 or FraC was measured in black 96-well microtiter plates (Costar) for 30 min (at 31 s intervals) at 25 ºC. Before the measurements, 100 μl of LUV suspensions were added to twofold serial dilutions of the protein in PBS (100 μl) in the wells. The final concentration of protein in the first well was 16 μM, and the final concentration of LUVs (lipid concentration) in all wells was 50 μM. After the measurements, 2 μl of Triton X-100 (200 mM) was added to the wells, and fluorescence was
measured for an additional 3 min. Data were presented as the percentage of calcein release, as defined in Equation 1, versus protein concentration.

\[
\text{Calcein release} \% = \left( \frac{F - F_B}{F_T - F_B} \right) \times 100
\]

Equation 1

Where F stands for fluorescence, \(F_B\) is the baseline fluorescence at the beginning of the assay, and \(F_T\) is total fluorescence after the addition of Triton X-100. In the case of POPG-containing LUVs, experiments for MYTP1 were also performed in the presence of 50 and 500 mM NaCl. All measurements were performed three times.

**Molecular Dynamics Simulations**

Molecular dynamics simulations were performed with NAMD (Phillips et al. 2005) using the Charmm36 force field (Best et al. 2012). The created model of the MYTP1 pore was embedded in an equilibrated POPC:SM 1:1 bilayer and solvated with a 0.15 M aqueous solution of NaCl using CHARMM-GUI (Lee et al. 2016). To obtain an appropriate position of the pore relative to the membrane, we constructed three independent all-atom models by varying the height of the vertical position of the MYTP1, which we used in the simulations. All three systems were subjected to energy minimization, followed by a series of 10 ns equilibration molecular dynamics simulations. Subsequently, a 300 ns production simulation was performed for each system, with no position constraints applied. Each system consisted of ~215,000 atoms and was confined in an orthogonal simulation box of 125.5 Å × 125.5 Å × 132.0 Å. The temperature was maintained at 300 K by the Nosé–Hoover thermostat, and the pressure was maintained at 1 bar by the Parrinello–Rahman barostat. Long-range electrostatic interactions were calculated with the particle-mesh Ewald method. A cut-off of 1.0 nm for short range electrostatic interactions and 1.4 nm for van der Waals interactions was used.

**SNP Mapping on Mytiporin-1**

The whole-genome resequencing Illumina data from the 15 *M. galloprovincialis* individuals described above and from the reference genome were individually mapped on the coding sequence of MYTP1 with the CLC Genomics Workbench, allowing the presence of mismatches (the mapping parameters were set to length fraction = 0.5 and similarity fraction = 0.95). Based on the observed read coverage of MYTP1 in the reference genome, which was approximately equal to double of the expectations, MYTP1 was assumed to be present with two nearly identical paralogous genes. This hypothesis was confirmed by the observations that (i) the coverage was similarly higher than expected in all the other resequenced *M. galloprovincialis* individuals and (ii) the *M. edulis* orthologs were present with two almost identical paralogous copies in the high-quality, chromosome-scale genome assembly of the sister species. Therefore, the presence of up to four different alleles was taken into account, and thereby the expected possible frequencies of observations for a SNP at any given position were assumed to be 0.25, 0.50,
0.75, or 1. Only SNPs resulting in non-synonymous substitutions within the mature peptide region were annotated, and the predicted amino acid change was reported, along with the predicted number of underlying alleles.
Supplementary Figures

Supplementary Figure 1

>MYTP1
MRVALLLCCLTGFARVTGLIDWAAAIAAGSAASAVTSAGVNLGLLGSGYVSVACTVEVE
NWTKYPLIESIYNGGIIQAPPVVRPGQREQVFAHVKTGNTATGTYGTA5WLSSTTNKRAV
VMWSCPYSFDDLHSNELGVLTDKGVTKHDWFQQMEAGTSGSLNRRGEYQHTKTIISIKD
SQFEVTGIMGTSKAKARIIVRPFLNLDASLKKQVEKVIPVG

Supplementary Figure 1. *Primary structure of mytiporin-1 (MYTP1)*. FASTA format of MYTP1. The primary structure was assembled from expressed sequence tags from *Mytilus galloprovincialis* (GenBank accession numbers FL493263.1, FL493260.1, FL493261.1, FL493258.1, FL493262.1, FL493264.1). The first 19 amino acids (highlighted in bold) were predicted as a signal peptide by using SignalP version 6.0 (Teufel et al. 2022).
Supplementary Figure 2. Sequence alignment of mollusc porins and actinoporin-like proteins. MYTP1, mytiporin-1; Col16, actinoporin-like protein coluporin-16 from Cumia reticulata; Col26, coluporin-26; Cand, actinoporin-like protein from Conus adremenezii; EqtII, equinatoxin II from the sea anemone Actinia equina; FraC, fragaceatoxin C from the sea anemone Actinia fragacea; StnII, sticholysin II from the sea anemone Stichodactyla helianthus; Bryo, bryoporin from the moss Physcomitrella patens; HALT_1, actinoporin-like protein from Hydra magnipapillata. Identical residues are highlighted in red, similar residues are written in red and framed in blue. The numbering is according to the MYTP1 sequence. The alignment was visualized with ESPript 3 (Robert et al. 2014). Schematic representations of the secondary structure of MYTP1 (based on the model structure of monomeric MYTP1 obtained by AlphaFold (Jumper et al. 2021)) are presented above the sequences.
Supplementary Figure 3. Pairwise distance matrix of mytiporins. (A) p-distances among the 27 mytiporin amino acid sequences, calculated for the region denoting the actinoporin-like domain only. (B) A worm representation of Alphafold (Jumper et al. 2021) model of mytiporin-1. The evolutionary conservation of amino acids in the mytiporin family is displayed in the polypeptide chain with color (red and blue represent conserved and nonconserved regions, respectively) and diameter (thicker tubes represent more conserved regions). The protein was visualized with UCSF Chimera.
Supplementary Figure 4. *Purification of mytiporin-1 (MYTP1), actinoporin-like protein from Conus andremenezi (Cand), coluporin-16 (Col16), and coluporin-26 (Col26).* The final purification step of MYTP1 (A), Cand (B), Col16 (C), and Col26 (D) was gel filtration. The gel filtration peaks corresponding to the porins are indicated by red frames in the chromatograms (left). The peaks were analyzed with sodium dodecylsulfate polyacrylamide gel electrophoresis (right), which showed high purity of the isolated porins. (E) Circular dichroism spectra were recorded at pH 5.0 for Col16 and Col26 and pH 7.0 for MYTP1 and Cand.
Supplementary Figure 5. Permeabilizing activity of mytoporin-1 (MYTP1). (A) Calcein release from large unilamellar vesicles (LUVs) induced by fragaceatoxin C (FraC) and MYTP1 (both 8 µM). Blue circles, LUVs composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG); red circles, LUVs composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC):sphingomyelin (SM) 1:1 (mol:mol); gray circles, LUVs composed of POPC. The final release of calcein from the vesicles was induced by adding Triton X-100. (B) The release of calcein induced by FraC (open circles) and MYTP1 (full circles). The colors for lipid composition are the same as in A. [NaCl] was 140 mM in phosphate-buffered saline. (C) Calcein release from POPG-containing LUVs induced by 4 µM MYTP1 at different [NaCl].
**Supplementary Figure 6. The structure of mytiporin-1 (MYTP1).** (A) A model of MYTP1 obtained by AlphaFold (Jumper et al. 2021), shown at different orientations. The secondary structures are labelled. (B) Superposition of Crystal structures of FraC (PDB:3L1M; grey) and predicted structures of MYTP1 (blue). C- and N-termini are labelled. Orange for MYTP1 and black for FraC. (C) Electrostatic surface potentials calculated with the APBS plugin in PyMol (Baker et al. 2001) for FraC and MYTP1 (proteins are in the same orientation as in fig. 7a). Red, negative potential; blue, positive potential. N- and C-termini are marked with orange letters. Rectangle marks the α2 of FraC and the structurally corresponding α3 of MYTP1.
Supplementary Figure 7. Purification of mytoporin-1 (MYTP1) pores. (A) The steps for MYTP1 pore isolation. (B) Chromatogram of MYTP1 pore purification using cation exchange chromatography (CEX) showing absorbance at 280 nm (black line) and [NaCl] (blue line). (C) Native polyacrylamide gel electrophoresis (PAGE) analysis of samples from the different pore preparation steps. M, monomeric MYTP1; BL, MYTP1 pores before loading onto the CEX column; UB, unbound fraction of CEX chromatography (B) containing Tween-80 detergent; B, bound fraction of CEX chromatography (B), eluted with [NaCl] gradient, containing MYTP1 pores; T, Tween-80 as a control to indicate its position on the gel.
Supplementary Figure 8. *Mass spectra of mytiorin-1 at varying energy potentials.* At 75 V, only the native hexamer complex is present at 5707.9 m/z (purple). At 120 V, dissociation of a monomer (23,775 Da) from the complex and the resultant pentamer are seen (118,888 Da).
Supplementary Figure 9. **Mass distribution histogram extracted from the mass photometry landing assay data.** The average mass of the mytiporin-1 pore was determined to be 143 kDa in solution, in line with the native mass spectrometry data.
Supplementary Figure 10. *Mytiporin-1 (MYTP1) low-resolution pore model.* (A) A cryo-electron microscope micrograph of MYTP1 pores. (B) 2D classes of the pores. (C) The pore model fitted according to electron density. (D) A single mytiporin-1 protomer fitted according to cryo-EM density map. (E) A Fourier shell correlation curve generated from the independent half maps of mytiporin-1 pore electron microscopy density.
Supplementary Figure 11. *Comparison of mytiporin-1 and fragaceatoxin C pore models.* (A) Cartoon representations of the two pores aligned to the protomer (framed). (B) Two pore models to better visualize the differences in the orientation and angle of the β-barrels. Mytiporin-1, blue; fragaceatoxin C, gray.
Supplementary Figure 12. Molecular dynamics simulation of the mytiporin-1 (MYTP1) pore. A top (A) and side view (B) of the equilibrated molecular dynamics structure of MYTP1 immersed in a membrane (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine:sphingomyelin 1:1). (C) The superimposed structures of all six protomers obtained by overlaying the β-sandwiches of protomers. (D) Root mean square fluctuation (RMSF) per residue for a reference protomer averaged for all protomers for three equivalent molecular dynamics simulations.
Supplementary Figure 13. **Non-synonymous single-nucleotide polymorphisms on mytiporin-1 monomers.** Polymorphisms (orange) at positions involved in protomer-to-protomer interactions (A), on the N-terminal membrane-penetrating helix (B), and at the predicted membrane-binding region (C). Residues located at areas of interest or involved in specific functions (grey); sidechains (sticks).
Table S1: List of accession IDs for the mytiporin sequences described in this study.

| sequence     | Genbank accession ID |
|--------------|----------------------|
| Mytiporin-2  | ON166504             |
| Mytiporin-3  | ON166505             |
| Mytiporin-4  | ON166506             |
| Mytiporin-5  | ON166507             |
| Mytiporin-6  | ON166508             |
| Mytiporin-7  | ON166509             |
| Mytiporin-8  | ON166510             |
| Mytiporin-9  | VDH90312.1           |
| Mytiporin-10 | ON166511             |
| Mytiporin-11 | ON166512             |
| Mytiporin-12 | ON166513             |
| Mytiporin-13 | ON166514             |
| Mytiporin-14 | ON166515             |
| Mytiporin-15 | ON166516             |
| Mytiporin-16 | VDI24517.1           |
| Mytiporin-17 | VDI15317.1           |
| Mytiporin-18 | ON166517             |
| Mytiporin-19 | ON166518             |
| Mytiporin-20 | VDI50654.1           |
| Mytiporin-21 | ON166519             |
| Mytiporin-22 | ON166520             |
| Mytiporin-23 | ON166521             |
| Mytiporin-24 | VDI29983.1           |
| Mytiporin-25 | ON166522             |
| Mytiporin-26 | ON166523             |
| Mytiporin-27 | ON166524             |
Table S2: List of the mytiporin-1-associated non-synonymous substitutions detected through the analysis of 16 resequenced individuals. The table reports the inferred number of alleles carrying any given mutation, assuming the presence of two paralogous gene copies in all individuals.

| position | reference genome sequence | LOLA | PURA | GALF1 | GALF2 | GALF3 | GALM1 | GALM2 | GALM3 | GALM6 | GALM11 | ITAF1 | ITAF2 | ITAF3 | ITAM1 | ITAM2 | ITAM3 |
|----------|--------------------------|------|------|-------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|-------|-------|
| 4        | Trp                      | Leu  (1/4) |      |       |       |       |       |       |       |       | STOP   (1/4) |       |       |       |       |       |       |
| 5        | Ala                      |       |      |       |       |       |       |       |       |       | Thr     (2/4) |       |       |       |       |       |       |
| 7        | Ala                      |       |      | Asp   (2/4) |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 13       | Ser                      | Thr   (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (2/4) | Thr (3/4) | Thr (2/4) |       |       |
| 20       | Ser                      | Phe   (1/4) |       |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 29       | Leu                      |       |      |       |       |       |       |       |       |       | STOP    (1/4) |       |       |       |       |       |       |
| 33       | Gly                      | Ala   (1/4) |       |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 37       | Ala                      | Val   (2/4) | Val (1/4) | Val (1/4) | Val (1/4) | Val (1/4) | Val (4/4) | Val (1/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (3/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) |
| 68       | Ala                      | Val   (3/4) | Val (4/4) | Val (4/4) | Val (2/4) | Val (2/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) | Val (4/4) |
| 71       | Pro                      |       |      |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 72       | Gly                      | His   (1/4) |       |       |       |       |       |       |       |       | Arg     (1/4) |       |       |       |       |       |       |
| 73       | Gln                      |       |      |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 78       | Val                      |       |      | Ile   (1/4) |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 83       | Gly                      |       |      | Pro   (1/4) | Pro (1/4) |       |       |       |       |       |         |       |       |       |       |       |       |
| 84       | Asn                      |       |      |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 86       | Ala                      |       |      |       |       |       |       |       |       | Val (1/4) |       |       |       |       |       |       |
| 91       | Gly                      |       |      |       |       |       |       |       |       | Arg     (1/4) |       |       |       |       |       |       |
| 98       | Ser                      |       |      |       |       |       |       |       |       | Pro     (1/4) |       |       |       |       |       |       |
| 100      | Thr                      |       |      | Ala   (1/4) |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 101      | Asn                      |       |      | Gly   (2/4), Ser (2/4) | Gly (2/4), Arg (1/4) | Gly (2/4), Arg (1/4) | Gly (2/4), Arg (1/4) | Gly (1/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) | Gly (3/4), Arg (1/4) |
| 116      | Leu                      |       |      | Phe   (1/4) | Phe (1/4) |       |       |       |       |       |         |       |       |       |       |       |       |
| 120      | Glu                      | STOP  (1/4) |       |       |       |       |       |       |       |       |         |       |       |       |       |       |       |
| 125      | Leu                      |       |      |       |       |       |       |       |       | Ile     (1/4) |       |       |       |       |       |       |
| 126      | Thr                      |       |      |       |       |       |       |       |       | Lys     (1/4) |       |       |       |       |       |       |
| 135      | Asp                      |       |      |       |       |       |       |       |       | Asn     (1/4) |       |       |       |       |       |       |
| 137      | Phe                      |       |      |       |       |       |       |       |       | Leu     (1/4) |       |       |       |       |       |       |
| 139      | Gln                      |       |      |       |       |       |       |       |       | Arg     (1/4) |       |       |       |       |       |       |
| 141      | Glu                      |       |      | Asp   (1/4) |       |       |       |       |       |         |       |       |       |       |       |       |
| 142      | Thr                      |       |      | Ala   (1/4) | Ala (1/4) | Ser (2/4) | Ser (1/4) | Ser (1/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) | Ser (2/4) |
| 151      | Phe                      |       |      | Leu   (2/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) | Leu (1/4) |
| 164      | Ser                      |       |      | Ala   (1/4) |       |       |       |       |       | Tyr     (1/4) |       |       |       |       |       |       |
| 173      | Thr                      |       |      |       |       |       |       |       |       | Ile     (1/4) |       |       |       |       |       |       |
| 185      | Arg                      |       |      |       |       |       |       |       |       | Cys     (1/4) |       |       |       |       |       |       |
| 202      | Lys                      |       |      | Val   (3/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (4/4) | Val (3/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) |
| 207      | Val                      |       |      | Gly   (2/4) | Val (2/4) | Val (1/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) | Val (2/4) |
References

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**:10037-10041.

Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD, Jr. 2012. Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J. Chem. Theory. Comput.* **8**:3257-3273.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.* 2021. Highly accurate protein structure prediction with alphafold. *Nature* **596**:583-589.

Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, Wei S, Buckner J, Jeong JC, Qi Y, *et al.* 2016. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *J. Chem. Theory. Comput.* **12**:405-413.

Li R, Zhang W, Lu J, Zhang Z, Mu C, Song W, Migaud H, Wang C, Bekaert M. 2020. The whole-genome sequencing and hybrid assembly of mytilus coruscus. *Front. Genet.* **11**:440.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. 2005. Scalable molecular dynamics with namd. *J. Comput. Chem.* **26**:1781-1802.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new endscript server. *Nucleic Acids Res.* **42**:W320-324.

Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. Signalp 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.*

Yang J-L, Feng D-D, Liu J, Xu J-K, Chen K, Li Y-F, Zhu Y-T, Liang X, Lu Y. 2021. Chromosome-level genome assembly of the hard-shelled mussel mytilus coruscus, a widely distributed species from the temperate areas of east asia. *GigaScience* **10**:giab024.