

Supplementary Information

A dynamic bio-interface in mussels mediated by a mechanoresponsive intermediate filament-based biopolymer

Lucia Youssef¹, Jenaes Sivasundarampillai¹, Emily N. P. Prowse², E. Deniz Eren¹, Franziska Jehle^{1,3}, Christopher Thibodeaux¹, Adam G. Hendricks², Daniel J. Jackson⁴, Matthew J. Harrington^{1*}

¹Department of Chemistry, McGill University, Montreal, QC H3A 0B8, Canada

²Department of Bioengineering, McGill University, 3480 University Street, Montreal, Quebec H3A 0E9, Canada

³Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Golm, Germany

⁴Department of Geobiology, Georg-August University of Göttingen, Goldschmidtstr. 3, 37077, Göttingen, Germany

***Corresponding Author:** matt.harrington@mcgill.ca

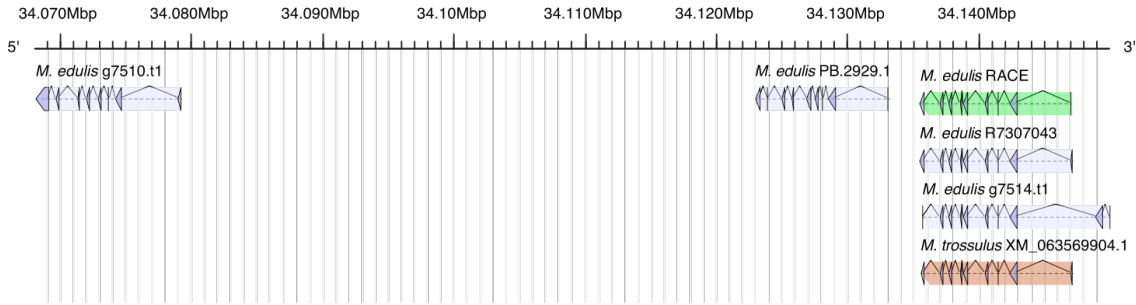
1 MFPMNPNSGRKQNPYDTSYKNFYNDKACRSCTYKPNKSNLYKSYDQPYQYKTTTNAFISPRTTTFR
 71 SPSYYPSTQQAWESYSSAGGDDFNTGVTNLSYSKSGGGNDYGIKRTIKLEDKAEHQHLNDRLESEYINKVR
 141 QLWEQRGQIDSSAFKSTKILEDEVLLKLDLYEYELANLRGQLQGSVQDKHSLEQQYGSFIQTSKDLEKR
 211 LALEVDKNRELVGNVTSYQQTISCLEQEIAKLRSAPQPLDEGPSLRKEIADLRDREDFKGRWERELQTR
 281 KDVEEAFAASTKKYEFDNNVLRQHVDLQKQRLKLEATTANILSLETRIRQLSKTDTNIPDILRQVRHAAEDE
 351 LKKHQSDIESRYGKSISVLKSQMENDADTIGRLEKEKSYAMGSGVDFKAKIAALESQVGAMAQQKQSLEA
 421 LLAQERSRAFGQIQEMEKRKFKEIQHVLVFKLEEAKVSKDSYIPLKAEIEAMKMLLAEEERRLSAPINYES
 491 TVSNVVSYPAMTRTHVASYPAMTSTKVASYNPGMTRSQTISYTPAVSRAQASSYSIPAVTRTYVGTGG
 561 HKNSYHQASYRPGYCRRYHNKYRY

— tryptic peptide detected

Supplementary figure 1. Peptide coverage map of msp-1 from mass spectrometric proteomic analysis. Proteomic analysis with mass spectrometry identified 23 individual tryptic peptides (underlined in blue) that cover 45% of the msp-1 sequence.

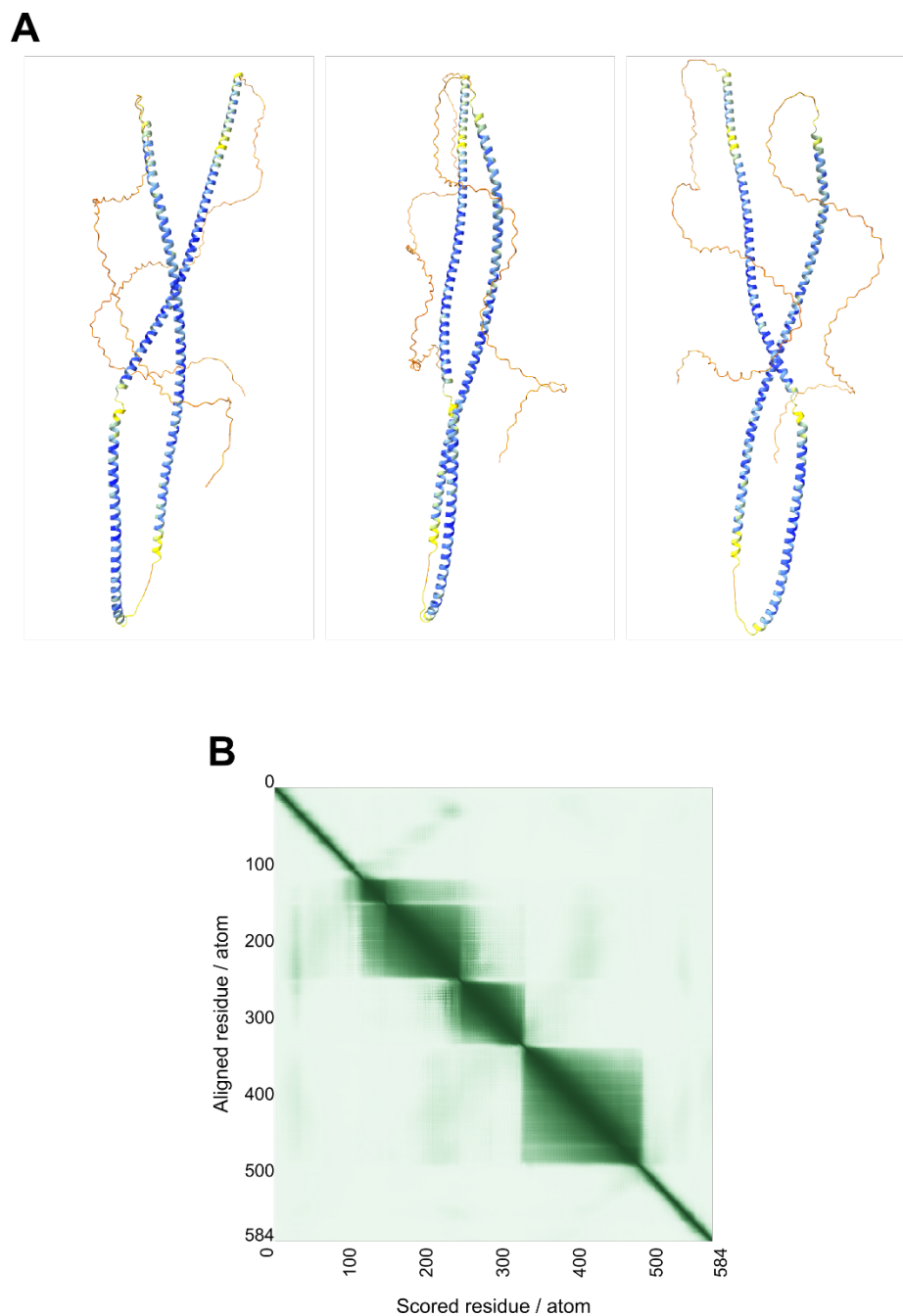


Supplementary figure 2. Proteins similar to MSP-1 found in other species. Sequences with similarity to *M. edulis mussel* stem protein 1 (MSP-1) from a broad phylogenetic range were identified via a BLASTp search against ClusteredNR (searching MSP-1 against standard NR returns hits only against the Autobranchia). Sequences were aligned using Clustal Omega102 and conserved blocks were extracted using Gblocks103. A MrBayes v3.2.7a 104 analysis was performed with the following parameters (lset rates=gamma; prset aamodelpr=mixed; mcmcpr nruns=4 ngen=2,000,000 printfreq=1000 samplefreq=1000 nchains=3 savebrlens=yes temp=0.2 stoprule=yes;) and a 50% consensus tree was generated. Sequences, alignment, and Gblocked files are provided in the supplementary material. Taxonomic lineage information was obtained from NCBI Taxondb.

A**B**

Medulis RACE	1	MTAKDFHNEAKANSRPWFLSKGFDTSCFVSDFIEKERVPNYRDLWLKVDGAMKQDDSTADMIFSVPEVTSY
g7514.t1		
PB.2929.1		
g7510.t1		
Medulis R7307043		
XP_063425974.1		
Medulis RACE	74	-----MF-----PW--MNPNSGRK--ONPRYDTSYKNFYNDKACRSCTYKPNPKSNLYKSYDO
g7514.t1		ISKYITIEEGDVVLIGIPDGGGPVTDGQIMEAGH--ADLVYDTSYKNFYNDKACRSCTYKPNPKSNLYLHKSQDO
PB.2929.1		-----MNYGMYQPSFNTTYRTYYGYPGKS-----LGSNSMSKSLM-----
g7510.t1		-----MF-----PW--MNPNSGRK--ONPRYDTSYKNFYNDKACRSCTYKPNPKSNLYKSYDO
Medulis R7307043		-----MF-----PW--MNPNSGRK--ONPRYDTSYKNFYNDKACRSCTYKPNPKSNLYKSYDO
XP_063425974.1		
Medulis RACE	147	PYQYKTTNNAFISPRTTTFRSPSYSPS--TQAWESYSSAGGDD--FNTGVTNLSYSKSGGGND---YGIKRTIK
g7514.t1		PYQYKTTNNAFISPRTTTFRSPSYSPS--TQAWESYSSAGGDD--FNTGVTNLSYSKSGGGND---YGIKRTIK
PB.2929.1		PGHKQSKIS-----VSDISKSLERSAT--LGLERPAHSHGLDRSATQKLGLOYSFGPTPTN--YRPOSSFFGR
g7510.t1		SSETKGNINAFSNPKTVAFQSFSSSSAEPKAWQN---GGGMT--S---SSSYTYTGKDIGNTYNQNSFTSTVR
Medulis R7307043		PYQYKTTNNAFISPRTTTFRSPSYSPS--TQAWESYSSAGGDD--FNTGVTNLSYSKSGGGND---YGIKRTIK
XP_063425974.1		PYQYKTTNNAFVSPRTTFRSPSYSPS--TQAWESYSSAGGDD--FNTGVTNLSYSKSGGGND---YGMKRTIK
Medulis RACE	220	LEDKAEQLHNLNDRLEYINKVROLWEORGQIDSSAFLKSTKILEDEVVLKLDLYEYELANLRGOLQGSVQDKH
g7514.t1		LEDKAEQLHNLNDRLEYINKVROLWEORGQIDSSAFLKSTKILEDEVVLKLDLYEYELANLRGOLQGSVQDKH
PB.2929.1		LEDKAEQLHNLNDRLEYINKVROLWEORGQIDSSAFLKSTKILEDEVVLKLDLYEYELANLRGOLQGSVQDKH
g7510.t1		LEDKAEQLHNLNDRLEYINKVROLWEORGQIDSSAFLKSTKILEDEVVLKLDLYEYELANLRGOLQGSVQDKH
Medulis R7307043		LEDKAEQLHNLNDRLEYINKVROLWEORGQIDSSAFLKSTKILEDEVVLKLDLYEYELANLRGOLQGSVQDKH
XP_063425974.1		
Medulis RACE	293	SLEQQYGSFIQTSKDLKRLALEVDKNRELVCNVTSYQQTISCLEQEIARLSAPQPLDEGPSLRKEIADLR
g7514.t1		SLEQQYGSFIQTSKDLKRLALEVDKNRELVCNVTSYQQTISCLEQEIARLSAPQPLDEGPSLRKEIADLR
PB.2929.1		KYQQQNGSYLQAIKDLENRLTQELSKNNELVNNINNYQVKTISCLEQELTRIKSAPKPVDESPTLRKQVADF
g7510.t1		NFOQQCGIYMOTIEDLKSLSLSDKNRLIERINGYEQMSYKQETSKMKSAAPADESPMLRQKQVADLR
Medulis R7307043		SLEQQYGSFIQTSKDLKRLALEVDKNRELVCNVTSYQQTISCLEQEIARLSAPQPLDEGPSLRKEIADLR
XP_063425974.1		SLEQQYGSFIQTSKDLKRLALEVDKNRELVCNVTSYQQTISCLEQEIARLSAPQPLDEGPSLRKEIADLR
Medulis RACE	366	DWDDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
g7514.t1		DREDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
PB.2929.1		EREDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
g7510.t1		EREDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
Medulis R7307043		DREDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
XP_063425974.1		DREDFKGRWERELQTRKDVEEAFAASTKKYEFDDNNVLRQHVADLKORLEATTANILSLETRIRQLSKDTNIP
Medulis RACE	439	DILRQVRHAAEDELKKHQSDIESRYGKSSISVLKSQMDADTIGRLEKEKSYAMGSVGFDFKAKIAALESQVGA
g7514.t1		DILRQVRHAAEDELKKHQSDIESRYGKSSISVLKSQMDADTIGRLEKEKSYAMGSVGFDFKAKIAALESQVGA
PB.2929.1		QLLRQVRLAAEEDMRKHQAETERKHNDSTILRVKESDGAALARLEKEKAGIFGSFGDLKAKIANLEAOVSS
g7510.t1		DLRQVRHAAEDELKKHQSDAELKYSKSSISALQSLQGDAGVIGKLEQEKTOIMGSFGDLKTRIAALEAQIGA
Medulis R7307043		DILRQVRHAAEDELKKHQSDIESRYGKSSISVLKSQMDADTIGRLEKEKSYAMGSVGFDFKAKIAALESQVGA
XP_063425974.1		DILRQVRHAAEDELKKHQSDIESRYGKSSISVLKSQMDADTIGRLEKEKSYAMGSVGFDFKAKIAALESQVGA
Medulis RACE	512	MAQQKQALEALLAQERSRAFGQIQEMEKRFKEIQHVLFVKLEEAQVSKDSYIPLKAEIEMKMLSAEEERRLS
g7514.t1		MAQQKQSLLEALLAQERSRAFGQIQEMEKRFKEIQHVLFVKLEEAQVSKDSYIPLKAEIEMKMLSAEEERRLS
PB.2929.1		LSQEKQSLGDLIAQERLMAAAQLREMERKRLQDMQNAVFAKLQELNVSKESYIPLKAEIEMKMLSAEEERRLS
g7510.t1		MGQEKRLLEALLAQERSMAAAQLREMERKRLQDMQNAVFAKLQELNVSKESYIPLKAEIEMKMLSAEEERRLS
Medulis R7307043		MAQQKQSLLEALLAQERSRAFGQIQEMEKRFKEIQHVLFVKLEEAQVSKDSYIPLKAEIEMKMLSAEEERRLS
XP_063425974.1		MAQQKQSLLEALLAQERSRAFGQIQEMEKRFKEIQHVLFVKLEEAQVSKDSYIPLKAEIEMKMLSAEEERRLS
Medulis RACE	585	APINYEYSTVSNVVSYPAMTRTHV--AS--YSPAMTSTKVASYNPGMTRSQTISYTPAVSRAQASSYSIPAVTRT
g7514.t1		APINYEYSTVSNVVSYPAMTRTHV--AS--YSPAMTSTKVASYNPGMTRSQTISYTPAVSRAQASSYSIPAVTRT
PB.2929.1		TPVNRNQVSTPY--GKRKY-----
g7510.t1		SPIKYTTIVKQTFHQADSSFTNQSCCENGGNWTIKSRGNVYAPGATTTTRTYYSK-----
Medulis R7307043		APINYEYSTVSNVVSYPAMTRTHV--AS--YSPAMTSTKVASYNPGMTRSQTISYTPAVSRAQASSYSIPAVTRT
XP_063425974.1		APINYEYSTVSNVVSYPAMTRTHV--AS--YAPAMTSTKVASYHNGMTRSQTMSYTPAVSRAQASSYSIPAVTRT

Supplementary Figure 3. Genomic locations, exon/intron structures and sequence similarity of *M. edulis* and *M. trossulus* MSP-1 orthologs and paralogs. **A)** 3 copies of *M. edulis* MSP-1 are located on chromosome 2 from the *M. edulis* genome reported by Regan et al. (2024). One gene model from this genome assembly (g7514.t1) closely matches the sequence we were able to RACE in both 5' and 3' directions, however our 5' RACE fragments do not match the 5' end predicted by the gene model g7514.t1, and the closest contig from our transcriptome assembly independently concurs with our RACE data (R7307043). In addition, an ortholog from *M. trossulus* (XM_063569904.1) agrees with the 5' end of our RACE sequence in contrast to g7514.t1. Interestingly no signal sequences could be detected in any of these orthologous or paralogous gene models. **B)** An alignment of all sequences located on this chromosome highlights their similarity to each other.



Supplementary Figure 4. Alternate viewing angles and PAE plot for AlphaFold3 structural prediction of MSP-1. A) 3 alternate viewing angles of an AlphaFold3 predicted structure for the full-length sequence of MSP-1. **B)** The corresponding PAE plot for the presented AlphaFold3 model shown in **A**.