PNAS RESEARCH ARTICLE | BIOCHEMISTRY
BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Invasive mussels fashion silk-like byssus via mechanical processing of massive horizontally acquired coiled coils

Miriam Simmons[a], Nils Horbelt[b], Tara Sverko[a], Ernesto Scoppola[b], Daniel J. Jackson[c,1], and Matthew J. Harrington[a,1]

Edited by Henrik Birkedal, Aarhus Universitet, Aarhus, Denmark; received July 12, 2023; accepted October 11, 2023 by Editorial Board Member Lia Addadi

Zebra and quagga mussels (*Dreissena spp.*) are invasive freshwater biofoulers that perpetrate devastating economic and ecological impact. Their success depends on their ability to anchor onto substrates with protein-based fibers known as byssal threads. Yet, compared to other mussel lineages, little is understood about the proteins comprising their fibers or their evolutionary history. Here, we investigated the hierarchical protein structure of *Dreissenid* byssal threads and the process by which they are fabricated. Unique among bivalves, we found that threads possess a predominantly β-sheet crystalline structure reminiscent of spider silk. Further analysis revealed unexpectedly that the *Dreissenid* thread protein precursors are mechanoresponsive α-helical proteins that are mechanically processed into β-crystallites during thread formation. Proteomic analysis of the byssus secretory organ and byssus fibers revealed a family of ultrahigh molecular weight (354 to 467 kDa) asparagine-rich (19 to 20%) protein precursors predicted to form α-helical coiled coils. Moreover, several independent lines of evidence indicate that the ancestral predecessor of these proteins was likely acquired via horizontal gene transfer. This chance evolutionary event that transpired at least 12 Mya has endowed *Dreissenids* with a distinctive and effective fiber formation mechanism, contributing significantly to their success as invasive species and possibly, inspiring new materials design.

mussel byssus | self-assembly | coiled coil | mechanobiology | horizontal gene transfer

Significance

Zebra and quagga mussels (*Dreissena spp.*) are invasive species that malign North American river and lake ecosystems as biofoulers. Their success depends on their ability to attach to surfaces with a tough, fibrous anchor known as a byssus. Here, we elucidated the multiscale fiber structure, formation process, and evolutionary history. Distinctive amongst mussels, Dreissenid fibers possess beta crystalline structure resembling spider silk. We identified the fiber precursors as massive coiled-coil (CC) proteins, which are mechanically converted into aligned beta crystallites during fiber formation. The precursor is the largest CC ever discovered and appears to have been horizontally transferred from a bacterium at least 12 Mya. Our findings have relevance for combatting biofouling, understanding mussel evolution, and inspiring novel polymer design.

Mussels have garnered attention both for the problems they may solve through bio-inspired design (1), as well as the problems they cause as biofoulers (2). The invasive freshwater mussel species *Dreissena polymorpha* and *Dreissena bugensis*, known more commonly as zebra and quagga mussels, respectively, are notorious biofoulers (Fig. 1 *A* and *B*). Introduced in the mid 1980s into the lakes and waterways of North America, these mussels have caused substantial damage to the delicate ecosystems in these environments as well as to aquatic recreational and industrial activities (3). The success of *Dreissena spp.* mussels as invasive biofouling species is largely due to their externally secreted holdfast known as a byssus (4, 5). The byssus is a collection of abiotic, protein-based fibers anchored into the mussel's soft tissue at one end and capped at the other by the adhesive plaque that fastens to underwater substrates (Fig. 1*C*). Byssal threads are produced by adult mussels of only certain bivalve lineages (4), and among these, there is a notable diversity of mechanics, morphologies and biochemical compositions (5, 6) (Fig. 1*D*).

Elucidating the molecular design of byssal threads has led directly to the development of biomedically and technically relevant materials, including self-healing polymers, wet adhesives, and tough coatings (1). However, thus far, most byssus research has focused on marine mussels (*Mytilus* spp.) (Fig. 1*B*). Knowledge of *Dreissena spp.* byssal threads is comparatively sparse and has focused almost exclusively on adhesion (8–13) with almost nothing known about the composition and structure of the fibers, although comparative mechanical analyses have demonstrated that zebra mussel threads are stronger, stiffer, and tougher than whole threads from the common blue mussel *Mytilus edulis* (6). Gaining a deeper understanding of *Dreissenid* byssus structure-function relationships is important both for combatting biofouling and for providing a new role model for bio-inspired design. Here, we performed a structural and compositional analysis of byssal threads from quagga and zebra mussels using a combination of spectroscopy, X-ray diffraction, and proteomics. Our findings reveal unexpected insights into the evolutionary history, hierarchical structure, and mechano-processing of the fibers, which are especially relevant given the potential for extracting design principles for sustainable production of high-performance polymers.

Dreissenid Thread Has Beta-Crystalline Structure. In previously studied mussel species, the conformational structure of byssus proteins and their hierarchical organization strongly influence the mechanical properties of the byssus fibers (14). However, there is

Author contributions: M.S., D.J.J., and M.J.H. designed research; M.S., N.H., T.S., E.S., and D.J.J. performed research; M.S., N.H., T.S., E.S., D.J.J., and M.J.H. analyzed data; and M.S., D.J.J., and M.J.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. H.B. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: djackso@gwdg.de or matt.harrington@mcgill.ca.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2311901120/-/DCSupplemental.

Published November 20, 2023.

considerable variation across different clades (Fig. 1*D*)—*Mytilid* mussels build their byssus fibers from modified collagen precursors (15), the giant clam *Tridacna maxima* produces fibers comprised of tetrameric coiled coils (CCs) (16), and Ostreida mussels (including *Pinna nobilis*) use globular proteins organized into superhelical nanofibrils (17). This diversity hints at a parallel or convergent evolution of the byssus (5). Currently, however, little is known about the identity and conformational structure of proteins in Dreissenid byssal fibers.

To fill this knowledge gap, we performed confocal Raman spectroscopy, ATR-FTIR (attenuated total reflectance Fourier transform infrared) microscopy, and synchrotron wide-angle X-ray diffraction (WAXD) on native threads from *D. polymorpha* (zebra mussel) and *D. bugensis* (quagga mussel). Confocal Raman spectra from both species were very similar and dominated by peaks typical of proteins, including amide I (1,630 to 1,700 cm^{-1}) and amide III bands (1,200 to 1,350 cm^{-1}) arising from protein backbone vibrations, as well as peaks arising from specific amino acid side chains, including tyrosine, phenylalanine, and tryptophan (Fig. 2*A*) (18), consistent with previous amino acid analysis of zebra mussel threads (11). For both species, the amide I band is sharp and centered around 1,672 cm^{-1} while the amide III band exhibited a maximum of around 1,251 cm^{-1}, suggestive of β-sheet conformation (19, 20). This is further supported by the large polarization dependence of the amide I band (Fig. 2*B*), which indicates that the proteins are well aligned with backbone carbonyls oriented perpendicular to the fiber axis—consistent with the presence of aligned β-sheets, as observed in spider silk (19, 20). Complementary ATR-FTIR spectra acquired from native fibers of both species were similar to one another (Fig. 2*C*). Negative second derivative analysis of the FTIR amide I bands (Fig. 2 *C*, *Inset*) revealed a single dominant component at 1,627 cm^{-1} and 1,631 cm^{-1} in quagga and zebra mussels, respectively, which, consistent with Raman measurements, indicates the presence of β-sheet structure (21).

As further support, synchrotron source WAXD was performed on native *Dreissena spp.* byssal threads. WAXD has been previously used to characterize the β-sheet crystalline structure and orientation of proteins in spider and insect silks based on well-characterized reflections from the orthorhombic unit cell typical of β-crystalline structure (22, 23). The diffraction patterns acquired from zebra and quagga mussel byssal threads were highly similar, indicating an oriented crystalline component based on prominent reflections along the equator and meridian, as well as an isotropic amorphous component based on the presence of a broad background ring (Fig. 2 *D* and *E*). Integrated radial profiles along the equatorial and meridional directions (Fig. 2 *F* and *G*) further highlight the similarity between species, both showing well-defined peaks at

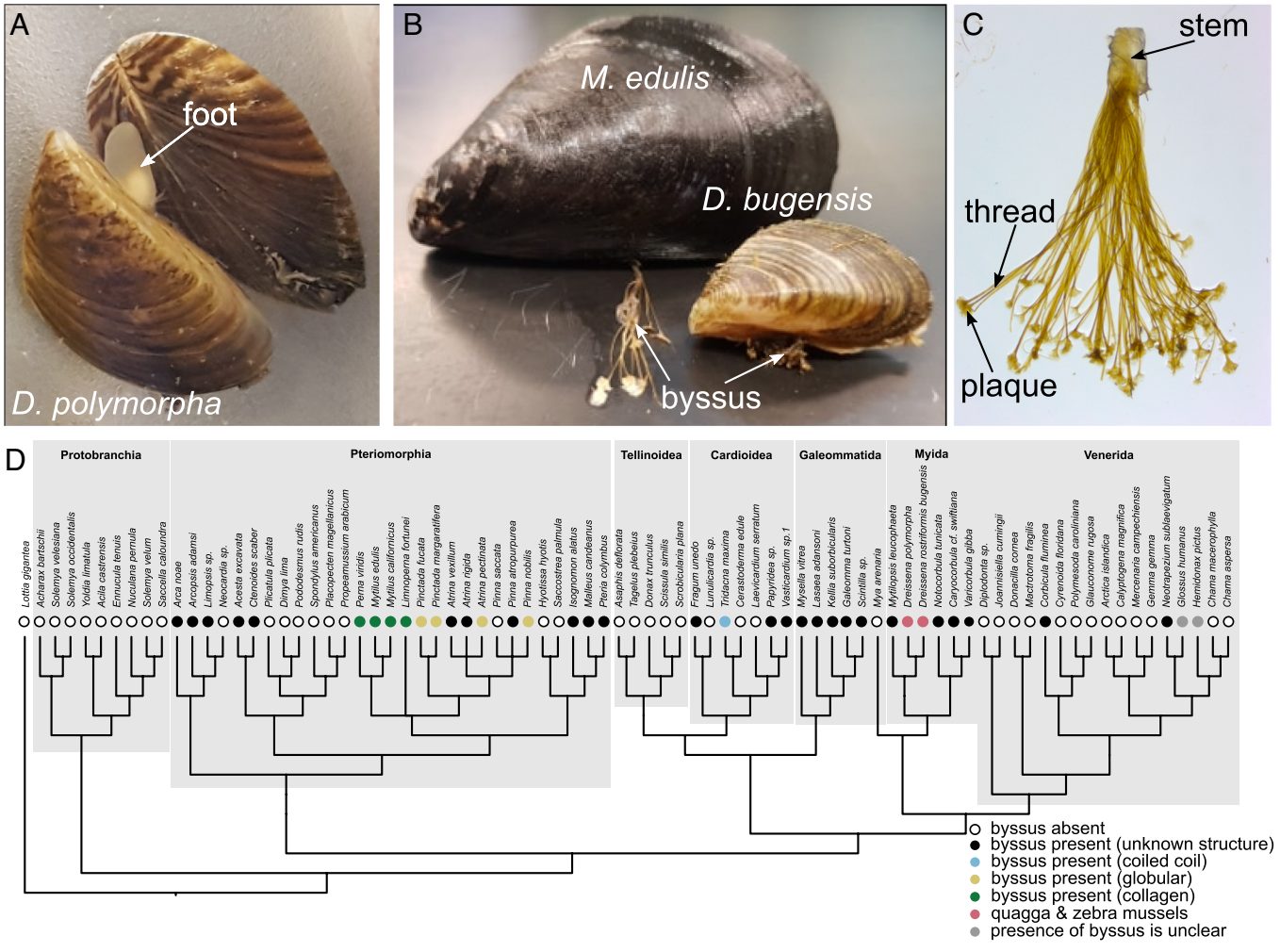

Fig. 1. Dreissenid byssus morphology. (*A*) Photograph of two zebra mussels (*D. polymorpha)* highlighting the semitransparent foot organ with which they fabricate byssal threads. (*B*) Comparison of a quagga mussel (*D. bugensis*) with a marine blue mussel (*M. edulis*). (*C*) Photograph of an extracted byssus from *D. polymorpha* indicating stem, thread, and plaque features. (*D*) A phylogeny [derived from McCartney (5) and Lemer et al. (7)] onto which the presence and absence of byssus threads is mapped. In cases in which the dominant conformation of the proteins comprising the byssus is known, it is indicated.

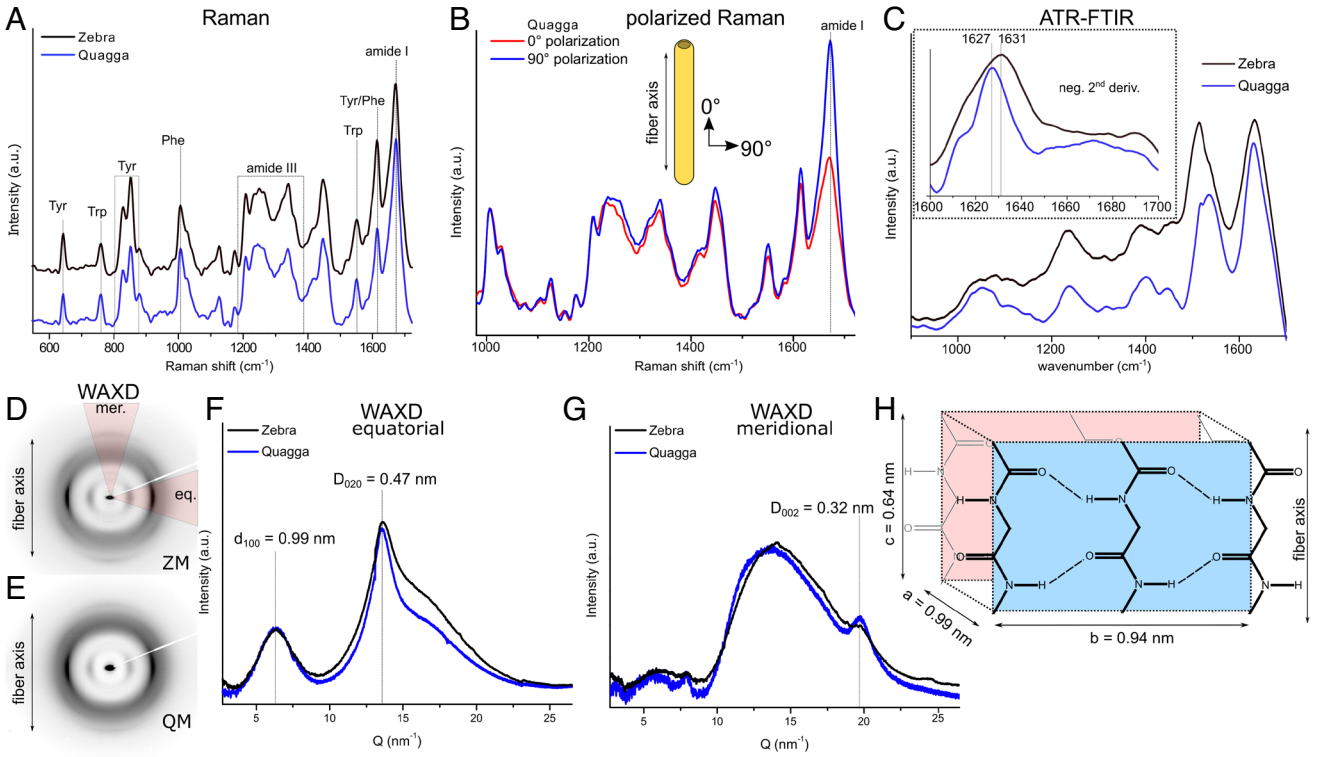

Fig. 2. Protein conformation characterization of native Dreissenid threads. (*A*) Raman spectra acquired from quagga and zebra mussel byssal threads highlighting prominent peak assignments. (*B*) Polarization-dependent Raman spectra acquired from quagga mussel byssal threads indicating strong alignment of protein chains based on amide I intensity variation. (*C*) ATR-FTIR spectra acquired from quagga and zebra mussel byssal threads. The *Inset* shows the negative second derivative spectrum of the amide I band, highlighting the primary component at 1,625 to 1,630 cm^{-1}. (*D* and *E*) WAXD diffractograms acquired from (*D*) zebra and (*E*) quagga mussels. (*F* and *G*) Radial intensity profiles of (*D*) and (*E*) as a function of the scattering vector q from sectors along the equatorial (*F*) and meridional axes (*G*) as indicated in (*D*). Background scattering due to air was subtracted. (*H*) Schematic of parallel β-sheet crystallite structure showing the proposed unit cell.

specific positions of q that correspond to the reflections visible in the diffraction patterns. Because q is inversely related to the spacing (d) between specific crystallographic planes, these peaks allow us to deduce the parameters of the crystallographic unit cell (22). Specifically, there are two distinct peaks along the equatorial direction at q = 13.6 nm^{-1} and 6.4 nm^{-1} on top of a broad amorphous background, corresponding to d-spacings of 0.47 nm and 0.99 nm, respectively (Fig. 2*F*). Along the meridian, there was one distinct peak at q = 19.7 nm^{-1} corresponding to d = 0.32 nm (Fig. 2*G*). Taken together, these three peaks provide clear indicators for the presence of β-crystallites in line with previous studies (22, 23) and were correspondingly indexed as the reflections originating from the (100), (020), and (002) crystallographic planes of the β-crystalline unit cell, respectively, following the unit cell convention originally used by Warwicker for spider silk (Fig. 2*H*) (23). Indeed, the reflection at d = 0.47 nm corresponds to the lateral β-sheet interchain distance dictated by the hydrogen bond length between β-strands along the b-axis of the unit cell (020), while the meridional peak at d = 0.32 nm arises from the translational rise per residues along the β-strand protein chain on the c-axis (002), which is roughly oriented along the fiber axis (Fig. 2*H*) (22, 23). The peak at d = 0.99 nm corresponds to the intersheet distance along the unit cell a-axis (100), dictated by the packing of amino acid side chains projecting off the β-sheet surface, indicating the presence of crystallites consisting of multiple stacked beta sheets

Consistent with polarized Raman measurements (Fig. 2*B*), the orientation of these reflections indicates that the protein chains comprising the β-sheets are primarily aligned along the thread axis, similar to spider and insect silks (24). However, the translational rise per residue of 0.32 nm predicts a parallel β-sheet configuration rather than silk's anti-parallel configuration (14). In addition to the β-sheet reflections, the broad amorphous ring indicates the additional presence of unstructured and randomly oriented protein structure. In summary, all three methods support the presence of parallel β-sheet crystallites in both zebra and quagga mussel byssal threads, which is distinctive among studied bivalve species (Fig. 1*D*). Given the close similarities between the two species, we focused our attention primarily on *D. bugensis* for subsequent investigation.

Dreissenid Thread Precursors Are Alpha-Helical Proteins. The finding that *Dreissenid* threads contain dominant parallel β-crystalline structure raises the question of how the threads are processed and fabricated, especially given the well-documented complexities of controlling β-sheet-rich fiber formation in vivo and in vitro, as exemplified in native and artificial spider silk processing (25), as well as pathological amyloid formation (26). During byssus formation, proteins synthesized and stored in secretory glands within the mussel foot are secreted into a narrow groove running along the ventral surface, where they are molded and formed into a new thread (11, 27) (Fig. 1*A* and Movie S1). While the glands responsible for storing the fiber proteins in *Dreissena spp.* have not been studied explicitly, we surmise that they are located surrounding the ventral groove, drawing analogy to the *Mytilus* byssus formation process (27). High-resolution confocal Raman spectroscopic measurements focused within the putative gland region from cryo-sectioned foot tissue from *D. bugensis* produced consistent spectra, showing similarity to native thread spectra with peaks corresponding to specific amino acid side chains (e.g.,

Trp, Tyr, and Phe), indicating a similar composition to the fibers (Fig. 3*C*). Surprisingly, however, analysis of the amide bands indicated a dominant α-helical rather than β-sheet conformation, based on the amide I maximum centered at 1,654 cm^{-1} and the amide III band maximum around 1,340 cm^{-1} (20).

To resolve this discrepancy, we acquired Raman spectra from artificially induced byssus material, which represents an intermediate stage in the byssus formation process. As previously established, induced fibers are formed by injecting the base of the mussel foot with a 0.56 M KCl solution, resulting in the paralysis of the foot and unregulated secretion of the gland proteins into the foot groove (8). The proteins coalesce without further processing, forming a roughly shaped fiber that superficially resembles a native thread (Fig. 3*B*). Native and induced threads are assumed to be nearly identical compositionally (8, 27); however, polarized light microscopy imaging reveals a much lower birefringence in induced threads indicating poor protein alignment (Fig. 3 *A* and *B*). Confocal Raman spectra acquired from freshly induced fibers closely resembled spectra acquired from the putative secretory glands in the foot indicating a dominant α-helical structure, in contrast to the β-crystalline structure of native threads (Fig. 3*C*). Additionally, polarized Raman measurements of induced threads show a weak amide I polarization dependence that is 90 degrees shifted compared to the native threads, consistent with the fact that the carbonyls in an α-helix backbone are roughly oriented along the induced thread axis but are perpendicular to the native thread in the β-crystalline conformation (Fig. 3*D*). These data strongly indicate that the thread precursor proteins are stored and secreted as α-helices, and that without active processing, remain as poorly aligned α-helices in induced threads. However, when properly secreted by the mussel foot, the proteins are apparently converted into well-aligned parallel β-crystallites. Similar measurements made on feet and threads of *D. polymorpha* show nearly identical results (*SI Appendix*, Fig. S1).

Dreissenid Thread Precursors Undergo a Mechanoresponsive α to β Transition. One compelling mechanism that may reconcile the conformational mismatch between the storage α-phase and the fiber β-phase is the possibility that the α-helices are mechanically converted into β-sheets through application of tensile stresses during thread formation (28). In fact, there is precedence for this mechanism in biology—mechanical α to β transitions occur reversibly in keratin (29), marine whelk egg cases (20), and fibrin (30), but also irreversibly in the case of hagfish slime (31) and in the amyloidogenesis of certain coiled-coil proteins (32, 33). We posit here that mechanical forces applied during the natural biological processing of *Dreissenid* byssal threads result in a permanent α-helical to β-sheet transition in the precursor proteins. Indeed, rhythmic contractions of the foot are observed during natural thread formation (Movie S1), which may produce tensile or shear forces that induce a conformational transition, as well as a higher degree of protein alignment, as evinced from the greater degree of birefringence observed in native vs. induced threads through crossed polarizers with polarized light microscopy measurements (Fig. 3 *A* and *B*).

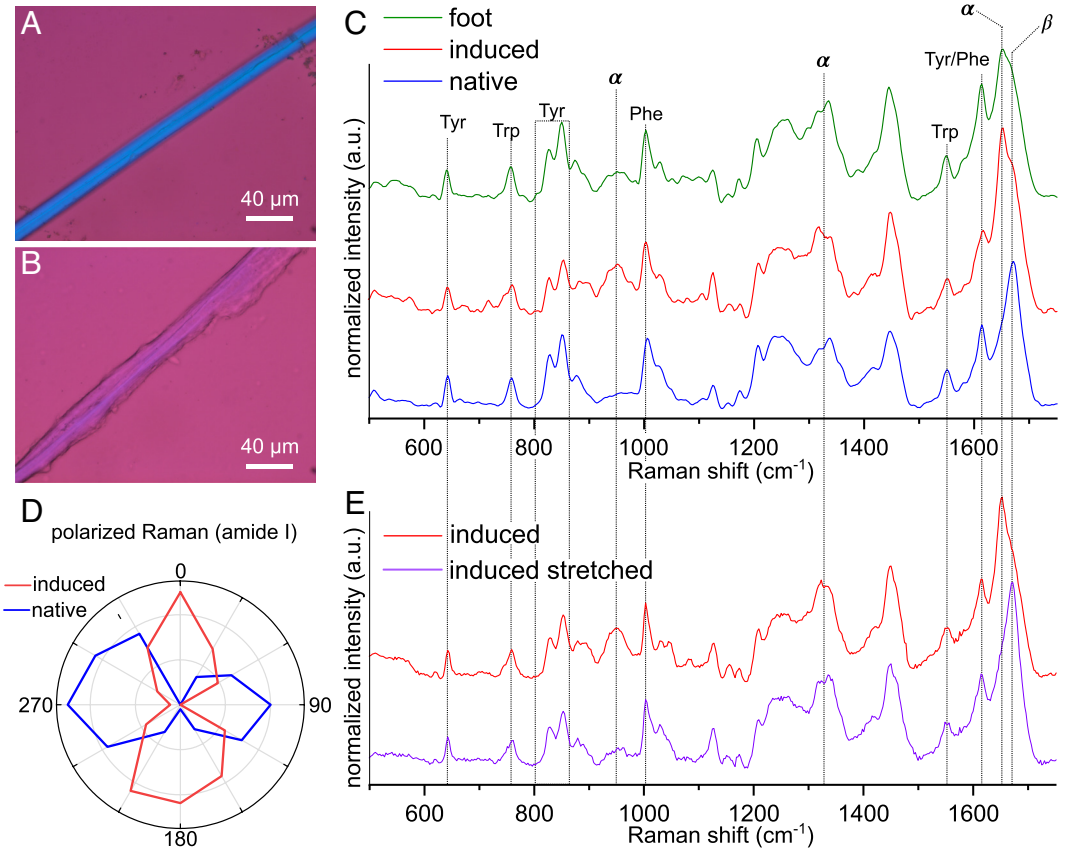

Fig. 3. Organization and conformation of proteins before, during, and after thread formation in *D. bugensis*. (*A* and *B*) Cross-polarized light microscopy images of a native (*A*) and induced (*B*) fiber from *D. bugensis* with fibers oriented 45° to polarizer with full-wave retardation plate. (*C*) Raman spectra acquired from the thread gland, induced thread, and native thread showing strong compositional similarity but clear conformational differences. (*D*) Plot of normalized amide I intensity as a function of laser polarization showing orientation of backbone carbonyls in native and induced threads (thread axis aligned along vertical axis—i.e., 0° to 180°). (*E*) Average confocal Raman spectra acquired from imaging of an induced thread before and after application of tensile strain showing clear transition in conformation. Dashed lines enable comparison between prominent spectral features between (*D*) and (*E*).

To test this hypothesis, freshly induced fibers were subjected to axially loaded tensile strains. Large area confocal Raman images were made before and after application of stress on the same threads. Averaged spectra were then extracted to monitor the global change in protein secondary structure (Fig. 3*E* and *SI Appendix*, Fig. S2). As predicted, the main differences between the samples before and after stress application were attributable to changes in secondary structure from α-helix to β-sheet conformation—there was a shift in the amide I band from ~1,652 cm^{-1} to ~1,670 cm^{-1} following stretching (Fig. 3*E*). A similar response was observed when induced threads were sheared during sample sectioning and with *D. polymorpha* induced threads subjected to axial stretching (*SI Appendix*, Fig. S3). We can thus confidently assert that *Dreissenid* fiber precursor proteins can undergo a mechanically induced conformational transformation from a dominant α-helical structure to a parallel β-sheet structure reminiscent of the native thread spectra.

Sequences of Thread Precursor Genes and Proteins in Dreissenid Mussels. Nearly all structural α-helical proteins found in biological materials are in the form of coiled coils (CCs), which consist of two or more α-helices wrapped around one another in a left-handed supercoil (34). CC protein sequences are characterized by tandemly repeated heptad motifs with polar, nonpolar, and charged residues in preferred positions (34), enabling bioinformatic prediction of CC structure based on primary sequence (35). Since our spectroscopic data strongly suggest an α-helical precursor, we adopted a proteomic approach to identify this putative CC precursor. Previous proteomic analysis of *D. polymorpha*- and *D. bugensis-induced* byssal threads revealed many byssal proteins; however, none were identified as CCs (8–10). We therefore performed mass spectrometry-based proteomic analysis on trypsin-digested protein extracts from the foot gland, as well as induced and native threads from *D. bugensis* and *D. polymorpha*, which were then searched against publicly available transcriptomes and genomes for both species (10, 36–38). The number of identified foot and induced thread proteins was quite high (around 1,500) while the number of proteins identified in the native threads was lower (383 for quagga and 66 for zebra), likely due to the cross-linking of proteins in native byssal threads (14). Nonetheless, we could identify many previously described byssal proteins in all samples (*SI Appendix*, Table S1), suggesting good proteomic coverage. We then analyzed protein candidates with high peptide coverage using a variety of CC prediction algorithms (35). Excluding obvious cellular contaminants (e.g., myosin, tubulin, and actin), the only candidate in native thread, induced thread and foot samples predicted to form CCs corresponded to a cluster of similar protein transcripts (*SI Appendix*, Table S2), a portion of which was previously reported in the quagga genome and transcriptome (37). We verified the full-length sequence of this candidate using 5′ and 3′ RACE and added over 2,500 bp of coding sequence to the previously reported transcript (*SI Appendix*, Tables S3 and S4). The final verified mRNA transcript is 13,285 bp long and encodes a massive protein of 4,068 residues with a predicted molecular weight of 440.3 kDa.

Upon inspection of the quagga genome, we were surprised to find that this >13K bp gene is apparently intronless. For reference, the average length of intronless genes in Deuterostomes ranges from 200 to 300 bp (39) (no such survey has been reported for protostomes). Three scaffolds (59, 2,947, and 1,234) could be identified in the quagga genome (37) that contained similar genes to the RACE sequence we generated; however, we focus specifically on scaffold 59 (GenBank accession VMBQ01000129.1) as this contains the gene corresponding to our RACE sequence. We name this apparent family of proteins "thread protein family 1" (tpf1), with species initials indicated before and members denoted by lower case roman numerals after (for example, Db-tpf1i). The previously reported annotation of scaffold 59 (37) identified one of three very similar, very large intronless genes (Fig. 4 and *SI Appendix*, Table S3). Additionally, we clearly identified three tpf1 orthologs in the genome of the closely related zebra mussel (38). However, we believe the loci that encode Dp-tpf1ii and Dp-tpf1iii may each have sequencing errors that interrupt the single exons of these genes. When these minor errors are adjusted, the open reading frames are restored and code for >4,050 residue proteins with features very similar to the *D. bugensis* tpf1 proteins (*SI Appendix*, Table S3) and to which peptides have been mapped by other workers. Notably, a relatively small fragment of a tpf1 protein was previously identified in the byssus of zebra mussels by Gantayet et al. (9) named Dpfp8 (AM230242) and in quagga mussels (10); however, the true sizes of the complete tpf1 proteins have until now escaped recognition.

Our efforts to model the structure of Db-tpf1iii (the RACE-verified isoform) using Alphafold2 (40) were significantly hampered by the size of the protein. After identifying a loose pattern of 20 repeats in Db-tpf1iii using RADAR (41) (*SI Appendix*, Fig. S4), we individually modelled the structures of these domains. In agreement with our initial predictions (see above) and with high confidence in the resulting structure, Alphafold2 predicts the majority of these domains to form (trimeric or tetrameric) CCs (Fig. 5 and *SI Appendix*, Fig. S5). Assuming the prediction accurately reflects the conformational structure of the tpf1 proteins, they would possess a length of ~600 nm and would measure more than 700 residues longer than the largest ever reported CC, Giantin (42). Additionally, tpf1 proteins have an unusually high asparagine content (19 to 20 mol%) compared with most coiled-coil structural proteins (Fig. 5 and *SI Appendix*, Table S3), which has been previously associated with increased likelihood of aggregation into β-crystalline amyloid aggregates in other proteins (32).

The sequences of tpf1 proteins are quite distinctive—we could find no other metazoan with similar sequences to these proteins (these searches included targeted queries against the complete genomes of 77 bivalve species), suggesting that this is a protein family unique to Dreissenids. Searches for conserved domains within the tpf1 proteins revealed only one very short motif (residues 1,249 to 1,269) with some similarity to a bacterial flagellar basal body rod protein (InterPro family IPR019776). These sequence-similarity search results, the large intronless nature of all 3 *tpf1* genes, and the apparent absence of tpf1 genes in any other metazoan prompted us to investigate the codon usage bias and GC content of Db-tpf1iii as these features can be indicative of horizontal gene transfer (HGT) events (44, 45). We assembled a "core" set of 1,300 full-length native protein-coding genes from the quagga mussel genome based on their homology with other conserved metazoan sequences. Plotting both the length of the coding sequence vs. the GC3 (GC% at the 3rd codon position) and overall GC% vs. GC3 revealed *tpf1* genes as markedly different from the 1,300 *D. bugensis* core proteins (*SI Appendix*, Fig. S6).

We also searched for other features on scaffold 59 that might shed light on the origin of the 3 *Db-tpf1* genes. We selected a 150 kb region of the quagga and zebra mussel genomes encompassing the 3 *tpf1* genes and divided these into 1,000 bp bins, then searched these against Genbank's nonredundant sequence database. The molecular machinery associated with mobile genetic elements such as transposases and reverse transcriptases (RTs), are in abundance throughout this 150 kb region and are near all *tpf1* genes (Fig. 4). Intact copies of the most prevalent RT sequence are also present in the quagga transcriptome, and it is abundant throughout the

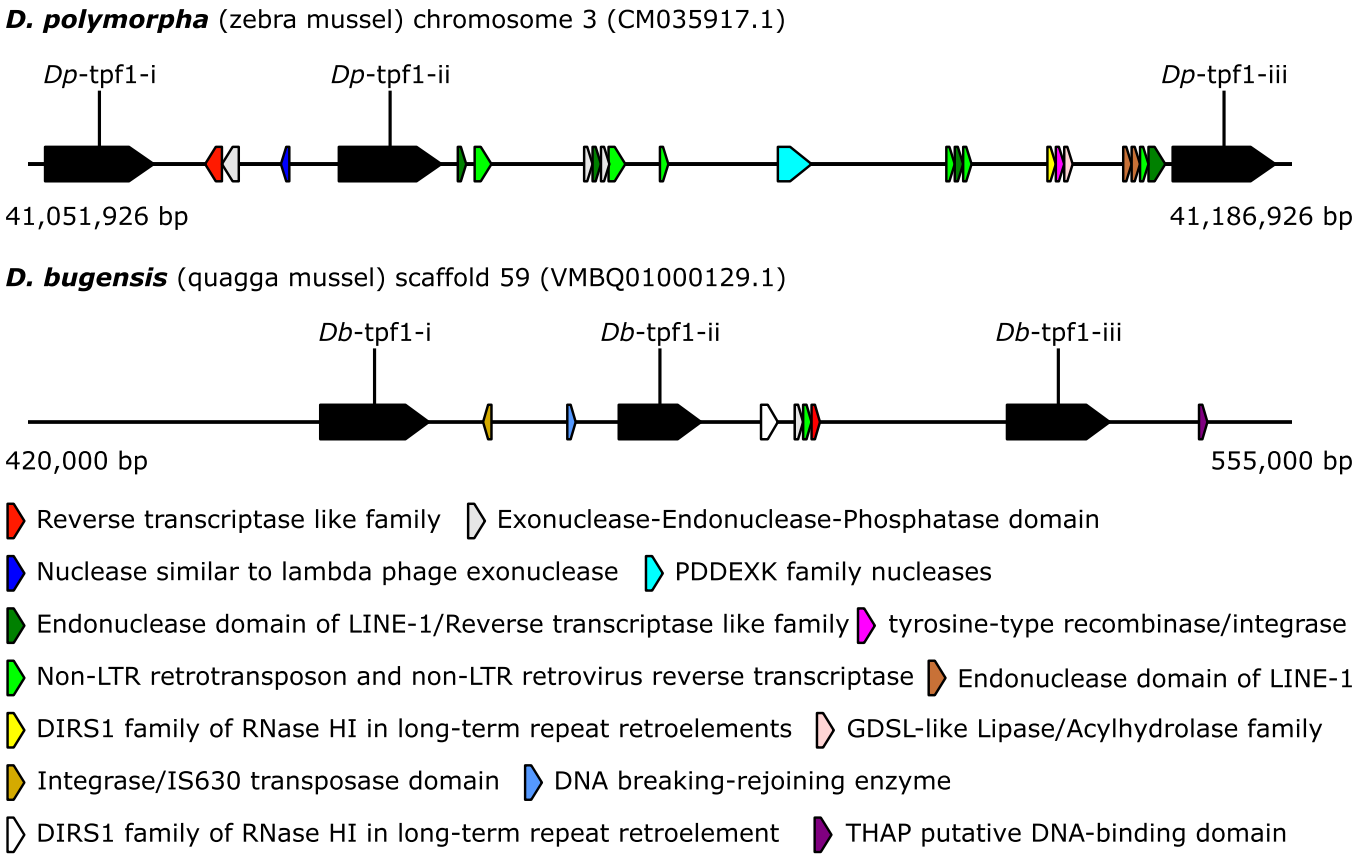

Fig. 4. Genomic landscape of tpf1 genes. Genomic arrangement of 3 zebra (*Upper*) and 3 quagga (*Lower*) tpf1 genes on chromosome 3 and scaffold 59 (respectively). Of note is the abundance and proximity of sequences with high similarity to mobile genetic elements such as transposons and RTs (colored arrows).

genome (50 different loci), suggesting that it has been (and likely still is) actively mobilizing DNA throughout the quagga genome. Interestingly, the most similar RT in GenBank (e-value 2e-154) is found in the bacterium (*Candidatus Thiodiazotropha endolucinida*), which is a symbiont of chemosymbiotic bivalves (family Lucinidae). It is noteworthy that this clade of bacteria was only recently identified (46). If we assume that the same bacterium that transferred this RT sequence into the ancestral *Dreissenid* genome also inserted the *tpf1* ancestral sequence, it becomes clear that we cannot currently identify the donor species. Despite the high sequence similarity between the intact quagga RT sequence and the *Candidatus* RT sequence, the overall sequence identity (33.12%) suggests that a bacterium from a clade distinct from *Candidatus,* that is yet to be fully sequenced, donated the ancestral *tpf1* sequence into the genome of an ancestral *Dreissenid*.

Conclusions

Our studies reveal a dominant β-sheet crystalline conformation in Dreissenid byssal threads that is distinctive from other bivalves. In spider dragline silks, β-crystallites are strengthening agents, and their combination with amorphous, extensible protein structure contributes to the exceptional toughness of silk (14, 47). We posit that a similar structure–function relationship exists in the *Dreissenid* fibers. These findings highlight the evolutionary diversity of byssus proteins across different phyla (Fig. 1*D*), adding further support to the hypothesis that the structural biochemistry of byssal threads is an instance of convergent or parallel evolution (5). Notably, our findings suggest that Dreissenid byssus evolution proceeded through the unusual route of HGT. This would not be the first example of a material functionality gained through HGT—e.g., ascidian cellulose production (48) and diatom adhesive proteins (49). However, this particular HGT event combined with human activity (e.g., aquatic transport) has led to devastating economic and environmental impacts on freshwater habitats worldwide (2, 3, 5).

Db-tfp1iii has several distinguishing biochemical features—it has an extremely high molecular weight, it can exist stably in both a CC and β-crystalline conformation, and it contains an unusually high content of asparagine (~20 mol%) (*SI Appendix,* Table S1). While bioinformatic analysis confirms that *Db*-tpf1iii will form highly stable CCs, we posit that the high asparagine content may play a key role in promoting β-crystalline structure as has been reported with certain Asn/Gln-rich CC proteins (32). For example, cytoplasmic polyadenylation element binding (CPEB) protein has a Gln-rich CC domain that can be mechanically stretched into a β-sheet conformation (33), β-sheet amyloid aggregation enhanced by polyGln expansions in specific proteins is associated with different neurological disorders (32, 33), and it was reported that CC propensity of Asn/Gln-rich proteins was directly correlated with their tendency to aggregate into β-sheet structure (32).

Computational studies of CPEB protein indicate that the α and β conformations each represent energy minima in the protein folding landscape and that mechanical activation provides the impetus to overcome the energy barrier between these minima (33). Similarly, we postulate that the *Db*-tpf1iii CC conformation provides a stable low-energy folded state that can be stored at high concentrations within the mussel foot prior to secretion and that mechanical forces during secretion align the CC precursors and provide the mechanical activation required to convert into a β-crystalline conformation, which is possibly stabilized through Asn-based interactions (e.g., H-bonding) (28, 32). Efficient

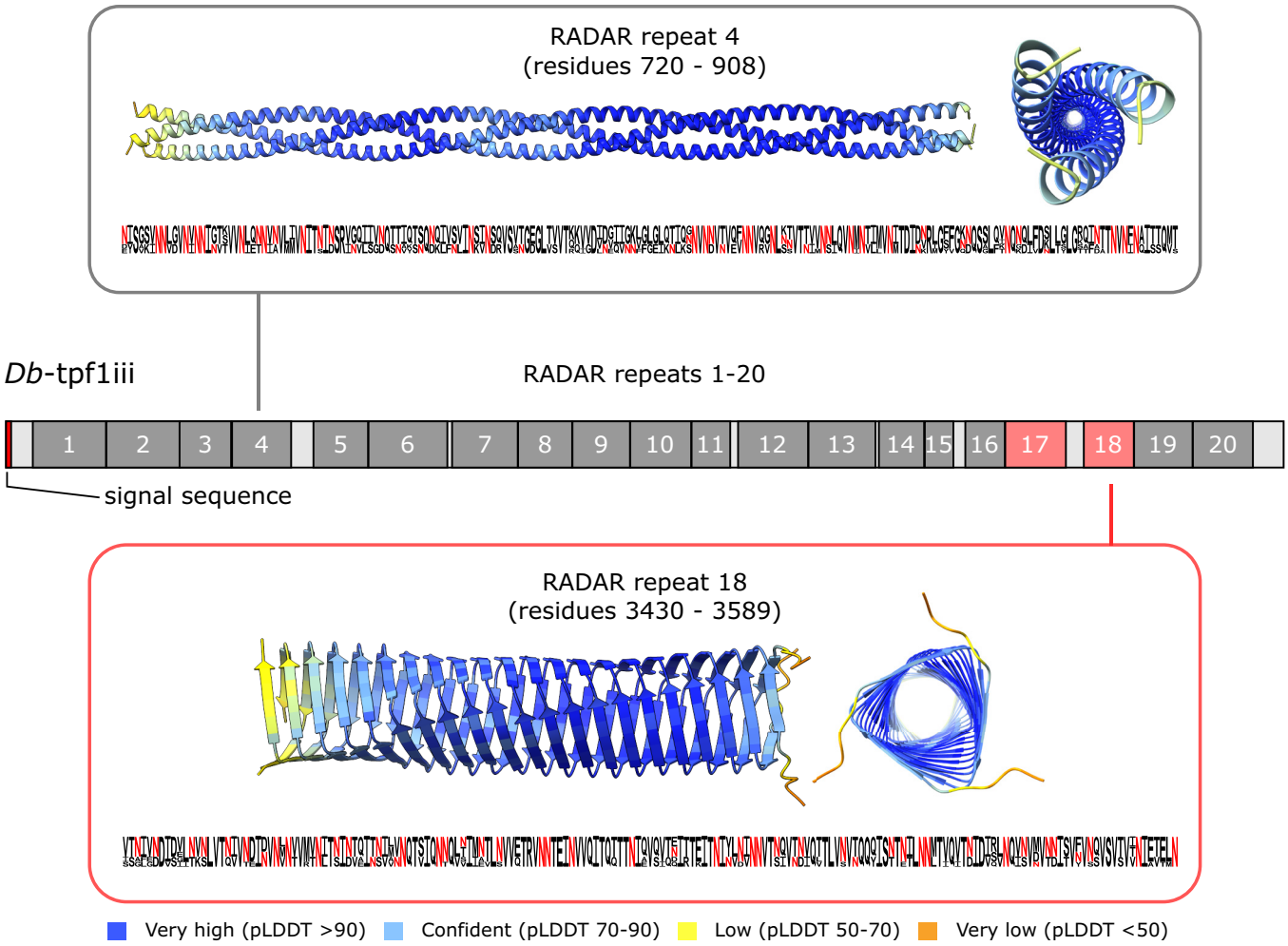

Fig. 5. Predicted conformational structure of Db-tpf1. A schematic representation of the quagga *Db*-tpf1iii protein and the 20 repetitive domains detected by RADAR (*SI Appendix*, Fig. S4). Structural modelling suggests 18/20 of these repeats form stable trimer or tetramer conformations (blue coloring indicates confident folding by Alphafold2). Repeat 18 (and to a lesser degree repeat 17) is predicted to fold as beta-barrels (see *SI Appendix*, Fig. S5 for structural predictions of all 20 repeats in trimeric and tetrameric conformations). Below each Alphafold prediction a WebLogo (43) of the respective repeat from all 6 aligned tpf1i-iii proteins from quagga and zebra mussel reveals the degree of sequence conservation (represented by the height of the letter) and the overall high asparagine content (highlighted in red). Alphafold CIs are indicated at the bottom.

processing of β-crystalline materials is notoriously challenging to control as emphasized in native and artificial spider silk processing (47) and pathological amyloid formation (26, 32). The formation of β-crystalline structure in *Dreissenid* fibers is rapid, controlled, and conceptually simpler since it seems to only depend on the application of mechanical forces rather than the combined acidification, ion exchange, shear stress, and water loss required during silk processing (25). This makes *Dreissenid* thread processing a compelling new role model for biotechnological production, upscaling, and commercialization of artificial protein-based fibers (25). It still remains to be seen how this mechanically induced transition is achieved in the living organism. While peristaltic contractions of the mussel foot during thread formation may play a role (Movie S1), we require a better understanding of the forces needed to drive the α to β transition. Indeed, it is conceivable that the transition could be initiated even after initial secretion due to forces experienced in the turbulent freshwater environment (e.g., flowing river water). Nonetheless, it is clear from our investigations that the precursor phase is α helical and is converted permanently to a β-crystalline structure in the mature thread.

In addition to the fascinating structural and mechanical features of tpf1 proteins, their evolutionary history is also highly unusual. While recognized HGT events from prokaryotes into metazoans are known (50), they are far less common than between prokaryotes (51). Nonetheless, the byssus of zebra and quagga mussels represents an ecologically impactful manifestation of an HGT into a metazoan—these fibers have directly enabled these species to attach to the hulls of ships and boats to spread into new environments and once transferred, have allowed the animals to find secure attachment where they prosper, outcompete local species, and cause billions of dollars in damage (2, 3). While there is still work to do in terms of identifying potential ancestral donors of the founding *tpf1* element, we can infer some details of when this HGT event may have occurred. According to the phylogeny of Lemer et al. (7) the closest sequenced bivalve genome outside of the Dreissenidae is that of *Mya arenaria* (52). This bivalve does not synthesize byssus threads and has no clear homolog to tpf1 proteins. The next closest bivalve with a sequenced genome that forms a byssus is the Asian freshwater clam *Corbicula fluminea* (53), and it too has no clear homolog to the tpf1 proteins. Therefore, the most constrained estimate according to the data currently available is that tpf1 genes are restricted to zebra and quagga mussels. As *D. bugensis* is estimated to have split from the other three *Dreissena* spp. 12.7 Mya (54), this is the earliest time at which an HGT could have inserted the ancestral tpf1 element into an ancestral Dreissenid genome. Alternatively, if all byssus-forming Myida members use *tpf1* genes to synthesize byssus

threads, this would push the HGT event significantly further back in time. Complete genomes from additional Myida members are needed to clarify this. From a broader perspective, we posit that numerous other such metazoan gain-of-function events via HGT await discovery. This is supported by the fact that commonly employed techniques to broadly survey genomes for HGT [phylogenetic incongruence and alien indices (55)] would not have detected the byssus thread case here nor others that we are aware of (45, 56). These examples were discovered through careful customized analyses of the respective protein-coding sequences and their genomic contexts because of their prominent functional roles in interesting biological processes.

Materials and Methods

Complete details can be found in *SI Appendix*.

Materials. Zebra (*D. polymorpha)* and quagga mussels (*D. bugensis)* were collected near Montreal (SEG Permit number 020-06-01-2831-06-16-S-P from Ministère des Forêts, de la Faune et des Parcs), separated by species and stored in room temperature source water. Native threads were removed from mussels and induced threads were produced by injecting the base of the mussel foot with 0.56 M KCl. Mussel feet were dissected and mounted in Optimal Cutting Temperature (OCT) medium, flash-frozen in liquid nitrogen, and sectioned.

Polarized Light Microscopy. Polarized light microscopy images of native and induced threads were recorded using a cross-polarized light microscope (Axio Scope.A1, Zeiss) equipped with a full-wave retardation plate and a 6-megapixel CCD camera (Axiocam 505 color, Zeiss).

Confocal Raman Spectroscopy. Raman spectra and images (including polarization-dependent measurements) from threads and foot sections were acquired with a confocal Raman microscope (Alpha 300R, WITec) and a 532 nm laser operating between 2 and 5 mW and focused with a 50× objective (NA = 0.7).

FTIR Spectroscopy. ATR-FTIR studies were conducted on a Digilab FTS 7000e FTIR with a UMA 600 microscope equipped with a Lancer detector with 32 × 32 pixels. Spectra were smoothed with a Savitzky-Golay filter before deconvoluting the amide I band by taking the second derivative.

Synchrotron WAXD. WAXD studies were conducted at the BESSY II synchrotron radiation facility at the mySpot beamline (Helmholtz-Zentrum Berlin, Adlershof). For each measurement, approximately 20 byssal threads were bundled together and mounted horizontally and perpendicular to the incident X-ray beam. The X-ray wavelength was set to 0.082656 nm (15 keV) using a MoBC multilayer monochromator (2 nm period), with a 30-μm beam size and 150 s measurement time. The diffraction patterns were collected with the Eiger 9M (Dectris) detector with a sample-to-detector distance of ~300 mm.

Protein Extraction and Purification. For protein extraction, foot extracts and induced threads were treated with an extraction buffer consisting of 20 mM phosphate buffer at pH 7, 4 M urea, 1 mM KCN, 1 mM EDTA, and 10 mM ascorbic acid. Samples were homogenized on ice with a glass hand-held tissue grinder, sonicated ~15 times with a microtip probe sonicator at 4 °C, and centrifuged (17,500 × g, 7 min, 4 °C). Freshly secreted native threads were collected for a maximum 3-d period, had the plaques removed, and were stored in a solution of 1% acetic acid with 10 mM N-phenylthiourea at 4 °C until a sufficient number of threads were collected. The threads were then treated as above but with the addition of 10 mM N-phenylthiourea and an increase of urea to 8 M in the extraction buffer.

Mass Spectrometry Sequencing. Gels containing quagga and zebra mussel foot, induced and native thread extracts were stained by Coomassie, and the bands excised for trypsin digestion were analyzed using standard procedures at the MUHC (McGill University Health Center) proteomics facility. Peptides were analyzed using a Thermo Orbitrap Fusion mass spectrometer operating at 120,000 full-width half maximum resolution in Mass Spectrometry mode with HCD sequencing. Data were analyzed using publicly available transcriptomic and genomic libraries.

RACE PCR, Sequence Analysis, and Alphafold2 Modelling. Gene-specific primers were initially designed to amplify short fragments from 5′ and 3′ RACE libraries using standard TA-cloning methods These 5′ and 3′ sequences were used to design gene-specific primers that amplified a 12.5-kb fragment. This PCR fragment was purified and directly Sanger sequenced with a series of walks using successively designed primers (*SI Appendix,* Table S4). Acquired sequences were analyzed with numerous bioinformatics tools, including Alphafold.

Data, Materials, and Software Availability. All data discussed in the paper are presented in the figures and supporting information. Materials will be made available to readers upon reasonable request. Db_tpf1iii has GenBank accession number OR105623 (57).

ACKNOWLEDGMENTS. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN-2018-05243), a Canada Research Chair award (CRC Tier 2 950-231953). T.S. was partially funded through a McGill University Science Undergraduate Research Award. D.J.J. was supported by a Deutsche Forshungsgemeinschaft grant (Project number 528314512). Research made use of the Laboratoire de Charactérisation des Matériaux facility at the Université de Montréal and the MUHC Proteomics and Molecular Analysis platform. We thank K. Blank, E. Sone, and L. Taylor for helpful discussions and A. Wähling and P. K. Siligam for assistance with installing Alphafold2 on the GWDG HPC. Portions of the paper were developed from the thesis of M.S.

Author affiliations: [a]Department of Chemistry, McGill University, Montreal, QC H3A 0B8, Canada; [b]Department of Biomaterials, Max Planck Institute of Colloids and Interfaces, Potsdam 14476, Germany; and [c]Department of Geobiology, Geoscience Center, University of Göttingen, Göttingen 37077, Germany

1. B. P. Lee, P. B. Messersmith, J. N. Israelachvili, J. H. Waite, Mussel-inspired adhesives and coatings. *Annu. Rev. Mater. Res.* **41**, 99–132 (2011).
2. T. F. Nalepa, D. W. Schloesser, *Quagga and Zebra Mussels: Biology, Impacts, and Control* (CRC Press, 2013).
3. N. A. Connelly, C. R. O'Neill, B. A. Knuth, T. L. Brown, Economic impacts of zebra mussels on drinking water treatment and electric power generation facilities. *Environ. Manage.* **40**, 105–112 (2007).
4. C. Yonge, On the primitive significance of the byssus in the Bivalvia and its effects in evolution. *J. Mar. Biol. Assoc. U. K.* **42**, 113–125 (1962).
5. M. A. McCartney, Structure, function and parallel evolution of the bivalve byssus, with insights from proteomes and the zebra mussel genome. *Philos. Trans. R. Soc. B* **376**, 20200155 (2021).
6. S. L. Brazee, E. Carrington, Interspecific comparison of the mechanical properties of mussel byssus. *Biol. Bull.* **211**, 263–274 (2006).
7. S. Lemer, R. Bieler, G. Giribet, Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proc. R. Soc. B: Biol. Sci.* **286**, 20182684 (2019).
8. A. Gantayet, L. Ohana, E. D. Sone, Byssal proteins of the freshwater zebra mussel, Dreissena polymorpha. *Biofouling* **29**, 77–85 (2013).
9. A. Gantayet, D. J. Rees, E. D. Sone, Novel proteins identified in the insoluble byssal matrix of the freshwater zebra mussel. *Mar. Biotechnol.* **16**, 144–155 (2014).
10. D. J. Rees, A. Hanifi, A. Obille, R. Alexander, E. D. Sone, Fingerprinting of proteins that mediate Quagga mussel adhesion using a De Novo assembled foot transcriptome. *Sci. Rep.* **9**, 1–14 (2019).
11. L. Rzepecki, J. Waite, The byssus of the zebra mussel, Dreissena polymorpha. I: Morphology and in situ protein processing during maturation. *Mol. Mar. Biol. Biotechnol.* **2**, 255–266 (1993).
12. L. Rzepecki, J. Waite, The byssus of the zebra mussel, Dreissena polymorpha. II: Structure and polymorphism of byssal polyphenolic protein families. *Mol. Mar. Biol. Biotechnol.* **2**, 267–279 (1993).
13. K. E. Anderson, J. H. Waite, A major protein precursor of zebra mussel (Dreissena polymorpha) byssus: Deduced sequence and significance. *Biol. Bull.* **194**, 150–160 (1998).
14. M. J. Harrington, P. Fratzl, Natural load-bearing protein materials. *Prog. Mat. Sci.* **120**, 100767 (2020).
15. J. H. Waite, X.-X. Qin, K. J. Coyne, The peculiar collagens of mussel byssus. *Matrix Biol.* **17**, 93–106 (1998).
16. A. Miserez, Y. Li, J. Cagnon, J. C. Weaver, J. H. Waite, Four-stranded coiled-coil elastic protein in the byssus of the giant clam, Tridacna maxima. *Biomacromolecules* **13**, 332–341 (2012).
17. D. Pasche *et al.*, A new twist on sea silk: The peculiar protein ultrastructure of fan shell and pearl oyster byssus. *Soft Matter* **14**, 5654–5664 (2018).
18. Z. Movasaghi, S. Rehman, I. U. Rehman, Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **42**, 493–541 (2007).
19. T. Lefèvre, M.-E. Rousseau, M. Pézolet, Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. *Biophys. J.* **92**, 2885–2895 (2007).
20. M. J. Harrington *et al.*, Pseudoelastic behaviour of a natural material is achieved via reversible changes in protein backbone conformation. *J. R. Soc. Interface* **9**, 2911–2922 (2012).
21. A. Barth, Infrared spectroscopy of proteins. *Biochim. Biophys. Acta* **1767**, 1073–1101 (2007).

22. R. E. Marsh, R. B. Corey, L. Pauling, An investigation of the structure of silk fibroin. *Biochim. Biophys. Acta* **16**, 1–34 (1955).
23. J. Warwicker, Comparative studies of fibroins: II. The crystal structures of various fibroins. *J. Mol. Biol.* **2**, 350–362 (1960).
24. R. Fraser, T. MacRae, D. Parry, E. Suzuki, The structure of β-keratin. *Polymer* **10**, 810–826 (1969).
25. A. Rising, M. J. Harrington, Biological materials processing: Time-tested tricks for sustainable fiber fabrication. *Chem. Rev.* **123**, 2155–2199 (2023), 10.1021/acs.chemrev.2c00465.
26. T. P. Knowles, M. Vendruscolo, C. M. Dobson, The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15**, 384–396 (2014).
27. T. Priemel, E. Degtyar, M. N. Dean, M. J. Harrington, Rapid self-assembly of complex biomolecular architectures during mussel byssus biofabrication. *Nat. Commun.* **8**, 14539 (2017).
28. A. Miserez, P. A. Guerette, Phase transition-induced elasticity of α-helical bioelastomeric fibres and networks. *Chem. Soc. Rev.* **42**, 1973–1995 (2013).
29. L. Kreplak, J. Doucet, F. Briki, Unraveling double stranded α-helical coiled coils: An x-ray diffraction study on hard α-keratin fibers. *Biopolymers* **58**, 526–533 (2001).
30. R. I. Litvinov, D. A. Faizullin, Y. F. Zuev, J. W. Weisel, The α-helix to β-sheet transition in stretched and compressed hydrated fibrin clots. *Biophys. J.* **103**, 1020–1027 (2012).
31. D. S. Fudge, K. H. Gardner, V. T. Forsyth, C. Riekel, J. M. Gosline, The mechanical properties of hydrated intermediate filaments: Insights from hagfish slime threads. *Biophys. J.* **85**, 2015–2027 (2003).
32. F. Fiumara, L. Fioriti, E. R. Kandel, W. A. Hendrickson, Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143**, 1121–1135 (2010).
33. M. Chen, W. Zheng, P. G. Wolynes, Energy landscapes of a mechanical prion and their implications for the molecular mechanism of long-term memory. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5006–5011 (2016).
34. D. A. Parry, R. B. Fraser, J. M. Squire, Fifty years of coiled-coils and α-helical bundles: A close relationship between sequence and structure. *J. Struct. Biol.* **163**, 258–269 (2008).
35. D. Simm, K. Hatje, M. Kollmar, Waggawagga: Comparative visualization of coiled-coil predictions and detection of stable single α-helices (SAH domains). *Bioinformatics* **31**, 767–769 (2015).
36. W. Xu, M. Faisal, Putative identification of expressed genes associated with attachment of the zebra mussel (Dreissena polymorpha). *Biofouling* **24**, 157–161 (2008).
37. A. D. Calcino *et al.*, The quagga mussel genome and the evolution of freshwater tolerance. *DNA Res.* **26**, 411–422 (2019).
38. M. A. McCartney *et al.*, The genome of the zebra mussel, Dreissena polymorpha: A resource for comparative genomics, invasion genetics, and biocontrol. *G3 (Bethesda)* **12**, jkab423 (2022).
39. M. Zou, B. Guo, S. He, The roles and evolutionary patterns of intronless genes in Deuterostomes. *Comp. Funct. Genomics* **2011**, 680673 (2011).
40. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
41. A. Heger, L. Holm, Rapid automatic detection and alignment of repeats in protein sequences. *Proteins.* **41**, 224–237 (2000).
42. L. Truebestein, T. A. Leonard, Coiled-coils: The long and short of it. *Bioessays* **38**, 903–916 (2016).
43. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
44. D. M. Matriano, R. A. Alegado, C. Conaco, Detection of horizontal gene transfer in the genome of the choanoflagellate Salpingoeca rosetta. *Sci. Rep.* **11**, 5993 (2021).
45. D. J. Jackson, L. Macis, J. Reitner, G. Wörheide, A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. *BMC Evol. Biol.* **11**, 238 (2011).
46. J. T. Osvatic *et al.*, Global biogeography of chemosynthetic symbionts reveals both localized and globally distributed symbiont groups. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104378118 (2021).
47. A. Rising, J. Johansson, Toward spinning artificial spider silk. *Nat. Chem. Biol.* **11**, 309–315 (2015).
48. K. Nakashima, L. Yamada, Y. Satou, J.-I. Azuma, N. Satoh, The evolutionary origin of animal cellulose synthase. *Dev. Genes Evol.* **214**, 81–88 (2004).
49. J. Zackova Suchanova *et al.*, Diatom adhesive trail proteins acquired by horizontal gene transfer from bacteria serve as primers for marine biofilm formation. *New Phytol.* **240**, 770–783 (2023).
50. H.-H. Zhang, J. Peccoud, M.-R.-X. Xu, X.-G. Zhang, C. Gilbert, Horizontal transfer and evolution of transposable elements in vertebrates. *Nat. Commun.* **11**, 1362 (2020).
51. E. Avni, S. Snir, A new phylogenomic approach for quantifying horizontal gene transfer trends in prokaryotes. *Sci. Rep.* **10**, 12425 (2020).
52. D. C. Plachetzki, M. S. Pankey, M. D. MacManes, M. P. Lesser, C. W. Walker, The genome of the softshell clam Mya arenaria and the evolution of apoptosis. *Genome Biol. Evol.* **12**, 1681–1693 (2020).
53. T. Zhang *et al.*, Dissecting the chromosome-level genome of the Asian Clam (Corbicula fluminea). *Sci. Rep.* **11**, 15021 (2021).
54. H. Bilandžija, B. Morton, M. Podnar, H. Ćetković, Evolutionary history of relict Congeria (Bivalvia: Dreissenidae): Unearthing the subterranean biodiversity of the Dinaric Karst. *Front. Zool.* **10**, 5 (2013).
55. C. Rancurel, L. Legrand, E. G. J. Danchin, Alienness: Rapid detection of candidate horizontal gene transfers across the tree of life. *Genes* **8**, 248 (2017).
56. C. A. Ettensohn, Horizontal transfer of the msp130 gene supported the evolution of metazoan biomineralization. *Evol. Dev.* **16**, 139–148 (2014).
57. M. Simmons *et al.*, Dreissena rostriformis bugensis isolate 1 tpf1iii mRNA, complete cds. Genbank. https://www.ncbi.nlm.nih.gov/nuccore/OR105623. Accessed 13 June 2023.