UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

Infiltration of chitin by protein coacervates defines the squid beak mechanical gradient

Permalink

https://escholarship.org/uc/item/3f21p9m3

Journal

Nature Chemical Biology, 11(7)

ISSN

1552-4450

Authors

Tan, YerPeng
Hoon, Shawn
Guerette, Paul A
et al.

Publication Date

2015-07-01

DOI

10.1038/nchembio.1833

Peer reviewed

Infiltration of chitin by protein coacervates defines the squid beak mechanical gradient

YerPeng Tan[1], Shawn Hoon[2,3], Paul A Guerette[4,5], Wei Wei[6], Ali Ghadban[5], Cai Hao[5], Ali Miserez[3,5]* & J Herbert Waite[1,6,7]*

The beak of the jumbo squid *Dosidicus gigas* is a fascinating example of how seamlessly nature builds with mechanically mismatched materials. A 200-fold stiffness gradient begins in the hydrated chitin of the soft beak base and gradually increases to maximum stiffness in the dehydrated distal rostrum. Here, we combined RNA-Seq and proteomics to show that the beak contains two protein families. One family consists of chitin-binding proteins (DgCBPs) that physically join chitin chains, whereas the other family comprises highly modular histidine-rich proteins (DgHBPs). We propose that DgHBPs play multiple key roles during beak bioprocessing, first by forming concentrated coacervate solutions that diffuse into the DgCBP-chitin scaffold, and second by inducing crosslinking via an abundant GHG sequence motif. These processes generate spatially controlled desolvation, resulting in the impressive biomechanical gradient. Our findings provide novel molecular-scale strategies for designing functional gradient materials.

Biomolecular gradients are critical to tissues that connect materials with different mechanical properties. The mechanical gradients identified in tissue transitions such as skin-fingernail[1], bone-tendon[2,3], tendon-muscle[4], and distal-proximal transitions in mussel byssal threads[5] and squid beak[6] are efficient at mitigating mechanical damage at interfaces[7–9]. Conversely, failures in engineering multicomponent systems are often due to mechanical property mismatches between various components[7]. Nature's gradient strategies for mitigating the mismatch problem thus offer considerable value to materials engineers. For example, bioengineering applications where stiff therapeutic devices are in immediate contact with soft biological tissues, such as glucose sensors for diabetics[10] and osseo-integrated prosthetic limbs for amputees[11], can benefit from a modulus buffer between the stiff implant and soft tissue[12]. Furthermore, in contrast to engineering manufacturing techniques, biological materials are processed with natural chemicals under mild processing conditions (ambient temperature and pressure), thus providing valuable lessons in 'green' chemical processing if their biosynthetic pathways can be elucidated.

The beak of the jumbo squid (*D. gigas*, **Fig. 1a**) has been proposed as a model system for the synthesis of bio-inspired functionally graded materials, with particular relevance to engineering and medical applications[6,12,13]. The squid beak is fully organic and composed of chitin, water, proteins and chemical crosslinks, the relative abundance of which varies spatially, from the soft base embedded within the buccal mass (**Fig. 1a**) to the very hard tip (rostrum). This design principle is key to mitigating the interfacial mechanical damage at the junction between the base of the beak and the buccal mass. The relative proportions of beak components indirectly define the stiffness and hardness gradient in the mature beak structure by controlling hydration levels[6,14]. A fundamental understanding of how a mechanical gradient forms in squid beak necessitates the identification and characterization of the beak proteins and their post-secretion processing. To date, our knowledge of squid beak proteins is limited because protein extraction methods have suffered from low yields due to the densely crosslinked nature of the beak. Classical protein extraction protocols for soft tissues and cells are not effective for the squid beak—or other highly sclerotized tissues in general, such as insect cuticles or arthropod fangs—because the majority of constituent proteins remain insoluble even under the most aggressive extraction procedures[15]. In this work, we employed non-enzymatic reagents that can cleave peptide bonds between specific residues to liberate beak proteins from the structure. In parallel, we assembled a transcriptome of beak tissue containing beccublast cells[16], which secrete the beak components, from transcripts obtained by RNA-Seq[17].

By combining RNA-Seq with proteomics techniques, we identified two important families of beak proteins. One family consists of chitin-binding proteins that can act as physical anchors to position and fix chitin polymers during secretion. The other comprises histidine (His)-rich proteins with repetitive tandem repeats and conserved GHG sequence motifs. One of the His-rich proteins was cloned, expressed and purified and was shown to undergo a fluid-fluid phase separation from equilibrium solution. Fluidic phase separation of proteins and other water-soluble polymers is known as coacervation[18], a process during which two distinct phases form, namely a concentrated and a dilute protein phase, respectively. Studies on His-rich protein–derived tandem repeat peptides identified the primary domains that promote coacervation and enable beak maturation by efficient delivery of His-rich proteins to their final location before quinone crosslinking of the conserved GHG sequence motifs. A model of the protein-protein and protein-chitin interactions that define the final biomolecular gradient is presented. In addition, the study suggests new bio-inspired avenues to process interpenetrating polymer networks based on aqueous-phase chemistry.

[1]Biomolecular Science and Engineering Program, University of California, Santa Barbara, California, USA. [2]Molecular Engineering Lab, Biomedical Sciences Institutes, Agency for Science, Technology and Research (A*STAR), Singapore. [3]School of Biological Sciences, Nanyang Technological University, Singapore. [4]Energy Research Institute at Nanyang Technological University (ERI@N), Nanyang Technological University, Singapore. [5]Biological & Biomimetic Material Laboratory, School of Materials Science & Engineering, Nanyang Technological University, Singapore. [6]Materials Research Laboratory, University of California, Santa Barbara, California, USA. [7]Department of Molecular Cellular and Developmental Biology, University of California, Santa Barbara, California, USA. *e-mail: herbert.waite@lifesci.ucsb.edu or ali.miserez@ntu.edu.sg
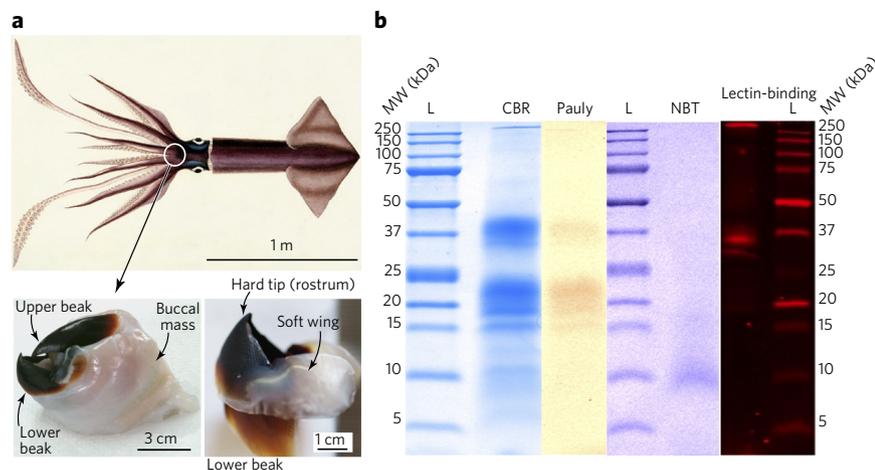
Figure 1 | *D. gigas* beaks and protein extraction from a beak. (**a**) Photographs and drawing of a whole animal (top), a dissected beak embedded within its soft buccal mass (bottom left) and a lower beak portion (bottom right). Drawing from ref. 45. (**b**) Tricine-SDS-PAGE of NH$_2$OH-extracted beak proteins stained with CBR (to detect all proteins), Pauly stain (for His-rich proteins) and NBT (for catechols). Lectin-binding assay (rightmost panel) was used to detect *N*-acetylglucosamine oligomers. Total amounts of protein loaded per lane in gels for the different stains are 10 μg, 30 μg, 30 μg and 60 μg, respectively. L, ladder.

RESULTS

Beak transcriptome library

An RNA-Seq transcriptome library was assembled from mRNAs isolated from the buccal mass tissue, which contains beccublast cells that secrete the beak proteins[16]. As no reference squid genomes are available, a transcript sequence database was generated from the short reads by *de novo* transcript assembly using the Trinity software suite[17]. The assembled transcript sequences were used to identify the full-length sequences of beak proteins either by sequence alignment, when a peptide fragment sequence was available, or by matching amino acid profiles. The relative abundance of transcripts described in this study is provided in **Supplementary Results**, **Supplementary Table 1**. Because of the highly repetitive nature of these proteins, we verified the assembly of these sequences by RACE-PCR followed by Sanger sequencing.

D. gigas chitin-binding beak proteins

Classical protein extraction methods for soft tissues and cells or enzymatic degradation are typically unsuitable for densely cross-linked tissues such as the squid beak. Instead, an alternative method was developed in which peptide bond–cleaving reagents were used to chemically cleave the constituent proteins so that proteins and peptide fragments were released from the beak. Hydroxylamine (NH$_2$OH), which cleaves between asparaginyl-glycyl peptide bonds[19], was used since *D. gigas* beak proteins were previously reported to be rich in asparagine and aspartic acid (Asx) and glycine (Gly) amino acid residues[6,20]. NH$_2$OH extraction resulted in distinct beak proteins and/or protein fragments, ranging in size from 5 kDa to more than 250 kDa, that could be visualized easily by SDS-PAGE analysis (**Fig. 1b**). Some of the protein bands were positive for Pauly's stain, indicating the presence of His-rich polypeptides[21]. In addition, nitroblue tetrazolium (NBT) staining showed that the smaller proteins contained catecholic residues, whereas a lectin-binding assay confirmed the presence of chitin oligomers in the higher-molecular-weight fractions. In-gel digestion followed by nanoscale liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS; for internal sequencing) or Edman (N-terminal) sequencing were then used to obtain partial sequences from these NH$_2$OH-extracted proteins. A family of three beak proteins was subsequently identified by using the obtained partial sequences

to query the beak transcriptome. These proteins were named *D. gigas* chitin-binding beak proteins (DgCBP-1, DgCBP-2 and DgCBP-3) because each protein in this family contains two or three chitin-binding type 2 domains (**Fig. 2** and **Supplementary Fig. 1a**) as determined by ScanProsite[22]. A fourth chitin-binding beak protein (DgCBP-4) was identified when the transcriptome was further searched for proteins with the same chitin-binding type 2 domains. All four cDNA sequences were confirmed by 3′ RACE even though DgCBP-4 was not among the NH$_2$OH-extracted native beak proteins (**Supplementary Table 2**). The chitin-binding characteristics of DgCBP-1–DgCBP-3 were also verified using a chitin-binding assay (**Supplementary Fig. 2**).

D. gigas histidine-rich beak proteins

Comparing the amino acid compositions of the DgCBPs with those previously reported for the beak[6] (**Table 1**) shows a clear trend: DgCBP amino acid profiles are similar to those of the untanned portions of the beak, but very different from those of the tanned regions, which include the rostrum. It is noteworthy that unlike the tanned beak portions, DgCBPs have low levels of His, Gly and alanine (Ala) residues. However, some of the NH$_2$OH-extracted proteins were Pauly positive (**Fig. 1b**), indicating that these extracts were His rich. Thus, we postulated that another set of beak proteins containing high levels of His, Gly and Ala was present in the beak. We therefore screened the beak transcriptome for proteins that contained a compositional bias (His, Gly and Ala compositions of 10–15%, 20–35% and 15–20%, respectively), and three proteins were readily identified (**Supplementary Table 3**). Their corresponding full-length sequences were confirmed by RACE-PCR, and an extensive search of the GenBank and Swiss-Prot databases indicated little to no homology to any known protein, suggesting that they represent a unique class of proteins which we have named *D. gigas* histidine-rich beak proteins (DgHBPs). As expected, all three proteins exhibited a heavy bias of Gly, Ala and His, which together accounted for 62%, 53% and 68% of DgHBP-1, DgHBP-2 and DgHBP-3, respectively, a composition that matches the overall composition of the beak's stiff tanned regions (**Table 1**). In addition, these proteins show 39–56% homology to one another at the primary sequence level. For DgHBP-2, the transcriptome assembly predicted multiple isoforms. Although this could arise from mis-assembly due to the repetitive nature of the proteins, we were able to confirm the presence of multiple isoforms by RACE cloning and sequencing of DgHBP-2 (**Supplementary Table 4**). Whether these are generated by alternative splicing or gene duplication is presently unclear.
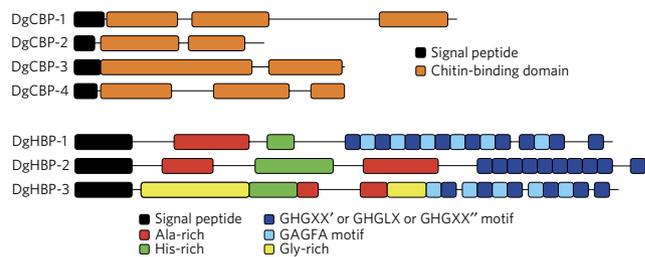
Figure 2 | Schematic representations of DgCBP-1–DgCBP-4 and DgHBP-1–DgHBP-3. The non-repetitive and modular domains present in each protein are indicated. X is a hydrophobic residue, while X′ is usually Y and X″ is mostly G or A.

Table 1 | Selected amino acid content of DgCBPs, DgHBPs and squid beak

	Amino acid			
	Asx	**Gly**	**Ala**	**His**
Tanned rostrum	7	27	15	12
Untanned (soft wing)	15	7	6	3
DgCBP-1	8.9	5.5	3.1	3.1
DgCBP-2	13.5	5.7	2.1	5.7
DgCBP-3	13.8	6.9	7.4	1.5
DgCBP-4	11.2	3.4	4.9	2.4
DgHBP-1	4.0	32.2	19.3	10.4
DgHBP-2	10.3	20.7	16.3	15.8
DgHBP-3	1.9	34.4	19.1	14.8

Shown are the content (in mol %) of selected amino acids for DgCBPs and DgHBPs, and a comparison with the known amino acid compositions of the untanned (soft wing) and fully tanned (rostrum) regions of the *D. gigas* squid beak[6].
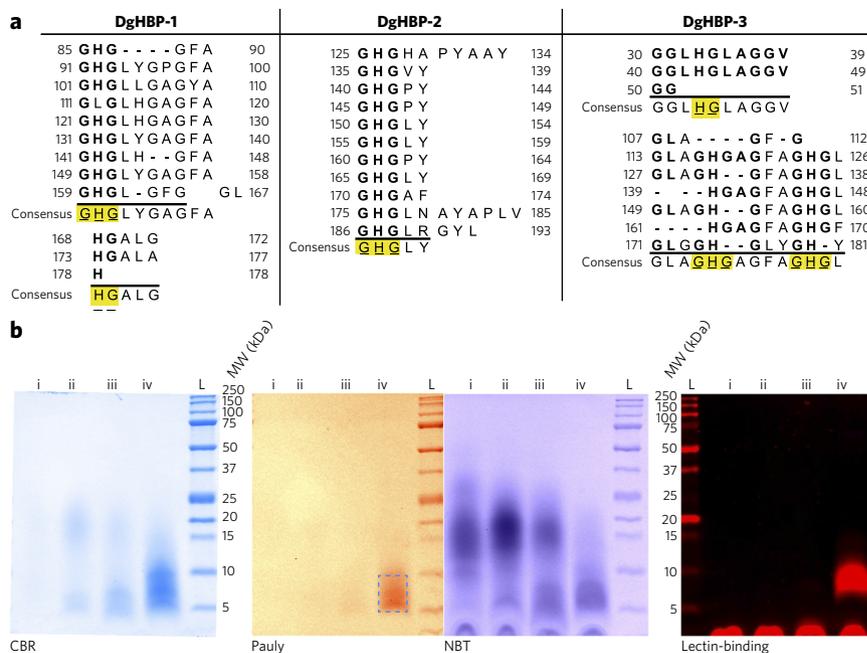
The aligned primary sequences and modular architectures of the DgHBPs are shown in **Supplementary Figure 1b** and **Figure 2**, respectively. All three proteins exhibit similar large-scale modular architectures, with a generalized two-domain organization in which the N-terminal regions of the proteins are relatively non-repetitive and contain stretches of Ala- and His-rich regions, whereas the C-terminal regions are dominated by tandem His-rich repeats. For example, the C-terminal halves of DgHBP-1 and DgHBP-3 comprise GAGFAGHGXX′/X″ repeats, whereas DgHBP-2 contains tandem repeats of GHGXY, where X is a hydrophobic residue, X′ is usually Y, and X′ is generally G or A (**Fig. 2**). The modular design of DgHBP-3 is similar to that of DgHBP-1; however, its N-terminal region contains two Gly-rich repeats not found in DgHBP-1. When the C-terminal tandem repeats were aligned, we observed that the His residues were flanked either on both sides or just on the C-terminal side by Gly residues in all three DgHBPs (**Fig. 3a**). Summing up—on the basis of composition similarity, peptide usage and their large-scale modular architectures—we conclude that DgHBPs are encoded by a unique gene family.

We then used our knowledge of the DgHBP sequences to corroborate their presence in the beak. Because the putative DgHBP transcripts contained noticeably more aspartate (Asp) than asparagine (Asn) residues, we performed DgHBP extraction with formic acid, which cleaves on either side of Asp residues[23]. The acid-liberated beak protein fragments were small (<10 kDa), with some weak bands in the 20 kDa range. A few of the bands contained *N*-acetylglucosamine oligomers and also reacted positively with Pauly's and NBT stains (**Fig. 3b**). These bands were digested in-gel and subjected to nanoLC-MS/MS. Given the regularity of tyrosine (Tyr) residues in the predicted DgHBP sequences, chymotrypsin was then used to produce peptide fragments. *De novo* partial sequences obtained from these fragments corresponded to the cDNA-deduced full-length sequences of DgHBP-1 and DgHBP-3 (**Supplementary Table 3**). The data therefore support the view that the tanned region of the beak is enriched in DgHBPs.

Self-coacervation of recombinant DgHBP-1

Previous work using poly-His and poly-Asp mimics had suggested that squid beak proteins with opposite charges can be processed as complex coacervates before coating and permeating the beak chitin network[24]. However, from the sequence analysis it appears highly unlikely that native DgHBPs engage in complex coacervation, because that would require a polyanionic partner. Indeed, the theoretical pIs of DgCBPs and DgHBPs (~8.8 and ~6.3, respectively) indicate that there is no pH window in which these proteins would be oppositely charged with a high enough charge density to form complex coacervates. On the other hand, we noticed that the modular domains of DgHBPs, comprising pentapeptides such as GHGXY (where **X** is often valine (Val; V) or leucine (Leu; L)), share intriguing homology with the Gly- and Val-rich hydrophobic sequence repeats of elastin[25], which are well-known to drive self-coacervation of tropoelastin proteins[26] via hydrophobic interactions, a mechanism that was also recently established for the mussel adhesive foot protein, *Mfp-3S*[27].

To test this hypothesis, we cloned and expressed DgHBP-1 in *Escherichia coli* and purified it to homogeneity (**Supplementary Fig. 3**) for subsequent characterization. We tested various conditions and found that a coacervate phase formed when the soluble *rec*DgHBP-1 (in 10 mM acetic acid) was introduced into a 100 mM PBS buffer solution at pH 6.5 at a buffer:protein solution volume ratio of 9:1. As illustrated in **Figure 4a**, pipetting the soluble *rec*DgHBP-1 into the buffer solution resulted in the spontaneous formation of a cloudy phase, hereafter referred to as 'condensed coacervate' (**Supplementary Video 1**). Optical micrographs confirmed the presence of fluidic droplets (**Fig. 4b**). Some irregular-shaped aggregates also formed over time, perhaps suggesting a metastable equilibrium between an insoluble fluid phase and a solid aggregate; however, fluidic droplets were clearly dominant when the solutions were first mixed (**Fig. 4b**). Circular dichroism (CD)

a

DgHBP-1		
85	GHG - - - - GFA	90
91	GHGLYGPGFA	100
101	GHGLLGAGYA	110
111	GLGLHGAGFA	120
121	GHGLYGAGFA	130
131	GHGLYGAGFA	140
141	GHGLH - - GFA	148
149	GHGLYGAGFA	158
159	GHGL - GFG GL	167
Consensus	GHGLYGAGFA	
168	HGALG	172
173	HGALA	177
178	H	178
Consensus	HGALG	

DgHBP-2		
125	GHGHA PYAAY	134
135	GHGVY	139
140	GHGPY	144
145	GHGPY	149
150	GHGLY	154
155	GHGLY	159
160	GHGPY	164
165	GHGLY	169
170	GHGAF	174
175	GHGLN AYAPLV	185
186	GHGLR GYL	193
Consensus	GHGLY	

DgHBP-3		
30	GGLHGLAGGV	39
40	GGLHGLAGGV	49
50	GG	51
Consensus	GGLHGLAGGV	
107	GLA - - - - GF - G	112
113	GLAGHGAGFAGHGL	126
127	GLAGH - - GFAGHGL	138
139	- - - HGAGFAGHGL	148
149	GLAGH - - GFAGHGL	160
161	- - - - HGAGFAGHGF	170
171	GLGGH - - GLYGH - Y	181
Consensus	GLAGHGAGFAGHGL	

b

CBR   Pauly   NBT   Lectin-binding

Figure 3 | Primary sequence and chemical characteristics of DgHBPs. (**a**) Tandem repeats of DgHBPs. Underlined amino acids in the consensus sequence are fully conserved within the repeats. Proposed crosslinking motifs GHG and HG are highlighted in yellow. (**b**) Formic acid–extracted proteins were separated using Tricine-SDS-PAGE and stained with the different stains after isolation by size-exclusion and reverse-phase HPLC. The blue rectangle corresponds to His-rich proteins that were subjected to in-gel digestion with chymotrypsin for protein sequencing by MS/MS. Latin numbers correspond to reverse-phase HPLC fractions: (i) 22–28 min, (ii) 28–32 min, (iii) 32–36 min and (iv) 36–44 min. L, ladder.
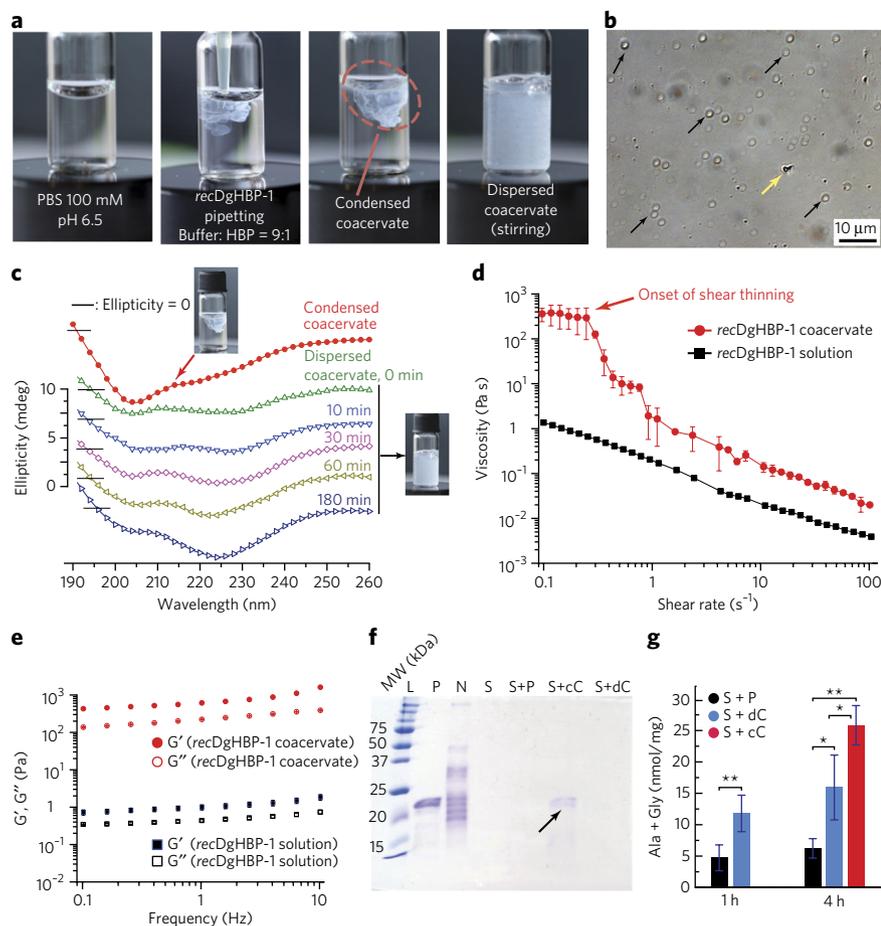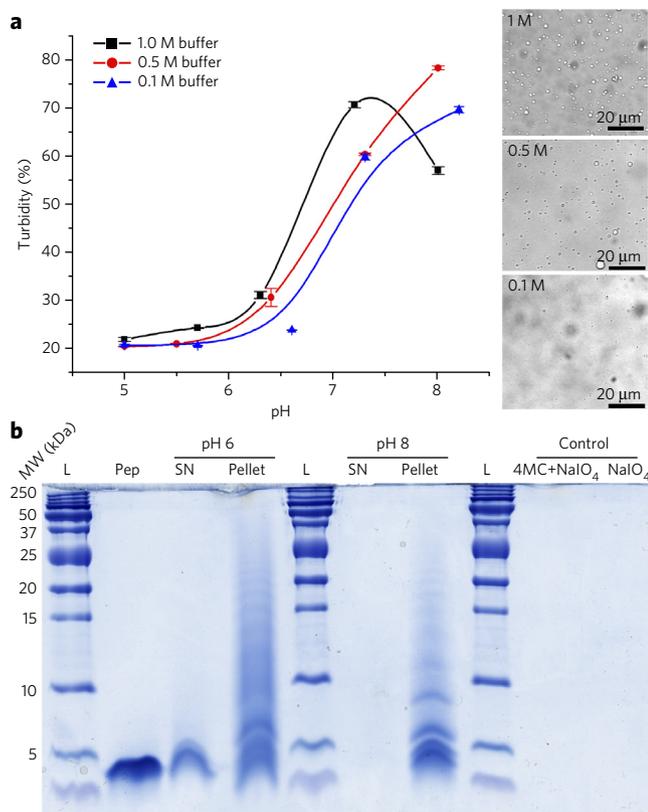
Figure 4 | Coacervation and characterization of *rec*DgHBP-1. (a) Coacervation of *rec*DgHBP-1 (see Online Methods for details). The cloudy phase corresponds to the "condensed coacervate," which upon stirring and vigorous mixing is referred as "dispersed coacervate." **(b)** Optical micrograph of the coacervate phase showing the characteristic fluidic droplets (black arrows) and an irregularly shaped aggregate (yellow arrow). **(c)** CD spectra of the condensed and dispersed coacervates. **(d)** Viscosity of the condensed coacervate and protein solution as a function of the shear rate, illustrating the characteristic shear thinning behavior of the coacervate phase. See Online Methods for statistical details here and throughout. **(e)** Frequency sweep measurements of storage (G'; closed symbols) and loss (G''; open symbols) moduli of the coacervate (red circles) and protein solution (black boxes). Error bars are of similar size to data points. **(f)** SDS-PAGE gel of infiltrated beak scaffolds with DgHBP-1 coacervates and protein solutions (see text for details; L, ladder; P, purified recombinant DgHBP-1; N, native beak sections before de-proteinization; S, beak sections after de-proteinization ("scaffolds"); cC, condensed coacervate; dC, dispersed coacervate). **(g)** Amino acid composition of infiltrated beak scaffolds expressed in nmol of (Ala + Gly) normalized by the dry weight of the scaffold pre-infiltration (N = 3, three technical replicates and two biological replicates; t-test, **$P < 0.001$, *$P < 0.05$). All error bars represent mean values ± s.d.

Another key attribute of coacervates is shear-thinning behavior[29], which is characterized by a decrease of viscosity above a critical shear rate and is regarded as beneficial for delivery of a concentrated polymer within a confined environment. Using rheometry, we subjected the *rec*DgHPB-1 condensed coacervates to dynamic torsional shear and measured their viscosity as a function of the shear rate (**Fig. 4d**). A distinct shear-thinning behavior was observed, with the viscosity sharply decreasing from about 300 Pa s to less than 10 Pa s above the critical shear rate of 0.3 s⁻¹. The viscosity then decreased monotonically with the shear rate above 1 s⁻¹. In contrast, for the *rec*DgHPB-1 solution, the viscosity was much lower at low shear rates and no shear thinning was noticed. We also measured the storage (G') and loss (G'') moduli of the *rec*DgHPB-1 coacervate and solution (**Fig. 4e**). Both moduli were observed to be more than two orders of magnitude higher for the coacervate phase, consistent with the higher protein concentration in the coacervate and with the rheological behavior of elastin-like polypeptide coacervates[30]. For a given initial protein concentration, no significant changes in G' and G'' values of the coacervates were observed over time (**Supplementary Fig. 4**).

We speculated that the specific physicochemical characteristics of the coacervates, as a concentrated fluid with low surface tension[31] and shear-thinning behavior, would allow them to infiltrate the initially soft scaffold of hydrated chitin nanofibers. To verify this hypothesis, soft sections from native beaks were cut and incubated with *rec*DgHBP-1 coacervates or solution and were then analyzed by SDS-PAGE in order to assess whether the coacervate phase efficiently diffused through the chitin scaffold. Because native chitin-binding proteins in the soft wing regions interfered with the assay, the samples were de-proteinized by alkali peroxidation treatment[6] before use. Protein extraction post-infiltration (following a washing step) indicated the presence of *rec*DgHBP-1 when the scaffold had been incubated with the condensed coacervate phase (S + cC), but not with the protein solution (S + P) or the dispersed coacervate (S + dC) (**Fig. 4f**). Since this outcome was also possible if the total protein content was below the SDS-PAGE detection limit for *rec*DgHBP-1 in the dispersed and solution states, we also quantitatively measured the amino acid composition of the scaffolds post-infiltration (**Fig. 4g**). In these experiments, the total *rec*DgHBP-1 concentration for the protein solution (in 10 mM acetic acid) and the dispersed coacervate were kept identical, thus providing a rigorous control to assess the influence of coacervation on the wicking behavior. The protein content measured in the scaffolds incubated with the dispersed coacervate was approximately threefold higher than that measured in the scaffolds incubated with the protein solution for 1 or 4 h, thus demonstrating that the coacervate phase is an efficient phase for infiltrating the nanoporous chitinous beak scaffold. We also measured the protein content for the scaffolds incubated with the condensed coacervate for 4 h. Not surprisingly, and in agreement

analyses of the condensed coacervate phase indicated a mostly helical conformation in the concentrated coacervate phase (**Fig. 4c**), with the two characteristic minima at 208 and 220 nm. However, vigorous dispersion of the condensed coacervate into its suspension state resulted in a different CD signature, with a decrease in the absolute value of the minimum at 208 nm and the appearance of a broad minimum at approximately 225 nm, indicative of β-sheets. These spectral features progressively evolved (**Fig. 4c**), with the ellipticity of the 225-nm minimum becoming more negative over time (3 h), suggesting the high mobility and gradual conformational transitions of protein chains within the coacervates. This behavior is consistent with the highly dynamic nature of the coacervates at the subnanometer scale[28] and with the notion that coacervates are in metastable equilibrium.

Figure 5 | Self-coacervation and crosslinking of hydrophobic DgHBP-*pep*. (**a**) Turbidity measurement of DgHBP-*pep* mixed in different buffer conditions (molar values represent the ionic strength of the buffer) at the protein concentration of 1 mg/ml (*N* = 3, technical replicates, error bars represent mean values ± s.d.). Optical micrographs show the coacervates settled onto the glass slides for different ionic strengths at pH 8. (**b**) SDS-PAGE gel showing crosslinked products of DgHBP-*pep* with 4MC. Multimers of the peptide were observed in the pellets at both pH 6 and 8. The trace is more intense for pH 6 because the pellet from pH 8 was partially insoluble in gel sample buffer. Monomers and dimers of DgHBP-*pep* were also observed in pH 6 supernatant. Pep, DgHBP-*pep* prior to crosslinking; SN, supernatant; L, ladder.

with the SDS-PAGE results (**Fig. 4f**), a significantly higher protein content was measured (25.8 ± 3.1 nmol/mg) than for scaffolds incubated with the dispersed coacervate (15.9 ± 5.2 nmol/mg) or the protein solution (6.2 ± 1.6 nmol/mg).

Coacervation domains of DgHBPs

To further test the hypothesis that the (GHG**XY**)$_n$ tandem repeats were responsible for the self-coacervation of DgHBPs, we chemically synthesized the peptide DgHBP-*pep* (GHGVY GHGVY GHGPY GHGPY GHGLY), which was designed from the consensus sequence of DgHBP-2 variants (**Supplementary Fig. 5**) and shared more than 90% homology with all DgHBPs. Based on its hydropathy plot, we noted that DgHBP-*pep* is mostly hydrophobic (**Supplementary Fig. 6**). Because turbidity increases when macromolecules associate to form phase-separated fluidic droplets[32,33], turbidity measurements at 600 nm were used to quantify DgHBP-*pep*'s self-coacervation. The data clearly demonstrated that under the appropriate microenvironmental conditions, DgHBP-*pep* exhibited self-coacervation behavior, with optimal coacervation occurring at pH 8 and a 0.5 M ionic strength (**Fig. 5a**). Notably, these conditions are close to the pH and salinity of seawater (pH 8, 0.3–0.7 M)[34]. Together with coacervation studies on the full-length DgHBP-1, these results support the notion that DgHBPs may be secreted as soluble precursors that phase separate into coacervates upon encountering seawater, and that this behavior is largely driven by the (GHG**XY**) repeats of the DgHBPs.

Interestingly, at pH 8 and 1.0 M ionic strength, although particulates could be observed settling during the UV-vis measurements (leading to lower turbidity than for other ionic strengths), no precipitation was observed by optical microscopy at all conditions tested. When DgHBP-*pep* (1 M ionic strength, pH 8 buffer) was tracked over a 24-h period, it was observed to undergo several phase transitions, first from solution to solid and then to coacervate (**Supplementary Fig. 7a**). Lastly, a coacervate film could clearly be seen at the bottom of the cuvette after the experiment (**Supplementary Fig. 7b**). Together these observations suggest that DgHBP-*pep* may be straddling the phase transition regime between coacervation and precipitation at conditions similar to those of seawater.

Crosslinking of DgHBP-*pep* with 4-methylcatechol

Previous work on native beak tissue provided direct evidence of extensive and complex crosslinking of the beak proteins involving catecholic moieties and His residues[14]. Therefore, we investigated the crosslinking potential of the DgHBPs with catechols using DgHBP-*pep*. Sodium periodate (NaIO$_4$) was used to oxidize 4-methylcatechol (4MC) to the quinone form before crosslinking with deprotonated peptidyl His residues via Michael addition reactions[35]. Crosslinking was performed on both solution and coacervate forms of DgHBP-*pep* (0.5 M ionic strength, pH 6 and 0.5 M ionic strength, pH 8, respectively) (**Supplementary Fig. 8a**). Incubation of DgHBP-*pep* with 4MC resulted in insoluble precipitates for both conditions tested (**Supplementary Fig. 8b,c**). However, at pH 6 (where DgHBP-*pep* was in the solution phase), the crosslinked products mostly adhered to the walls of the microcentrifuge tube, whereas at pH 8 (where initial peptides were in the coacervate phase), the particulates were observed at the bottom of the tube. This suggests that both forms of DgHBPs are readily amenable to crosslinking but that the more concentrated coacervate form allows for a higher number of condensed particles to be formed.

Gel electrophoresis detected monomeric and multimeric forms of DgHBP-*pep* in the pellets for both the pH 6 and 8 conditions, but not in the supernatant at pH 8 (**Fig. 5b**). This shows that coacervate-promoting conditions are more conducive to crosslinking. Hydrolyzing the crosslinked products gave rise to crosslinking adducts with m/z = 278 (**Supplementary Fig. 9**) that corresponded to crosslink adducts previously identified in hydrolyzed squid beak[14], which were proposed to arise from His-4MC Michael-type couplings[36]. This demonstrates that the DgHBP-*pep*, and by extension DgHBPs, are readily coupled by His-catechol crosslinking. Although a small molecule, 4MC, was used in this assay, it should be noted that DgHBPs contain Tyr residues that can be post-translationally modified into the catecholic side chain 3,4-dihydroxyphenylalanine (Dopa) by monophenol hydroxylase, an enzyme that was indeed detected in the beak transcritpome (**Supplementary Fig. 10**).

DISCUSSION

Squid beak contains one of the most wide-ranging functional stiffness gradients (200-fold) known to exist within any single material. This study provides the first glimpse of the molecules and processes involved in creating such a gradient. By combining chemical fragmentation strategies with transcriptome analysis of the buccal tissue surrounding the beak[37], which is key in the absence of a genomic database for *D. gigas*, we have successfully deconstructed a mature squid beak into its component proteins. By further preparing full-length recombinant proteins as well as synthetic peptides and investigating their phase behavior in solution, we have discovered a probable fabrication process for the squid beak. These insights and methodologies will undoubtedly help with the
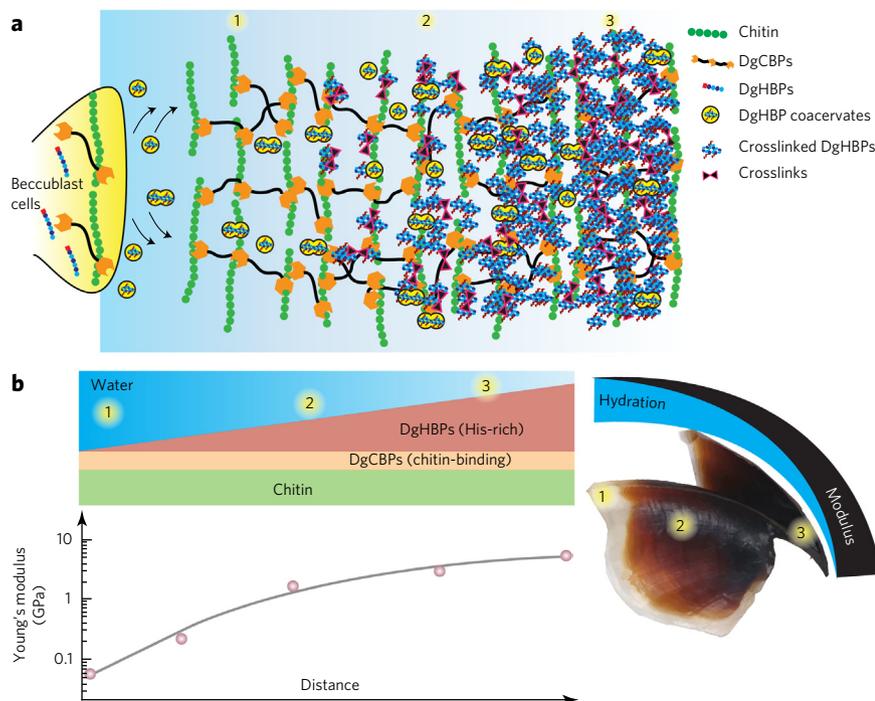
Figure 6 | Proposed model for jumbo squid beak gradient formation. (a) Squid beak is processed via the following steps: (1) DgCBPs are secreted together with chitin fibers and form a hydrophilic network via bi- and trivalent physical bonds in the untanned region. (2) As the beak ages, increasing levels of hydrophobic DgHBPs are secreted and phase separate into coacervates to impregnate the chitin network, resulting in dehydration of the hydrophilic network. (3) Crosslinking of His residues with catecholic moieties ensues and further dehydrates the DgCBP-chitin network, leading to higher beak modulus and hardness, which also correlates with a more tanned coloration. A modulus gradient can thus be created and maintained by differential titration of DgHBPs into the DgCBP-chitin network and then crosslinked to produce a hydration gradient along the beak. **(b)** Left, graphical representation comparing the squid beak's hydration gradient to its stiffness (pink circles). Distance, distance from proximal end of beak. Processing steps from the model are also noted in a squid beak picture (right) to highlight their equivalence in a native squid beak.

characterization of other fascinating sclerotized biomaterials such as arthropod fangs or marine worm jaws[15].

The two distinct families of structural proteins in squid beaks are the DgCBPs, which are likely to bind chitin via multiple physical interactions, and the DgHBPs, which become chemically crosslinked to one another[14] and probably also to the chitin-binding proteins and chitin, as manifested by gel stains of the extracted fragments. DgCBPs were detected with higher relative abundance in the untanned regions of the beak (**Table 1**), suggesting that these proteins are probably co-secreted with chitin fibers early on in beak development and serve as physical anchors (via hydrogen bonds) to hold the chitin fibers together as a scaffolding template. At this stage, the chitin-DgCBP network dominates the untanned portion of the squid beak (**Table 1**) where it is not chemically crosslinked, but is compliant and well hydrated (**Fig. 1a**). Next, hydrophobic DgHBPs are secreted into the beak and diffuse through the chitin-DgCBPs network to reach their final destination, where their physicochemical characteristics endow the beak with high mechanical strength. Primary amino acid sequence analysis and biophysical characterization of the DgHBPs suggest that this stepwise process is made possible by the modular design of these proteins and results in their multifunctionality.

In order to reach their final destination, DgHBPs must easily wet and diffuse through the chitin-DgCBP network, which raises at least one important physicochemical obstacle. The proteins must be fluidic enough to flow to their intended destination, a process

that could, however, be limited by the high viscosity of a concentrated fluidic protein phase. We propose that this obstacle is overcome by the formation of coacervates. Coacervates are implicated in the processing of other marine biomaterials, for example, mussel threads[27] and sandcastle worm cement[38]. The properties of coacervates[39] include high polymer concentration, relatively low viscosity, very low interfacial energy and shear-thinning behavior[40]. These characteristics make coacervates an ideal modality for spontaneous spreading and wicking of concentrated proteins into a nanoporous network. Notably, viscous flow at a fixed flow rate within a porous network is characterized by a r^{-3} dependence of the shear rate on the pore size r (ref. 41). Therefore, for small nanoscale porous scaffolds, the shear rate is high, resulting in a low viscosity of the coacervate phase due to shear thinning, which would facilitate the delivery of the highly concentrated phase to its final location. In addition, coacervates are less solvated, a characteristic that enables local dehydration simply by varying the coacervate supply to the different beak regions. This is important for beak processing as dehydration was significantly correlated to stiffness in the beak[6]. Our data demonstrated that *rec*DgHBP-1 is capable of self-coacervation, with suitable physicochemical properties, notably shear thinning, for impregnation of a nanoporous network (**Fig. 4d**). Furthermore, studies on DgHBP-*pep* strongly suggest that the C terminus modular repeats (GHGXY) drive the coacervation. Turbidity measurements showed that self-coacervation of DgHBP-*pep* was optimal at conditions close to those of seawater (0.5–1 M ionic strength, pH 8). Given that DgHBP-*pep* is strongly hydrophobic (**Supplementary Fig. 6**) and that optimal self-

coacervation happens when the peptide is uncharged, it follows that coacervation of the peptide is strongly influenced by hydrophobic interactions in a manner similar to that of tropoelastin[26,42]. Furthermore, optical microscopy detected a tendency toward precipitation over time, perhaps indicating that DgHBP coacervates are in a metastable state that the animal may control kinetically in order to trigger precipitation before final curing.

The second functional role of DgHBPs is to mechanically stabilize and strengthen the beak, which we suggest is made possible by two features of their primary sequence. At their N termini, DgHBPs are nonrepetitive and contain poly-Ala or Ala-rich domains (**Fig. 2**). Ala-rich sequences are known to mediate the formation of nanoconfined β-sheets, whereby cooperative hydrogen bonding enhances load-bearing properties, as observed in various biomolecular structures such as dragline silks[43,44]. *rec*DgHBP-1 forms β-sheets as coacervates over time (**Fig. 4c**), which could suggest the formation of similar stiff nanocrystals upon precipitation. Another striking feature of DgHBPs is the strong conservation of multiple copies of GHG or HG motifs, which are well-suited for chemical coupling because the flanking Gly residues result in localized high flexibility of the peptide backbone and minimal steric hindrance for His reactivity with *o*-quinones. Our crosslinking experiments on DgHBP peptides clearly demonstrate that these domains form crosslinks identical to adducts isolated from native beak[14] and that the coacervated environment clearly facilitates efficient crosslinking to form highly insoluble particulates. In native tissue, His residues

in DgHBPs could crosslink with 4MC and/or post-translational Dopa produced by monophenol hydroxylase detected in the beak transcriptome (**Supplementary Fig. 10**). However, more characterization of native tissue will be needed to identify the *in vivo* crosslinking partners.

In summary, we propose that processing and curing of the biocomposite beak structure occurs by a three-step bioprocessing pathway (**Fig. 6**). In the first step, chitin and chitin-binding DgCBPs are co-secreted to form the initial scaffold. DgCBPs may also alter the hydrophobic characteristics of chitin. Indeed, although chitin is quite hydrophobic, the soft regions made of the chitin-DgCBP complex are well hydrated in the native state (approximately 75% water per unit weight). Thus, DgCBPs may be able to enhance water adsorption of the soft parts in order to ensure a high flexibility at the attachment point to the soft buccal tissue, thereby avoiding mechanical mismatch at the beak/buccal tissue interface. Next, hydrophobic DgHBPs are secreted and form coacervates upon exposure to seawater, and this highly concentrated phase may also be conducive for the formation of stiff β-sheets by the Ala-rich domains of the proteins. Although the precise distribution of each beak protein in the hard region has not been determined, the relative weight fraction of each family of proteins can be estimated from the known dry weight fractions of chitin and total protein[6]. For this calculation, we note that DgHBPs are virtually absent from the soft regions. We also make the assumption that chitin and DgCBPs are first secreted and that any additional protein found in the hard region comes from the diffusion of DgHBPs. We find that in the hard rostrum, the weight ratio of DgHBPs to DgCBPs is about 9.5:0.5 (soft regions 0:10), as schematically illustrated in **Figure 6b**. Coacervate permeation of the chitin-DgCBP network then results in local dehydration and stiffening of the hydrophilic network. In the final step, dense crosslinks between His and catecholic molecules, such as 4MC and/or Dopa, results in further and irreversible water desolvation from the network. In effect, differential secretion of coacervated DgHBPs and subsequent crosslinking generate a hydration gradient that results in the hardness and stiffness gradients of the mature squid beak. Although the time scales associated with both processes (transport and crosslinking) remain unknown, transport is more probably the kinetically limited factor. In synthetic polymers and composites, hardening and stiffening via covalent crosslinking usually come at the expense of the fracture tolerance. In contrast, *D. gigas* beaks maintain a high fracture energy in their hardened rostra, which can be attributed to micro- and nanostructural toughening originating from the beak lamellar structure and to chitin fibril bridging, respectively[20].

The elucidation of beak protein sequences, demonstration of the processing of proteins by coacervation, and the identification of crosslinking motifs provide a wide range of new insights for biomimetic engineering using both peptide analogs and recombinant proteins, which could be exploited in the formulation of mechanically graded biomaterials, structural biocomposites, bioadhesives or smart and tough polymer networks and hydrogels. From a biological perspective we propose that these mechanisms may be more widespread in nature than currently recognized, notably in the bioprocessing of other extracellular biological structures of interest such as insect cuticles, which are also chitin- and His-rich protein biocomposites that undergo chemical stabilization through dehydration and His-catechol crosslink formation.

METHODS

Methods and any associated references are available in the online version of the paper.

References

1. Farren, L., Shayler, S. & Ennos, A.R. The fracture properties and mechanical design of human fingernails. *J. Exp. Biol.* **207**, 735–741 (2004).
2. Benjamin, M. *et al.* Where tendons and ligaments meet bone: attachment sites ('entheses') in relation to exercise and/or mechanical load. *J. Anat.* **208**, 471–490 (2006).
3. Lu, H.H. & Thomopoulos, S. Functional attachment of soft tissues to bone: development, healing, and tissue engineering. *Annu. Rev. Biomed. Eng.* **15**, 201–226 (2013).
4. Zajac, F.E. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Crit. Rev. Biomed. Eng.* **17**, 359–411 (1989).
5. Sun, C.J. & Waite, J.H. Mapping chemical gradients within and along a fibrous structural tissue, mussel byssal threads. *J. Biol. Chem.* **280**, 39332–39336 (2005).
6. Miserez, A., Schneberk, T., Sun, C., Zok, F.W. & Waite, J.H. The transition from stiff to compliant materials in squid beaks. *Science* **319**, 1816–1819 (2008).
7. Suresh, S. Graded materials for resistance to contact deformation and damage. *Science* **292**, 2447–2451 (2001).
8. Waite, J.H., Lichtenegger, H.C., Stucky, G.D. & Hansma, P. Exploring molecular and mechanical gradients in structural bioscaffolds. *Biochemistry* **43**, 7653–7662 (2004).
9. Mikos, A.G. *et al.* Engineering complex tissues. *Tissue Eng.* **12**, 3307–3339 (2006).
10. Helton, K.L., Ratner, B.D. & Wisniewski, N.A. Biomechanics of the sensor-tissue interface-effects of motion, pressure, and design on sensor performance and the foreign body response-part I: theoretical framework. *J. Diabetes Sci. Technol.* **5**, 632–646 (2011).
11. Fleckman, P. & Olerud, J.E. Models for the histologic study of the skin interface with percutaneous biomaterials. *Biomed. Mater.* **3**, 034006 (2008).
12. Fox, J.D., Capadona, J.R., Marasco, P.D. & Rowan, S.J. Bioinspired water-enhanced mechanical gradient nanocomposite films that mimic the architecture and properties of the squid beak. *J. Am. Chem. Soc.* **135**, 5167–5174 (2013).
13. Zvarec, O., Purushotham, S., Masic, A., Ramanujan, R.V. & Miserez, A. Catechol-functionalized chitosan/iron oxide nanoparticle composite inspired by mussel thread coating and squid beak interfacial chemistry. *Langmuir* **29**, 10899–10906 (2013).
14. Miserez, A., Rubin, D. & Waite, J.H. Cross-linking chemistry of squid beak. *J. Biol. Chem.* **285**, 38115–38124 (2010).
15. Rubin, D.J., Miserez, A. & Waite, J.H. Diverse strategies of protein sclerotization in marine invertebrates: structure-property relationships in natural biomaterials. *Adv. Insect Physiol.* **38**, 75–133 (2010).
16. Dilly, P.N. & Nixon, M. The cells that secrete the beaks in octopods and squids (Mollusca, Cephalopoda). *Cell Tissue Res.* **167**, 229–241 (1976).
17. Grabherr, M.G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
18. de Kruif, C.G., Weinbreck, F. & De Vries, R.K. Complex coacervation of proteins and anionic polysaccharides. *Curr. Opin. Colloid Interface Sci.* **9**, 340–349 (2004).
19. Smith, B.J. in *The Protein Protocols Handbook* (ed. Walker, J.M.) 507–510 (Humana Press Inc., 2002).
20. Miserez, A., Li, Y., Waite, J.H. & Zok, F. Jumbo squid beaks: inspiration for design of robust organic composites. *Acta Biomater.* **3**, 139–149 (2007).
21. Sahal, D. *et al.* Specific and instantaneous one-step chemodetection of histidine-rich proteins by Pauly's stain. *Anal. Biochem.* **308**, 405–408 (2002).
22. de Castro, E. *et al.* ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34**, W362–W365 (2006).
23. Li, A.Q. *et al.* Chemical cleavage at aspartyl residues for protein identification. *Anal. Chem.* **73**, 5395–5402 (2001).
24. Tan, Y., Yildiz, U.H., Wei, W., Waite, J.H. & Miserez, A. Layer-by-layer polyelectrolyte deposition: a mechanism for forming biocomposite materials. *Biomacromolecules* **14**, 1715–1726 (2013).
25. Mithieux, S.M. & Weiss, A.S. in *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers* (*Advances in Protein Chemistry* vol. 7) (eds. Parry, D.A.D. & Squire, J.M.) 437–461 (Elsevier Academic Press Inc., 2005).
26. Yeo, G.C., Keeley, F.W. & Weiss, A.S. Coacervation of tropoelastin. *Adv. Colloid Interface Sci.* **167**, 94–103 (2011).

27. Wei, W. *et al.* A mussel-derived one component adhesive coacervate. *Acta Biomater.* **10**, 1663–1670 (2014).

28. Ortony, J.H., Hwang, D.S., Franck, J.M., Waite, J.H. & Han, S. Asymmetric collapse in biomimetic complex coacervates revealed by local polymer and water dynamics. *Biomacromolecules* **14**, 1395–1402 (2013).

29. Kizilay, E., Kayitmazer, A.B. & Dubin, P.L. Complexation and coacervation of polyelectrolytes with oppositely charged colloids. *Adv. Colloid Interface Sci.* **167**, 24–37 (2011).

30. Betre, H., Setton, L.A., Meyer, D.E. & Chilkoti, A. Characterization of a genetically engineered elastin-like polypeptide for cartilaginous tissue repair. *Biomacromolecules* **3**, 910–916 (2002).

31. Spruijt, E., Sprakel, J., Cohen-Stuart, M.A. & van der Gucht, J. Interfacial tension between a complex coacervate phase and its coexisting aqueous phase. *Soft Matter* **6**, 172–178 (2010).

32. Weinbreck, F., de Vries, R., Schrooyen, P. & de Kruif, C.G. Complex coacervation of whey proteins and gum arabic. *Biomacromolecules* **4**, 293–303 (2003).

33. Kaibara, K., Okazaki, T., Bohidar, H.B. & Dubin, P.L. pH-induced coacervation in complexes of bovine serum albumin and cationic polyelectrolytes. *Biomacromolecules* **1**, 100–107 (2000).

34. Lin, J. & Brown, C.W. Spectroscopic measurement of NaCl and seawater salinity in the near-IR region of 680–1230 nm. *Appl. Spectrosc.* **47**, 239–241 (1993).

35. Liu, B., Burdine, L. & Kodadek, T. Chemistry of periodate-mediated cross-linking of 3,4-dihydroxylphenylalanine-containing molecules to proteins. *J. Am. Chem. Soc.* **128**, 15228–15235 (2006).

36. Kerwin, J.L. *et al.* Mass spectrometric analysis of catechol-histidine adducts from insect cuticle. *Anal. Biochem.* **268**, 229–237 (1999).

37. Guerette, P.A. *et al.* Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. *Nat. Biotechnol.* **31**, 908–915 (2013).

38. Zhao, H., Sun, C., Stewart, R.J. & Waite, J.H. Cement proteins of the tube-building polychaete *Phragmatopoma californica. J. Biol. Chem.* **280**, 42938–42944 (2005).

39. Oparin, A.I. & Synge, A. *The Origin of Life on the Earth* 3rd edn. (Academic Press, 1957).

40. Hwang, D.S. *et al.* Viscosity and interfacial properties in a mussel-inspired adhesive coacervate. *Soft Matter* **6**, 3232–3236 (2010).

41. Fournier, R.L. *Basic Transport Phenomena in Biomedical Engineering* 3rd edn. (CRC Press, Taylor & Francis Group, 2012).

42. Vrhovski, B. & Weiss, A.S. Biochemistry of tropoelastin. *Eur. J. Biochem.* **258**, 1–18 (1998).

43. Giesa, T. & Buehler, M.J. Nanoconfinement and the strength of biopolymers. *Annu. Rev. Biophys.* **42**, 651–673 (2013).

44. Gosline, J.M., Guerette, P.A., Ortlepp, C.S. & Savage, K.N. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* **202**, 3295–3303 (1999).

45. d'Orbigny, A.D. *et al. Voyage dans l'Amérique méridionale: Tome Neuvième* (Pitois-Levrault, Paris, 1847).

Author contributions

Y.T. carried out beak protein extraction and MS/MS experiments and prepared figures and tables. P.A.G. extracted and purified RNA samples, conducted 3′ RACE and designed DgHBP-1 cloning, expression and purification experiments. S.H. performed and analyzed RNA-Seq, did transcriptome analysis and RACE sequence confirmation and prepared figures. P.A.G., S.H., Y.T. and A.M. analyzed protein sequence data. C.H. expressed, purified and analyzed DgHBP-1. A.G. helped to purify DgHBP-1 and conducted rheological experiments. W.W. designed and carried out coacervation and infiltration experiments on *rec*DgHBP-1. A.G. and C.H. conducted the characterization and infiltration of coacervates under the supervision of P.A.G. and A.M. Y.T. and W.W. carried out and analyzed coacervation and crosslinking experiments on DgHBP-*pep*. Y.T., J.H.W., P.A.G., S.H. and A.M. wrote the manuscript. J.H.W. and A.M. designed and supervised the research.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to J.H.W. or A.M.

ONLINE METHODS

Beak protein extraction. *D. gigas* beaks (including the surrounding buccal mass) were dissected from animals collected off the coast of Ventura, California, in June 2007 (samples were waste products from recreational fishing) and stored at −80 °C until processing. To extract the beak proteins, either a whole freeze-dried *D. gigas* upper beak or beak pieces from different parts of the upper beak were triturated using a ceramic mortar and pestle under liquid nitrogen. The obtained beak powder was divided into multiple 5 ml microreaction vials such that each vial contained approximately 20–50 mg of beak powder. 1 ml of hydroxylamine (NH_2OH) extraction buffer was then added to each vial and the mixture incubated capped at 45 °C, with stirring, for 4 h. The buffer consisted of 2 M hydroxylamine hydrochloride (Sigma-Aldrich, ≥99%), 2 M guanidine hydrochloride (Sigma-Aldrich, ≥98%) and 0.2 M potassium carbonate (Sigma-Aldrich, ≥99%), adjusted to pH 9 with 10 N sodium hydroxide (Fisher, 30% w/w) solution per the protocol of Smith[19]. The reaction was stopped at 4 h with 3 ml of 2% trifluoroacetic acid (Pierce, ≥99.5%) (TFA) solution. The extraction can be repeated up to 3 times on the same beak powder before the process is futile. The NH_2OH-extracted proteins were then dialyzed against 5% acetic acid (Fisher Chemical, ≥99.7%) (HOAc) solution (200× volume) in 1 kDa MWCO dialysis tubing (Spectrum Laboratories, Inc.) for 12 h, with a change of 5% HOAc every 4 h.

To obtain His-rich proteins, beak powders and 1 ml of 2% formic acid (Sigma-Aldrich Fluka, >98%) solution were sealed under vacuum in 2 ml vacuum ampules and then incubated in an air-circulated oven at 108 °C for 2 h. Removal from high temperature and freeze-drying stopped the process. To estimate the quantity of proteins liberated by the chemical cleavage methods, the protein solutions (dialyzed or otherwise) were concentrated by freeze-drying and redissolved in 1 ml of 5% HOAc prior to estimation of protein concentrations on a Thermo Scientific NanoDrop 2000c (ThermoFisher Scientific, Inc.) using the built-in mass extinction coefficient of BSA ($6.67 \, l \, mol^{-1} \, cm^{-1}$).

Gel electrophoresis. Proteins were separated by gel electrophoresis by Tricine-SDS-PAGE using 4% stacking and 16% resolving gels, prepared as previously described[46], and then stained either with Coomassie Blue R-250 (CBR) stain[46], Pauly's stain[21] or nitroblue tetrazolium (NBT) stain[47], which stain for generic proteins, His-rich and catecholic proteins, respectively. Approximately 10 μg of proteins were loaded in each lane for CBR stain and 30 μg for Pauly's stain and NBT, unless otherwise stated.

Transcriptome library construction. RNA-Seq analysis was performed as previously described[17,37]. Briefly, total RNA from a *D. gigas* buccal mass (tissue that was in close proximity to the upper and lower beak were combined and used) was extracted with the RNeasy Mini Kit (Qiagen). Poly-A mRNA was then enriched from 10 μg of total RNA with oligo-dT beads (Invitrogen). The buccal mass sample was collected on R/V New Horizon (Scripps Institute of Oceanography, San Diego, CA, USA), June 2010, at Guayamas Basin, Mexico, with Permit No. DAPA/2/100510/1640.

~100 ng of recovered poly-A mRNA was used to construct an RNA-seq library using the ScriptSeq mRNA-Seq library kit v1 (Epicenter, Illumina) according to manufacturer's instructions. Phusion PCR polymerase (ThermoFisher Scientific, Inc.) was used for the final library amplification (12 cycles). PCR cleanup was performed with the MinElute PCR purification kit (Qiagen) and the library quality was assessed with an Agilent 21000 Bioanalyzer. The library was then sequenced on an Illumina GA IIx and 2 lanes of 2 × 76 bp paired-end reads were collected. *De novo* transcript assembly was performed on quality-filtered FASTQ sequences with the Trinity software suite using standard parameters on a computational cluster.

The final assembled transcript sequences were used for subsequent analysis. Putative coding regions were extracted with the 'transcripts_to_best_scoring_ORFS.pl' script provided in the Trinity software suite[17]. Alternatively, short peptide sequences generated from proteomic analysis were queried with translated blast (tblastn). Amino acid profiles for each predicted protein was generated using a custom Bioperl[48] script (**Supplementary Table 5**) in order identify proteins which match the amino acid composition of the beak proteins.

RACE-PCR. Rapid amplification of cDNA ends (RACE) was used to verify the complete sequence of Trinity-assembled transcripts. 3′ RACE cDNA libraries were generated using Invitrogen's Generacer Kit. All predicted transcripts were full length (encoding entire protein coding sequence) and gene-specific forward primers were designed (**Supplementary Table 6**) to amplify the entire coding sequence with the GenRace 3′ primer. KOD Xtreme Hot Start DNA Polymerase (Merck Millipore) was used for PCR and the amplicons were cloned into pCR2.1 by TA cloning and subjected to Sanger sequencing.

Partial peptide sequencing. Lyophilized proteins, extracted from the squid beak using NH_2OH or formic acid, were resuspended in 5% HOAc and separated first by size-exclusion HPLC (Shodex KW 802.5, 300 mm × 8 mm) and then reverse-phase C18 HPLC (PerkinElmer, Aquapore OD-300, 250 mm × 4.6 mm) using a linear gradient of 0–100% solvent B from 5 to 75 min. The elution buffers used were 5% HOAc for size exclusion, and A: Milli-Q water with 0.1% TFA and B: acetonitrile (Fisher Chemical, ≥99.9%) (ACN) with 0.1% TFA for reverse phase. Individual 1 ml fractions were collected at a rate of 1 per min and freeze-dried before the concentration was determined by NanoDrop. Approximately 10 μg from each fraction was visualized on Tricine-SDS-PAGE gels using CBR stain. Fractions showing the highest concentration of individual bands were then selected for in-gel protease digestion.

About 100 μg of the selected fractions were then loaded into multiple lanes of Tricine-SDS-PAGE gels to ensure strong bands that were distinct from the surrounding bands. These bands were excised with a razor blade and chopped into small pieces, approximately 1 mm × 1 mm in size, and in-gel digested with either trypsin or chymotrypsin per the protocol of Shevchenko *et al.*[49].

The digested products were then sequenced using nanoLC-MS/MS. HPLC was performed using an Agilent Technologies 1100 nanoLC series, equipped with a Zorbax 300SB-C18 (3.5 μm size, 0.075 × 150 mm) Nano column (Agilent Technologies). Sample aliquots (8 μl) were first injected using an autosampler into the trap (Agilent Technologies Zorbax 300SB-C18; 5 μm size, 0.3 × 5 mm) at a flow rate of 100 μl/min for 2.5 min. After a 20 min column wash at 400 μl/min with 2.5% solvent B, the peptides were eluted at 300 μl/min with linear gradients of (i) 2.5–50.0% B from 20.1–38.0 min, (ii) then 50.0–90.0% B from 38.0–40.0 min, (iii) maintain at 90% B for 10 min (40.0–50.0 min), (iv) 90.0–2.5% B from 50.0–51.0 min and (v) stop flow at 65.0 min. The solvents used were A: aqueous 0.1% formic acid and B: acetonitrile with 0.1% formic acid. Eluted peptides were directly sprayed into a Q-TOF-2 (Waters Micromass) working in positive-ion mode. Other settings were: electrospray (ESI) capillary voltage +3.8 V; sample cone voltage +45 V; and sample cone temperature 80 °C. Eluting peptides were subjected to MS scans (*m/z* 300–2,000) at a rate of 1.1 s/scan during the time period of 10–66 min. The three most abundant double-, triple- and quadruple-charged peptide ions were then subjected to collision-induced dissociation (CID) by alternating between the MS and MS/MS (*m/z* range of 50–2,000, 2.5 s/scan) scans. MS/MS spectra were post processed in MassLynx 4.1 (Waters Micromass) using MaxEnt3 and copied to the PepSeq module within MassLynx 4.1 for sequence prediction. Predicted sequences were confirmed by manual analysis.

To obtain N-terminal sequences, selected fractions were blotted to PVDF membrane (Millipore Immobilon-P^SQ) and stained with CBR stain after Tricine-SDS-PAGE separation. The bands were then excised and subjected to Edman sequencing (Iowa State University Proteomics Facility).

***In silico* analysis.** Predicted full cDNA sequences of the beak proteins from RNA-Seq were scanned against the PROSITE collection of motifs using the web-based ScanProsite tool with default selections[22]. Of the proteins that returned hits, all were of the "chitin-binding type 2" domain. The identified domains in these chitin-binding proteins were next aligned against a known chitin-binding type 2 domain of the peritrophin-44 protein (Uniprot Q25255|28-85), using the MUSCLE multiple alignment server[50] and visualized by JalView 2.8 (ref. 51) with ClustalX color scheme[52]. The remaining His-rich beak proteins with no hits to protein databases were analyzed using the XSTREAM *de novo* tandem repeat detection tool[53] to identify tandem repeats and additional fine adjustment of the repeats was done manually. These His-rich proteins were also screened for conserved residues using the ClustalO multiple sequence alignment program for proteins[54].

Chitin-binding assay (CBA). To verify the chitin-binding capability of the extracted beak proteins, a chitin-binding assay was adopted from ref. 55. One ml of commercially available chitin beads (New England Biolabs 6651S) was washed 3 times with 10 ml of binding buffer (syringe-filtered 0.5 M NaCl, 10 mM Tris-HCl, pH 7.0, 0.05% Triton X-100) in a new 15 ml centrifuge tube.

Individual 100 μg protein aliquots extracted chemically from the squid beaks were resuspended in 500 μl of binding buffer and then transferred to the washed chitin beads for a total volume of 3 ml binding buffer per aliquot. Each mixture was gently shaken at room temperature for 4 h, after which the mixture was transferred to individual polypropylene columns capped with a two-way stopcock. The flow-through was collected and each column was washed 4 times with 4 ml of binding buffer, which was individually collected in 15 ml centrifuge tubes. Bound proteins were then eluted with 5 ml of elution buffer (syringe-filtered 8 M urea, 5 mM Tris-HCl, pH 8.6) into individual 15 ml centrifuge tubes. The chitin beads were then stripped by boiling in 2 ml of 1% SDS.

The collected fractions were then dialyzed against 5% HOAc (200× volume) in 1 kDa MWCO dialysis tubing for 12 h, with a change of 5% HOAc every 4 h. Next, the dialyzed fractions were lyophilized and then resuspended in 1 ml of 5% HOAc prior to determining their concentration by NanoDrop. Tricine-SDS-PAGE gels were used to visualize the fractions, using the same volume for each fraction.

Lectin-binding assay. A lectin-binding assay adapted from ref. 56 was used to screen for *N*-acetylglucosamine oligomers in the extracted proteins. After gel electrophoresis as described above (~60 μg and ~100 μg per lane for NH₂OH-extracted proteins and formic acid-extracted proteins, respectively), the separated proteins were horizontally transferred to PVDF membrane (Millipore Immobilon-P^SQ) by electrophoresis (first 30 min at 200 mA, next 90 min at 400 mA) using a Genie blotter (Idea Scientific, 4017). The membrane was then blocked in TBS buffer (Tris-buffered saline, pH 7.5) with 3% BSA for 1 h, followed by an overnight 4 °C incubation in TBS buffer with 0.3% BSA, 0.1% Tween 20 and 5 μg/ml biotinylated *Datura stramonium* lectin (DSL) (Vector Laboratories, B-1185). After three washes with washing buffer (TBS with 0.3% BSA, 0.1% Tween 20), the membrane was incubated in washing buffer with 0.2 μg/ml Pierce streptavidin protein, DyLight 680 conjugated (Thermo Scientific 21848) for 1 h. The membrane was then washed three times with washing buffer and visualized in a LI-COR Odyssey Classic Imager. Ovalbumin (0.5 μg) was used as a positive control.

Cloning, expression and purification of recombinant DgHBP-1. A codon-optimized gene encoding the full-length sequence of DgHBP-1 (without signal peptide and with an M starting residue) was purchased from DNA 2.0 (Menlo Park, CA, USA). The gene was provided as an insert cloned into the multiple cloning cite of the expression ready pJexpress vector. 5 ng of the vector/insert construct was transformed into TOP 10 cells, spread onto Luria-Bertani plates containing 100 μg/ml ampicillin (LB-amp), and incubated overnight at 37 °C. The following day, individual colonies were inoculated into 4 ml LB-amp and grown for 6 h with shaking at 250 rpm. The resulting cultures were used to establish glycerol stocks. 10 μL of cells from the glycerol stock were then streaked onto LB-amp plates and grown overnight. Single colonies were inoculated into LB-amp medium and the cultures grown to an OD₆₀₀ of ~1.0. The plasmid was then purified using a QIAprep Spin Miniprep kit according to the manufacturer's instructions and transformed into *E. coli* BL-21(DE3) cells. Glycerol stocks were established as described above. 10 μL of cells from this glycerol stock were then inoculated into selective LB-amp medium for overnight shaking cultures at 37 °C.

The next day, stationary phase cell cultures were diluted 125 fold into two 1 L Erlenmeyer flasks containing LB-amp and grown under the same conditions as above until an OD₆₀₀ of ~0.6-1 was obtained. Protein expression was induced by the addition of 0.5 mM IPTG with shaking for 3–4 h before cell harvesting by centrifugation at 10,000 rpm for 15 min at 4 °C. The cell pellets were frozen at −20 °C for subsequent use. Cell pellets were resuspended in lysis buffer (20 mM Tris-HCl, pH 7.4), lysed by sonication (40% amplitude, 1 s pulse, 7.5 min sonication) and clarified by centrifugation at 12,000 rpm for 10 min at 4 °C. Small aliquots (10 μl) of the supernatant and urea-solubilized pellets were loaded into SDS-PAGE gels to verify protein expression and confirm that the protein was predominantly present in the soluble supernatant fraction. The supernatant was dialyzed against 5% acetic acid, and the dialysate was freeze-dried, resolubilized in 5% acetic acid and centrifuged at 12,000 rpm. The resulting supernatant contained recombinant DgHBP-1 at relatively high purity as indicated by SDS-PAGE and Matrix Assisted Laser Desorption Time of Flight mass spectrometry (MALDI-TOF) (**Supplementary Fig. 3a,b**). MALDI-TOF spectra were obtained by mixing

1 μl of a 1 mg/ml preparation of recombinant DgHBP-1 in 5% acetic acid with 1 μl sinapinic acid dissolved in a mixture of 50:50 Q-water:ACN with 0.1% TFA. The sample was applied to the MALDI plate and allowed to dry for 30 min before irradiation. Experiments were conducted on a Kratos Axima TOF2 (Kratos-Shimadzu Biotech) equipped with a N₂ laser (337 nm, 4 ns pulse width). An accelerating voltage of 20 kV was used and spectra were recorded in linear mode by averaging at least 100 laser shots at a power of 120 system units.

Purified recombinant DgHBP-1 was then subjected to solution phase Circular Dichroism (CD) for structural analysis. CD was conducted with a Chriascan spectropolarimeter (Model 420, AVIV Biomedical Inc.). 1 mg/ml of recombinant DgHBP-1 in 5% acetic acid was used for the measurements, which were conducted in triplicate at wavelengths ranging from 190 to 260 nm, with a 1 nm step size and 1 nm bandwidth. Spectra were smoothed by the Savitzky-Golay method with a 2nd order polynomial.

Coacervation of DgHBP-1 and characterization. 10 mg of DgHBP-1 powder was weighed and dissolved in 1 ml of 10 mM acetic acid solution. 100 μl of the solution was then pipetted gently into 900 μl of 0.06 M phosphate buffer (Na₂HPO₄, ionic strength = 0.1 M, pH = 6.5). Upon contact with the buffer, a visible turbidity developed, and optical microscopy was conducted to observe the fluidic droplets, i.e., the coacervate phase. Two types of coacervate concentrations were assessed. In the first type, about 700 μl of the liquid buffer was gently removed from the glass vial, such that only the cloudy phase remained within the vial, defined herein as the "condensed coacervate". In the second type, the coacervate phase was vigorously stirred with the buffer, thus leading to a uniform dispersion of droplets in the vial, designated as "dispersed coacervate".

The secondary structure of DgHBP-1 coacervates and solution were characterized by CD (as mentioned above). The condensed and dispersed coacervates were characterized ~3–5 min after their formation at wavelengths ranging from 190 to 260 nm, with a 1 nm step size and 1 nm bandwidth. The spectra were smoothed by the Savitzky-Golay method with a polynomial order of 2. To detect whether there were any changes in the secondary structure of the coacervates, spectra of the dispersed coacervates were acquired at regular time intervals.

The rheological properties of DgHBP-1 coacervate and solution phases were recorded on Physica MCR501 from Anton Paar at 20 °C with a protecting cell to prevent solvent evaporation. The condensed coacervate was analyzed 5 min after formation. For shear thinning assessment, torsional experiments were conducted by a step-wise increase of the shear rate from 0.1 to 100 s⁻¹ using a parallel plate geometry (diameter = 10 mm) with an inter-plate gap less than 1 mm. Frequency sweep experiments in the dynamic mode were recorded using the same geometry in the viscoelastic regime at a strain of 0.5%. As a control, the rheological properties of the protein solution (8 mg/ml) were recorded under similar conditions using a cone-plate geometry (d = 60 mm) with an inter-plate gap of 56 μm. All rheological measurements were done in triplicate. The frequency sweep measurements were also obtained at regular time intervals, chosen to match the time intervals of the CD experiments.

For infiltration experiments, untanned sections of squid beaks were cut into ~0.5 cm × 0.5 cm slices. To remove residual proteins from these sections, slices were immersed overnight at 80 °C in a solution of 92.5% v/v H₂O, 5% v/v of 30% wt/wt H₂O₂ (Sigma-Aldrich), and 2.5% v/v NaOH. The per-oxidation treatment was repeated until full depigmentation was achieved and the sections (hereafter designated as "scaffolds") were then freeze-dried. Dried scaffolds were immersed in the dispersed or condensed coacervate phases overnight. As a control, sections were also immersed within a solution of DgHBp-1 in 10 mM acetic acid, which was diluted tenfold in order to reach the same total protein concentration as in the dispersed coacervate. Infused scaffolds were then rinsed with distilled water to remove adsorbed proteins from the surfaces, dried and ground in liquid nitrogen with a mortar and pestle. The particles were then dissolved with 50 μl of a 5% acetic acid in 8 M urea buffer overnight. The solutions were centrifuged at 14,000 rpm for 3 min, and 10 μl of the supernatant was analyzed by SDS-PAGE. In parallel, scaffold sections were incubated with DgHBP-1 coacervate phases or solution for 1 or 4 h and analyzed by amino acid analysis (AAA) using a Hitachi L8900 AAA system based on ninhydrin-based chemistry. Samples were hydrolyzed *in vacuo* overnight in 100 μL of 6 M HCl with 8% phenol at 110 °C. After being washed with water and methanol, the hydrolyzed products were dissolved in

0.02 M HCl and injected into the AAA. The results were expressed in terms of nmol of (Ala + Gly) normalized to the dried weights of the scaffolds.

DgHBP-*pep* self-coacervation and turbidity measurements. A peptide (DgHBP-*pep*) derived from the repeating domains of the *D. gigas* His-rich beak protein DgHBP-2 was custom ordered in desalted form from Genscript USA (Piscataway, NJ). DgHBP-*pep* was verified to be pure by reverse-phase C18 HPLC and amino acid analysis. A stock solution of DgHBP-*pep* dissolved in MilliQ water (10 mg/ml) was added to buffers at a volume ratio of 1:9 (stock:buffer) for a final concentration of 1 mg/ml. Coacervation of the peptide under different buffer conditions was assessed through turbidity measurements at 600 nm by UV-vis spectrophotometry on a M220Pro plate reader (Tecan Group Ltd). DgHBP-*pep* absorbance was negligible at 600 nm. The relative turbidity is defined as $\ln(T/T_0)$ where T and T_0 are light transmittance with and without sample, respectively[57]. Higher turbidity indicated an increased degree of coacervation due to increased amounts and/or larger sizes of spherical coacervate droplets in the solution (**Fig. 4b**). Precipitation could also result in higher turbidity values, but this outcome can be readily differentiated from coacervation by the visible settling of the precipitate particulates and also by their optical micrographs. Coacervates appear as round droplets when viewed with an optical microscope whereas the precipitates are irregularly shaped particles[24]. The phase changes during the coacervation process at 1 M salt in a pH 8 buffer were also tracked over 24 h with photographs from a Nikon D3100 equipped with a AF MICRO NIKKOR 60 mm lens.

Optical microscopy. The turbidity associated with coacervate droplet formation was visually inspected by light microscopy in the reflection mode, using a Zeiss "Axio Scope A1" microscope (Carl Zeiss Pte Ltd, Oberkochen, Germany) and differential interference contrast (DIC) filters. Protein distribution was also investigated by fluorescence microscopy (Olympus DSU model IX81, Olympus, Tokyo, Japan). Images were taken with an ImageEM camera (C9100-13, Hamamatsu, Shizuoka, Japan) under the control of MetaMorph software (Olympus).

Crosslinking of DgHBP-*pep* with 4-methylcatechol (4MC). Approximately 200 μl (2 mg, ~0.8 μmole) of 10 mg/ml peptide stock solution was diluted in 780 μl of 0.5 M buffers (pH 6 and 8) individually. These were then mixed with 10 μl of 80 μmole/ml (equimolar) 4MC (Sigma-Aldrich, ≥95%) solution and 10 μl of 160 μmole/ml (2× molar quantity) sodium periodate (NaIO$_4$, Acros Organics ≥99.8%) solution for a final volume of 1 ml. Twice the molar quantity NaIO$_4$ was used so that each 4MC could be oxidized twice for crosslinking to two different His residues, thus creating a network of insoluble peptides. Two negative controls were used for each buffer condition: one without the peptide (4MC+NaIO$_4$ only), and one without the peptide and 4MC (NaIO$_4$ only). The mixtures were incubated overnight at room temperature on a vertical rotator at a speed of 40 rpm. The resulting mixtures were then separated by microcentrifugation for 30 min at 14,000 rpm. The supernatants from the end products were transferred to new microcentrifuge tubes and the pellets washed 3× with 1 ml of Milli-Q water. The pellets were then resuspended in 50:50 ACN:H$_2$O with 0.1% TFA and transferred into a 1 ml hydrolysis glass vial for freeze-drying to obtain foam-like structures (**Supplementary Fig. 8d**). The crosslinked products were visualized on a Tricine SDS-PAGE gel as described above. Briefly, supernatant and pellet fractions obtained after lyophilization were resuspended in 10 μl aliquots. These aliquots, together with an aliquot of DgHBP-*pep* (10 μg), were then separated electrophoretically in a Tricine gel.

Next, aliquots of the lyophilized crosslinked products (pellets) were hydrolyzed at 155 °C with 6 N HCl (Sigma-Aldrich, 19.7–20.7 w/w %) (phenol was added to protect the catechols from oxidation) and washed as previously reported by Miserez *et al.*[14] to obtain and characterize the crosslinked adducts. Hydrolyzed and washed samples were then analyzed by electrospray ionization (ESI) MS/MS using a Micromass QTOF2 quadrupole/time-of-flight mass spectrometer (Waters Corp., Milford, MA). Samples were injected into the electrospray source through a fused silica tubing attached to a syringe pump that continuously injected the sample at a rate of 5 μl/min. Typical settings were a capillary tension of 3.5 kV, a cone tension of 45 V, and a collision voltage of 10 V for the MS mode. *m/z* values of the adducts found in the squid beak[14] were used to search for the corresponding adducts from the *in vitro* crosslinking experiment. Once crosslinks of interest were detected, they were subsequently fragmented at collision voltages of 20 or 35 V to obtain large or small molecular weight fragments, respectively. Fragmentation patterns were processed with the MassLynx software package and later manually analyzed for final crosslink determination.

46. Schägger, H. Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
47. Paz, M.A., Fluckiger, R., Boak, A., Kagan, H.M. & Gallop, P.M. Specific detection of quinoproteins by redox-cycling staining. *J. Biol. Chem.* **266**, 689–692 (1991).
48. Stajich, J.E. *et al.* The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**, 1611–1618 (2002).
49. Shevchenko, A., Tomas, H., Havliš, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2006).
50. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
51. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. & Barton, G.J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
52. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
53. Newman, A.M. & Cooper, J.B. XSTREAM: A practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinformatics* **8**, 382 (2007).
54. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
55. Rebers, J.E. & Willis, J.H. A conserved domain in arthropod cuticular proteins binds chitin. *Insect Biochem. Mol. Biol.* **31**, 1083–1093 (2001).
56. Goodarzi, M.T., Fotinopoulou, A. & Turner, G.A.A. in *The Protein Protocols Handbook Springer Protocols Handbooks* (ed. Walker, J.M.) 1205–1214 (Humana Press, 2009).
57. Hwang, D.S., Waite, J.H. & Tirrell, M. Promotion of osteoblast proliferation on complex coacervation-based hyaluronic acid—recombinant mussel adhesive protein coatings on titanium. *Biomaterials* **31**, 1080–1084 (2010).