Biomimetic studies on *Dosidicus gigas* Histidine-rich beak proteins

Cai, Hao

2018

Cai, H. (2018). Biomimetic studies on *Dosidicus gigas* Histidine-rich beak proteins. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73239>

<https://doi.org/10.32657/10356/73239>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**BIOMIMETIC STUDIES ON DOSIDICUS GIGAS HISTIDINE-
RICH BEAK PROTEINS**

CAI HAO

SCHOOL OF MATERIALS SCIENCE AND ENGINEERING

2017

BIOMIMETIC STUDIES ON DOSIDICUS GIGAS HISTIDINE- RICH BEAK PROTEINS

CAI HAO

SCHOOL OF MATERIALS SCIENCE AND ENGINEERING

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2017

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....

Date

蔡昊 Cai Hao

.....

Cai Hao

Abstract

Over 4 billion years of evolution on earth, Nature has presented countless examples of adaptations towards contextual restraints, which have inspired humankind facing practical engineering problems throughout the entire history of civilization. One intriguing example is *Dosidicus gigas* (*D. gigas*) or Humboldt squid, a predatory species of squids often found in the Eastern Pacific Oceans. Ranking among the most formidable hunters in Ocean ecosystems, *D. gigas* relies largely upon its hunting biotools, one of which being *D. gigas* beak, an intriguing biological model structure for biomimetic studies. As sharp and tenacious as it is, *D. gigas* beak is non-mineralized and highly proteinaceous, demonstrating a 200-fold stiffness gradient evolving from soft buccal tissues to the rigid tip. Its constituting proteins, namely *D. gigas* beak proteins (DgBPs), are believed to play central roles in the formation and sclerotization of the beak material. Among the various members of DgBPs revealed by recent sequencing work, the DgHBP (*D. gigas* histidine-rich beak protein) family has attracted most interest due to its unique molecular composition and architecture. This thesis therefore focuses on biomimetic studies of DgHBPs, employing a bottom-up approach to gain insights into their functions *in vivo*.

Three parts are included in the thesis. The first part describes the establishment of an optimized production and purification system for DgHBPs. Recombinant expression using *E. coli* cells carrying engineered gene vectors, followed by purification steps that were developed based on primary solubility studies on DgHBPs, enabled sufficient proteins of high purity to be prepared for subsequently studies. Basic structural characterization of DgHBPs in lyophilized phase and aqueous solution phase are also presented in this part, indicating an important phase-transition feature, which was further explored in the second part.

The second part of this thesis highlights the biophysical properties of DgHBPs, in particular their coacervation behavior, which is a liquid-liquid phase separation of a macromolecular solution process under specific conditions, also found in other

biological systems such as tropoelastin. Thorough investigation into different aspects (morphology, kinetics, affecting parameters, mechanical properties, thermodynamics, etc.) of the coacervation behavior helped clarifying the roles of DgHBPs during beak formation, and also pointed out the need to study their sequence-coacervation ability relationship.

The third part of this thesis examines the sequence-coacervation ability relationship by constructing and studying DgHBPs-based peptides and variants. Different parts from the full sequences of DgHBPs were isolated and tested for their ability to form coacervates, and previous hypotheses about which sequence domains has led to this ability was verified.

This thesis highlights important implications of DgHBPs' phase behavior (coacervation) on *D. gigas* beak bio-processing. By conducting *in vitro* studies of DgHBPs, their coacervation ability was examined comprehensively, from the role of the peptide domain aspects to structural and mechanical aspects. A deeper understanding about *D. gigas* beak formation was obtained, providing inspirations for the design and manufacturing of biomimetic composite materials with tailored mechanical properties, which have relevant potential in medical and engineering applications.

Acknowledgements

The author would like to express his sincere appreciation and gratitude towards his thesis supervisor, A/P Ali Miserez for his guidance and support throughout the project, as well as the offered opportunity to join this group in the first place. His suggestions and questions raised in each group talk and lab meeting has also deeply influenced the author in terms of scientific taste, professional habit, logical thought and communication skills.

The author would also like to extend the appreciation to the collaborators who offered help on instrument using, including Prof. Gerhard Grüber (School of Biological Science, Nanyang Technological University) for small angle X-ray scattering, and Assoc Prof Konstantin Pervushin (School of Biological Science, Nanyang Technological University) for *E. coli* recombinant expression.

Sincere thanks also goes to several predecessors in this research. A/P Ali Miserez and Dr. Daniel Rubin conducted initial compositional and chemical analysis on *D. gigas* beak with outstanding achievements. Mr. Abbas Bacha and Ms. Ye Kehan passed valuable experience to the author regarding DgHBPs production. Dr. Yerpeng Tan, Dr. Shawn Hoon, and Dr. Paul Guerette has laid the ground for this thesis with their marvelous work on DgHBPs identification and sequential analysis. Dr. Paul Guerette also provided his most insightful suggestions to the author during the stage of purifying and characterizing DgHBPs. His brilliant experience, as well as incomparable passion in the area of biomimicry has profoundly affected the author. Dr. Ali Ghadban has dedicated his expertise in polymer chemistry to this project without reservation. He has also set an example for the author on proper record keeping. The author benefited greatly from his harsh training and generous aid. Dr. Bartosz Gabryelczyk has provided enormous help on the preparation and characterization of DgHBPs and DgHBPs-based variants. The author appreciate his help on all matters, including the revision of his manuscripts. Special thanks also goes to Dr. Luigi Petrone and Dr. Ping Yuan for their kind help on the author's qualifying examination.

To all the colleagues and friends in Biological and Biomimetic Materials Laboratory (BBML) and Centre for Biomimetic Sensor Science (CBSS), the author would also like express his gratitude for their amazing friendship and camaraderie in the past four years, which the author shall always cherish in his life.

The author also wishes to deliver his appreciation to Nanyang Technological University and school of Material Science and Engineering for providing Research Scholarship and other financial support for this project and the author.

Last but not the least, the author offers his deepest gratitude to his family and beloved ones, who has always been an indispensable part of his life, and offered unconditional support and understanding to him.

Table of Contents

Abstract.....	i
Acknowledgements	iii
Table of Contents	v
Table Captions	xiii
Figure Captions.....	xv
Abbreviations	xxvii
1 Chapter 1 Introduction.....	1
1.1 Background and Hypothesis.....	2
1.1.1 Biomimicry, Biological Materials and Protein-based Materials	2
1.1.2 <i>D. gigas</i> and <i>D. gigas</i> Beak (Proteins)	3
1.1.3 Hypotheses	3
1.2 Objectives and Scope	4
1.3 Dissertation Overview.....	5
1.4 Findings and Outcomes/Originality	6
References.....	7
2 Chapter 2 Literature Review	9
2.1 Biomimetic Materials.....	10
2.2 <i>D. gigas</i> and <i>D. gigas</i> Sucker Ring Teeth.....	12
2.2.1 <i>D. gigas</i>	12

2.2.2	<i>D. gigas</i> Sucker Ring Teeth.....	13
2.3	<i>D. gigas</i> Beak: Composition, Structure and Properties.....	15
2.3.1	<i>D. gigas</i> Beak: Composition.....	15
2.3.2	<i>D. gigas</i> Beak: Structure.....	18
2.3.3	<i>D. gigas</i> Beak: Mechanical Properties	20
2.3.4	<i>D. gigas</i> Beak: Composition-Mechanical Property Relationships	22
2.3.5	<i>D. gigas</i> Beak Inspired Materials	23
2.4	<i>D. gigas</i> Beak Proteins (DgBPs).....	25
2.4.1	DgBPs Extraction and Identification.....	27
2.4.2	DgCBPs: Sequence Features, Homology and Conservation	29
2.4.3	DgHBPs: Sequence Features, Homology and Conservation.....	29
2.5	Coacervation & Self-coacervating Proteins	35
2.5.1	Coacervation and Complex Coacervation	35
2.5.2	Self-coacervation: Mussel Foot Protein (Mfp-3S)	38
2.5.3	Self-coacervation: Tropoelastin.....	42
2.5.4	Sequential Homology between DgHBPs and Self-coacervating Proteins..	46
2.6	Scope of Project	47
	References.....	48
3	Chapter 3 Experimental Methodology.....	55
3.1	Rationale for Selection	56
3.2	Cloning and Expression of Recombinant DgHBPs.....	56
3.2.1	Plasmid Preparation.....	57
3.2.2	Protein Expression.....	57
3.3	Purification of Recombinant DgHBPs	58
3.3.1	Lyses of Cells by Microfluidization	58

3.3.2	Primary Protein Extraction by Acetic Acid.....	59
3.3.3	Size Exclusion Chromatography	60
3.3.4	Reverse Phase Chromatography	61
3.4	Self-coacervation of DgHBPs	62
3.5	Characterization of DgHBPs at Solid and Solution State	63
3.5.1	SDS-PAGE	63
3.5.2	MALDI-ToF	64
3.5.3	Fourier Transform Infrared Spectroscopy	65
3.5.4	Circular Dichroism Spectroscopy.....	66
3.6	Characterization of DgHBPs at Coacervates State	68
3.6.1	Optical Microscopy	69
3.6.2	Turbidity Measurement	69
3.6.3	Dynamic Light Scattering.....	70
3.6.4	Isothermal Titration Calorimetry	71
3.6.5	Small Angle X-ray Scattering.....	72
3.6.6	Rheology.....	75
3.6.7	Contact Angle and Surface Tension Measurement	78
3.6.8	Amino Acid Composition Analysis (AAA)	80
3.7	Preparation and Characterization of DgHBPs-based Peptides and Variants	81
3.7.1	Preparation of DgHBPs-based Peptides	81
3.7.2	Recombinant Expression of DgHBPs-based Variants	81
3.7.3	Trypsin Cleavage Experiment of DgHBPs-based Variants.....	82
3.7.4	Western Blot.....	83
3.7.5	Characterization for DgHBPs-based Peptides and Variants	84
	References.....	84

4	Chapter 4 Recombinant Expression, Purification and Characterization of DgHBPs.....	89
4.1	Introduction	90
4.2	Experimental Methods	91
4.2.1	Recombinant Expression of DgHBP-1 and DgHBP-2	91
4.2.2	Purification Strategies for (<i>rec</i>)DgHBP-1 and (<i>rec</i>)DgHBP-2.....	91
4.2.3	Further Sequence Analysis	91
4.2.4	SDS-PAGE	92
4.2.5	MALDI-ToF	92
4.2.6	Fourier Transform Infrared Spectroscopy	92
4.2.7	Circular Dichroism Spectroscopy.....	92
4.3	Results and Discussion.....	92
4.3.1	Recombinant Expression of DgHBP-1 and DgHBP-2	92
4.3.2	Solubility Test and Purification of (<i>rec</i>)DgHBP-1 and (<i>rec</i>)DgHBP-2	95
4.3.3	Primary Sequence Analysis	98
4.3.4	Structural Characterization: Solid State FTIR.....	98
4.3.5	Structural Characterization: Solution State CD.....	100
4.4	Conclusions	103
	References.....	104
5	Chapter 5 Self-coacervation of DgHBP-1 and DgHBP-2	107
5.1	Introduction	108
5.2	Experimental Methods	108
5.2.1	Self-coacervation of DgHBPs	109
5.2.2	Optical Microscopy	109
5.2.3	Turbidity Measurements.....	109

5.2.4	Dynamic Light Scattering.....	109
5.2.5	Isothermal Titration Calorimetry.....	109
5.2.6	Circular Dichroism Spectroscopy.....	110
5.2.7	Small Angle X-ray Scattering.....	110
5.2.8	Rheology.....	110
5.2.9	Contact Angle/Surface Tension Measurement.....	111
5.2.10	Chitin Scaffold Infiltration Experiment of DgHBP-1 coacervates.....	111
5.3	Results and Discussion.....	111
5.3.1	Morphology of DgHBP-1 and DgHBP-2 Coacervates.....	111
5.3.2	Kinetics of DgHBP-1 and DgHBP-2 self-coacervation: Size	115
5.3.3	Kinetics of DgHBP-1 and DgHBP-2 Self-coacervation: Structure.....	116
5.3.4	Parameters affecting DgHBP-1 and DgHBP-2 Self-coacervation: Temperature	120
5.3.5	Parameters affecting DgHBP-1 and DgHBP-2 Self-coacervation: pH and Ionic Strength	121
5.3.6	Mechanical Properties of DgHBP-1 and DgHBP-2 Coacervates: Viscosity and G'/G''	124
5.3.7	Thermodynamics of DgHBP-1 and DgHBP-2 self-coacervation.....	127
5.3.8	Wettability of DgHBP-1 and DgHBP-2 Coacervates: Contact Angle and Surface Tension.....	129
5.3.9	Infiltration Ability of DgHBP-1 Coacervates.....	130
5.4	Conclusions	132
	References.....	134
6	Chapter 6 DgHBPs-based Peptides and Variants	137
6.1	Introduction	138

6.2	Experimental Methods	140
6.2.1	Preparation of Short Peptides	141
6.2.2	<i>E. coli</i> Recombinant Expression.....	141
6.2.3	Trypsin Cleavage of DgHBP-1-V11 to DgHBP-1-V15 and DgHBP-2 ...	142
6.2.4	Reverse Phase Chromatography	143
6.2.5	SDS-PAGE and Western Blot	143
6.2.6	MALDI-ToF	143
6.2.7	Optical Microscopy	143
6.2.8	Turbidity Measurements.....	144
6.2.9	Dynamic Light Scattering.....	144
6.2.10	Circular Dichroism Spectroscopy.....	144
6.2.11	Rheology.....	144
6.3	Results and Discussion.....	145
6.3.1	Expression and Purification of DgHBP-1-V3-C and DgHBP-1-V4-C	145
6.3.2	Expression, Cleavage and Purification of DgHBP-1-V11-C/N to DgHBP-1-V15-C/N	147
6.3.3	Cleavage and Purification of DgHBP-2-V1-C/N	150
6.3.4	Self-coacervation Abilities of DgHBPs-based Peptides and Variants	152
6.3.5	Morphology of Coacervates by DgHBPs-based Peptides and Variants...	155
6.3.6	Rheological Properties of Coacervates by DgHBPs-based Peptides and Variants	155
6.3.7	Structure of DgHBPs-based 23/26 Peptides.....	157
6.3.8	Variables Affecting Coacervation of GW-26 and DgHBP-1-V3-C: pH and Ionic Strength	158
6.4	Conclusions	160
	References	160

7	Chapter 7 Discussion and Future Work	163
7.1	General Discussion and Conclusion.....	164
7.1.1	Preparation of DgHBPs and DgHBPs-based Variants	164
7.1.2	Characterization of DgHBPs and Self-coacervation	165
7.1.3	Sequence Studies on DgHBPs Self-coacervation.....	166
7.1.4	Proposed Model for <i>D. gigas</i> beak formation	166
7.2	Reconnaissance Work for Future Studies	168
7.2.1	Mixed Coacervation by DgHBP-1 and DgHBP-2.....	168
7.2.2	His-catechol Cross-linking of DgHBP-1 Coacervates	169
7.2.3	Chitin Network Preparation: Histidine Functionalization of Chitosan	170
7.3	Outlook and Future Work	170
7.3.1	Further Studies on DgHBPs	170
7.3.2	Studies on DgCBPs	172
7.3.3	Studies on Chitin Network	173
7.3.4	<i>D. gigas</i> Beak Bio-inspired Materials	173
	References.....	174

Table Captions

Table 2.1 cDNA deduced sequences of DgCBPs with predicted molecular weight (Mw) and isoelectric point (pI). Underlined: chitin binding domains. Bold Italic: signal peptide sequences. Blue: MS/MS deduced sequences. Red: Edman sequence. Bold purple: potential NH₂OH cleavage sites.

Table 2.2 cDNA deduced sequences of DgHBPs with predicted molecular weight (Mw) and isoelectric point (pI). Underlined: tandem repeated domains (first amino in each repeat marked as bold). Bold Italic: signal peptide sequences. Blue: MS/MS deduced sequences. Bold purple: potential NH₂OH cleavage sites.

Table 2.3 cDNA deduced sequences of DgHBP-2 variants with predicted molecular weight (Mw) and isoelectric point (pI). Bold Italic: signal peptide sequences.

Table 6.1 Sequences of DgHBPs-based peptides and/or their positions on full protein sequences. “A. A.” are short for “amino acids”.

Table 6.2 Coacervation test results summary of DgHBPs-based peptides and variants. Red highlighted: no coacervation. Green highlighted: coacervation. At least 2 of the motifs “**GHGXY**” must be present to induce coacervation.

Figure Captions

Figure 2.1 Illustrations of biomimetic materials engineering. (a) *Nelumbo nucifera* lotus leaf. (b) SEM image of the lotus leaf surface micro-papillae-structures. (c) SEM images of superhydrophobic polystyrene film with structures inspired by lotus leaf surface. Inset is a water droplet on the film. (d) Riblet-structures of fast-swimming shark skin, at different body regions. Scale bar: 0.5 mm. (e) Speedo *Fastskin*® swimsuit, adapted from <http://www.popsci.com/technology/article/2012-07/speedos-super-fast-sharkskin-inspired-swimsuit-actually-nothing-sharks-skin>. (f) Flow chart of biomimetic materials research process. Arrow 1, 2 and 3 indicates the priorities of different steps.

Figure 2.2 Drawing of *D. gigas* and its “hard tissues” appendages. By French naturalist, Alcide D’Orbigny.

Figure 2.3 Structural and mechanical properties and primary structure features of *D. gigas* SRT. (a) A representative *D. gigas* SRT. (b) SEM image of the SRT nanotubular structure. (c) Ashby plot of common engineering materials (polymers, metals, and ceramics) in comparison to dried native SRT. (d) Highly modular architecture of a representative *D. gigas* suckerin (suckerin-19) showing alternating Ala- and Tyr- rich domains.

Figure 2.4 Beak of *Dosidicus gigas*. (a) Beak on a captured *Dosidicus gigas* squid. Adapted from <http://www.fadedindustry.com/extremely-terrifying-facts-about-squids/>. (b) Schematic illustration of *D. gigas* beak on buccal mass. (c) Upper and lower beak, after removal from the buccal mass.

Figure 2.5 Compositional gradient in *D. gigas* beak. (a) Side view of a split upper beak highlighting the transition from the wing to the rostrum. (b) Series of specimens showing different pigmentation levels. (c) Chitin, protein, pigment and water contents from chemical and gravimetric analysis results.

Figure 2.6 *D. gigas* beak structures. (a) Computer rendering of the upper (blue) and lower (red) beaks of *Octopus bimaculoides* (cephalopod) in a front right quarter view. Left: the upper beak fits within the lower beak; Right: the upper beak separates from the lower beak. (b) Side view diagram showing movement of the upper (blue) beak to anchored lower beak (red) during a stereotypical bite cycle. Position **i**: resting; **ii**: opening; **iii**: fully open; **iv**: closing; **v**: closed and retracted. (c) Cross-section SEM images of the beak showing the lamellae alignment, featuring an external surface layer. (d) Lamellar microstructure SEM micrographs of fractured beak surfaces. Inset shows an increased magnification of the rectangle area. (e) SEM micrographs of Chitin leaflet microstructure in the beak after alkaline peroxide treatment. (f) SEM micrographs of intertwined chitin nanofibers.

Figure 2.7 Mechanical properties of *D. gigas* beak. (a) Ashby plot (H against E) of *D. gigas* beak tip part with typical engineering materials. (b) $H^3/E^2 - G_c$ map of different materials yielding/cracking showing abrasion resistance properties. Critical loads for yielding and cracking are represented by horizontal and vertical lines, respectively. Families of L-shaped curves represent different failure loads. (c) Young's modulus and tensile strength gradients in *D. gigas* beak at different regions, expressed as modulus against (protein and cross-link content). (d) Typical tensile stress/strain curves for hydrated beak specimens.

Figure 2.8 Catechol-functionalized chitosan/ γ -Fe₂O₃ composite material inspired by *D. gigas* beak. (a) Synthetic strategy of the composites. (b) Young's modulus of dehydrated and hydrated composites (with and without MNPs). (c) Magnetization curves for mnp-CAT_{CS} films. Inset shows a typical hysteresis curve of control super-paramagnetic material. (CS: chitosan; CAT: catecholic moieties; DOPA: 3,4-dihydroxyphenylalanine; DHA: 3,4-dihydroxyphenethylamine. MNP: magnetic nano-particle.)

Figure 2.9 Allyl-CNCs-PVAc nanocomposite material inspired by *D. gigas* beak. (a) Schematic representation of mechanical gradient model in *D. gigas* beak and in the Allyl-CNCs-PVAc film. (b) Storage modulus gradient of the CNCs-PVAc film at different UV exposure time. Red rectangles on the top part represents the different sample locations that

undergoes the photo-initiated cross-linking; bottom part shows the relative modulus at different locations on the film, under wet and dry conditions, respectively. (c) Twisting of a nanocomposite film that underwent gradient UV exposure (top picture, gradient as shown in panel (b) and one that underwent uniform UV exposure (bottom picture).

Figure 2.10 *D. gigas* beak proteins identification. (a) Tricine-SDS-PAGE of extracted DgBPs stained/binded with CBR, Pauly stain, NBT stain and Lectin-binding assay. CBR: to show all proteins; Pauly stain: to show His-rich proteins; NBT stain: to show catecholic proteins; Lectin-binding assay: to show *N*-acetylglucosamine oligomers in the beak. L: Ladder. Loaded protein amounts per lane for each assay were 10 µg, 30 µg, 30 µg and 60 µg, respectively. (b) Schematic sequential representations of DgCBPs and DgHBPs. Modular domains in each protein are indicated. X: hydrophobic residue; X': tyrosine (mostly); X'': glycine or alanine.

Figure 2.11 Test and analysis of DgCBPs. (a) Tricine-SDS-PAGE of NH₂OH-extracted proteins eluting from chitin beads and separated as non-binding fraction (“flow-through” and “wash”) and binding fraction (“elute”). DgCBP-1, DgCBP-2 and DgCBP-3 are labeled in the elute fraction. (b) Aligned consensus residue comparison (10 % ClustalO threshold) between chitin-binding domains of DgCBPs and peritrophin-44, a known chitin-binding protein.

Figure 2.12 Sequence analysis of DgHBPs. (a) Aligned consensus residue comparison (30 % ClustalO threshold) among DgHBPs and their highly conserved consensus logo. (b) Tandem repeats of C-terminus domain sequences of DgHBPs. Underlined: fully conserved amino acids in the consensus sequences. Yellow highlighted: proposed cross-linking motifs GHG and HG.

Figure 2.13 Complex coacervation of protein-polysaccharide. (a) Microscopic image of gelatin-gum arabic coacervates. (b) Cartoon of gum arabic (white ribbon)- β -lactoglobulin (orange spheres) complex. The approximate diameter of proteins is 4 nm, whereas gum arabic has an average diameter of 50 nm. (c) Summary of physicochemical

factors influencing complex coacervation. Scale bar: 50 μm . (d) Encapsulation of krill oil in gelatin-gum arabic coacervates, monitored by stereomicroscopy.

Figure 2.14 Complex coacervation by cement proteins of *P. californica*. (a) Micrograph of *P. californica* tube with globular sand beads (0.5 mm diameter). Right panel: zoom of cement plaques between two beads, as indicated by the arrow. (b) Sequences of Pc-1, Pc-2 and two variants of Pc-3. Italic: signal peptide. Underlined: repeated domains. (c) Schematic representation of the coacervation and cross-linking of *P. californica* cement proteins.

Figure 2.15 mfp-3S from *M. californianus* byssal plaques. (a) *M. californianus* mussel. Inset: Close view of adhesive protein residues secreted on the plaque-substrate interface and within the plaque bulk. (b) Representative sequence of mfp-3S with predicted molecular weight and isoelectric point. Red: positively charged residues. Blue: negatively charged residues. Green: aromatic residues. (c) Schematic representation of the Hydrogen bonds formed between Dopa and mica surface. (d) Hydropathy plot (Hopp and Woods) of mfp-3S as compared to another mussel foot protein, mfp-3f, highlighting the hydrophobicity for mfp-3S.

Figure 2.16 Self-coacervation of mfp-3S from *M. californianus* byssal plaques. (a) Top panel: mfp-3S in coacervates state (right) compared with solution state (left). Bottom panel: light microscopic image of mfp-3S coacervates. (b) Coacervates-associated turbidity results of 0.1 mg/ml mfp-3S, indicating dependence on pH and ionic strength. Buffer 1: phosphate buffer. Buffer 2: acetate buffer. Buffer 3: phosphate buffer. (c) mfp-3S (0.1 mg/ml) coacervates turbidity change with temperature. (d) Schematic comparison between complex coacervates (gelatin-GA) and mfp-3S coacervates.

Figure 2.17 Elastin and tropoelastin. (a) Illustration of elastin fibers in aorta (left), lung (center) and ear cartilage (right). (b) Domain structure of human tropoelastin. Orange regions: hydrophobic domains. Light blue regions: cross-linking domains rich in lysine. (c)

SEM graph of tropoelastin molecules (~ 15 nm) in the form of nanoparticle assemblies prior to coacervation. **(d)** SEM image of full-sized tropoelastin coacervates (~ 2 to $6\ \mu\text{m}$).

Figure 2.18 Influencing factors and mechanism of tropoelastin self-coacervation. **(a)** Turbidity-time plot of tropoelastin solution (40 mg/ml, 10 mM sodium phosphate, pH 7.4, 150 mM NaCl) at different temperatures. **(b)** Transition temperature-tropoelastin concentration plot, using temperatures required to achieve 50 % maximal coacervation at different pH values as an indicator of coacervation propensity. Same in (c) and (d). **(c)** Transition temperature-salt concentration plot. **(d)** Transition temperature-pH plot. **(e)** Schematic representation of tropoelastin reacting to increased temperature, showing the disruption of H-bonds between clathrate water molecules that shields the hydrophobic domains of tropoelastin due to heating, leaving the domains exposed.

Figure 3.1 Schematic representation of recombinant expression and subsequent purification procedures for DgHBP-1 (red) and DgHBP-2 (blue).

Figure 3.2 CD spectra of polypeptides and proteins with some representative secondary structures.

Figure 3.3 SAXS solution scattering curves containing different information at different size ranges.

Figure 3.4 Schematic representation of geometric parameters in a pendant drop measurement.

Figure 4.1 Recombinant expression of DgHBP-1 and DgHBP-2. **(a)** and **(b)** Vector maps of DgHBP-1 and DgHBP-2 gene-encoding plasmids (red fragment are the DgHBPs-encoding genes), respectively. **(c)** and **(d)** DNA gel electrophoresis of NdeI and SmaI digestion products of DgHBP-1 and DgHBP-2, respectively. Circled bands are the cleaved gene inserts. **(e)** and **(f)** SDS-PAGE of DgHBP-1 and DgHBP-2 expression after cell lysis. L: ladder.

Figure 4.2 Solubility test and acetic acid extraction of expressed DgHBP-1 and DgHBP-2. (a) and (b) SDS-PAGE of DgHBP-1 and DgHBP-2, respectively, dissolved in various buffers then centrifuged. L: ladder. T: total fraction (before centrifugation). S: supernatants after centrifugation. (c) and (d) SDS-PAGE of acetic acid extraction of the proteins after cell lysis.

Figure 4.3 Chromatographical purification of DgHBP-1 and DgHBP-2. (a) and (b) RPC spectra of DgHBP-1 and DgHBP-2, respectively. (c) and (d) SDS-PAGE of RPC purification fractions. (e) and (f) MALDI-ToF spectra of RPC-purified proteins.

Figure 4.4 Sequential information and hydropathy plots of DgHBP-1 and DgHBP-2. (a) and (b) Sequences of DgHBP-1 and DgHBP-2, respectively. Red: positively charged residues. Blue: negatively charged residues. Green: aromatic residues. (c) Molecular weight (Mw), amino acid number (A. A.), isoelectric point (pI), and acidic/basic/aromatic residues count ($R^{\text{acidic}}/R^{\text{basic}}/R^{\text{arom}}$). (d) Hydropathy plots of DgHBP-1 and DgHBP-2, using hydropathy values from Hopp T. P. et al.

Figure 4.5 FTIR of lyophilized, solid-state (*rec*)DgHBP-1 and (*rec*)DgHBP-2. (a) and (b) FTIR spectra. (c) and (d) Deconvolution of the Amide I region spectra.

Figure 4.6 Conformational transitions of DgHBP-1 and DgHBP-2. (a) and (b) CD spectra of DgHBP-1 and DgHBP-2 of different concentrations in 10 mM acetic acid, showing concentration-dependent conformation. (c) and (d) CD spectra of DgHBP-1 and DgHBP-2 at different temperatures, showing temperature-dependent conformation.

Figure 5.1 Macro-photos and optical microscopic images of DgHBP-1 and DgHBP-2 coacervates. (a) and (b) Self-coacervation process of DgHBP-1 and DgHBP-2, respectively. **i:** injection of stock protein solution into PBS buffer. **ii:** formation of a distinct liquid phase (coacervates phase) **iii:** dispersion of the prior coacervates phase within the buffer solution. **iv:** 20 min after dispersion of coacervates phase. Red circle in ii: condensed

coacervates phase (for DgHBP-1 only). White dashed rectangle in iii: dispersed coacervate phase (for DgHBP-1 and DgHBP-2) (c) and (d) Optical micrographs of coacervate phase before dispersion. (e) and (f) Optical micrographs of coacervates phase after dispersion.

Figure 5.2 Micrographs of DgHBP-1 and DgHBP-2 coacervates after different time after initial formation, illustrating the “merging” behavior of DgHBP-1 coacervates. Red dashed ellipse indicates ongoing merging of two micro-droplets into a larger one.

Figure 5.3 DLS size measurements of DgHBP-1 and DgHBP-2 coacervates at different time after dispersion. (a) DLS profiles for DgHBP-1 (left panels) and DgHBP-2 (right panels) coacervates showing the intensity distributions of micro-droplet diameters in the coacervation systems. (b) Effective particle diameters (Z-average mean values) of DgHBP-1 and DgHBP-2 coacervates vs. dispersion time, respectively. (c) and (d) Correlation functions for the calculated distribution profiles shown in (a).

Figure 5.4 CD spectra of coacervates phase after different time after intentional dispersion in DgHBP-1 and DgHBP-2 self-coacervation system, respectively. (a) and (b) Dispersed coacervates phase (open symbols) at different time compared to protein solution phase (dash lines) and condensed coacervates phase (close symbols, only for DgHBP-1).

Figure 5.5 SAXS for DgHBP-1 and DgHBP-2 dispersed coacervates at different dispersion time. (a) and (b) SAXS scattering curves for dispersed coacervates at different time as compared to solution phase and condensed coacervates phase (only for DgHBP-1) in DgHBP-1 and DgHBP-2 self-coacervation system, respectively. (c) and (d) Kratky plots calculating from scattering curves in (a) and (b), respectively, revealing the folding signature of coacervates.

Figure 5.6 Turbidity measurement for DgHBP-1 and DgHBP-2 coacervates at different temperatures. Inset: optical image of DgHBP-1 and DgHBP-2 coacervates at 80 °C.

Figure 5.7 Turbidity measurement for DgHBP-1 and DgHBP-2 dispersed coacervates, respectively. (a) and (b) Turbidity as a function of pH at 100, 500 and 1000 mM ionic strength for DgHBP-1 and DgHBP-2 dispersed coacervates, respectively.

Figure 5.8 SAXS for DgHBP-1 and DgHBP-2 dispersed coacervates at different dispersion time. (a) and (b) SAXS scattering curves for dispersed coacervates at different pH at controlled ionic strength (100 mM) for DgHBP-1 and DgHBP-2, respectively. (c) and (d) Kratky plots calculating from scattering curves in (a) and (b), respectively, revealing the folding signature of coacervates. (e) Power law fitting coefficient x values for DgHBP-1 and -2 at pH 4 and pH 10.

Figure 5.9 Rheological studies on DgHBP-1 and DgHBP-2. (a) Viscosity of DgHBP-1 and DgHBP-2 coacervates as a function of shear rate, suggesting that only DgHBP-1 coacervates have shear-thinning behavior. (b) Frequency sweep measurements of storage moduli (G' , closed symbols) and loss moduli (G'' , open symbols) of DgHBP-1 and DgHBP-2 coacervates, suggesting that only DgHBP-1 coacervates have gel properties. (c) Viscosity of *pep*DgHBP-1 and *pep*DgHBP-2 coacervates as a function of the shear rate. (d) Frequency sweep measurements of storage (G' ; closed symbols) and loss (G'' ; open symbols) moduli of *pep*DgHBP-1 (blue boxes) and *pep*DgHBP-2 (pink circles) coacervates.

Figure 5.10 ITC measurement of DgHBP-1 and DgHBP-2 self-coacervation. (a) and (b) Thermogram (top) and isotherm (bottom) of DgHBP-1 and DgHBP-2 self-coacervation, respectively. Inset: Thermodynamic parameters of self-coacervation. The experiments were performed at pH 6.5 (phosphate buffer, ionic strength 100 mM) and 25 °C.

Figure 5.11 Contact angle measurement for DgHBP-1 and DgHBP-2 coacervates. (a) Control buffer (10 mM acetic acid diluted into pH 6.5 phosphate buffer with 100 mM ionic strength at dilution ratio of 9). (b) and (c) DgHBP-1 and DgHBP-2 coacervates, respectively.

Figure 5.12 Surface tension measurement by pendant drop method for DgHBP-1 and DgHBP-2. (a) Surface tension diagram for phosphate buffer, solution and coacervates phases of DgHBP-1 and DgHBP-2, respectively. Insets are real-time captured images for each sample. (b) Summary chart of surface tension values for different samples. ($N = 3$, technical replicates, error bar represents mean values $\pm$ s.d.)

Figure 5.13 Infiltration experiment of DgHBP-1 coacervates. (a) and (b) Schematic representation of the experimental and characterization procedures, respectively. (c) SDS-PAGE of infiltrated beak with coacervates phase and solution phase of DgHBP-1. L: ladder. P: purified DgHBP-1 for control. N: native *D. gigas* beak sections before de-proteinization. S: beak “scaffold” after de-proteinizing the native beak sections. cC: condensed coacervates. dC: dispersed coacervates. (d) Amino acid composition of (Ala + Gly) in infiltrated beak “scaffold” normalized by the “scaffold” dry weight before infiltration. $N = 3$; ** $P < 0.001$; * $P < 0.05$.

Figure 6.1 Schematic representation of DgHBPs-based peptides and variants (25 amino acids and above). (a) DgHBP-1 full protein, GA-25, GH-25, GY-23, V3-C, V4-C, V11-C to V15-C, and V11-N to V15-N. (b) DgHBP-2 full protein, GW-26, V1-C and V1-N.

Figure 6.2 Expression and purification of DgHBP-1-V3-C and DgHBP-1-V4-C. (a) SDS-PAGE (left panel) and Western blot (right panel) of expressed DgHBP-1-V3-C and DgHBP-1-V4-C fractions after cell lysis. L: ladder. S: supernatants after cell lysis. P: pellets after cell lysis. (b) SDS-PAGE of DgHBP-1-V3-C and DgHBP-1-V4-C fractions after reverse phase HPLC purification (shown in (c) and (d)). (c) and (d) Reverse phase HPLC chromatograms of DgHBP-1-V3-C and DgHBP-1-V4-C, respectively. (e) and (f) MALDI-ToF spectra of purified DgHBP-1-V3-C and DgHBP-1-V4-C, respectively.

Figure 6.3 Expression and purification of DgHBP-1-V11 to DgHBP-1-V15. (a) SDS-PAGE of expressed DgHBP-1-V11 to DgHBP-1-V15 after cell lysis. (b) Acetic acid extraction of DgHBP-1-V11 to DgHBP-1-V15 after cell lysis. L: ladder. T: total fraction

(before extraction). S1: supernatants after first acetic acid extraction. P1: pellets after first acetic acid extraction. S2: supernatants after second acetic acid extraction of P1. P2: pellets after second acetic acid extraction of P1. **(c)** Reverse phase chromatogram of DgHBP-1-V15 (representative of V11 to V15). **(d)** SDS-PAGE of DgHBP-1-V15 fractions after reverse phase purification (representative of V11 to V15).

Figure 6.4 Trypsin cleavage and purification of DgHBP-1-V11-C/N to DgHBP-1-V15-C/N. **(a)** SDS-PAGE of trypsin cleavage experiments. L: ladder. BC: protein variant before cleavage. AC: products after cleavage. **(b)** Reverse phase HPLC chromatograms of DgHBP-1-V11 cleavage products (representative of V11 to V15). **(c)** SDS-PAGE of cleaved V11 to V15 after RPC. **(d)** and **(e)** MALDI-ToF of DgHBP-1-V11-C and DgHBP-1-V11-N after reverse phase HPLC separation (representative of V11-C/N to V15-C/N), respectively.

Figure 6.5 Trypsin cleavage and purification of DgHBP-2-V1-C/N. **(a)** Reverse phase HPLC chromatogram of DgHBP-2 cleavage products. **(b)** Formation of coacervates by adding the dissolved HPLC fractions into phosphate buffer. **(c)** SDS-PAGE of DgHBP-2 after cleavage, and fractions after reverse phase HPLC. L: ladder. BC: before cleavage. AC: after cleavage. **(d)** SDS-PAGE of DgHBP-2-V1-C/N separation by the coacervation method. L: ladder. P: pellets after coacervation. S: supernatants after coacervation. **(e)** and **(f)** MALDI-ToF spectrum of DgHBP-2-V1-C and DgHBP-2-V1-N, respectively, after separation by coacervation.

Figure 6.6 Optical microscopy images of coacervates formed by DgHBPs-based peptides and variants. **(a)** to **(f)** Optical micrographs of coacervates formed by GY-23, GW-26, DgHBP-1-V3-C, DgHBP-1-V4-C, DgHBP-1-V11-C and DgHBP-1-V15-N, respectively.

Figure 6.7 Rheological studies on coacervates of DgHBPs-based peptides GY-23 and GW-26. **(a)** Viscosity of coacervates formed by GY-23 and GW-26. **(b)** Frequency

sweep measurements of storage moduli (G' , closed symbols) and loss moduli (G'' , open symbols) of coacervates formed by GY-23 and GW-26.

Figure 6.8 CD spectra of DgHBPs-based peptides. (a) GY-23 at solution phase (1 mg/ml, 10 mM acetic acid) and coacervates phase (final concentration: 1 mg/ml). (b) GW-26 at solution phase (1 mg/ml, 10 mM acetic acid) and coacervates phase (final concentration: 1 mg/ml).

Figure 6.9 Turbidity measurement for (a) GW-26 and (b) DgHBP-1-V3-C coacervates, respectively. Turbidity as a function of pH at 100, 500, and 1000 mM ionic strength.

Figure 7.1 Proposed model for *D. gigas* beak formation. Three steps in beak bioprocessing are (i) DgCBPs and chitin nano fibers are secreted to form the beak scaffold; (ii) coacervation of DgHBPs enables protein infiltration within the chitin network; (iii) aggregation and cross-linking by DgHBPs causes the dehydration and stiffening of the beak material.

Figure 7.2 Optical microscopy of DgHBP-1 and DgHBP-2 mixed coacervates. (a) Presence of uniform micro-droplets in coacervates form. (b) Dendritic microstructures formed by proteins, possibly due to self-assembly.

Figure 7.3 DgHBP-1 cross-linking experiment. (a) Experimental design. Samples taken for subsequent SDS-PAGE are marked with number 1 to 4. (b) SDS-PAGE of samples taken at different experimental stages. L: ladder. 1-4: different samples taken during the experiments with each number annotated in (a).

Figure 7.4 Histidine functionalization of chitosan. (a) Substituting reaction of the amide group of chitosan with Boc-Histidine. (b) NMR spectra of the reaction product indicating successful substitution. DS %: degree of substitution.

Abbreviations

4MC	4-Methyl-Catechol
AA	Acetic Acid
AAA	Amino Acid Composition Analysis
AFM	Atomic Force Microscopy
Ala/A	Alanine
Asp/D	Aspartic Acid/Aspartate
ATR	Attenuated Total Reflection
ATR-FTIR	Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy
Au	Gold
Boc	Tert-butyloxycarbonyl
BSA	Bovine Serum Albumin
CBD	Chitin-binding Domains
CBR	Coomassie Blue R250
CCD	Charge-Coupled Device
CD	Circular Dichroism Spectroscopy
CHCA	α -Cyano-4-Hydroxycinnamic Acid
CMOS	Complementary Metal-Oxide-Semiconductor
CNC	Cellulose Nanocrystals
Cys/C	Cysteine
<i>D. gigas</i>	<i>Dosidicus gigas</i>
Da	Dalton
DgBP	<i>D. gigas</i> Beak Protein
DgCBP	<i>D. gigas</i> Chitin-binding Beak Protein
DgHBP	<i>D. gigas</i> Histidine-rich Beak Protein
DIC	Differential Interference Contrast
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic Acid
DS %	Degree of Substitution
DTT	Dithiothreitol

<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetraacetic Acid
FESEM	Field Emission Scanning Electron Microscopy
FFT	Fast Fourier Transform Algorithm
FPLC	Fast Protein Liquid Chromatography
FTIR	Fourier Transform Infrared Spectroscopy
GA	Gum Arabic
Gln/Q	Glutamine
Glu/E	Glutamic Acid/Glutamate
Gly/G	Glycine
His/H	Histidine
HPLC	High Performance Liquid Chromatography
IEX	Ion-Exchange Chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IR	Infrared
ITC	Isothermal Titration Calorimetry
kDa	Kilo Dalton
LB	Lysogeny Broth
Leu/L	Leucine
MALDI	Matrix-Assisted Laser Desorption/Ionization
MALDI-ToF	Matrix Assisted Laser Desorption Ionization-Time of Flight
MES	2-(N-Morpholino)ethanesulfonic Acid
ML	Mantle Length
MNP	Magnetic Nano-particle
NaOH	Sodium Hydroxide
NBT	Nitroblue Tetrazolium
NMR	Nuclear Magnetic Resonance
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDMS	Polydimethylsiloxane

Phe/F	Phenylalanine
PMSF	Phenylmethane Sulfonyl Fluoride/Phenylmethanesulfonyl Fluoride
PPP	Poly-penta-peptides
PRIMUS	Pedigree Reconstruction and Identification of a Maximum Unrelated Set
Pro/P	Proline
PVAc	Polyvinyl Acetate
RNA	Ribonucleic Acid
RPC	Reverse Phase Chromatography
SAXS	Small Angle X-ray Scattering
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Scanning Electron Microscope
Ser/S	Serine
SRT	Sucker Ring Teeth
TEM	Transmission Electron Microscopy
Tyr/Y	Tyrosine
UV	Ultra-violet
Val/V	Valine

Chapter 1

Introduction

*Through millions of years of evolution, Nature has provided countless inspirations for humankind with contextual tools and solutions to tackle emerging challenges. The science of biomimicry focuses on unveiling and utilizing the principles and patterns behind biological systems for engineering and technological applications. One example that has attracted recent attention is the Jumbo (Humboldt) squid beak, from the predatory species *Dosidicus gigas* (*D. gigas*). This chapter briefly introduces *D. gigas* and their beak. The hypotheses and objectives of this project are accordingly presented.*

1.1 Background and Hypothesis

1.1.1 Biomimicry, Biological Materials and Protein-based Materials

Biomimicry is the imitation of the models, systems and elements from Nature for the purpose of solving human problems.^[1] Evolution on earth has been going on for over nearly 4 billion years, whereby all creatures have developed delicate designs and engineering features adaptive to specific functions and environments. To mimic these specific designs and features, the mathematical, physical, chemical, or even philosophical principles underneath need to be understood as well.^[2]

Biological (or natural) materials fulfill various functions in living organisms, and constitute the organs internal skeletons or external load-bearing structure of animals and plant life. These structures are usually complex organic/inorganic composite materials, with mechanical characteristics optimized for their biological function.^[3] The building blocks contain inorganic elements (such as calcium, sodium, chlorine, sulfur, etc.) in ionized or mineralized form, as well as natural polymers (polysaccharides, proteins, nucleic acids, etc.), which play vital roles during biological processes.^[4] Among all natural polymers, proteins are arguably the most versatile members. Formed by covalent linking of amino acids, they often undergo multiple levels of structure to perform complicated functions. The diversity of their side-chains (20 different types of canonical amino acids) allows them to play a wide range of biological function,^[5] which can be further enhanced by post-translation modifications.^[6]

Structural proteins in Nature use secondary structure feature, as α -helices and β -sheets, as nano-scale load-bearing elements which translates into robust structures at the macroscopic scale.^[7] Classical examples of structural proteins found in vertebrates include collagen (in skin, bone, tendon, etc.), elastin (aorta, lungs, skin, bladder, etc.) or keratin (skin, hair, nails, etc.). With the rapid development and improvement in genetic and protein engineering techniques, there is an increasing interest to manufacture protein-based materials using fermentation processes.^[8–10] By doing so, a high level of control over materials'

macromolecular architecture is achieved since these techniques allow to control the chemical structure with single molecular precision, a potential envision by material scientists more than a decade ago.^[7]

1.1.2 *D. gigas* and *D. gigas* Beak (Proteins)

Dosidicus gigas, also known as Jumbo squid or Humboldt squid, is a member of the squid family *Ommastrephidae*.^[11] Commonly encountered in Eastern Pacific Ocean, they are known as highly efficient Ocean predators, with impressive size (up to 2 m in length) and speed (24 km/h). *D. gigas* has drawn increased interest in the biomimetic materials community due to their unique biotools, namely their toothed suckers and sharp beaks. The teeth embedded with the suckers, known as Sucker Ring Teeth (SRT), are non-mineralized protein-based materials that exhibit remarkable mechanical properties and that function as a grappling tool in predation.^[12–14] The squid beak is a composite, protein-based structure with an internal gradient of hardness and stiffness: its soft proximal region is attached within soft muscular buccal mass and protrudes toward the distal rostrum.^[15] The spatial gradient of the beak components (protein, chitin, water) have been proposed to be regulated by the constituting two protein families (DgCBPs and DgHBPs) to achieve the chemical gradient.^[16–18] The first protein family, DgCBPs (*D. gigas* chitin-binding beak proteins), binds to the polysaccharide chitin nanofibers to form the structural scaffold of the beak, possibly by reverting the intrinsic hydrophobic nature of chitin. The second protein family DgHBPs (*D. gigas* histidine-rich beak proteins) are believed to play important roles during beak “curing” (sclerotization process) and strengthening by forming a dense cross-linked network within the chitin/DgCBPs scaffold, the details to which still remain to be explored.

1.1.3 Hypotheses

To date, there have been no studies about the biochemistry and specific functions of DgHBPs and how it related to beak development, let alone relationships between DgHBPs’ sequences and their functions. Based on limited prior knowledge about DgHBPs, the following hypotheses are proposed:

- a. After secretion from the bucciblast gland, large amounts of DgHBPs go through a permeation process within the chitin/DgCBPs scaffold to reach the rostrum region of the beak.
- b. The highly modular design of DgHBPs' sequences (described in Chapter 2) are playing important roles during formation of the beak material.
- c. After permeation of DgHBPs, cross-linking reactions among proteins sclerotize the beak and result in a controlled gradient in material properties.

1.2 Objectives and Scope

This thesis aims to better understand the biological and physical roles that DgHBPs play during the beak material production, as well as to identify the relationship between protein sequences and the mechanical properties of the beak material. To achieve the above targets, two most abundant variants from the DgHBP family, DgHBP-1 and DgHBP-2 were selected. In their final form within the sclerotized beak material, they are densely cross-linked and extremely difficult to extract and purify. Therefore, a recombinant expression system has been set up to express, purify and study these proteins in an efficient manner. With that in mind, the project objectives are the following:

- a. To set up an efficient recombinant expression and purification system for DgHBP-1 and DgHBP-2.

The DNA sequences of full-length DgHBP-1 and DgHBP-2 were identified by combining RNA-Seq with proteomics techniques. The genes encoding the above sequences were designed and cloned into pJexpress vectors and then transferred into *E.coli* cells for subsequent recombinant expression. The crude proteins were purified mainly by reverse phase chromatography (RPC) with a high performance liquid chromatography (HPLC) system. The purified proteins were then examined by Matrix-Assisted Laser Dispersion

Ionization: Time-of-Flight (MALDI-ToF) mass spectroscopy, as well as by Sodium Dodecyl Sulfate Polyacrylamide Gel (SDS-PAGE) methods.

b. To perform comprehensive studies on the biophysical features of (*rec*)DgHBP-1 and (*rec*)DgHBP-2, and to shed light on the way by which they are transported within the beak material scaffold.

(*rec*)DgHBP-1 and (*rec*)DgHBP-2 were examined with a variety of characterization methods, including optical microscopy, UV-vis spectroscopy, dynamic light scattering (DLS), circular dichroism (CD) spectroscopy, Fourier transform infrared (FTIR) spectroscopy, isothermal calorimetry (ITC), rheology, small angle X-ray scattering (SAXS), etc.

c. To identify the sequence domains of DgHBP-1 and DgHBP-2 that contribute to their biophysical properties.

Multiple peptides were designed and synthesized based on the selection of different sequence domains of DgHBP-1 and DgHBP-2, then examined via methods including optical microscopy, UV-vis spectroscopy, dynamic light scattering (DLS), and circular dichroism (CD) spectroscopy.

1.3 Dissertation Overview

This thesis is focused on biomimetic investigation of *D. gigas* beak by examining two protein members from the *D. gigas* histidine-rich beak protein (DgHBPs) family, DgHBP-1 and DgHBP-2.

Chapter 1 provides a brief introduction of biomimicry, biological materials and protein-based materials, followed by basic knowledge about *D. gigas* and the *D. gigas* beak proteins. Hypotheses, objectives and scopes of this project are also outlined.

Chapter 2 reviews the literature concerning previous studies of *D. gigas* beak and *D. gigas* beak proteins. Proteins that are similar to DgHBPs in terms of primary structure and share common biophysical and structural features, namely tropoelastin and mussel foot protein (mfp-3S) are presented. All these proteins share the ability to be processed via the “coacervation” behavior, which is also a key hypothesis in this thesis.

Chapter 3 presents the methodologies and underlying principles of the experimental procedures employed during the preparation and the biophysical characterization of DgHBPs,

Chapter 4 describes the results of recombinant expression, purification and primary characterization of DgHBP-1 and DgHBP-2.

Chapter 5 present the bio-physical studies of DgHBP-1 and DgHBP-2, with attention to their phase separation behavior (“coacervation”) and the unique properties demonstrated by the coacervates phase.

Chapter 6 elaborates on DgHBPs-based peptides and their variants, from their preparation to their characterization, with the aim to clarify the sequence-coacervation ability relationship of DgHBPs peptide domains.

Chapter 7 draws together the previous chapters by summarizing the main findings of this work, and by proposing an outlook for future research and prospects on *D. gigas* beak biomimetic research.

1.4 Findings and Outcomes/Originality

This project led to several novel outcomes by:

- a.** Establishing an efficient platform to study *D. gigas* histidine-rich beak proteins (DgHBPs) through recombinant expression, purification, and subsequent characterization.

- b.** Revealing the profound implications of DgHBPs' phase behavior (coacervation) on *D. gigas* beak bio-processing
- c.** Clarifying the sequence-coacervation ability relationship in DgHBPs.
- d.** Providing inspirations for the design and manufacturing of biomimetic biocomposite materials with graded mechanical properties.

References

- [1] J. F. V Vincent, O. A. Bogatyreva, N. R. Bogatyrev, A. Bowyer, A.-K. Pahl. *J. R. Soc. Interface* **2006**, 3, 471–82.
- [2] J. Aizenberg, P. Fratzl. *Adv. Mater.* **2009**, 21, 187–188.
- [3] M. A. Myers, P. Chen, Y. L. Albert, Y. Seki. *Prog. Mater. Sci.* **2008**, 53, 1–206.
- [4] R. C. Hermand. **1993**.
- [5] G. Walsh. *Proteins: Biochemistry and Biotechnology*, John Wiley & Sons, **2002**.
- [6] B. P. Lee, P. B. Messersmith, J. N. Israelachvili, J. H. Waite. *Annu. Rev. Mater. Res.* **2011**, 41, 99–132.
- [7] J. C. M. van Hest, D. A. Tirrell. *Chem. Commun.* **2001**, 1897–1904.
- [8] B. D. Olsen. *AIChE J.* **2013**, 59, 3558–3568.
- [9] D. Sengupta, S. C. Heilshorn. *Tissue Eng. Part B Rev.* **2010**, 16, 285–293.
- [10] R. L. DiMarco, S. C. Heilshorn. *Adv. Mater.* **2012**, 24, 3923–3940.
- [11] C. M. Nigmatullin, K. N. Nesis, A. I. Arkhipkin. *Fish. Res.* **2001**, 54, 9–19.
- [12] S. H. Hiew, A. Miserez. *ACS Biomater. Sci. Eng.* **2016**, acsbiomaterials.6b00284.
- [13] P. A. Guerette, X. S. Hoon, X. D. Ding, S. Amini, A. Masic, et al. *ACS Nano* **2014**, 8, 7170–7179.
- [14] D. Ding, P. A. Guerette, S. Hoon, K. W. Kong, T. Cornvik, M. Nilsson, A. Kumar, J. Lescar, A. Miserez. *Biomacromolecules* **2014**, 15, 3278–3289.
- [15] T. A. Uyeno, W. M. Kier. *J. Morphol.* **2005**, 264, 211–222.
- [16] A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite. *Science* **2008**, 319, 1816–1819.

- [17] A. Miserez, D. Rubin, J. H. Waite. *J. Biol. Chem.* **2010**, 285, 38115–38124.
- [18] Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez, J. H. Waite. *Nat. Chem. Biol.* **2015**, 11, 488–495.

Chapter 2

Literature Review

Biomimetic materials research has become a rapidly growing and highly promising field; however successful translation of Nature's inspirations into the technical world requires thorough studies on natural materials. D. gigas beak has drawn interest due to its unique properties, composition, and structure. In this chapter, a comprehensive review of D. gigas beak is presented, with an emphasis on its constituting structural proteins, D. gigas Beak Proteins (DgCBPs and DgHBPs). Sequential homology between DgHBPs and other structural proteins (mussel foot proteins and tropoelastin) will also be discussed, together with an introduction of "coacervation", a special phase separation behavior demonstrated by those proteins. Finally the scope of this project is described.

2.1 Biomimetic Materials

As a major source of inspiration, Nature has served human endlessly in developing new technologies. In the Material Science community, there has been a growing aspiration in the past few decades to create advanced, biomimetic materials that could reformulate the “smart” designs from biological structures into man-made products. For example, studies on *Nelumbo nucifera* lotus leaf (**Figure 2.1a**) have revealed that the combination of hierarchical micro-nanostructure and hydrophobic epicuticular wax led to its super-hydrophobicity and self-cleaning ability (**Figure 2.1b**), and efforts have been made to construct similar structures (**Figure 2.1c**).^[1,2] Another example is shark skin, which possesses optimal riblet geometrical micro-structures that decrease the friction drag of water flows (**Figure 2.1d**).^[3] Based on this design, racing swimsuits have been produced by Speedo, helping athletes set records in the Olympics (**Figure 2.1e**).^[4]

However, it has been pointed out that mere observation and direct transferring of macro-scale structural principles into engineering materials is often insufficient, and that the understanding of their underlying design principles and physical/chemical mechanisms at the molecular, cellular, tissue, and organism levels must precede the engineering of bio-inspired materials.^[5] A complete circle of biomimetic materials research, therefore, ought to start with elucidating structure-function relations in biological materials, starting from the molecular level (**Figure 2.1f**, arrow 1), followed by studying physical/chemical principles behind the biological materials’ properties (**Figure 2.1f**, arrow 2), and finalized with material development under the discovered principles (**Figure 2.1f**, arrow 3).^[5] It is also noteworthy to mention, though, that biomimetic production must also account for realistic engineering and economic viability constraints.

Figure 2.1 Illustrations of biomimetic materials engineering. (a) *Nelumbo nucifera* lotus leaf.^[2] (b) SEM image of the lotus leaf surface micro-papillae-structures.^[2] (c) SEM images of superhydrophobic polystyrene film with structures inspired by lotus leaf surface.^[2] Inset is a water droplet on the film. (d) Riblet-structures of fast-swimming shark skin, at different body regions.^[3] Scale bar: 0.5 mm. (e) Speedo *Fastskin*® swimsuit, adapted from <http://www.popsci.com/technology/article/2012-07/speedos-super-fast-sharkskin-inspired-swimsuit-actually-nothing-sharks-skin>. (f) Flow chart of biomimetic materials research process.^[5] Arrow 1, 2 and 3 indicates the priorities of different steps.

2.2 *D. gigas* and *D. gigas* Sucker Ring Teeth

As a predatory species, *D. gigas* (Jumbo squid, or Humboldt squid) squids are equipped with two effective hunting hard biotools with robust mechanical properties that has attracted the attention from biomimetic materials researchers, namely the sucker ring teeth (SRT), and the squid beak. Reviews on *D. gigas* and SRT will be covered in this Sub-chapter, while studies on squid beak will be thoroughly presented in *Sub-chapter 2.3*.

2.2.1 *D. gigas*

Dosidicus gigas, commonly known as the jumbo squid, is the largest and most abundant nektonic squid species.^[6] *D. gigas* belongs to the family Ommastrephidae, and the sub-family Ommastrephinae, which is the most phylogenetically advanced sub-family in the family.^[6] *D. gigas* is the largest among all five genera of Ommastrephinae, with a dorsal mantle length (ML) of up to 1000 to 1200 mm, and weight up to 30 to 50 kg.^[7] Different from the other four ommastrephins, *D. gigas* possesses only minor mantle skin photophores instead of large ones; it lacks eye and intestinal photophores in paralarvae; it has elongated, attenuated arm ends; and it has much a higher number of suckers (up to 200) on the arms than others (32 to 35).^[7]

The occurring areas of *D. gigas* mainly lie within the region of the current deformation field of the eastern Pacific, in the range between 30 °N and 25 °S. Its frequent habitats are the transformed waters of the California and Peru currents.^[8]

D. gigas is characterized by a fast juvenile growth rate, up to 5 to 8 % mantle length (ML) per day. Typically, it attains 700 to 750 mm ML within one year after hatching.^[9]

D. gigas is an active predator, known for its aggressive hunting behavior. Its common prey species include amphipods, pelagic shrimps, red crabs (*Pleuroncodes planipes*), various fishes, heteropod mollusks, etc.^[10] Cannibalism behavior is also observed for *D. gigas*, accounting for 5 to 9 % of the adult squid's body weight.^[11]

D. gigas anatomic features were first accounted and described in the 19th century by Alcide D'Orbigny, including the hard appendages, namely the beaks and the toothed suckers^[12] (Figure 2.2), both of which have attracted the attention of material scientists.

Figure 2.2 Drawing of *D. gigas* and its “hard tissues” appendages. By French naturalist, Alcide D’Orbigny.^[12]

2.2.2 *D. gigas* Sucker Ring Teeth

The *D. gigas* sucker ring teeth (SRT) are embedded in the hundreds of sucker cups aligned along the long, strong arms. Used as grappling utensil by *D. gigas*, SRT match requirements to sustain both high dynamic compressive and shear stresses.^[12,13] SRT also

Figure 2.3 Structural and mechanical properties and primary structure features of *D. gigas* SRT. (a) A representative *D. gigas* SRT.^[12] (b) SEM image of the SRT nanotubular structure.^[12] (c) Ashby plot of common engineering materials (polymers, metals, and ceramics) in comparison to dried native SRT.^[12,14] (d) Highly modular architecture of a representative *D. gigas* suckerin (suckerin-19) showing alternating Ala- and Tyr- rich domains.^[15]

exhibit a higher elastic modulus (up to 7.5 GPa) than many synthetic polymers such as polyamides, polymethyl methacrylate, etc. (Figure 2.3c), despite the absence of common strengthening phases found in other hard biological structures (e.g. minerals, metal ions, and chitin).^[12] SRT are also found to be fully proteinaceous, composed of a protein family “suckerins”, which is biased towards 9 amino acids (Gly, Tyr, His, Leu, Ala, Val, Pro, Thr and Ser), most abundant being glycine (37 mol%), tyrosine (14 mol%) and histidine (13 mol%).^[13,16] Extensive sequencing studies have revealed the highly modular design of suckerins primary structures, specifically the GGY repeats with alternating poly-Ala motifs (Figure 2.3d), which holds resemblance to silk proteins.^[17–19] This modular feature of protein sequences is believed to favor the formation of β -sheet conformation, and subsequently leads to the self-assembly of suckerins into the tenacious, intertwined nanotubular structure of SRT, as observed by Scanning Electron Microscopy (SEM) (Figure 2.3b).^[12,13,16,20]

2.3 *D. gigas* Beak: Composition, Structure and Properties

In addition to SRT, the *D. gigas* beak represents another wholly organic biological material known to have remarkable hardness and stiffness. Deeply embedded within the soft buccal envelope of *D. gigas* (Figure 2.4a), the beak serves as an effective feeding tool. Controlled by mandibular muscles, the upper beak and lower beak (Figure 2.4b) can easily crush crab shells, or sever the nerve cord of a captured fish by a single notch.^[21] Initially selected as a model biological material due to their relatively large size (on the order of 3 to 7 cm, Figure 2.4c) that facilitates sample preparation for mechanical testing^[22], it turned out to possess unique biochemistry and impressive biomechanical properties, which will be covered in the following sections.

2.3.1 *D. gigas* Beak: Composition

Compositional analysis has revealed that *D. gigas* beak is a biocomposite material, composed of four components: α -Chitin, proteins, black pigments, and water.^[21–23] α -Chitin is a polysaccharide of *N*-acetylglucosamine arranged with β -1,4 linkages, resulting

Figure 2.4 Beak of *Dosidicus gigas*. (a) Beak on a captured *Dosidicus gigas* squid. Adapted from <http://www.fadedindustry.com/extremely-terrifying-facts-about-squids/>. (b) Schematic illustration of *D. gigas* beak on buccal mass.^[21] (c) Upper and lower beaks, after dissection from the buccal mass.^[22]

in strong, linear fibers.^[24] As a common reinforcing natural substance, it is also present in exoskeletons of many arthropods such as crabs, lobsters and shrimps.^[24] Full protein extraction from the beak proved futile, due to its resistance to most aggressive extraction cocktails, which has been attributed to its highly sclerotized and cross-linked nature.^[22] However, overall amino acid composition analysis of the beak was made possible after degradation treatments by acid hydrolysis and alkaline peroxidation.^[23] Glycine (Gly), Alanine (Ala), and Histidine (His) were found to be the most abundant amino acid in the beak.^[22,23] Tanning pigments were also found in the beak, causing the dark color of the

Figure 2.5 Compositional gradient in *D. gigas* beak. (a) Side view of a dissected upper beak highlighting the transition from the wing to the rostrum.^[23] (b) Series of specimens from the beak showing different pigmentation levels.^[23] (c) Chitin, protein, pigment and water contents from chemical and gravimetric analysis results.^[23]

rostral regions.^[23] Those insoluble pigments were demonstrated to be catecholic-rich cross-links based on the catechol-specific Arnow stain^[25], the origin of which was initially unclear.^[23,26] The water content in beak was found to vary greatly in different regions of the beak, about 70 wt% in the proximal region embedded within the buccal mass, but only 15 wt% at the distal rostrum. Considering the hydrophobic nature of Chitin, which is present

throughout the beak, this change of water content was considered to be controlled by other factors, including pigmentation level and the hydrophobicity of proteins present in the beak.^[23]

Further examination of the beak composition unveiled significant biomolecular gradients. It was first noticed by the visible incremental differences in pigmentation, ranging from a

semi-transparent color along the wing edge to a dark brownish one at the beak tip (Figure 2.5a and Figure 2.5b).^[23] After cutting out specimens from regions of different pigmentation level and carrying out chemical analysis, it was found that there was a consistent compositional trend among all four components. As shown in Figure 2.5c, from the untanned wing region to fully tanned rostrum region, the chitin content decreased from about 25 wt% to less than 15 wt%; the protein content increased significantly from 5 wt% to 60 wt%; the pigment level increased from 2 wt% to 20 wt%; and the water content decreased from 70 wt% to 15 wt%.

Notably, mineral phases were found to be absent in *D. gigas* beaks, in contrast to most vertebrate and invertebrate hard tissues, e.g. human teeth, mollusk shells, etc.^[27–29] This indicates that instead of biomineralization, *D. gigas* employs a different strategy to achieve a strong, abrasion-resistant tooth-like structure.

2.3.2 *D. gigas* Beak: Structure

The macroscopic, anatomical structure of *D. gigas* beak has been shown to be typical of cephalopod beaks (Figure 2.6a and Figure 2.6b).^[30,31] Composed of two sub-structures (upper beak and lower beak, Figure 2.4c and Figure 2.6a), the beak movement is controlled by four major muscles in the buccal mass. During feeding actions, the upper beak and lower beak slide by each other, conducting high biting power and stability despite of no direct connection with each other (Figure 2.6b). The two beak portions differ in shape yet share same morphology, exhibiting a three-dimensional V-shape structure (**Figure 2.4c**). The proximal wing region of both beak portions are embedded within the muscular buccal mass tissues (**Figure 2.4a** and **Figure 2.4b**). At the beak/buccal mass interface, a single layer of elongated cells (the “beccublast”) exist, acting as connective tissues between muscle and beak, as well as secreting glands of beak precursors.^[32]

The microstructure of *D. gigas* beak as revealed by electron microscopy indicated a multi-layered structure with 2 to 5 μm -thick lamellae (**Figure 2.6d**).^[21] These lamellae are aligned in parallel with the long axis of the beak pointing to the tip, yet are oblique to the beak

surface plane by 50 ° to 90 ° (**Figure 2.6c**).^[21] At the outmost external surface of the beak,

Figure 2.6 *D. gigas* beak structures. (a) Computer rendering of the upper (blue) and lower (red) beaks of *Octopus bimaculoides* (cephalopod) in a front right quarter view.^[31] Left: the upper beak fits within the lower beak; Right: the upper beak separates from the lower beak. (b) Side view diagram showing movement of the upper (blue) beak to anchored lower beak (red) during a stereotypical bite cycle.^[31] Position i: resting; ii: opening; iii: fully open; iv: closing; v: closed and retracted. (c) Cross-section SEM images of the beak showing the lamellae alignment, featuring an external surface layer.^[21] (d) Lamellar microstructure SEM micrographs of fractured beak surfaces.^[21] Inset shows an increased magnification of the rectangle area. (e) SEM micrographs of Chitin leaflet microstructure in the beak after alkaline peroxide treatment.^[23] (f) SEM micrographs of intertwined chitin nanofibers.^[23]

however, the lamellae are covered with a protective layer oriented in parallel to the surface plane (**Figure 2.6c**), possibly serving as hydrophobic sealant.^[21]

The chitin scaffold microstructure in *D. gigas* beak was revealed by removing other components of the beak using alkaline peroxide treatment.^[23] An entangled fibrillar network was observed, containing leaflets sub-microstructures made of smaller chitin fibers with a diameter of about 30 nm (**Figure 2.6e** and **Figure 2.6f**).^[23] Further SAXS studies on different regions of the beak showed an anisotropic pattern at the tip part, and an isotropic pattern at the proximal region, indicating an elongated and orientated alignment of chitin fibers in the first region whereas a more random organization in the latter.^[21]

2.3.3 *D. gigas* Beak: Mechanical Properties

Stiffness, hardness, tensile strength and fracture toughness of *D. gigas* beak were measured by nanoindentation, tensile testing, and fracture toughness measurements, respectively. Results demonstrated a large disparity in mechanical properties for different regions of the beak. In the near-tip region for hydrated beak, the average Young's modulus (E) and hardness (H) values are 5.1 GPa and 0.43 GPa, respectively, higher than most synthetic polymers (**Figure 2.7a**).^[21] At the untanned, proximal wing region, the Young's modulus drops drastically to about 0.05 GPa (**Figure 2.7c**). Further examination revealed a stiffness gradient between these two sections along the beak axis (**Figure 2.7c**). This 100-fold stiffness gradient buffers the mechanical mismatch and minimizes contact damage between the hard beak and the soft buccal mass, which provides inspiration for orthopedic materials that connect soft and hard tissues, such as hip or cartilage. The stiffness gradient was shown to be closely related to hydration levels. As shown in **Figure 2.7c**, freeze-dried samples from the different regions exhibited relatively small differences in the Young's modulus values (factor 2) compared to the values in hydrated conditions. Similarly, hardness was also correlated to water content, with large differences between dry and hydrated samples (**Figure 2.7a**).

Figure 2.7 Mechanical properties of *D. gigas* beak. (a) Ashby plot (H against E) of *D. gigas* beak tip part with typical engineering materials.^[21] (b) H^3/E^2 - G_c map of different materials yielding/cracking showing abrasion resistance properties.^[21] Critical loads for yielding and cracking are represented by horizontal and vertical lines, respectively. Families of L-shaped curves represent different failure loads. (c) Young's modulus and tensile strength gradients in *D. gigas* beak at different regions, expressed as modulus against (protein and cross-link content).^[22,23] (d) Typical tensile stress/strain curves for hydrated beak specimens.^[23]

Tensile testing for various regions of the beak also illustrated the mechanical gradient. As shown in **Figure 2.7c** and **Figure 2.7d**, untanned proximal sections showed only 2 MPa stress increase with 3 % strain, whereas heavily tanned section showed over 30 MPa stress increase.

The tip region was also predicted to have a high resistance to abrasion. By plotting material's critical yielding load H^3/E^2 against critical cracking load G_c (**Figure 2.7b**), where H is hardness (GPa), E is Young's modulus (elastic modulus), and G_c is fracture energy, a comparative diagram of various materials assessing their abrasion resistance is obtained. In the diagram, a family of horizontal lines represents failure by plastic deformation; another family of vertical lines represents failure by cracking.^[21,33] As the vertical lines meet the horizontal lines, L-shaped curves are illustrated, each representing failure load for one abrasion. Materials with property data above and to the right of one of the L-curves, therefore indicates higher abrasion resistance. As shown in **Figure 2.7b**, the beak's resistance to plastic deformation when in contact with a blunt abrasive is comparable to enamel, dentin, and most engineering metals and polymers.^[21] Abrasion resistance also appeared to be linked to water content, as shown by the different property regions by wet and dried beak materials (**Figure 2.7b**).

2.3.4 *D. gigas* Beak: Composition-Mechanical Property Relationships

Composition analysis and mechanical characterization of the beak demonstrated biomolecular and mechanical gradients along the beak axis, from the proximal wing to the rostral tip. Thus, a central question can be raised: are these gradients correlated with each other? Based previous findings from reminiscent materials, such as the jaw of the bloodworm (*Glycera*) jaw^[34] and insect cuticles^[35], the answer most likely is yes. This would mean that the change in composition (chitin, proteins, pigments and water) would be correlated to the gradual sclerotization of the beak, leading to a steep stiffness gradient from untanned to fully tanned region. Although a complete mechanism still awaits to be unveiled, it has been proposed that two mechanisms play a critical role in this hardening process: the dehydration of α -chitin fibers, as well as a high inter-protein cross-linking density between the squid beak proteins.^[26] The first mechanism is related to the strong hydrated-dependent mechanical properties of α -chitin. Dried α -chitin has an elastic modulus of 41 GPa^[36], while that of hydrated α -chitin is only about 10 to 15 MPa^[37]. Also, as previously described, chitin functions as the scaffold for the beak material, therefore the

porosity content of its network would affect the compactness of the material, leading to large stiffness differences.^[22] Thus, the mechanical gradient of the beak is very likely to be tuned by fine control of hydration in the tissue. Given that Chitin is naturally hydrophobic, one could only assume that its wettability could be mediated by other components in the beak, namely proteins or pigments.

To date, the dehydration process of chitin fibers in squid beak and other similar biological materials such as insect cuticle remains unclear, the mechanism of which has long been debated.^[26,35] A common view is that protein cross-linking occurs via aromatic residues, which leads to the formation of aromatic pigments, causing the chitin network pore size to decrease, and stiffness of the beak to increase. Another argument was suggested, which is that the impregnation of the chitin-rich scaffold by hydrophobic proteins could also lead to chitin dehydration. The cross-linking mechanism involves the nucleophilic additive reaction between imidazole and a 4-methyl-catechol ring, existence of both has been verified by NMR and MS/MS.^[26] The imidazole comes from the His residues of DgHBPs, whereas the source of catechol ring is uncertain. One option is that it comes as a low-molecular-weight catechol secreted from the “beccublast” cells (Section 2.3.2).^[26] Another possibility is that it comes from the non-natural amino acid L 3,4-dihydroxyphenylalanine (L-Dopa), which is formed by the post-translation modification of Tyrosine (Tyr)^[38] that has been detected by amino acid analysis of squid beaks^[23]. However, further work is needed to elucidate the source of the catechols involved in cross-linking, though both possibilities mentioned above may co-exist.

In summary, the understanding of *D. gigas* beak sclerotization mechanisms necessitates comprehensive studies of the proteins forming the beak, as well as the interactions of proteins with other components, including water and chitin.

2.3.5 *D. gigas* Beak Inspired Materials

As illustrated in the previous sections, *D. gigas* beak displays a delicate design of organic composite material with intercorrelated compositional and mechanical gradients, which

Figure 2.8 Catechol-functionalized chitosan/ γ -Fe₂O₃ composite material inspired by *D. gigas* beak. (a) Synthetic strategy of the composites.^[39] (b) Young's modulus of dehydrated and hydrated composites (with and without MNPs).^[39] (c) Magnetization curves for mnp-CAT_{CS} films.^[39] Inset shows a typical hysteresis curve of control super-paramagnetic material. (CS: chitosan; CAT: catecholic moieties; DOPA: 3,4-dihydroxyphenylalanine; DHA: 3,4-dihydroxyphenethylamine. MNP: magnetic nano-particle.)

provides inspiration for the design of synthetic materials. Here two examples of artificial composite materials inspired by *D. gigas* beak are presented.

The sclerotization of *D. gigas* beak was proposed to derive from desolvation of the chitin scaffold coupled with His-catechol cross-linking. Following the same paradigm, a chitosan/ γ -Fe₂O₃ composite^[39] was designed as shown in **Figure 2.8a**. A chitosan scaffold was firstly modified with catecholic moieties with the goals of: (i) reducing the hydrophilicity of chitosan by replacing amine motifs along the backbone, and (ii) providing the chitosan with side-chain functionality that enables subsequent cross-linking.^[40] Next, by adding γ -Fe₂O₃ nanoparticles, catechol-Fe³⁺ coordination cross-links were triggered, a

mechanism inspired by the mussel thread coating materials.^[41,42] Thus, the chitosan scaffold here imitates the chitin network in the beak; the modified catechols mimics the 4-methyl-catechols in the beak, and the catechol-Fe³⁺ coordination cross-linking mimics His-catechol cross-links of the beak. The resulting film material displayed high stiffness (over 1 GPa in hydrated conditions, **Figure 2.8b**), proving successful stiffening effect of chitosan network dehydration and cross-linking. Besides providing cross-linking sites, the addition of γ -Fe₂O₃ nanoparticles also introduced stimuli-responsive super-paramagnetic properties into the material (**Figure 2.8c**), which is desirable for many bio-medical applications.

The gradient design of *D. gigas* beak also inspired nanocomposite material studies. By mimicking the beak model, where fibrous network (chitin) is filled with biopolymer (protein) matrix stiffened by cross-links (His-catechol), a similar system was constructed, where alkene-modified cellulose nanocrystals (Allyl-CNCs) functioned as filler, poly-vinyl acetate (PVAc) acted as matrix, and photo-induced thiol-ene cross-links fulfilled the role of sclerotizing the material with gradient, by varying UV exposure time (**Figure 2.9a**).^[43] The resulting film material successfully reproduced the mechanical gradient in the beak (**Figure 2.9c**). For hydrated, 15 wt% allyl-CNCs-PVAc nanocomposite film, an 5-fold increase of storage modulus from 60 to 300 MPa was observed (**Figure 2.9b**).^[43] Also, due to the water-responsive nature of CNCs, the composite film also demonstrated a “switchable” behavior, where the gradient only occurs in wet state (**Figure 2.9b**), exhibiting great potential for biomedical applications.

2.4 *D. gigas* Beak Proteins (DgBPs)

As summarized previously, the elusive roles of proteins in *D. gigas* beak hindered further biomimetic studies. Thus, identification and sequencing of *D. gigas* beak proteins was carried out shortly before the author started this research, providing critical background information for the present thesis work, which is reviewed in this Sub-chapter.

Figure 2.9 Allyl-CNCs-PVAc nanocomposite material inspired by *D. gigas* beak. (a) Schematic representation of mechanical gradient model in *D. gigas* beak and in the Allyl-CNCs-PVAc film.^[43] (b) Storage modulus gradient of the CNCs-PVAc film at different UV exposure time.^[43] Red rectangles on the top part represents the different sample locations that undergoes the photo-initiated cross-linking; bottom part shows the relative modulus at different locations on the film, under wet and dry conditions, respectively. (c) Twisting of a nanocomposite film that underwent gradient UV exposure (top picture, gradient as shown in panel b) and one that underwent uniform UV exposure (bottom picture).^[43]

2.4.1 DgBPs Extraction and Identification

To better understand the mechanical gradient formation in *D. gigas* beak, it is necessary to identify and characterize the beak proteins, as well as to unveil their post-secretion process. However, due to the densely cross-linked nature of the beak, extraction and identification of beak proteins employing conventional protein extraction procedures or enzymatic degradation methods mostly failed. Therefore a new approach was adopted, whereas non-enzymatic reagents (NH_2OH ^[44] and formic acid^[45]) were used to cleave peptide bonds between specific peptide bonds (Asx-Gly, Asp- and -Asp), therefore liberating fragmented proteins from the beak structure.^[46] This was complemented with various protein detections assays (Tricine-SDS-PAGE^[47], Pauly's stain^[48], NBT stain^[49], Lectin-binding assay^[50]), confirming that: *(i)* molecular weights of protein fragments range from 5 to 250 kDa; *(ii)* His-rich polypeptides are present in the fragments; *(iii)* Smaller fragments contain catecholic residues: and *(iv)* chitin oligomers exist in the higher-molecular-weight fractions (**Figure 2.10a**).^[46] Partial sequences of those fragments were obtained by tandem mass spectrometry (nanoLC-MS/MS) and Edman sequencing^[51]. Thus, sequential profiles of fragmented DgBPs' pieces was established, yet this was still far from full-length protein sequencing.

What helped completing this puzzle was another crucial technique called "RNA-Seq" ("whole transcriptome shotgun sequencing")^[52,53]. As an emerging approach to transcriptome profiling, it provides a far more precise assessment of levels of transcripts and their isoforms than other methods.^[52] Here, an RNA-Seq transcriptome library was established, based on the sequencing results from mRNAs isolated from the buccal tissue, a biological tissue comprised of "beccublast" cells that presumably secreted the beak proteins (*Section 2.3.2*).^[46] Since there was no available reference genome to DgBPs, a *de novo* method to assemble the full-length transcripts of DgBPs was employed using Trinity software.^[54] Then, by matching the assembled transcript sequences to DgBPs fragment sequences, the full-length sequences of different DgBPs were acquired and further confirmed by RACE-PCR^[55] and Sanger sequencing^[56]. As a result, two families of proteins were identified.^[46] The first family is comprised of 4 variants, most of which

Figure 2.10 *D. gigas* beak proteins identification. **(a)** Tricine-SDS-PAGE of extracted DgBPs stained/binded with CBR, Pauly stain, NBT stain and Lectin-binding assay.^[46] CBR: to show all proteins; Pauly stain: to show His-rich proteins; NBT stain: to show catecholic proteins; Lectin-binding assay: to show *N*-acetylglucosamine oligomers in the beak. L: Ladder. Loaded protein amounts per lane for each assay were 10 μ g, 30 μ g, 30 μ g and 60 μ g, respectively. **(b)** Schematic sequential representations of DgCBPs and DgHBPs.^[46] Modular domains in each protein are indicated. X: hydrophobic residue; X': tyrosine (mostly); X'': glycine or alanine.

containing two to three chitin-binding type 2 domains based on the sequence alignment against peritrophin-44, a known chitin-binding protein, as determined by ScanProsite tool^[57]. Therefore, the first protein family was proposed to function as chitin-binding agents, named as “*D. gigas* chitin binding beak proteins”, or “DgCBPs” (**Figure 2.10b**).^[46] The second family of DgCBPs was ascertained by the knowledge that DgCBPs amino acid compositions deviate from the amino acid composition of tanned beak regions. At the same time, DgCBPs amino acid composition was close to that of the untanned parts, where His, Gly and Ala content are relatively low. This indicated that in the tanned beak portions, another class of proteins containing high levels of His, Gly and Ala exists. By screening the transcriptome of DgBPs with a compositional bias, this family comprised of three variants was identified (**Figure 2.10b**) and named “*D. gigas* histidine-rich beak proteins”, or “DgHBPs”.^[46]

2.4.2 DgCBPs: Sequence Features, Homology and Conservation

The RACE-PCR confirmed sequences of DgCBPs are shown in **Table 2.1**. Among them, DgCBP-1, DgCBP-2 and DgCBP-3 were also discerned in NH₂OH extracts by a chitin-binding assay (**Figure 2.11a**), while DgCBP-4 was identified using transcriptome search as having the same chitin-binding type 2 domains. Conservation analysis of these sequences using Clustal software indicated high resemblance to peritrophin-44, a known chitin-binding protein from insect *Lucilia cuprina*, for all 4 proteins (**Figure 2.11b**). Those chitin-binding domains (CBDs) accounts for over 60 % of the full sequence for all 4 variants (**Figure 2.11b**). No repetitive features in the sequences was detected from the analysis. In summary, it is with a low possibility that DgCBPs could play other roles other than as a fixing and stabilizing component for chitin fibers present in the beak.

2.4.3 DgHBPs: Sequence Features, Homology and Conservation

The RACE-PCR confirmed sequences of DgHBPs are shown in **Table 2.2**. Like DgCBPs, these sequences were also sent to several protein sequence database (GenBank and SwissProt) to search for homogeneous proteins, yet none appeared. This indicated that

DgHBPs represent an original class of proteins present in nature, encoded by a unique gene family. As expected, all three proteins showed a high content of Gly, Ala and His, whose combined composition contributes to 62 %, 53 % and 68 % to the sequence of DgHBP-1, DgHBP-2 and DgHBP-3, respectively, which is in accordance with previous squid beak compositional studies as introduced in *Section 2.3.1*.^[22,23]

Table 2.1 cDNA deduced sequences of DgCBPs with predicted molecular weight (Mw) and isoelectric point (pI).^[46] Underlined: chitin binding domains. Bold Italic: signal peptide sequences. Blue: MS/MS deduced sequences. Red: Edman sequence. Bold purple: potential NH₂OH cleavage sites.

Protein	cDNA Deduced Sequences	Predicted Mw / pI
DgCBP-1	<i>MGTNGLPNVG</i> <i>LILFLWGLTF</i> <i>CVAFG</i> <u>QSFDI</u> CAGGLTGVEE DKTDRCNLFLE CLSGRLKHFR CHNNKVFSTK TK <i>TCVTLNEA</i> <u>AGQCSTDFSR</u> KVHEIVNPYS YCQLHPTAVL PHPDSCALFY NCSRPTSLGK <i>YFSPYVDECP</i> <i>YPELFSK</i> TTK RCEPYRTVHC ERRPEPRDVC QYRQNWCGQS HCVPCHIRFP SCRGKSEGLN SVPYVK <i>WGPQ</i> <i>YIECR</i> HGRLI TKSKCIGKPG QPAVFSPQSR SCISVFSVSN PPTNPISCRY KSDGFYANPT NCKKYFQCLN <u>QNFIDMLECP</u> ENLVFDSRAA SCVRPSNSCP PCGTKIGC	33.1 kDa/8.9
DgCBP-2	<i>MYRTAAIFIV</i> <i>VIVGVVICQD</i> <i>NLISICK</i> NVK RRVGTNLVRD <u>QGNCSVFYSC</u> RHPNPRLEFPM ACGYTTVFSQ <u>SQQVCVMQNS</u> <u>IYDDCKRSIY</u> <i>GGTANDPLCR</i> PHPSGLNRDP ADCTRFIPCF <u>NHTLILHVCQ</u> <i>DSLHLLSKE</i> CCRTGDLMPH NSHTEHTF	15.8 kDa/8.3
DgCBP-3	<i>MYIKALIVIG</i> <i>LVSVLAGLAT</i> <i>NAFTCPPKGT</i> <i>SWIRD</i> PRNCS LNHWCIN <i>NGNY</i> SGSFYCGPGY VAGGSAALCM VAPQEYDWN <u>CKDLMKINIE</u> <u>NDPRCAHK<i>NG</i></u> SNPEPTDCAR FVECWQGTSM <u>SVKWCPKGLL</u> YNATKRICMW PTRVNCANRP NKMADSAKIS <u>KFRTECNTAR</u> GKGHMLIAYP GDSCQYISCD <u>SNDQSPQQCA</u> <u>SGTVFNSEKQ</u> RCDFRANVPS CKVPA	22.6 kDa/8.7
DgCBP-4	<i>MFLYATCVFL</i> <i>AAGLVLSTAQ</i> IATNPCEVAG VHTQLMADPT SCSNFFFCVG KIPRRMECGP GTLFNARIKM CDHAENVNCD <u>DTKRPIVPPT</u> KPSPSVRPTV PIHDTKLVEQ YYERWYKFYC <u>SEYGSNNVVR</u> WPHPNKCDRF YECRQQKLSA VVCPTY <i>PQQM</i> <u>LYNRTLQNC</u> D LPQVRCEKS LPFGNNFASA SEEELKYRAL <u>CKPGTMSFNR</u> WANSHDCCRFF YQCLN	23.9 kDa/8.8

Figure 2.11 Test and analysis of DgCBPs. (a) Tricine-SDS-PAGE of NH_2OH -extracted proteins eluting from chitin beads and separated as non-binding fraction (“flow-through” and “wash”) and binding fraction (“elute”).^[46] DgCBP-1, DgCBP-2 and DgCBP-3 are labeled in the elute fraction. (b) Aligned consensus residue comparison (10 % ClustalO threshold) between chitin-binding domains of DgCBPs and peritrophin-44, a known chitin-binding protein.^[46]

Among the three DgHBPs, DgHBP-2’s transcriptome assembly showed multiple isoforms (Table 2.3), which either arose from gene duplication, or alternative-splicing-caused misassembly due to the repetitive pattern of the sequences.

Conservation analysis was also conducted for DgHBPs. As shown in Figure 2.12, all three proteins exhibit highly conserved, large-scale modular architectures. In general, all

DgHBPs are characterized by a bipolar domain organization: the N-terminal regions are non-repetitive and contain stretched Ala- and His-rich domains, while the C-terminal regions contain tandem His-rich repeats (**Figure 2.12a**). For instance, C-terminus of DgHBP-1 and DgHBP-3 include GAGFAGHGXX'/X'' repeats, and DgHBP-2 contains tandem repeats of GHGXY, where X means hydrophobic residue, X' means Tyr (mostly), and X'' is Gly or Ala (**Figure 2.10b**). Also noted is the high resemblance of DgHBP-3's modular design to DgHBP-1, though its Gly-rich repeats in N-terminus are absent in DgHBP-1 (**Figure 2.10b**). After alignment of the C-terminus tandem repeats for all three proteins, it was also noticed that His is almost always flanked by Gly followed by hydrophobic or aromatic residues (**Figure 2.12b**), which is an interesting pattern that will be discussed in following chapters.

Table 2.2 cDNA deduced sequences of DgHBPs with predicted molecular weight (Mw) and isoelectric point (pI).^[46] Underlined: tandem repeated domains (first amino in each repeat marked as bold). Bold Italic: signal peptide sequences. Blue: MS/MS deduced sequences. Bold purple: potential NH₂OH cleavage sites.

Protein	cDNA Deduced Sequences	Predicted Mw / pI
DgHBP-1	<i>MKCVSFAIFL</i> <i>VVAIFGANSQ</i> <i>LYGAPAVGGV</i> <i>VENAVNAAES</i> <i>GAAATHDAQG</i> <i>AYAEADTAGV</i> <i>LDVNHAEHHD</i> <i>GVHDASGYGF</i> <i>GGLAGHGGA</i> <i>GHGLYGPGFA</i> <i>GHLLGAGYA</i> <i>GLGLHGAGFA</i> <i>GHGLHGAGFA</i> <i>GHGLYGAGFA</i> <i>GHGLHGFA</i> <i>GLYGAGFAGH</i> <i>GLGLGGLHGA</i> <i>LGHGALAHY</i>	15.1 kDa/6.0
DgHBP-2	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> <i>FFGAGPFNTA</i> <i>HHSVSDAAA</i> <i>AHHDAAGEYA</i> <i>QNAATGLLDT</i> <i>HHNENHDMTH</i> <i>DLANGYGLHE</i> <i>HDEQHHGLAD</i> <i>GLHQEYAARA</i> <i>AQGANAVHND</i> <i>AAQSHSALAA</i> <i>ANTFGHGHAP</i> <i>YAAYGHGVY</i> <i>HGPYGHGPY</i> <i>HGLYGHGLY</i> <i>HGPYGHGLY</i> <i>HGAFGHGLNA</i> <i>YAPLVGHGLR</i> <i>GYL</i>	18.1 kDa/6.1
DgHBP-3	<i>MNCYSFAIFL</i> <i>AVAILGARSQ</i> <i>YYGGVGGGLG</i> <i>GLHGLAGGVG</i> <i>GLHGLAGGVG</i> <i>GFAGSSGFHE</i> <i>HDEEHHGVAD</i> <i>GVHQAYATHA</i> <i>AQGAHAEHDA</i> <i>AQRHSAALAA</i> <i>ASAAHGGLAG</i> <i>FGGLAGHGAG</i> <i>FAGHGLGLAG</i> <i>HGFAGHGLHG</i> <i>AGFAGHGLGL</i> <i>AGHGFAGHGL</i> <i>HGAGFAGHGF</i> <i>GLGGHGLYGH</i> <i>Y</i>	15.1 kDa/6.5

Table 2.3 cDNA deduced sequences of DgHBP-2 variants with predicted molecular weight (Mw) and isoelectric point (pI).^[46] Bold Italic: signal peptide sequences.

Variant	cDNA Deduced Sequences	Predicted Mw / pI
DgHBP-2 α	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFNNA HSAVSDAAA SHHDAAGEYA QNAATGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGVDD GLLREYAASA AQGANAVHDD AAQRHAASAA ANNGHFGHGF GPYAAFHGV YHGVYGHGF YHGPGYGHGV YHGPGYGHAL YHGGLRAPLV GHGPGYGL	17.6 kDa/6.0
DgHBP-2 β	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFHNA HSAVSDAAA AHHDAAGEYA QNAATGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGLAD GLHQEYAARA AQGANAVHND AAQSHSALAA ANTFGHGHAP YAAIGHGVY HGPYGHGPGY HGLYGHGLY HGPYGHGLY HGAFHGLNA YAPLVGHGLR GYL	18.1 kDa/6.2
DgHBP-2 γ	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFHNA HSAVSDAAA AHHDAAGEYA QNSAFGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGVDD GLHREYAASA AQGANAVHDD AAQSHSALAG ANTFGHGYRP YAAIGHGVY HGPYGHGLY HGQYGHGPGY HGLYGHGAFG HGLNAYAPLV GHGLRGYL	17.8 kDa/6.1
DgHBP-2 δ	<i>MKFCLSVALL</i> <i>AVAMVIVNGQ</i> FLGAGPFNAA HSAVSDAAA AHHDAAGGFA EEAATGLLDT HHNENHDMTH DLADGYGLHE HDEQHHGVDD GLHREYAASA AQGANAVHDD AAQRHFASAA ANTFGHGFPGP YAAFHGVHGV PYGHGPGYGHG YHGFYGHGP YHGVYGHGV YHGHGLHAL YHGLRAPLV GHGAYGYL	18.5 kDa/6.0
DgHBP-2 ϵ	<i>MKFCLSVALL</i> <i>AVAMVIVNGQ</i> FLGAGPFNAA HSAVSDAAA AHHDATGGFA EEAATGLLDT HHNENHDMTH DLADGYGLHE HDEQHHGVDD GLHREYAASA AQGANAVHDD AAQRHSASAA ANTFGHGFPGP YAAFHGVHGV PYGHGPGYGHG YHGFYGHGP YHGVYGHGV YHGHGLHAL YHGLRAPLV GNAAYGYL	18.5 kDa/6.0
DgHBP-2 ζ	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFNTA HSAVSDAAA AHHDAAGEYA QNAATGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGLAD GLHQEYAARA AQGANAVHND AAQSHSALAA ANTFGHGHAP YAAIGHGVY HGPYGHGPGY HGLYGHGLY HGPYGHGLY HGAFHGLNA YAPLVGHGLR GYL	18.1 kDa/6.2
DgHBP-2 η	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFHNA HSAVSDAAA AHHDAAGEYA QNSAFGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGVDD GLHREYAASA AQGANAVHDD AAQSHSALAG ANTFGHGYRP YAAIGHGVY HGPYGHGLY HGQYGHGPGY HGLYGHGAFG HGLNAYAPLV GHGLRGYL	17.8 kDa/6.1
DgHBP-2 θ	<i>MKFCLSVALL</i> <i>VVAMVMVNGQ</i> FFGAGPFHNA HHAAGFGHGL YHGGYGHGL YHGGLAGHGL YHGGLSGHGL YHGGLGYGH GYAPFARHY	7.2 kDa/8.4
DgHBP-2 ι	<i>MKFCLSVALL</i> <i>AVAMVIVNGQ</i> FLGAGPFNAA HSAVSDAAA AHHDAAGGFA EEAATGLLDT HHNENHDMTH DLADGYGLHE HDEQHHGVDD GFHREYAASA AQGANAVHDD AAQRHFASAA ANTFGHGFPGP YAAFHGVHGV PYGHGPGYGHG YHGFYGHGP YHGVYGHGV YHGHGFHAL YHGLRAPLV GHGAYGYL	18.6 kDa/6.0
DgHBP-2 κ	<i>MKFCLSVALL</i> <i>AVAMVMVNGQ</i> FFGAGPFNSA HSAVSDAAA SNHDAATGFA EEAVSGLLDT HHNENHDMTH DLANGYGLHE HDEQHHGLAD GLHQEYAARA AQGANAVHND AAQSHSALAA ANTFGHGHAP YAAIGHGVY HGPYGHGPGY HGLYGHGLY HGPYGHGLY HGAFHGLNA YAPLVGHGLR GYL	18.0 kDa/6.1

Figure 2.12 Sequence analysis of DgHBPs. (a) Aligned consensus residue comparison (30 % ClustalO threshold among DgHBPs and their highly conserved consensus logo.^[46] (b) Tandem repeats of C-terminus domain sequences of DgHBPs.^[46] Underlined: fully conserved amino acids in the consensus sequences. Yellow highlighted: proposed cross-linking motifs GHG and HG.

The GHGXX'/GHGXY pentapeptide repeats in DgHBPs' sequences discussed above sparked the author's interest in studying the biophysical properties of DgHBPs. This is because in several other reported structural proteins, similar sequences were featured. For example, in tropoelastin, which is the precursor of elastin, hydrophobic pentapeptides sequence repeats enriched with Gly and Val are found. Details regarding such homology is further discussed in *Section 2.5.4*.

2.5 Coacervation & Self-coacervating Proteins

As mentioned in the end of *Sub-chapter 2.4*, the modular structure of DgHBPs' sequences drew considerable attention from the author and his colleagues, as similar design was discovered in other proteins (mussel foot protein and tropoelastin), both of which shared a liquid-liquid phase separation behavior from their initial aqueous solutions. This behavior, also known as "coacervation", constitutes the main hypothesis of research in this thesis, for that it is proposed that DgHBPs possess the same property to enable their permeation within the beak scaffold, thereby endowing the beak with composition and mechanical gradients. In this *Sub-chapter* the concept of "coacervation" ("complex coacervation" and "self-coacervation") is reviewed with specific examples. Then, sequence comparison analysis is presented between DgHBPs and proteins that exhibit self-coacervation.

2.5.1 Coacervation and Complex Coacervation

Interaction-induced phase behaviors of biopolymer-biopolymer-water solutions is of fundamental importance in biological systems, as reported in the case of DNA-replication enzymes and immunoglobulins-exocellular proteins/polysaccharides/virus systems, where they played functional roles in surface and colloid chemistry.^[58] In most cases, a mixture of biopolymers in solution is unstable, leading to two-types of phase separation. The first type, namely "segregative phase separation", originates from repulsive interactions, which causes two distinct phases, each enriched with one biopolymer.^[59] The second type, on the contrary, involves attractive interactions between biopolymers, causing a two-liquid-phase system where the two bio-polymers are concentrated in one phase ("coacervates phase"), and a biopolymer-depleted solution in the another phase ("aqueous phase"). This liquid-liquid phase separation is often referred to as "coacervation".^[58,60] In most studied cases, it is electrostatic attractions between two oppositely charged polymers that induced coacervation, which is thus named as "complex coacervation".^[60,61] In recent years, however, there has been increasing awareness that coacervation driven by other interactions (H-bonds, hydrophobic interactions, etc.) can occur without a counter poly-

electrolyte. This refers to “self-coacervation” and constitutes the focus of this thesis, as discussed in *Section 2.5.2*.

The term “complex coacervation” was first reported in the work on gelatin-gum arabic (GA) complex, a protein-polysaccharide system and the most extensively studied complex coacervates system to date.^[60,62–64] In gelatin-GA mixture, proteins and polysaccharides are oppositely charged, thereby attracted by Coulombic force. This attractive force dominates over the entropy effects that tend to disperse the molecules in solution, causing two polymers to associate with each other (**Figure 2.13b**) and leading to the reduction in the total free energy and to the spontaneous formation of protein-polysaccharide complex (the coacervates) in the form of spherical micro-droplets with 1 to 100 μm diameter (**Figure 2.13a**).^[58] It was later discovered that the formation and stability of (complex) coacervates is affected by many physicochemical factors, such as pH^[61], ionic strength^[65], biopolymer ratio^[66], charge density^[67,68], molecular mass^[69], temperature^[70,71] and mechanical forces^[72] (**Figure 2.13c**), yet an established theory regarding its mechanism and biological relevance still remain elusive.^[58] On the other hand, due to their specific structure, size and composition, the protein-polysaccharide coacervates exhibit a wide range of interesting functional properties, including hydration^[73], viscosity^[74], emulsification^[75], as well as foaming abilities^[76]. This has led to the wide applications of protein-polysaccharide coacervates in the food industry, one instance being microencapsulation of krill oil, as shown in **Figure 2.13d**.^[77]

A wide variety of complex coacervates systems other than protein-polysaccharide complex have also been reported, e.g. protein-protein^[78], protein-nucleic acid^[79], and nucleic acid-dendrimer^[80]. In the marine sandcastle worm (*Phragmatopoma californica*), protein-protein complex coacervation was reported as a delivery strategy for the underwater glue used by the animal to bind sediment particulates together. The worm secretes three highly polar proteins (Pc-1, Pc-2 and Pc-3) that undergo protein-protein complex coacervation and form a bioadhesive cement to glue sand grains (**Figure 2.14a**), in order to build tubular shelter for itself.^[81] Pc-1 and Pc-2 are basic (pI $\sim$ 10), Gly-rich proteins with the consensus repeats VGGYGYGGKK and HPAVXHKALGGYG, where X is a non-repeating residue,

and Y is Dopa (**Figure 2.14b**).^[81] Pc-3 is a Ser-rich protein (80 mol%, shown in **Figure 2.14b**) that is likely to play the role of binding calcium phosphate and hydroxyapatites.^[81] Upon co-secretion, Pc-1 and Pc-2 carry positive charge, whereas phosphorylated Pc-3 is dominated by negative charge, thereby forming coacervates that attaches to Ca^{2+} -abundant sand grains. Subsequent Dopa oxidation and Cys-Dopa cross-linking further solidifies the “tubular sand building” (**Figure 2.14c**).^[81]

Figure 2.13 Complex coacervation of protein-polysaccharide. (a) Microscopic image of gelatin-gum arabic coacervates.^[62] (b) Cartoon of gum arabic (white ribbon)- β -lactoglobulin (orange spheres) complex.^[60] The approximate diameter of proteins is 4 nm, whereas gum arabic has an average diameter of 50 nm. (c) Summary of physicochemical factors influencing complex coacervation.^[64] Scale bar: 50 μm . (d) Encapsulation of krill oil in gelatin-gum arabic coacervates, monitored by stereomicroscopy.^[77]

Figure 2.14 Complex coacervation by cement proteins of *P. californica*. **(a)** Micrograph of *P. californica* tube with globular sand beads (0.5 mm diameter).^[81] Right panel: zoom of cement plaques between two beads, as indicated by the arrow. **(b)** Sequences of Pc-1, Pc-2 and two variants of Pc-3.^[81] *Italic*: signal peptide. Underlined: repeated domains. **(c)** Schematic representation of the coacervation and cross-linking of *P. californica* cement proteins.^[81]

2.5.2 Self-coacervation: Mussel Foot Protein (Mfp-3S)

In stark contrast to complex coacervation, coacervation of single-biopolymer system (“self-coacervation”) is rarely reported, with only two known examples in nature: mussel foot protein (mfp-3S) and tropoelastin.

Figure 2.15 mfp-3S from *M. californianus* byssal plaques. (a) *M. californianus* mussel.^[82] Inset: Close view of adhesive protein residues secreted on the plaque-substrate interface and within the plaque bulk. (b) Representative sequence of mfp-3S with predicted molecular weight and isoelectric point.^[82] Red: positively charged residues. Blue: negatively charged residues. Green: aromatic residues. (c) Schematic representation of the Hydrogen bonds formed between Dopa and mica surface.^[83] (d) Hydropathy plot (Hopp and Woods^[84]) of mfp-3S as compared to another mussel foot protein, mfp-3f, highlighting the hydrophobicity for mfp-3S.^[83]

Mfp-3S is a protein extracted from the adhesive plaques of California mussels, *Mytilus californianus* (Figure 2.15a), which is a marine mussel known for its versatile, rapid and permanent underwater adhesion to diverse solid surfaces including woods, rocks, and vessels.^[82,83,85] As shown in Figure 2.15b, mfp-3S is a relatively small (molecular weight ~ 5 kDa), zwitterionic protein with a high content (25 mol%) of tyrosine (about half was

modified to Dopa) and strong hydrophobicity (**Figure 2.15d**).^[82,83] Amongst a large family of mussel foot proteins (mfps) found on the mussel foot plaques^[83], mfp-3S is proposed to have dual functions^[82,83]: binding to hydrophilic surfaces (e.g. mica, hydroxyapatite), as well as cohesion of adhesive proteins present in mussel plaques. The first role mainly attributes to hydrogen bonds (**Figure 2.15c**) formed by high content of Dopa in its sequence^[82,83], whereas the second ability come from its unique self-coacervation behavior^[82,83,85].

Under test conditions identical to the micro-environment of mussel foot, mfp-3S was found to undergo spontaneous self-coacervation process and form micro-droplets with 0.1 to 5 μm diameter (**Figure 2.16a**), demonstrating efficient phase-separation as well as surface spreading abilities.^[82] The self-coacervation of mfp-3S was found to be highly dependent on a number of parameters, including pH, ionic strength (**Figure 2.16b**) and temperature (**Figure 2.16c**), indicating a mixed self-coacervation mechanism.^[82] As shown in **Figure 2.16d**, in a classical polycation-polyanion complex (such as gelatin-gum arabic) system, the optimal coacervation occurs at a certain pH where the opposite charges carried by two polymers neutralize each other, causing zero net charge of the complex.^[82] For mfp-3S, however, maximal self-coacervation happens at pH (7.5) < pI (8.3), where the inter-molecular coulombic repulsions still exist, which implies the involvement of hydrophobic interactions during the protein coalescence process.^[82,83]

The self-coacervation of mfp-3S is believed to play crucial roles in the underwater adhesion behavior of *M. californianus*.^[82,83,85–87] First, it creates an insulated microenvironment from the surrounding aqueous solution, preventing Dopa-abundant proteins within the coacervates from premature oxidation, thereby maintaining the adhesive ability to substrates.^[82] Second, it holds the mfps together through electrostatic and hydrophobic interactions, thereby enhancing the mechanical integrity of byssal plaques, which reinforces the bonded adhesions.^[83] In summary, the self-coacervation of mfp-3S demonstrates a great example of nature-adapted strategy for underwater polymer protein delivery as well as interfacial adhesion, through delicate control over protein phase

separation behaviors, which provides molecular level bio-inspiration for dental, orthopedic and other biomedical applications.^[5,42,82,83,85–89]

Figure 2.16 Self-coacervation of mfp-3S from *M. californianus* byssal plaques. (a) Top panel: mfp-3S in coacervates state (right) compared with solution state (left).^[82] Bottom panel: light microscopic image of mfp-3S coacervates. (b) Coacervates-associated turbidity results of 0.1 mg/ml mfp-3S, indicating dependence on pH and ionic strength.^[82] Buffer 1: phosphate buffer. Buffer 2: acetate buffer. Buffer 3: phosphate buffer. (c) mfp-3S (0.1 mg/ml) coacervates turbidity change with temperature.^[82] (d) Schematic comparison between complex coacervates (gelatin-GA) and mfp-3S coacervates.^[82]

2.5.3 Self-coacervation: Tropoelastin

Tropoelastin self-coacervation is the most extensively studied one-component coacervation system so far. Tropoelastin (~ 60 to 70 kDa) is the soluble precursor protein of elastin, a highly insoluble protein that constitutes 90 % of elastic fibers in all vertebrates (present in organs, skin, tendons, and vessels, shown in **Figure 2.17a**), with a half-life of up to 70 years, able to undergo billions of extension-recoil cycles without mechanical failure.^[90–92] Tropoelastin matures into elastin through a process called “elastogenesis”, where tropoelastin was expressed, secreted, delivered, aligned and finally cross-linked and deposited onto the growing extracellular, filamentary elastin network.^[92]

Although there are some species variation among different vertebrates, elastin/tropoelastin generally contains over 30 % Gly, and about 75 % of the entire sequence is made up of hydrophobic residues (Gly, Val, Ala, and Pro).^[90] The amino acid sequences of tropoelastin from various sources showed significant homology, specifically the conservation of functionally distinct yet sequentially alternating hydrophobic and cross-linking domains.^[90–93] As shown in **Figure 2.18b**, human tropoelastin (~ 72 kDa) consists of 34 domains encoded by 36 exons in gene, among which exon 33 and 34 were lost during translation and exon 1 encodes signal peptide.^[90–92] In these 34 domains, the hydrophobic ones are comprised of repeats of non-polar amino acid peptides like GVGVP, GGVP, and GVGVP, and the spacing cross-linking domains contains stretches of lysine separated by alanine residues such as AAKAAKAA.^[90–92,94]

Self-coacervation of tropoelastin occurs when temperature of aqueous tropoelastin solutions was raised towards the physiological range.^[90–93] Below the transition temperature, tropoelastin exists as soluble monomers of about 15 nm diameter (**Figure 2.17c**).^[91] Upon heating, the aqueous solution rapidly becomes turbid with the formation of progressively growing 2 to 6 μm diameter micro-droplets (**Figure 2.17d**), which assembles into a distinct visco-elastic, high-concentrated liquid phase known as the coacervates phase.^[91] Besides being dominated by temperature, the self-coacervation of tropoelastin is susceptible to a variety of other factors, such as protein concentration, solution pH and

ionic strength (**Figure 2.18a** to **Figure 2.18d**).^[90–92,94–97] Interestingly, the optimal conditions (37°C, 150 mM NaCl, pH 7 to 8) for tropoelastin coacervation is identical to those found in the extracellular matrix, which suggests that coacervates phase is the favored form of tropoelastin after secretion.^[90–92]

Some controversies about the mechanism of tropoelastin self-coacervation exist, yet agreed upon by all theories is the dominant driving factor, *i.e.* hydrophobic interactions.^[90–92,94] As shown in **Figure 2.18e**, the triggering factor –temperature increase– is believed to induce the loss of clathrate-like water molecule shielding the hydrophobic regions in the proteins, thereby exposing them to associate with each other and leading to the ordered assembly of tropoelastin monomers.^[91,98–100] Therefore, at the molecular level, tropoelastin self-coacervation directly derives from its primary sequence, where hydrophobic domains, GVGVP, GGVP, and GVGVP drive coacervation.^[91,92,94,95,98]

The studies on tropoelastin self-coacervation are often linked to its conformation shift.^[90,91,94,99,101] It is proposed that associated with the tropoelastin coacervation is a transition of its secondary structure from a disordered random one into one with evidently improved helical, β -sheet and β -turn/spiral content.^[90,91,94,99,102]

Tropoelastin self-coacervation is of fundamental importance to the elastogenesis process and during elastic fiber formation. First, it delivers a large amount of protein monomers efficiently to the extracellular sites within a highly concentrated phase, allowing them to assemble with an increased intermolecular and intramolecular order, which is a prerequisite for a densely intertwined elastin network.^[90–92,94,98,103] Second, it aligns the monomer proteins with each other, closely positioning the Lys-rich domains by the formation of helical and β -structures, which enables Lys cross-linking catalyzed by lysyl oxidase, an essential step for elastin synthesis.^[90–92,94] Investigation studies on tropoelastin self-coacervation also provides sufficient experience and inspiration for other one-component coacervation systems in nature, in particular *D. gigas* beak proteins, as will be discussed throughout this thesis.

Figure 2.17 Elastin and tropoelastin. (a) Illustration of elastin fibers in aorta (left), lung (center) and ear cartilage (right).^[92] (b) Domain structure of human tropoelastin.^[91,104] Orange regions: hydrophobic domains. Light blue regions: cross-linking domains rich in lysine. (c) SEM graph of tropoelastin molecules (~ 15 nm) in the form of nanoparticle assemblies prior to coacervation.^[105] (d) SEM image of full-sized tropoelastin coacervates (~ 2 to 6 μm).^[105]

Figure 2.18 Influencing factors and mechanism of tropoelastin self-coacervation. (a) Turbidity-time plot of tropoelastin solution (40 mg/ml, 10 mM sodium phosphate, pH 7.4, 150 mM NaCl) at different temperatures.^[94] (b) Transition temperature-tropoelastin concentration plot, using temperatures required to achieve 50 % maximal coacervation at different pH values as an indicator of coacervation propensity.^[94] Same in (c) and (d). (c) Transition temperature-salt concentration plot.^[94] (d) Transition temperature-pH plot.^[94] (e) Schematic representation of tropoelastin reacting to increased temperature, showing the disruption of H-bonds between clathrate water molecules that shields the hydrophobic domains of tropoelastin due to heating, leaving the domains exposed.^[98]

2.5.4 Sequential Homology between DgHBPs and Self-coacervating Proteins

As one examines the sequences of DgHBPs, it is hard not to notice the striking similarities to both mfp-3S and tropoelastin.

For mfp-3S, the sequence is dominated by aromatic and hydrophobic residues, which leads to its coacervation behavior, as well as the Dopa-involved cross-links with various surfaces. Similarly in DgHBPs, the majority of sequence is also highly hydrophobic with the non-negligible existence of Tyr and Phe, which could be the source of catechol-His cross-links detected in the beak. As an underwater adhesive, mfp-3S functions as a binding agent to maintain the mechanical integrity of mussel byssal plaques, while self-coacervation fulfills the task of transporting and insulating highly concentrated proteins to the solid substrate in the form of micro-droplets. This made us wonder whether the *D. gigas* could utilize the same strategy: could it help transport DgHBPs in a high concentration within the chitin network prior to the final cross-linking?

For tropoelastin, the analogy is even more obvious. Besides the overall sequence hydrophobicity, the spacing and alternation pattern of functionally and sequentially distinct domains in tropoelastin is also to be found in DgHBP-1, where the GHGXX' domain alternates with GAGFA/GAGYA in DgHBP-1's sequence. In addition, the flanking of Gly on both sides of a bulky residue in tropoelastin GVGVP/GVGVP repeats is believed to make the middle residues available for interaction, which is also noticed in DgHBPs' GHG/GAG repeats. Further, the abundance of Ala and Pro in tropoelastin's cross-linking domains is proposed to favor the partially helical and β -turn conformation, which facilitates the alignment of Lys residues prior to lysine cross-linking. In DgHBP-1 and DgHBP-2, abundance of Ala and Pro is also noticed. Could this be a sign of similar conformation which might favor cross-linking in beak?

Another similarity of tropoelastin to the *D. gigas* beak is the existence of a micro-fibrillar scaffold, which likely provides an initial "skeleton" for the structure and subsequently guides the deposition of proteins within the inter-fibrillar spacing. In tropoelastin, this

scaffold is formed by structural glycoproteins (fibrillin-1, fibrillin-2, and glycoprotein-1).^[90–92] In *D. gigas* beak, the scaffold is proposed to be the micro-fibrillar network of chitin and DgBPs.

In summary, the sequence similarities between DgHBPs with two self-coacervating proteins in Nature, as well as the parallel contextual and functional requirements for the respective materials begs the question as whether self-coacervation also occurs for DgHBPs, as well as the possible significance of self-coacervation during *D. gigas* beak bio-processing *in vivo*.

2.6 Scope of Project

As a wholly organic composite material in nature, *D. gigas* beak presents an effective way to cushion the mechanical mismatch between hard biotool and soft organism tissue, which is done by tuning the 4 components (chitin, proteins, pigments, water) in its composition to create a stiffness gradient of the material. Among these 4 components, the function of proteins (DgBPs) remains most elusive yet simultaneously intriguing. Preliminary identification work of DgBPs revealed two families of proteins carrying different functions: DgCBPs and DgHBPs. The former is hypothesized to mainly bind to chitin and fixes the beak scaffold, whereas role of the latter is unclear. By comparing the sequences of DgHBPs to other structural proteins in Nature, the propensity for DgHBPs to be involved in the biophysical process called “coacervation” is noticed, which was reported to have profound significance in the formation of many other biological structures, especially sclerotized and adhesive ones.

Hence, the scope of this project is to conduct comprehensive studies on the DgHBPs (DgHBP-1 and DgHBP-2) found in *D. gigas* beak. Specifically, the author aims to explore the biophysical properties and molecular interactions of DgHBPs with a focus on their self-coacervation behavior, as well as its implications to the mechanical gradient of *D. gigas* beak. Thereby, a deeper understanding of the roles of proteins in *D. gigas* beak sclerotization could be established, which will provide inspiration for designing

structurally, mechanically and functionally biomimetic materials in various applications such as prosthetics or tissue engineering.

References

- [1] L. Jiang, Y. Zhao, J. Zhai. *Angew. Chemie* **2004**, 116, 4438–4441.
- [2] K. Liu, L. Jiang. *Nano Today* **2011**, 6, 155–175.
- [3] B. Dean, B. Bhushan. *Philos. Trans. A. Math. Phys. Eng. Sci.* **2010**, 368, 4775–806.
- [4] M. Phelps, I. De, L. Benko. *Science (80-.)*. **2004**, 305, 636–637.
- [5] J. Aizenberg, P. Fratzl. *Adv. Mater.* **2009**, 21, 187–188.
- [6] K. N. Nesis. *Cephalop. life cycles* **1983**, 1, 215–231.
- [7] C. M. Nigmatullin, K. N. Nesis, A. I. Arkhipkin. *Fish. Res.* **2001**, 54, 9–19.
- [8] Y. A. Aleksandronets, Y. I. Magaras, C. M. Nigmatullin. *Syst. Ecol. Cephalopods. Zool. Inst. Acad. Sci. USSR, Leningr.* **1983**, 99–102.
- [9] N. M. Ehrhardt, P. S. Jacquemin, F. Garcia-B, G. Gonzales-D, J. M. Lopez, J. Ortiz-C, A. Solis-N. *Mem. Natl. Museum Victoria* **1983**, 44, 305–311.
- [10] U. Markaida. *Fish. Res.* **2006**, 79, 16–27.
- [11] N. M. Ehrhardt. *Bull. Mar. Sci.* **1991**, 49, 325–332.
- [12] S. H. Hiew, A. Miserez. *ACS Biomater. Sci. Eng.* **2016**, acsbiomaterials.6b00284.
- [13] A. Miserez, J. C. Weaver, P. B. Pedersen, T. Schneeberk, R. T. Hanlon, D. Kisailus, H. Birkedal. *Adv. Mater.* **2009**, 21, 401–406.
- [14] P. A. Guerette, S. Hoon, Y. Seow, M. Raida, A. Masic, et al. *Nat. Biotechnol.* **2013**, 31, 908–15.
- [15] P. A. Guerette, X. S. Hoon, X. D. Ding, S. Amini, A. Masic, et al. *ACS Nano* **2014**, 8, 7170–7179.
- [16] D. Ding, P. A. Guerette, S. Hoon, K. W. Kong, T. Cornvik, M. Nilsson, A. Kumar, J. Lescar, A. Miserez. *Biomacromolecules* **2014**, 15, 3278–3289.
- [17] R. V. Lewis. *Chem. Rev.* **2006**, 106, 3762–3774.
- [18] J. Gatesy. *Science (80-.)*. **2001**, 291, 2603–2605.
- [19] P. A. Guerette, D. G. Ginzing, B. H. F. Weber, J. M. Gosline. *Science (80-.)*. **1996**, 272, 112.
- [20] V. Latza, P. A. Guerette, D. Ding, S. Amini, A. Kumar, et al. *Nat. Commun.* **2015**, 6,

8313.

- [21] A. Miserez, Y. Li, J. H. Waite, F. Zok. *Acta Biomater.* **2007**, 3, 139–149.
- [22] D. J. Rubin, A. Miserez, J. H. Waite, C. Jérôme, J. S. Stephen. in *Adv. In Insect Phys.*, **2010**, pp. 75–133.
- [23] A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite. *Science* **2008**, 319, 1816–1819.
- [24] R. A. A. Muzzarelli. *Chitin*, Elsevier, **2013**.
- [25] L. E. Arnow, B. Y. L. E. Arnow. *J. Biol. Chem* **1937**, 118, 531–537.
- [26] A. Miserez, D. Rubin, J. H. Waite. *J. Biol. Chem.* **2010**, 285, 38115–38124.
- [27] S. Mann. *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*, Oxford University Press On Demand, **2001**.
- [28] S. Weiner, P. Zaslansky. in *Learn. from Nat. How to Des. New Implant. Biomater. From Biominer. Fundam. to Biomim. Mater. Process. Routes Proc. NATO Adv. Study Institute, Held Alvor, Algarve, Port. 13--24 Oct. 2003* (Eds.: R.L. Reis, S. Weiner), Springer Netherlands, Dordrecht, **2005**, pp. 3–13.
- [29] L. Addadi, D. Joester, F. Nudelman, S. Weiner. *Chem. Eur. J.* **2006**, 12, 980–987.
- [30] A. J. Kear. *J. Mar. Biol. Assoc. United Kingdom* **1994**, 74, 801–822.
- [31] T. A. Uyeno, W. M. Kier. *J. Morphol.* **2005**, 264, 211–222.
- [32] P. N. Dilly, M. Nixon. *Cell Tissue Res.* **1976**, 167, 229–241.
- [33] B. R. Lawn. *J. Am. Ceram. Soc.* **1998**, 81, 1977–1994.
- [34] H. C. Lichtenegger, T. Scho, M. H. Bartl, H. Waite, G. D. Stucky, T. Schöberl, M. H. Bartl, H. Waite, G. D. Stucky. *Science (80-.)*. **2002**, 298, 389–392.
- [35] J. F. V Vincent, S. Ablett. *J. Insect Physiol.* **1987**, 33, 973–979.
- [36] T. Nishino, R. Matsui, K. Nakamae. *J. Polym. Sci. Part B Polym. Phys.* **1999**, 37, 1191–1196.
- [37] T. Wu, S. Zivanovic, F. A. Draughon, W. S. Conway, C. E. Sams. *J. Agric. Food Chem.* **2005**, 53, 3888–3894.
- [38] C. Giulivi, N. J. Traaseth, K. J. A. Davies. *Amino Acids* **2003**, 25, 227–232.
- [39] O. Zvarec, S. Purushotham, A. Masic, R. V. Ramanujan, A. Miserez. *Langmuir* **2013**, 29, 10899–10906.
- [40] K. Kim, J. H. Ryu, D. Y. Lee, H. Lee. *Biomater. Sci.* **2013**, 1, 783.

- [41] M. J. Sever, J. T. Weisser, J. Monahan, S. Srinivasan, J. J. Wilker. *Angew. Chemie Int. Ed.* **2004**, 43, 448–450.
- [42] N. Holten-Andersen, T. E. Mates, M. S. Toprak, G. D. Stucky, F. W. Zok, J. H. Waite. *Langmuir* **2009**, 25, 3323–3326.
- [43] J. D. Fox, J. R. Capadona, P. D. Marasco, S. J. Rowan. *J. Am. Chem. Soc.* **2013**, 135, 5167–5174.
- [44] B. J. Smith, J. M. Walker. *The Protein Protocols Handbook*, Humana Press Inc.: Totowa, **2002**.
- [45] D. L. Crimmins, S. M. Mische, N. D. Denslow. *Curr. Protoc. Protein Sci.* **2005**, Chapter 11, Unit 11.4.
- [46] Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez, J. H. Waite. *Nat. Chem. Biol.* **2015**, 11, 488–495.
- [47] H. Schägger. *Nat. Protoc.* **2006**, 1, 16–22.
- [48] D. Sahal, R. Kannan, A. Sinha, V. Babbarwal, B. Gnana Prakash, G. Singh, V. . Chauhan. *Anal. Biochem.* **2002**, 308, 405–408.
- [49] M. A. Paz, R. Flückiger, A. Boak, H. M. Kagan, P. M. Gallop, et al. *J. Biol. Chem.* **1991**, 266, 689–692.
- [50] J. M. Walker. *Protein Protocols Handbook*, Humana Press, New Jersey, **2002**.
- [51] P. Edman, E. Högfeldt, L. G. Sillén, P.-O. Kinell. *Acta Chem. Scand.* **1950**, 4, 283–293.
- [52] Z. Wang, M. Gerstein, M. Snyder. *Nat. Rev. Genet.* **2009**, 10, 57–63.
- [53] Y. Chu, D. R. Corey. *Nucleic Acid Ther.* **2012**, 22, 271–4.
- [54] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, et al. *Nat. Biotechnol.* **2011**, 29, 644–652.
- [55] M. A. Frohman, M. K. Dush, G. R. Martin. *Proc. Natl. Acad. Sci.* **1988**, 85, 8998–9002.
- [56] F. Sanger, S. Nicklen, a R. Coulson. *Proc. Natl. Acad. Sci.* **1977**, 74, 5463–5467.
- [57] E. de Castro, C. J. A. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, N. Hulo. *Nucleic Acids Res.* **2006**, 34, W362–W365.
- [58] C. Schmitt, C. Sanchez, S. Desobry-Banon, J. Hardy. *Crit. Rev. Food Sci. Nutr.* **1998**, 38, 689–753.

- [59] V.B.Tolstoguzov. *Food Hydrocoll.* **1991**, 4, 429–468.
- [60] C. G. De Kruif, F. Weinbreck, R. De Vries. *Curr. Opin. Colloid Interface Sci.* **2004**, 9, 340–349.
- [61] F. W. Tiebackx. *Zeitschrift für Chemie und Ind. der Kolloide* **1911**, 8, 198–201.
- [62] H. Bungenberg de Jong, HG and Kruyt. *Proc K. Ned. Akad. Wet.* **1929**, 32, 849–856.
- [63] H. G. de Jong. *Colloid Sci.* **1949**, 2, 335–432.
- [64] C. Schmitt, S. L. Turgeon. *Adv. Colloid Interface Sci.* **2011**, 167, 63–70.
- [65] D. J. Burgess, J. E. Carless. *J. Colloid Interface Sci.* **1984**, 98, 1–8.
- [66] K. W. Mattison, I. J. Brittain, P. L. Dubin. *Biotechnol. Prog.* **1995**, 11, 632–637.
- [67] H. Noguchi. *Biochim. Biophys. Acta* **1956**, 22, 459–462.
- [68] H. Noguchi. *J. Phys. Chem.* **1960**, 64, 185–187.
- [69] J. Shieh, C. E. Glatz. in *Macromol. Complexes Chem. Biol.*, Springer Berlin Heidelberg, Berlin, Heidelberg, **1994**, pp. 273–284.
- [70] R. Kelly, E. S. Gudo, J. R. Mitchell, S. E. Harding. *Carbohydr. Polym.* **1994**, 23, 115–120.
- [71] S. Le Henaff, S. L. Turgeon, C. Sanchez, P. Paquin. *J. Dairy Sci.* **1997**, 80, 103.
- [72] S. Tirkkonen, L. Turakka, P. Paronen. *J. Microencapsul.* **1994**, 11, 615–626.
- [73] P. M. T. Hansen, J. Hidalgo, I. A. Gould. *J. Dairy Sci.* **1971**, 54, 830–834.
- [74] V. E. Sánchez, G. B. Bartholomai, a M. R. Pilosof. *LWT-Food Sci. Technol.* **1995**, 28, 380–385.
- [75] S. Suzuki, T. Kondo. *Colloids and Surfaces* **1982**, 4, 163–171.
- [76] P. M. T. Hansen, D. H. Black. *J. Food Sci.* **1972**, 37, 452–456.
- [77] S. Aziz, J. Gill, P. Dutilleul, R. Neufeld, S. Kermasha. *J. Microencapsul.* **2014**, 31, 774–784.
- [78] M. Brouwer, R. Cashon, J. Bonaventura. *Biotechnol. Bioeng.* **1990**, 35, 831–836.
- [79] K. Rawat, V. K. Aswal, H. B. Bohidar. *J. Phys. Chem. B* **2012**, 116, 14805–14816.
- [80] X. Liu, C. Liu, C. V. Catapano, L. Peng, J. Zhou, P. Rocchi. *Biotechnol. Adv.* **2014**, 32, 844–852.
- [81] H. Zhao, C. Sun, R. J. Stewart, J. H. Waite. *J. Biol. Chem.* **2005**, 280, 42938–42944.
- [82] W. Wei, Y. Tan, N. R. Martinez Rodriguez, J. Yu, J. N. Israelachvili, J. H. Waite. *Acta Biomater.* **2014**, 10, 1663–1670.

- [83] W. Wei, J. Yu, C. Broomell, J. N. Israelachvili, J. H. Waite. *J. Am. Chem. Soc.* **2013**, 135, 377–383.
- [84] T. P. Hopp, K. R. Woods. *Mol. Immunol.* **1983**, 20, 483–489.
- [85] W. Wei, L. Petrone, Y. Tan, H. Cai, J. N. Israelachvili, A. Miserez, J. H. Waite. *Adv. Funct. Mater.* **2016**, 26, 3496–3507.
- [86] L. Petrone, A. Kumar, C. N. Sutar, N. J. Patil, S. Kannan, A. Palaniappan, S. Amini, B. Zappone, C. Verma, A. Miserez. *Nat. Commun.* **2015**, 6, 8737.
- [87] D. S. Hwang, H. Zeng, A. Srivastava, D. V. Krogstad, M. Tirrell, J. N. Israelachvili, J. H. Waite. *Soft Matter* **2010**, 6, 3232.
- [88] J. Sedó, J. Saiz-Poseu, F. Busqué, D. Ruiz-Molina. *Adv. Mater.* **2013**, 25, 653–701.
- [89] S. Kim, J. Huang, Y. Lee, S. Dutta, H. Y. Yoo, Y. M. Jung, Y. Jho, H. Zeng, D. S. Hwang. *Proc. Natl. Acad. Sci.* **2016**, 113, E847–E853.
- [90] B. Vrhovski, A. S. Weiss. *Eur. J. Biochem.* **1998**, 258, 1–18.
- [91] G. C. Yeo, F. W. Keeley, A. S. Weiss. *Adv. Colloid Interface Sci.* **2011**, 167, 94–103.
- [92] S. M. Mithieux, A. S. Weiss. in *Fibrous Proteins: Coiled-Coils, Collagen and Elastomers*, Academic Press, **2005**, pp. 437–461.
- [93] M. Miao, E. Sitarz, C. M. Bellingham, E. Won, L. D. Muiznieks, F. W. Keeley. *Biopolymers* **2013**, 99, 392–407.
- [94] B. Vrhovski, S. Jensen, A. S. Weiss. *Eur. J. Biochem.* **1997**, 250, 92–98.
- [95] S. A. Jensen, B. Vrhovski, A. S. Weiss. *J. Biol. Chem.* **2000**, 275, 28449–28454.
- [96] B. A. Cox, B. C. Starcher, D. W. Urry. *J. Biol. Chem.* **1974**, 249, 997–998.
- [97] L. D. Muiznieks, S. E. Reichheld, E. E. Sitarz, M. Miao, F. W. Keeley. *Biopolymers* **2015**, 103, 563–573.
- [98] L. D. Muiznieks, S. A. Jensen, A. S. Weiss. *Arch. Biochem. Biophys.* **2003**, 410, 317–323.
- [99] A. M. Tamburro, V. Guantieri, D. Daga-Gordini, G. Abatangelo. *J. Biol. Chem.* **1978**, 253, 2893–2894.
- [100] D. W. Urry, T. L. Trapane, K. U. Prasad. *Biopolymers* **1985**, 24, 2345–2356.
- [101] D. W. Urry. in *Methods Enzymol.*, **1982**, pp. 673–716.
- [102] D. W. Urry, M. M. Long, E. Gross. *CRC Crit. Rev. Biochem.* **1976**, 4, 1–45.
- [103] S. G. Wise, G. C. Yeo, M. A. Hiob, J. Rnjak-Kovacina, D. L. Kaplan, M. K. C. Ng,

A. S. Weiss. *Acta Biomater.* **2014**, 10, 1532–1541.

[104] A. W. Clarke, S. G. Wise, S. A. Cain, C. M. Kielty, A. S. Weiss. *Biochemistry* **2005**, 44, 10271–10281.

[105] Y. Tu, S. G. Wise, A. S. Weiss. *Micron* **2010**, 41, 268–272.

Chapter 3

Experimental Methodology

A wide range of experimental approaches have been employed to prepare D. gigas Histidine-rich Beak Proteins (DgHBPs) and investigate their biophysical, structural and mechanical properties, as well as to prepare and investigate DgHBPs-based peptides and variants. These techniques include: (i) recombinant protein/peptide expression and purification, which lays the ground for subsequent coacervation studies; (ii) various structural characterization methods (circular dichroism, Fourier transform infrared spectroscopy, etc.), to study proteins/peptides' structures in different forms; (iii) mechanical and biochemical characterization methods, to study proteins/peptides in their coacervates form. In this chapter, all sample preparation methods and experimental techniques including their general theoretical principles will be introduced.

3.1 Rationale for Selection

A range of synthesis, sample preparation and characterization methods have been used in the project. In order to conduct thorough investigation on DgHBPs, sufficient amount of (*rec*)DgHBPs were produced through recombinant protein expression method, together with downstream purification techniques. Subsequently, structural characterization methods were applied on DgHBPs at solution state. In addition, thermal, mechanical and biochemical characterization methods were used to study the special form of DgHBPs – coacervates.

In this chapter, these methods will be described in detail and presented as six main sections: (i) Cloning and expression of recombinant DgHBPs; (ii) Purification of DgHBPs; (iii) Self-coacervation of DgHBPs; (iv) Characterization of DgHBPs at solid and solution state; (v) Characterization of DgHBPs at coacervates state; and (vi) Preparation and characterization of DgHBPs-based peptides. Details regarding sample preparation and data analysis will be described within each section, according to different characterization technique employed.

3.2 Cloning and Expression of Recombinant DgHBPs

Recombinant protein expression is a widely used technology in the field of molecular biology.^[1] Along with further scale-up expression, it enables massive production of target proteins for subsequent studies such as molecular and structural analysis, compounds production in pharmaceutical industry, as well as biomaterials fabrication.^[2]

There are various choices of host in a recombinant expression system, such as prokaryotes, yeasts and filamentous fungi, as well as insect and mammalian cells, among which the *E. coli* bacterial system is most common used, because of ease of handling, short generation time, and low cost.^[3] In a typical *E. coli* recombinant expression system, the selected gene is first inserted in a plasmid expression vector, which is generally a larger sequence of DNA that transfers the inserted genes to bacterial cells to multiply and express themselves. Apart from the inserted gene, a plasmid usually also contains an antibiotic-resistant (e.g.

ampicillin-resistant) gene which enables the subsequent bacteria selection process with the corresponding antibiotics.^[2] A plasmid also has a promoter region, which enables the binding of the RNA polymerase and transcription factor, thereby facilitating the induction of the targeted gene.^[4] Various plasmids are commercially available nowadays for different expression systems, for the specific needs (antibiotic type, promoter type, etc.) of each project, customized plasmids are also available.^[5]

3.2.1 Plasmid Preparation

In this work, an *E. coli* BL-21(DE3) cell system was used to carry out the recombinant expression of DgHBP-1/2. A pJ express 414 plasmid ampicillin-resistant vector containing full-length sequence insert of DgHBP-1 and -2 (without signal peptide; with M as starting residue) and a T7 promoter provided by DNA 2.0 (Menlo Park, CA, USA). The plasmid vector was first transformed into cells, spread onto LB plates containing with 100 µg/ml ampicillin, and then incubated at 37 °C overnight. The following day mini-prep cultures were prepared, by inoculating individual colonies into 4 ml LB medium supplemented with ampicillin (LB-amp). The cultures were then incubated overnight at 37 °C in a shaking incubator (250 rpm).

The following day glycerol stocks were prepared (50 µL of autoclaved 50 % Glycerol in each tube and add 100 µL of bacterial cultures), and the remaining volume was used for plasmid isolation.

The plasmids were purified using QIAprep Spin Miniprep kit according to instructions, and then transformed into *E. coli* BL-21(DE3) cells. Glycerol stocks were established as previously described.

3.2.2 Protein Expression

a. Pre-cultures

E. coli BL-21(DE3) cells containing the plasmids with DgHBP-1 or DgHBP-2 genes were

stored in glycerol stocks. 10 μ L of cells from this glycerol stock were then inoculated into selective LB-amp media (4 ml each tube of culture) for overnight shaking at 37 °C.

b. Dilution of pre-cultures

The next day 125-fold dilution of the stationary phase cell cultures were done by transferring precultures into 2 L Erlenmeyer flasks containing 500 ml main LB-ampicilin medium, which were cultivated under the same conditions as above until an OD600 value of 0.8 was obtained. Expression of protein was induced with 0.5 mM IPTG. Protein was expressed for 4 hours Cells were harvested by centrifugation at 10,000 rpm, 15 min, and 4 °C. The cell pellets, which contains the target protein were frozen at -20 °C for further usage.

3.3 Purification of Recombinant DgHBPs

A schematic representation of DgHBPs expression and purification steps is shown in Figure 3.1, summarizing the main procedures involved in the process, and also the differences between the purification of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 after cell lysis, the details to which shall be elaborated in *Chapter 4*.

3.3.1 Lyses of Cells by Microfluidization

Microfluidization is an emerging method to lyse cells in protein expression.^[6] By high pressure and shear rate, cells entering the micro-channels in an microfluidization instrument would break, and therefore release proteins.^[6] Compared to traditional methods such as sonication and bead-beating homogenization, microfluidization demonstrates higher efficiency and better recovery of proteins.^[7–10]

Here, the harvested cell pellets were re-suspended in lysis buffer (50 mM Tris, 200 mM NaCl, 1 mM PMSF, pH 7.4) and lysed using microfluidizer (M-110P model, Microfluidics Corp., USA). Cells were lysed under 25000 psi pressure, and the procedure was repeated for 6 times. Then, cell lysates were centrifuged at 12,000 rpm at 4 °C for 10 min. Small

aliquots (10 μ l) of the supernatant (containing soluble proteins), and insoluble fractions (pellets) were collected for SDS-PAGE to confirm protein expression.

3.3.2 Primary Protein Extraction by Acetic Acid

a. Dialysis (DgHBP-1)

Lysed cells contain released target protein, buffer salts, as well as cell debris, therefore dialysis method was applied to remove impurities. The supernatant containing soluble protein was dialyzed against 5 % acetic acid, and the dialysate was lyophilized.

Figure 3.1 Schematic representation of recombinant expression and subsequent purification procedures for DgHBP-1 (red) and DgHBP-2 (blue).

b. Extraction from inclusion bodies (DgHBP-2)

Pellets were first washed twice with wash buffer 1 (100 mM Tris, 5 mM EDTA, 2 M urea, 2 vol% Triton X-100, 5 mM DTT, pH 7.4) to cleave disulfide bonds and solubilize insoluble impurities such as cell membranes, and then washed twice with wash buffer 2 (100 mM Tris, 5 mM EDTA, 5 mM DTT, pH 7.4) to solubilize soluble impurities. Inclusion bodies were collected after centrifugation (5,000 g for 15 min at 4 °C) between each washing steps, and subsequently washed in 5 % acetic acid, 20 % acetic acid and 5 % acetic acid with 8 M urea to retrieve maximum amount of target protein.

3.3.3 Size Exclusion Chromatography

SEC (Size Exclusion Chromatography) is a chromatographical method used to separate different biological macromolecules based on their size (molecular weight).^[11] During SEC a mixture of biomolecules is passed through a column that contain beads usually made of agarose which serve as a matrix to separate molecules.

The beads contain pores of different sizes and provide a separation range.^[12] In a typical SEC column run, smaller bio-macromolecules will be trapped in the pores, while larger molecules easily pass through column and elute first.^[12–14] Fraction containing different sizes of molecules can be collected during the run.

FPLC (Fast Protein Liquid Chromatography) is a form of liquid chromatography system on which various types of columns are installed to separate proteins, including SEC column.^[15] In FPLC system, a buffer acts as mobile phase, of which the flow rate is controlled by a positive-displacement pump. A column acts as stationary phase.

Here, in order to purify the proteins after cell lyses and initial extraction (described, point 3.3.2), an AKTA-purifier FPLC system (GE Healthcare, Life Sciences) equipped with SEC column (Superose 6 10/300 gel filtration column) was applied. The SEC chromatography was carried out using SEC Buffer (20 mM Tris-HCl, 150 mM NaCl, 50 mM EDTA, 1 mM dithiotreitol), at flow rate of 0.4 ml/min. Sample was dissolved in SEC Buffer at 5 mg/ml,

then went into the purification by SEC-FPLC. UV detector was set as $\lambda = 280$ nm, at which abundant aromatic amino acids (tyrosine and histidine) in DgHBP-1 or DgHBP-2 show distinctive peaks on the chromatogram. Collection of 1 ml fractions was conducted during the whole filtration run.

3.3.4 Reverse Phase Chromatography

RPC (Reverse Phase Chromatography) is a chromatographic method that uses a hydrophobic stationary phase to separate biological macromolecules in solutions based on their hydrophobicity and polarity.^[16] In RPC columns, the non-polar carbon-chain functionalized resins (e.g. C4, C8, or C18 resins) function as hydrophobic stationary phases, with a strong affinity to hydrophobic molecules in the mobile phase.^[17] Hydrophobic molecules in the mobile phase tend to bind to the hydrophobic stationary phase, where hydrophilic molecules in the mobile phase pass through the column and elute first. By decreasing the mobile phase's polarity using an organic solvent (usually acetonitrile), hydrophobic molecules elute later. The gradient of polarity decrease could be set at different liquid chromatography systems.

Here, RPC was used to further purify the proteins after Crude protein extraction (described, point 3.3.2). The experiments were performed using an Agilent Infinity 1260 HPLC system (High Performance Liquid Chromatography). Two RPC columns were used, Agilent Zorbax 300SB-C8 (semi-preparative, pore size 300 Å, resin diameter 5 µm) and Kromasil 300SB-C8 (preparative, pore size 300 Å, resin diameter 5 µm). Semi-preparative Agilent Zorbax 300SB-C8 was used for initial test runs and small-scale purifications; Preparative Kromasil 300SB-C8 was used for established, larger scale purifications.

The sample was dissolved in 10 mM acetic acid at 5-10 mg/ml, then went into the purification by RPC-HPLC. UV detector was set as $\lambda = 210$ nm, 220 nm, 254 nm and 280 nm. A gradient of Acetonitrile (ACN) and water was used to elute the protein from the column, programed as followed, where collection of 1.5 ml fractions was conducted during 10-60 % ACN:

0-4 min, 10 % ACN

4-7 min, 10-30 % ACN

7-25 min, 30-60 % ACN

25-30 min, 60-100 % ACN

30-35 min, 100 % ACN

35-40 min, 100-10 % ACN

3.4 Self-coacervation of DgHBPs

For DgHBP-1, 10 mg/ml (0.65 mM) stock solutions were prepared in 10 mM acetic acid buffer (pH 2.5). Coacervates phase was created by adding stock solution to phosphate buffer (60 mM Na_2HPO_4 , ionic strength = 100 mM, pH = 6.5) at a volume ratio of 1:9 (stock : buffer). Two types of coacervates phases were assessed. The first type was prepared as following: upon contact with the buffer, a visible cloudy phase was floating within the buffer. By pipetting, about 50-80 % of the surround buffer was removed from the system, leaving only the cloudy phase, herein defined as the “condensed coacervates” phase. For the second type, instead of being separated, the “condensed coacervates” phase was vigorously mixed with the buffer by pipetting, leading to a uniform dispersion of coacervates droplets, defined as the “dispersed coacervates phase”.

For DgHBP-2, 11 mg/ml stock (0.65 mM) solution were prepared in 10 mM acetic acid buffer (pH 2.5). Coacervates phase was created by adding stock solution to 0.06 M phosphate buffer (Na_2HPO_4 , ionic strength = 0.1 M, pH = 6.5) at a volume ratio of 1:9 (stock : buffer). Only one type of coacervates (“dispersed coacervates”) phase were assessed, prepared in the same way as DgHBP-1.

For turbidity and SAXS measurement of coacervates at different buffer pH, the coacervation buffers used were as below:

pH 4-5.5, Acetate buffer (60 mM NaAc, 100, 500 and 1000 mM ionic strength)

pH 6, MES buffer (60 mM MES, 100, 500 and 1000 mM ionic strength)

pH 6.5-8, Phosphate buffer (60 mM Na_2HPO_4 , 100, 500 and 1000 mM ionic strength)

pH 9, Tris buffer (60 mM Tris, 100, 500 and 1000 mM ionic strength)

pH 10, CAPS (60 mM CAPS, 100, 500 and 1000 mM ionic strength)

3.5 Characterization of DgHBPs at Solid and Solution State

In order to analyze the DgHBPs during and after expression and purification, different characterization techniques were applied. SDS-PAGE and MALDI-TOF were used to verify the purity and molecular weight of DgHBPs. FTIR measurements were conducted to determine the dominant secondary structure of DgHBPs in lyophilized solid state. CD measurements were conducted to determine the secondary structure of DgHBPs in solution state at different conditions. Combining them all provides a clearer picture of DgHBPs under different conditions in terms of structure, and also reveal the difference between DgHBP1/2.

3.5.1 SDS-PAGE

SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) is a most commonly used technique in molecular biology and protein chemistry. It could reveal the molecular weight or a sample containing single or multiple proteins (or peptides), as well as separate them and assess the purity of a certain protein.^[18]

The process of a SDS-PAGE run, during which protein molecules passes through the polyacrylamide gel, is described as followed:

- a.** A detergent (usually sodium dodecyl sulfate (SDS)) denatures proteins and imparts a surface charge to them.
- b.** Further denaturation can be achieved by adding reducing agent such as DTT (to cleave disulfide bonds) and heating the sample.
- c.** Denatured sample is loaded onto the gel, voltage is applied and the sample migrates

through the gel network.

In SDS-PAGE, the distribution of charge per unit mass is even, so the mobility of protein molecules would relate only to its own molecular mass.^[19] Molecules with a greater molecular weights would therefore migrate slower than those with smaller ones, causing separation of protein bands. Protein or protein mixture with known molecular weight is often used as a reference.

Here, SDS-PAGE was applied to examine the molecular weight and verify the purity of DgHBPs after purification. Protein samples were mixed with Laemmli buffer (Bio-Rad, California, USA) at 1:1 ratio and DTT (final concentration 5 mM) (Bio-Rad), followed by heating at 90 °C for 3 min. Then samples were loaded SDS-PAGE gel (4 % stacking gel and 12 % resolving gel), which was then put on constant electrical charge at 70 V for 20 min, then 220 V for 50 minutes. The gel was then stained in Coomassie Blue 250 staining buffer (CBR-250, AppliChem, Germany), then destained in 5 % AA, 7 % methanol.

3.5.2 MALDI-ToF

MALDI-ToF (Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry) mass spectrometry is a technique frequently used in biochemistry to determine the accurate molecular weight of a protein sample.^[20] In an MALDI-ToF measurement, sample containing protein(s) is first mixed with a matrix (e.g. 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid), α -cyano-4-hydroxycinnamic acid (CHCA, alpha-cyano or alpha-matrix) and 2,5-dihydroxybenzoic acid (DHB)) that serves as a proton donor. The mixture are then put on a metal plate and left to dry. Upon installation of the plate into the instrument, a laser initiate the ionization, and charged molecules would be further accelerated through the electric field, and go to the mass detector.^[21] Two detection modes could be chosen, namely linear mode and reflector mode. In linear mode, the ions travel down a linear flight path and their mass/charge (m/z) ratio is determined by their flying time.^[22] The mass of a sample X is measured as $[X+H]^+$ due to protonization. In reflector mode, an ion mirror at its end which reflects the ions back to

a detector. Reflector mode permits limited mass spectrometric sequencing and a higher mass accuracy.

In this project, 1 μ l of 1 mg/ml (*rec*)DgHBP(1/2) dissolved in 5 % acetic acid was mixed with 1 μ l of saturated solution of sinapinic acid (50 % MilliQ-water, 50 % ACN and 0.1 % TFA). The sample was then pipetted to the MALDI plate and left for 30 min to dry prior to analysis. Experiments were carried out on a Kratos Axima TOF2 instrument (Kratos-Shimadzu Biotech). The N₂ laser was set at 337 nm bandwidth and 4 ns pulse width. 20 kV voltage (accelerating mode) was used, spectra recording was set at linear mode, and 100 laser shots was done at each sample position at 100-120 W power.

3.5.3 Fourier Transform Infrared Spectroscopy

Fourier Transform Infrared (FTIR) spectroscopy is a commonly used technique to analyze the infrared absorption spectrum of organic and biological compounds. In FTIR, different molecules absorb infra-red wavelength region lights in its unique pattern, due to different molecular structures.^[23] This is because different intra-molecular bonds have different vibration or oscillation mode towards light, and combining them all could provide an “fingerprint” of an molecule. Absorbance spectrum recorded in FTIR can help identify sample, by relating signature peaks to certain bonds (a process also named “peak assignment”).^[24] Covalent bond peaks usually have most strong absorbance and best signal, therefore often used as standards for peak assignment. For covalent bonds, vibrational modes includes stretching (symmetric/asymmetric) and bending (rocking/scissoring/wagging/twisting).^[25] For proteins, amide bonds give quite strong absorbance peaks which also contain information about their secondary structures, so it is also used for protein structure characterization.^[23–25] Those peaks usually show up in three regions, namely amide I region (1600 – 1690 cm⁻¹), amide II region (1480 – 1575 cm⁻¹) and amide III region (1229 – 1301 cm⁻¹).^[25] For Amide I bands, 80 % of them are contributed by C=O bond stretching; For amide II bands, 60 % are from N-H bond bending, and 40 % are from C-N bond stretching; For amide III bands, 40 % are from C-N bond stretching, and 30 % are from N-H bond bending.^[24] Further de-convolution of the Amide

I band usually reveals secondary structures present (such as alpha-helix, beta-sheets, random coil, etc.), of which relative contents could be by area integration of the de-convoluted peaks.

In this project, FTIR was a method to assess the DgHBPs' secondary structures at solid state after purification and lyophilization. The measurements were carried on Bruker Vertex 70 FTIR spectrometer. Prior to the experiments, a water vapor spectra was recorded as blank. The Attenuated Total Reflection (ATR) mode was on to enable further quantitative data analysis. Spectra recording range was $800 - 4000 \text{ cm}^{-1}$. Resolution was 4 cm^{-1} , and each spectrum was averaged from over 280 scans.

After the measurement, OPUS 6.5 and Origin 8.6 program was applied to plot the spectra. After water compensation by subtraction, inverted secondary derivative of the original spectra was obtained. By counting maxima peaks from these derivative spectra, one could note down the de-convoluted Infra-red peak values. Then, manual peak-fitting was carried on with the programs in amide I region of $1600\text{-}1700 \text{ cm}^{-1}$ with Gaussian mode, based on the de-convoluted IR peak values from secondary derivative spectra. The assignment of de-convoluted peaks to specific secondary structures were then carried out after fitting the peaks.^[26]

3.5.4 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy is a technique used to assess secondary structure of the proteins.^[27] It is based on the measurement of a difference in absorbance between the left-handed and right-handed circularly polarized light by protein sample.^[28] The secondary structure of protein sample can be determined by comparing of its CD spectra with characteristic spectra of protein secondary structure elements such as beta sheets, alpha helixes and random coils.^[29]

In a CD measurement, absorbance difference data (ΔA) are usually described as delta absorbance or ellipticity (θ), in the unit of millidegrees (mdeg). It can be defined as

below:^[28]

Figure 3.2 CD spectra of polypeptides and proteins with some representative secondary structures.^[29]

$$\Delta A = A_L - A_R$$

It could be transformed as below according to Beer-Lambert Law:^[28]

$$\Delta A = (\varepsilon_L - \varepsilon_R) \times C \times l$$

In above equation, ε means the molar extinction coefficient, C means the molar concentration of sample solution, and l means the path length value in cm unit. After removal of parameters C and l , and conversion into standard unit degrees*cm²*dmol⁻¹, one could get molar ellipticity ($[\theta]$) as below:^[28]

$$[\theta] = 3298.2 \times (\varepsilon_L - \varepsilon_R)$$

Molar ellipticity spectra could be used to evaluate and compare secondary structures of different solutions.

In this work, DgHBPs' secondary structure under various conditions in solution state were examined on an Aviv 420 spectrometer (Aviv Biomedical Inc., New Jersey, USA) equipped with Xenon monochromatic light source, a 50 Hz photoelastic modulator and a temperature controller. QS quartz cuvette (Hellma GmbH & Co., Muellheim, Germany) with a pathlength of 0.2 mm was used. Data was acquired in wavelength steps of 0.5/1 nm from 190 to 260 nm over 3 scans, with an averaging time of 1 s. The obtained spectra were averaged and background (CD spectra of respective buffer) was subtracted. Spectra were smoothed with Savitzky-Golay method for 3-5 points on OriginPro 8.6 to reduce the noise. The data are expressed in terms of the mean residue molar ellipticity in units of $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$.

Further analysis of the spectra was done via software BeStSel.^[30] It gives a rough deconvolution result of CD spectra based on common possible secondary structures: α -helix, β -sheet, random coil, β turn I, II and III, γ -turn, etc. It enables semi-quantitative analysis of acquired CD spectra.

3.6 Characterization of DgHBPs at Coacervates State

In order to investigate the unique properties of DgHBPs at coacervates state, a variety of physical/mechanical/structural techniques were applied. Optical microscope was used to detect and observe the coacervates of DgHBPs at micro-meter scale. UV-vis spectrometer was used to measure the turbidity level of DgHBPs coacervates system. DLS (dynamic light scattering) was chosen to investigate the size distribution of coacervates particles of DgHBPs. ITC (Isothermal titration calorimetry) device was used to determine the thermodynamic nature of the self-coacervation process. CD measurements were conducted to determine the protein secondary structure of DgHBPs at coacervates state, and to compare it to solution state. SAXS (small angle X-ray scattering) experiments were conducted to further explore the molecular structure of DgHBPs coacervates. Rheology

studies were also performed, in order to test the rheological properties of DgHBPs coacervates. Contact angle measurements were also performed to confirm the infiltration ability of DgHBPs coacervates. All of the above methods combined provided a better understanding of coacervates phase of DgHBPs in different terms.

3.6.1 Optical Microscopy

Optical microscope, one of the oldest designs of microscope invented in the 17th century, It is a type of microscope which uses visible light and a series of lenses to magnify images of a sample.^[31] An object is enlarged in the eye of the viewer as a virtual image, through angular magnification of a set of optical lenses.^[32] Most of microscopes share nine basic components: ocular lens, objective turret, objective lenses, focus knobs, stage, light source, condenser and mechanical stage.^[33]

The image from an optical microscope can be captured by normal light-sensitive cameras to generate a micrograph, captured originally by films. Modern developments in CMOS and charge-coupled device (CCD) cameras nowadays enables the capture of digital images, such as a purely digital microscopes which uses a CCD camera to examine a sample and shows the resulting image directly on a computer screen.^[34]

Here, the coacervates droplets of DgHBP-1/2 were first applied on a glass plate in the stage, then visually inspected by a Zeiss “Axio Scope A1” microscope (Carl Zeiss Pte Ltd, Oberkochen, Germany) in the reflection mode, with differential interference contrast (DIC) filters. Images and videos were taken with an AxioCam MRc 5 camera under the control of AxioVision software.

3.6.2 Turbidity Measurement

Turbidity is the cloudiness or haziness of a fluid caused by light scattering of large numbers of individual particles in it. Turbidity level of colloid systems is often measured by Ultraviolet–visible spectroscopy. Ultraviolet–visible spectroscopy (UV-Vis or UV/Vis)

refers to the spectroscopy of absorption or reflectance in the ultraviolet-visible (10–700 nm) wavelength region. The absorption/reflectance in this range reflects the perceived color of the chemicals involved.^[35]

Here, Turbidity of coacervates of DgHBP-1 and DgHBP-2 were measured at 600 nm by a Lambda 35 UV-vis spectrophotometer (PerkinElmer Inc., Massachusetts, USA). The measurement was carried out in the following steps:

- a. Coacervation was induced and incubated at room temperature for 10-30 seconds.
- b. The coacervates were dispersed with buffer by vigorous pipetting.
- c. The dispersed coacervates were put into cuvette (pathlength 10 mm) and start being measured.

The measurement were done in triplicate at 25 °C. Phosphate buffer (60 mM, 100 mM ionic strength with NaCl, pH 6.5) was used as blank. Light transmittance value was taken at the wavelength of 600 nm. Relative turbidity (T_R) is thus calculated using the below equation:^[36]

$$T_R = \ln(T/T_0)$$

Where T is the light transmittance with the coacervates sample; T_0 is the light transmittance without the coacervates sample.

3.6.3 Dynamic Light Scattering

Dynamic Light Scattering (DLS) measures the fluctuations of light scattering intensity of a colloid solution containing scattering particles/matters. According physical law, Particles of size 1 ~ 1000 nm dispersed in colloid solution undergo Brownian motion.^[37] Scattered light from those particles provides information about the size of particles, and this is

described as the hydrodynamic diameters.^[38] The average size of particles measured by DLS, can be calculated for a given monodispersed system and it is usually presented as a distribution profile which shows detected intensity of all particles at all hydrodynamic diameters.^[38]

In this project, Samples were analyzed with a Zetasizer Nano-ZS instrument (Malvern Instruments Ltd, Worcs., UK) with a He/Ne ion laser (633 nm). The effective diameter (also called Z-average mean diameter) was measured after equilibration of samples. Coacervates were dispersed with pipetting to ensure homogeneity of particle size. Relevant buffers (for DgHBPs solutions - 10 mM acetic acid; for DgHBPs coacervates - mixture of acetic acid to phosphate buffer (1:9)) were used as negative control. Measurements were collected at 175 ° detector angle, at 25 °C. The results are reported as the mean with standard deviation calculated by the Malvern software by averaging 10-15 measurements, depending on the sample. Results were plotted with OriginPro 8.6.

3.6.4 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a quantitative physical technique to determine the thermodynamic parameters (binding affinity K_a , enthalpy changes ΔH , and binding stoichiometry n) of physical and chemical interactions in solutions.^[39]

ITC measures thermodynamic parameters by titration. The essential part of an isothermal titration calorimeter is two identical cells made from thermally conducting and chemically inert material (e.g. gold), surrounded with an adiabatic jacket.^[40] One of the cells is reference cell, usually filled with water; the other is a sample cell, which contains the macromolecules of interest. The temperature differences between the two cells are constantly monitored by sensitive thermopile/thermocouple circuit in the instrument. A spinning syringe is utilized for injecting and subsequent mixing of reactants in the sample cell.^[41] During the titration, ligand solution is titrated through a channel into the sample cell, in precisely known aliquots, causing change of temperature due to heat increase/loss. This change then could be translated into real time energy/heat change on the screen, which

then gives information about the nature of the reaction (exothermic/endothermic reaction).^[39,40]

Here, ITC measurements were carried out with a Nano-ITC calorimeter (TA instruments Inc., Delaware, USA) to measure the enthalpic and entropic changes due to coacervation of DgHBP-1 and DgHBP-2. Before titration, the protein stock solutions (0.65 mM, 10 mM acetic acid) were prepared in acetic acid buffer (10 mM, pH 2.5). The dispersion of acetic acid buffer into phosphate buffer was used as a blank reference. All solutions and buffers were degassed under vacuum for 3 min by means of device provided with the ITC apparatus. The injector-stirrer syringe (50 μ l) was loaded with stock solutions of DgHBP-1 or DgHBP-2. Portions of 2.5 μ l were injected sequentially into the titration cell (volume = 300 μ l) initially containing phosphate buffer. The equilibration time between consecutive injections was 300 s. During the titration, the stirring speed was 350 rpm. The heat of dilution from the blank titration of acetic acid buffer (10 mM, pH 2.5) into phosphate buffer (60 mM NaH₂PO₄, 100 mM ionic strength, pH 6.5) was measured, and the dilution heat was subtracted from the raw data to measure corrected enthalpy changes. The results are reported as the change in enthalpy per mole of protein injected into the reaction cell. The first injection was not taken into account for analysis. Calorimetric data analysis was carried out with NanoAnalyze Software (v.3.7, TA instrument Inc.). Thermodynamic parameters including binding stoichiometry (*N*), affinity constant (*K*), enthalpy (ΔH) and entropy (ΔS) changes were calculated by iterative curve fitting of the binding isotherms. The “single binding site” model from the software was used. The Gibbs free energy change (ΔG) was calculated from the equation below: ^[41]

$$\Delta G = \Delta H - T\Delta S$$

3.6.5 Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) is a commonly used technique for structure analysis for biological molecules.^[42] It is based on the deflection of collimated X-ray particles from the original straight direction after interacting with structure with similar or larger size than

the X-ray wavelength.^[43] “Small angle” suggests the small deflection (0.1 – 10 °) angle. Able to detect detailed sample structure information on their sizes and shapes in partially or even completely disordered systems, SAXS is applied widely to measure solutions of proteins, nucleic acids and their complexes.^[44]

In a typical SAXS measurement, the straight electron beam first hit the sample, and then scattering deviate the original direction of the electrons, causing them to shoot in other specific ways. This is then sensed and recorded by the detector at the end of the beam route. The recorded data is called SAXS patterns. In a typical SAXS pattern, data are shown as scattered intensity (“I(q)” or “Ig”) against the magnitude of the scattering vector (“q” or “s”), which is a parameter relevant to scattering angle θ , and could also be interpreted as the resolution with which the sample is observed.. Their relationship can be expressed as below.^[45]

$$q = 4\pi \times \sin(\theta)/\lambda$$

Where θ is the half angle between the X-ray beam and the detector; λ is the wavelength of the X-rays.

A lot of information could be extracted from an X-ray solution scattering pattern. For particles existent in the solution, random positions and orientations lead to an isotropic distribution of scattering intensity. For a monodispersed system, scattering intensity is proportional to that from a single particle, averaged over all orientations. In a scattering pattern as shown in Figure 3.3, one could get information about particle shape at low angle region (2–3 nm resolution); particle folding at medium angle range (2 – 0.5 nm); atomic structure (usually very similar) at above 0.5 nm.^[46] Combining them all could provide information about the gross structural features of particles in a monodispersed system (shape, quaternary and tertiary structure). What is more, SAXS could reveals the folding extent of the molecules in the system most clearly, which could be very useful in the study of DgHBPs coacervation process.^[44]

Figure 3.3 SAXS solution scattering curves containing different information at different size ranges.^[46]

In this project, SAXS data on the proteins and coacervates were measured with the Bruker NANOSTAR SAXS instrument (Bruker GmbH & Co., Bremen, Germany) equipped with a Metal-Jet X-ray source (Excillum AB, Kista, Sweden) and Bruker VÅNTEC-2000 detector system. The X-ray radiation was generated from a liquid gallium alloy with microfocus electron source ($K\alpha = 0.13414$ nm with a potential of 70 kV and a current of 2.857 mA). The X-ray beams were filtered through Montel mirrors and collimated by two pinhole system. Samples (volume of 80 μ l) were pipetted into a vacuum-tight quartz capillary. Solid glassy carbon was used as a standard, and was fixed into the beam path after the sample. Both sample chamber and beam path was evacuated and all the measurements were performed at 15 °C. Sample-to-detector distance was set at 0.67 m, which covers a range of momentum transfer of $0.16 < q < 4$ nm⁻¹ ($q = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle).^[47] The data were collected for 30 min. A total of six frames at

5 min intervals were recorded for each measurement. Detected scattered two-dimensional X-rays by area detector were flood-field and spatially corrected. The flood-field correction rectifies the intensity distortions arising from the non-uniformity in the pixel-to-pixel sensitivity differences in the detector using a radioactive source (^{55}Fe) (Bruker AXS, Germany). These two-dimensional data were then converted to one-dimensional scattering as a function of momentum transfer by radial averaging using the built-in SAXS software (Bruker AXS, Germany), and normalized by the incident intensity and transmission of the sample using a strongly scattered glassy carbon of known X-ray transmission. The transmission of X-ray through the sample was determined using an indirect method, in which the integrated counts per unit time for glassy carbon is connected to the relative intensity of the direct beam and hence the transmission of the sample.^[48] Possible radiation damage were examined by comparing the six frames detecting changes. The scattering of the relevant buffer was subtracted. Data processing steps were carried out using the program package PRIMUS as well as Origin8.6. The scattering intensity curves were plotted against scattering variable q . To assess protein folding and molecular flexibility, the SAXS data was plotted in the so called Kratky plot ($q^2 (I(q)/I(0))$ vs. q).^[49]

3.6.6 Rheology

Rheology refers to the study of liquid or “soft solid” matters at the conditions where they respond with plastic flow under applied force.^[50] Frequently used in the polymer study, Rheology often applies to substances with complex microstructures that demonstrate complex micro-mechanical responses.^[51] Therefore, rheology is suitable in this project to measure the micro-mechanical properties of DgHBPs at coacervates state, specifically their dynamic viscosity and storage/loss moduli (G'/G'').

Viscosity of a fluid means the resistance of the substance to gradual deformation by shear stress (applied to the fluid by a plate in oscillatory rheology measurements). Usually measured as dynamic viscosity of a fluid, expressed in the Newton’s law of friction.^[51]

$$\tau = \mu \times \frac{dc}{dy}$$

Where τ (N/m²) is the shear stress; μ (N*s/m² or Pa * s) is the dynamic viscosity; dc (m/s) is unit velocity; dy (m) is unit distance between layers.

Storage and loss moduli (G' / G'') in viscoelastic materials reflect the elastic energy storage property and the viscous energy dissipation property of the material, respectively.^[50] Their deduction could be shown in the following process:^[50,52,53]

For materials, moduli (G) means the ratio of shear stress (σ) to strain (ε) as below:

$$G = \frac{\sigma}{\varepsilon}$$

However, for viscoelastic materials measured in rheology experiments, there is a phase lag existing for strain (ε) in response to applied shear stress (σ). This means that the calculation of moduli (G) could not be done by calculating strain (ε) value and shear stress (σ) value recorded at the same moment, but by first analyzing sinusoidal curve and equation of both values, followed by transformation:

$$\begin{aligned}\varepsilon &= \varepsilon_0 \times \sin(\omega t) \\ \sigma &= \sigma_0 \times \sin(\omega t + \delta)\end{aligned}$$

Where ε_0 and σ_0 are constants referring to the maximum value of each parameter during the oscillation; $\omega = 2\pi f$ where f is frequency of strain oscillation; t is time; δ is phase lag between stress and strain.

Dividing strain (ε) by applied shear stress (σ) gives the following:

$$G = \frac{\sigma}{\varepsilon} = \frac{\sigma_0}{\varepsilon_0} \times \frac{\sin(\omega t + \delta)}{\sin(\omega t)}$$

While the following transformation is given:

$$\frac{\sin(\omega t + \delta)}{\sin(\omega t)} = \frac{e^{i(\omega t + \delta)}}{e^{i\omega t}} = \cos \delta + i \times \sin \delta$$

Where i is the imaginary unit; δ is phase lag between stress and strain. Therefore shear moduli could be expressed as following:

$$G = \frac{\sigma}{\varepsilon} = \frac{\sigma_0}{\varepsilon_0} \times (\cos \delta + i \times \sin \delta) = G' + iG''$$

Where isolated parameter G' and G'' reflecting the elastic and viscous portion of the overall moduli G , respectively, could be isolated and evaluated properly. By combining and comparing G' and G'' , one could get informed knowledge about the studied material. Often in hydrogel samples, G'' is higher than G' at low oscillation frequencies. When the frequency increases to a critical point, collapse of the hydrogel's internal network leads to liquid-like behavior ($G' > G''$). Above could be important criterion to determine gel-like materials.

In this project, rheology experiments were conducted to test the viscoelastic properties of DgHBPs coacervates. A Physica MCR 501 (Anton Paar with TruGap™) system with appropriate plates (PP-10, CP-25 and CP-60) at RT (25 °C) were applied. For measurements of coacervates phase, the coacervation was induced as described earlier (point 3.4) and it was incubated 5 min, at room temperature. Then, for DgHBP-1, “condensed coacervates” phase was applied. For DgHBP-2, “dispersed coacervates” phase was applied. Upon formation and equilibration, Coacervates were gently pipetted onto the measuring plate.

For viscosity measurements, torsional experiments were conducted by a step-wise shear rate from 0.1 to 100 s⁻¹ using a parallel plate geometry (diameter = 10 mm) with an inter-

plate gap about 0.5 mm. For storage/loss moduli measurements, frequency sweep mode were conducted using the same geometry in the viscoelastic regime at a strain of 0.5 %. For control, the rheological properties of the DgHBPs solutions (~10 mg/mL) and dispersed coacervates were recorded under similar conditions using a cone-plate geometry ($d = 25$ or 60 mm) with an inter-plate gap of $56 \mu\text{m}$. Acquired data point were then plotted as viscosity against shear rate and moduli against frequency, respectively.

3.6.7 Contact Angle and Surface Tension Measurement

Contact angle measurement is a way of assessing the wettability of liquid/hydrogel samples.^[54] Contact angle refers to the liquid-solid angle when a liquid-vapor interface meets a solid interface. It is determined by Young's equation as followed:^[55,56]

$$\gamma_{sv} = \gamma_{sl} + \gamma_{lv} \times \cos \theta_c$$

Where γ_{sv} , γ_{sl} , and γ_{lv} are the interfacial energy of solid-vapor, solid-liquid and liquid-vapor, respectively; θ_c is the contact angle measured.

The contact angle of different liquids or solutions to the same solid surface in stable air environment could be used to compare wettability between them. In ideal situations, if the liquid molecules are attached to the solid molecules most strongly, contact angle would be 0° , in which case liquid would be completely spread out on a solid surface. On the other hand, if the liquid molecules repels the solid molecules most strongly, this would lead to a contact angle of 180° , where liquid would stay spherical shape even on the solid surface. Generally, for the same liquid, a low contact angle indicates a low hydrophobicity for the plate surface, and a high contact angle indicates a high hydrophobicity for the plate surface. For the same solid surface, however, a low contact angle indicates a liquid with good wettability, whereas a high one indicates a liquid with bad wettability.^[57]

Surface tension (surface free energy) is an another important parameter that could be obtained from the measurement.^[54] Surface tension refers to the elastic tendency of a fluid

Figure 3.4 Schematic representation of geometric parameters in a pendant drop measurement.

surface which drives it to acquire the least possible surface area. It results from the attraction of liquid molecules to each other, which is stronger than them to the vapor (usually air) molecules.

In this project, together with contact angle, surface tension gives information about DgHBPs coacervates wetting and infiltration abilities into solid chitin networks in squid beak material. Static contact angle measurements of DgHBPs coacervates, were carried out using a goniometer (Attention Theta, Biolin Scientific, Sweden). 3 μL of DgHBP-1/2 coacervates (dispersed phase) was loaded onto a flat, hydrophobic surface of parafilms (made of olefin waxes). Measurements were done under sessile drop mode. Average contact angle was acquired by three to five repeats for each sample. DgHBPs solution samples were also measured in the same way as control.

Surface tension measurement (Figure 3.4) for DgHBPs coacervates were done under pendant drop mode using the same goniometer. 6 μL volume of coacervates was dispensed, then was measured for diameters of the droplets. Subsequently, surface tension values were calculated by the software as following:

$$\gamma = \frac{\rho \times g \times R_0}{\beta}$$

Where γ is surface tension; ρ is the density of measured liquid, in this project measured by weighing a certain volume of DgHBPs coacervates; g is gravitational constant; R_0 is radius of liquid drop curvature at apex; β is shape factor, defined through Young-Laplace equation^[58] expressed through 3 dimensionless first order equations as following:

$$\frac{dx}{ds} = \cos \phi$$

$$\frac{dz}{ds} = \sin \phi$$

$$\frac{d\phi}{ds} = 2 + \beta \times z - \sin \frac{\phi}{z}$$

Where x , s , z and ϕ are geometric parameters of the liquid drop shown in Figure 3.4. As shown in the above equations, γ could also be estimated from γ value of water as a control, given that the surface tension of water (72.8 mN/m, 20 °C) is already well known, and one just need to measure the diameters of water drop and coacervates drop. As control, surface tension of DgHBPs solution samples were also measured in this project in the same way.

3.6.8 Amino Acid Composition Analysis (AAA)

Hydrolysis followed by ion-exchange chromatography (IEX) is a classical quantitative method to analyze the amino acid composition in studied samples, also known as amino acid composition analysis.^[59] As an indispensable technique for protein quantification, it has broad applications in medical and biological science fields.^[60]

In a typical amino acid analysis process, the sample (usually proteins) substrate is first hydrolyzed using either chemical methods or enzymes, then the hydrolyzed sample in liquid state are injected into IEX with a pH gradient, resulting in elution of different amino acids at different time.^[59] After elution from the IEX column, the free amines in amino

acids of the sample reacts with ninhydrin, a most frequently used residue for post-column residue derivatization. During the reaction, a deep blue or purple color is produced, known as Ruhemann's purple.^[61] The color absorbs visible light at wavelength 440 nm and 470 nm, both used as quantification of the respective amino acids.^[62]

In this project, AAA was used as a way to quantify the proteins during mocking "infiltration" experiments of DgHBPs coacervates, details to which will be described in *Sub-chapter 5.2.10*. The analysis was conducted using a Hitachi L8900 AAA system. The "infiltrated" chitin scaffolds were hydrolyzed overnight *in vacuo* in 100 μ L of 6 M HCl with 8 μ L phenol (Sigma) at 110 °C. After rinsing by MilliQ water 3 times and ethanol once, the hydrolyzed products were dissolved in 100 μ L sample dilution buffer (Sykam GmbH) and spun at 13,000 g for 5 min to remove insoluble fractions, then injected into the AAA. The results were expressed as nmol of (Ala + Gly) normalized to scaffold dry weights.

3.7 Preparation and Characterization of DgHBPs-based Peptides and Variants

In order to study the impact of different sequence domains on coacervation ability of DgHBPs, DgHBPs-based peptides and variants were designed, produced and characterized.

3.7.1 Preparation of DgHBPs-based Peptides

Preparation of DgHBPs-based pentapeptides (GAGFA, GHGLY, GHGVY and GHGPY), decapeptides (GAGFA GHGLY and GAGFA GHGPY), and polypeptides (GA-25, GH-25, GY-23 and GW-26) was done through contracted solid phase peptide synthesis (> 95 % purity) by GL Biochem (Shanghai) Ltd.

3.7.2 Recombinant Expression of DgHBPs-based Variants

Preparation of DgHBPs-based variants (DgHBP-1-V3-C, DgHBP-1-V4-C, and DgHBP-1-V11 to DgHBP-1-V15) was done as following:

For DgHBP-1-V3-C and DgHBP-1-V4-C, mutated DgHBP-1 genes encoding respective variant sequence in pJ express 414 plasmid ampicillin-resistant vectors were transcribed and translated in *E. coli* cells, then expressed. The expressed protein variants existed mainly in inclusion body of the cells. Subsequent cell lysis and purification procedures followed the same protocols as full sequence DgHBPs.

For DgHBP-1-V11 to DgHBP-1-V15, full protein genes of DgHBP-1 were mutated by inserting a lysine-encoding base sequence at different sites corresponding to the designed trypsin-cleavage sites that splice each C- and N- terminus variant. Then, the mutated genes for each variant were transcribed, translated and expressed in *E. coli* cells as mutated proteins (DgHBP-1-V11 to DgHBP-1-V15) with an extra lysine after Y102, Y107, Y112, Y117 and Y122, respectively. The expressed mutated proteins existed in both inclusion body of the cells and the soluble fractions. Subsequent cell lysis and purification procedures followed the same protocols as full sequence DgHBPs.

3.7.3 Trypsin Cleavage Experiment of DgHBPs-based Variants

Preparation of DgHBPs-based C- and N- terminus variants (DgHBP-1-V11-C/N to DgHBP-1-V15-C/N, and DgHBP-2-V1-C/N) was done using trypsin, a lysine/arginine-specific digestive enzyme to cleave the Lys/Arg within the sequences of DgHBP-1-V11 to DgHBP-1-V15 and DgHBP-2, and acquire target C-/N- terminus variants of purpose. Details of trypsin cleavage experiment are as following:

Lyophilized protein (variant) after recombinant expression and primary acetic acid purification was dissolved in 50 mM Tris-HCl buffer (pH 8.8) at a concentration of 1 mg/ml. Then, Trypsin (“Trypsin Gold”, MS grade, Promega) dissolved in 50 mM acetic acid was added into protein (variant) solution at volume dilution ratio of 1/1000. The mixed solution was incubated at 37 °C for 1 to 3 h. 10 mg/ml PMSF stock solution (isopropanol) was then added to terminate the cleavage at volume dilution ratio of 1/100. SDS-PAGE was conducted for post-cleavage products to confirm the result. RPC was subsequently employed on the products to separate N-terminus product and C-terminus product.

3.7.4 Western Blot

During the expression of DgHBP-1-V3-C and DgHBP-1-V4-C, it was discovered that over-expression of impurity protein obscured the target protein variant bands on SDS-PAGE gel, therefore a protein-specific detection method, western blot was used to confirm the expression of DgHBP-1-V3-C and DgHBP-1-V4-C.

Western blot is a method that transfer the proteins electrophoretically from bands on SDS-PAGE gels onto a membrane, then detect the target proteins through specific antibody-antigen binding.^[63] It usually contains 5 steps as followed:

- a.** Gel electrophoresis, where the proteins are separated using normal SDS-PAGE gels.
- b.** Transfer, where separated proteins are moved from within the gel onto a membrane made of nitrocellulose or polyvinylidene difluoride (PVDF), usually by an electrophoretic method called “electroblotting”.
- c.** Blocking, where non-specific binding between the membrane parts without transferred target proteins and the specific antibody for target proteins are prevented using dilute solutions of non-target proteins (usually non-fat dry milk) to block the non-transferred room on the membrane.
- d.** Incubation, where the binding between target protein and specific antibody linked to colorimetric enzyme happens.
- e.** detection, where the membrane is exposed to an appropriate substrate, produces a color and detects the target protein.

Here, western blot for DgHBP-1-V3-C and DgHBP-1-V4-C after expression was conducted using PVDF membrane with TE 77 PWR semi-dry transfer system (Hoefer, Inc.

USA). The blot was incubated overnight in a 1/1000 diluted Anti-6X His tag® antibody (Abcam, UK), then stained using SIGMAFAST BCIP/NBT Substrate Staining Solution (Sigma-Aldrich Inc., USA).

3.7.5 Characterization for DgHBPs-based Peptides and Variants

Characterization for DgHBPs-base peptides and variants mostly involved techniques that was also conducted for full sequence DgHBPs, including SDS-PAGE, MALDI-ToF, optical microscopy, turbidity measurement, dynamic light scattering, circular dichroism spectroscopy, as well as rheology. Therefore same protocols was applied as previously described for DgHBPs.

References

- [1] S. Gräslund, P. Nordlund, J. Weigelt, J. Bray, O. Gileadi, et al. *Nat. Methods* **2008**, 5, 135–146.
- [2] G. L. Rosano, E. A. Ceccarelli. *Front. Microbiol.* **2014**, 5, 341.
- [3] F. R. Schmidt. *Appl. Microbiol. Biotechnol.* **2004**, 65, 363–372.
- [4] G. M. Cooper. **2000**.
- [5] C. M. Thomas, D. Summers, C. M. Thomas, D. Summers. in *Encycl. Life Sci.*, John Wiley & Sons, Ltd, Chichester, UK, **2008**.
- [6] I. Agerkvist, S. Enfors. *Biotechnol. Bioeng.* **1990**, 36, 1083–1089.
- [7] E. Dumay, D. Chevalier-Lucia, L. Picart-Palmade, A. Benzaria, A. Gràcia-Julià, C. Blayo. *Trends Food Sci. Technol.* **2013**, 31, 13–26.
- [8] T. Sauer, C. W. Robinson, B. R. Glick. *Biotechnol. Bioeng.* **1989**, 33, 1330–1342.
- [9] I. Theodossiou, I. J. Collins, J. M. Ward, O. R. T. Thomas, P. Dunnill. *Bioprocess Eng.* **1997**, 16, 175–183.
- [10] N. Xia, J.-M. Wang, Q. Gong, X.-Q. Yang, S.-W. Yin, J.-R. Qi. *J. Cereal Sci.* **2012**, 56, 482–489.
- [11] H. G. Barth, C. Jackson, B. E. Boyes. *Anal. Chem.* **1994**, 66, 595R–620R.
- [12] M. B. Handbook. **1989**.
- [13] S. Mori, H. G. Barth. *Size Exclusion Chromatography*, Springer Science & Business

Media, **2013**.

[14] C. Wu. *Handbook of Size Exclusion Chromatography and Related Techniques: Revised and Expanded*, CRC Press, **2003**.

[15] A. Madadlou, S. O’Sullivan, D. Sheehan. **2011**, pp. 439–447.

[16] R. L. Heinrikson, S. C. Meredith. *Anal. Biochem.* **1984**, 136, 65–74.

[17] A. Ducret, I. Van Oostveen, J. K. Eng, J. R. Yates, R. Aebersold. *Protein Sci.* **1998**, 7, 706–719.

[18] A. L. Shapiro, E. Viñuela, J. V Maizel. *Biochem. Biophys. Res. Commun.* **1967**, 28, 815–820.

[19] A. K. Dunker, R. R. Rueckert. *J. Biol. Chem.* **1969**, 244, 5074–5080.

[20] U. Pieses, W. Zürcher, M. Schär, H. Moser. *Nucleic Acids Res.* **1993**, 21, 3191–3196.

[21] F. Hillenkamp, M. Karas, R. C. Beavis, B. T. Chait. *Anal. Chem.* **1991**, 63, 1193A–1203A.

[22] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo. *Rapid Commun. mass Spectrom.* **1988**, 2, 151–153.

[23] D. M. Byler, H. Susi. *Biopolymers* **1986**, 25, 469–487.

[24] M. Jackson, H. H. Mantsch. *Crit. Rev. Biochem. Mol. Biol.* **1995**, 30, 95–120.

[25] P. I. Haris, F. Severcan. *J. Mol. Catal. B Enzym.* **1999**, 7, 207–221.

[26] J. Siekmann. *Advanced Lectures on Machine Learning*, **2003**.

[27] R. D. Gillard. *Analyst* **1963**, 88, 825.

[28] N. Berova, K. Nakanishi. *Circular Dichroism: Principles and Applications*, John Wiley & Sons, **2000**.

[29] N. J. Greenfield. *Nat. Protoc.* **2007**, 1, 2876–2890.

[30] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112, E3095–103.

[31] S. J. Gould. *The Lying Stones of Marrakech: Penultimate Reflections in Natural History*, Random House, **2001**.

[32] A. (Ariel) Lipson, S. G. (Stephen G. . Lipson, H. (Henry) Lipson. *Optical Physics*, Cambridge University Press, **2011**.

[33] G. Seward, SPIE (Society). *Optical Design of Microscopes*, SPIE, **2010**.

[34] J. Di, J. Zhao, H. Jiang, P. Zhang, Q. Fan, W. Sun. *Appl. Opt.* **2008**, 47, 5654–5659.

- [35] G. Britton. *ChemInform* **1995**, 26.
- [36] H. K. Lichtenthaler, C. Buschmann. *Curr. Protoc. food Anal. Chem.* **2001**.
- [37] T. Hida. in *Brownian Motion*, Springer, **1980**, pp. 44–113.
- [38] W. I. Goldburg. *Am. J. Phys.* **1999**, 67, 1152–1160.
- [39] E. Freire, O. L. Mayorga, M. Straume. *Anal. Chem.* **1990**, 62, 950A--959A.
- [40] A. Velázquez-Campoy, H. Ohtaka, A. Nezami, S. Muzammil, E. Freire. *Curr. Protoc. Cell Biol.* **2004**, 17.8. 1-17.8. 24.
- [41] S. Leavitt, E. Freire. *Curr. Opin. Struct. Biol.* **2001**, 11, 560–566.
- [42] O. Kratky. in *Small Angle X-Ray Scatt.*, **1982**.
- [43] H. P. Klug, L. E. Alexander. *X-Ray Diffraction Procedures*, Wiley New York, **1954**.
- [44] P. Bernadó, E. Mylonas, M. V Petoukhov, M. Blackledge, D. I. Svergun. *J. Am. Chem. Soc.* **2007**, 129, 5656–5664.
- [45] I. Pilz, O. Glatter, O. Kratky. *Methods Enzymol.* **1979**, 61, 148–249.
- [46] W. contributors. “Biological small-angle scattering,” **2016**.
- [47] W. G. Saw, G. Tria, A. Grüber, M. S. Subramanian Manimekalai, Y. Zhao, et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, 71, 2309–2327.
- [48] C. A. Dreiss, K. S. Jack, A. P. Parker. *J. Appl. Crystallogr.* **2006**, 39, 32–38.
- [49] D. Durand, C. Vivès, D. Cannella, J. Pérez, E. Pebay-Peyroula, P. Vachette, F. Fieschi. *J. Struct. Biol.* **2010**, 169, 45–53.
- [50] R. G. Larson. *The Structure and Rheology of Complex Fluids*, Oxford University Press New York, **1999**.
- [51] Y. Doi, F. M. Gray, F. L. Buchholz, A. T. Graham, A. Striegel, W. W. Yau, J. J. Kirkland, D. D. Bly, M. J. Gordon Jr, R. D. Archer. **1994**.
- [52] M. MA, K. K. Chawla. **1999**.
- [53] H. A. Barnes, J. F. Hutton, K. Walters. *An Introduction to Rheology*, Elsevier, **1989**.
- [54] D. Y. Kwok, A. W. Neumann. *Adv. Colloid Interface Sci.* **1999**, 81, 167–249.
- [55] A. B. D. Cassie, S. Baxter. *Trans. Faraday Soc.* **1944**, 40, 546–551.
- [56] E. L. Decker, B. Frank, Y. Suo, S. Garoff. *Colloids Surfaces A Physicochem. Eng. Asp.* **1999**, 156, 177–189.
- [57] T. Sun, L. Feng, X. Gao, L. Jiang. *Acc. Chem. Res.* **2005**, 38, 644–652.
- [58] A. W. Adamson, A. P. Gast, others. **1967**.

- [59] F. D. Macchi, F. J. Shen, R. G. Keck, R. J. Harris. *Amin. Acid Anal. Protoc.* **2000**, 9–30.
- [60] M. Fountoulakis, H.-W. Lahm. *J. Chromatogr. A* **1998**, 826, 109–134.
- [61] S. Ruhemann. *J. Chem. Soc., Trans.* **1910**, 97, 2025–2031.
- [62] J. Csapó, C. Albert, K. Lóki, Z. Csapó-Kiss, others. *Acta Univ. Sapientiae Aliment.* **2008**, 1, 5–29.
- [63] W. N. Burnette. *Anal. Biochem.* **1981**, 112, 195–203.

Chapter 4

Recombinant Expression, Purification and Characterization of DgHBPs

As a fully organic hard structure, D. gigas beak has drawn interest due to its unique mechanical gradient linked to biomolecular gradients. Amongst all components, proteins, especially DgHBPs are arguably the most intriguing ones, likely to play sophisticated roles during beak formation and sclerotization in vivo. In this Chapter, the recombinant expression system of two DgHBPs, DgHBP-1 and DgHBP-2, has been established based on previous sequencing work described in Chapter 2. Subsequent purification strategies for both proteins have also been designed and carried out. Primary biophysical characterizations of the two proteins in various forms and conditions have been conducted by a number of techniques, featuring the interesting conformation shift for both proteins under varying conditions.

4.1 Introduction

As an example of seamlessly-built mechanically mismatched materials in Nature, the beak of *D. gigas* has intrigued researchers with its 200-fold stiffness gradient increasing gradually from the soft base to the dehydrated distal rostrum, correlated highly to its constituting proteins.^[1–4] Previous work combining RNA-Seq and proteomics, mentioned in *Chapter 2*, has revealed the existence and unveiled the sequences of two protein families in the beak, namely DgCBPs (chitin-binding proteins) and DgHBPs (histidine-rich proteins), the latter believed to play crucial roles in beak sclerotization.

A bottom-up approach to investigate roles of DgHBPs during *D. gigas* beak formation has been established in this Chapter. First, vectors containing the genes encoding DgHBP-1 and DgHBP-2, two sequentially different members in DgHBPs family, have been cloned and transformed into *E. coli* cells for scaled-up recombinant production of DgHBP-1 and DgHBP-2. Then, specific purification strategies combining chromatography and solvent exchange for DgHBP-1 and DgHBP-2 have been designed based on their unique polarity, as well as their high solubility in acidic buffers. Finally, secondary structure characterization of purified (*rec*)DgHBP-1 and (*rec*)DgHBP-2 by FTIR and CD featured conformation transitions under between disordered random conformations and ordered β -structures and some helical content, which provides insights into the dynamic nature of the proteins and paving the way for studies on their self-coacervation behaviors presented *Chapter 5*.

One of the hypotheses of this thesis is that DgHBPs are capable of self-coacervation due to their sequence similarities to mfp-3S and tropoelastin. In order to further unveil these similarities, a closer scrutiny of DgHBPs' sequential characteristics was also conducted. In addition, the comparison between the detected conformational transitions of DgHBPs and tropoelastin is also discussed in this Chapter.

4.2 Experimental Methods

A range of experiments, techniques and methods were employed in order to acquire sufficient (*rec*)DgHBP-1 and (*rec*)DgHBP-2 and carry out studies on their biophysical properties

4.2.1 Recombinant Expression of DgHBP-1 and DgHBP-2

Recombinant DgHBP-1 and DgHBP-2 were separately expressed in *E. coli* BL-21 (DE3) cell system, where inserted genes encoding their sequences in pJ express 414 plasmid ampicillin-resistant vector were transcribed and translated in cells. The expressed proteins were mainly found in the inclusion body of the cells. Lysis of cells was conducted by microfluidization methods, followed by buffer washing steps, after which the purity of proteins was checked by SDS-PAGE. More details are described in *Sub-chapter 3.2*.

4.2.2 Purification Strategies for (*rec*)DgHBP-1 and (*rec*)DgHBP-2

Lysed cells after expression of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 contained target proteins with considerable amounts of impurities, which necessitate initial purification steps prior to chromatographic methods. For DgHBP-1, this was done by dialysis of the supernatants after cell lysis against acetic acid. For DgHBP-2, this was accomplished by stepwise washing and solubilization of the pellets after cell lysis. Afterwards, size exclusion chromatography and reverse phase chromatography on proteins were conducted for further purification. Purity and exact molecular weight of purified proteins were examined by SDS-PAGE and MALDI-ToF, respectively.

4.2.3 Further Sequence Analysis

The online bioinformatics tool ExPaSy was used to produce the Hopp and Woods^[5] hydropathy plots of DgHBP-1 and DgHBP-2 based on their sequences. The calculation of molecular weights, amino acid numbers, and isoelectric points, as well as annotations of

charged and aromatic residues were also conducted using ExPaSy tools.

4.2.4 SDS-PAGE

SDS-PAGE was used to confirm the rough molecular weight of expressed and purified (*rec*)DgHBP-1 and (*rec*)DgHBP-2, as well as to examine their purity. Details are given in *Section 3.5.1*.

4.2.5 MALDI-ToF

MALDI-ToF was used to characterize the accurate molecular weight of expressed and purified (*rec*)DgHBP-1 and (*rec*)DgHBP-2. Details are given in *Section 3.5.2*.

4.2.6 Fourier Transform Infrared Spectroscopy

Fourier transform infrared (FTIR) spectroscopy was used to characterize the secondary structures of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 in lyophilized solid state. Measurements details were provided in *Section 3.5.3*.

4.2.7 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was used to study the secondary structures of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 in their buffer solution state. Various conditions were tested for both proteins, for example temperature, protein concentration, buffer pH, etc. Measurements were done as described in *Section 3.5.4*.

4.3 Results and Discussion

4.3.1 Recombinant Expression of DgHBP-1 and DgHBP-2

Prior to expression, verification for the presence of genes encoding DgHBP-1 and DgHBP-

2 sequence in the pJ express 414 plasmid ampicillin-resistant vectors (purchased from DNA 2.0, vector maps shown in Figure 4.1a and Figure 4.1b) was carried out. Restriction enzymes (NdeI and SmaI) were used to cleave the gene inserts of DgHBP-1 and DgHBP-2, and the results (Figure 4.1c and Figure 4.1d) confirmed their existence in the vectors.

After the expression of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 in *E. coli* system, 10 µl samples from lysed cells were loaded into SDS-PAGE and run. As shown in Figure 4.1e and Figure 4.1f, both proteins were expressed abundantly, although the presence of impurity bands indicated the need for further purification. SDS-PAGE results also indicated that for DgHBP-1, the majority of expressed target protein was in the soluble fractions (supernatants) after cell lysis, whereas for DgHBP-2, most expressed proteins were in the insoluble inclusion body fractions (pellets). This indicated that the solubility of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 differ vastly in lysis buffer (50 mM Tris, 200 mM NaCl, 1 mM PMSF, pH 7.4), requiring one to optimize their downstream purification steps for each case, and also indicating first evidences of physico-chemical differences between both proteins.

It is also noteworthy to mention that both DgHBP-1 and DgHBP-2 protein band migrated at a slower speed than expected, since the molecular weight values from the gel were found to be around 20 kDa and 25 kDa for DgHBP-1 and DgHBP-2, respectively, whereas their theoretical values are 15.2 kDa and 18.2 kDa. This is likely due to the rich abundance of repeating hydrophobic residues in their sequences, which affected the protein's binding to SDS molecules, thus causing the “gel shifting” behavior.^[6] However MALDI-ToF spectra confirmed their correct MWs as discussed later.

Yield of expression was about 100 to 150 mg per liter culture for both proteins after purification, which is quite substantial for *E. coli* recombinant expression system, especially considering that little optimization of expression conditions was carried out.

Figure 4.1 Recombinant expression of DgHBP-1 and DgHBP-2. (a) and (b) Vector maps of DgHBP-1 and DgHBP-2 gene-encoding plasmids (red fragment are the DgHBPs-encoding genes), respectively. (c) and (d) DNA gel electrophoresis of NdeI and SmaI digestion products of DgHBP-1 and DgHBP-2, respectively. Circled bands are the cleaved gene inserts. (e) and (f) SDS-PAGE of DgHBP-1 and DgHBP-2 expression after cell lysis. L: ladder.

4.3.2 Solubility Test and Purification of (*rec*)DgHBP-1 and (*rec*)DgHBP-2

To separate (*rec*)DgHBP-1 and (*rec*)DgHBP-2 from impurities after cell lysis, separate purification strategies were developed for DgHBP-1 and DgHBP-2.

For DgHBP-1, as it exists in the soluble fractions (supernatants) after cell lysis, large amounts of Tris, NaCl and other impure proteins need to be removed prior to further study. Therefore, the supernatants were first dialyzed against water to remove the containing Tris and NaCl, and then lyophilized. The lyophilized protein mixture was dissolved in various buffer solvents (pH 7.4 TBS buffer, 5 % acetic acid buffer, 5 % acetic acid with 8 M urea, 2 M guanidine buffer, and 2 M guanidine with 8 M urea), in order to select an optimal solvent that dissolved the maximum amount of DgHBP-1 while at the same time minimizing the amount of impurities. As shown in Figure 4.2a, 5 % acetic acid showed the best solubility as well as purification ability. Therefore, 5 % acetic acid was chosen as the main extraction agent for DgHBP-1 after cell lysis. In the following batches of DgHBP-1 recombinant expressions, supernatants after cell lysis were all dialyzed against 5 % acetic acid, which removed Tris and NaCl and allowed initial “semi-purification” of DgHBP-1 (Figure 4.2c). The remaining product in the dialysis bag was then centrifuged, and the supernatant liquid underwent chromatography for further purification.

For DgHBP-2, the majority of expressed proteins were insoluble after cell lysis. Therefore, solubility test was first carried out, using different solvents (5 % acetic acid buffer, 5 % acetic acid buffer with 8 M urea, pH 7.4 TBS buffer, pH 7.4 PBS buffer, 1 M pH 5.8 phosphate buffer, 1 M pH 8 phosphate buffer) to dissolve the inclusion body. 5 % acetic acid also showed both good solubilizing and purifying effect (Figure 4.2b). Therefore, 5 % acetic acid was chosen as the main agent for inclusion body solubilization of DgHBP-2 after cell lysis prior to chromatography. In the following batches of DgHBP-2 recombinant expressions, pellets after cell lysis were solubilized with 5 % acetic acid subsequently 20 % acetic acid (for complete extraction), during which impurity proteins remained insoluble (Figure 4.2d). The solubilized mixture was then centrifuged, and the supernatant liquid underwent chromatography for further purification.

Figure 4.2 Solubility test and acetic acid extraction of expressed DgHBP-1 and DgHBP-2.

(a) and (b) SDS-PAGE of DgHBP-1 and DgHBP-2, respectively, dissolved in various buffers then centrifuged. L: ladder. T: total fraction (before centrifugation). S: supernatants after centrifugation. (c) and (d) SDS-PAGE of acetic acid extraction of the proteins after cell lysis.

Chromatographic methods proved to be efficient in obtaining pure products for both proteins. SEC was initially conducted for both proteins, yet the separation failed due to the existence of impurity proteins with a similar MW range. Therefore reverse phase chromatography (RPC) with H₂O (0.1 % TFA) and ACN (0.1 % TFA) as mobile phases was instead employed and delivered adequate separation (Figure 4.3a to Figure 4.3d) for both DgHBP-1 and DgHBP-2. The MWs RPC-purified proteins were further confirmed by MALDI-ToF analysis, showing values matching their expected MWs (15.2 kDa and 18.2 kDa) as shown in Figure 4.3e and Figure 4.3f. Thus, RPC was established as the final step for protein purification.

Figure 4.3 Chromatographical purification of DgHBP-1 and DgHBP-2. (a) and (b) RPC spectra of DgHBP-1 and DgHBP-2, respectively. (c) and (d) SDS-PAGE of RPC purification fractions. (e) and (f) MALDI-ToF spectra of RPC-purified proteins.

In summary, based on solubility test results, the purification strategies for DgHBP-1 and DgHBP-2 were established as dialysis (against 5 % acetic acid)-centrifugation-RPC, and solubilization (with 5 % acetic acid)-centrifugation-RPC, respectively, which provided yields of about 50 mg proteins per liter of culture for both proteins. In both cases, the final protein purity was estimated to be above 90 % (based on protein RPC peak integration analysis), which is sufficient for subsequent characterization experiments.

4.3.3 Primary Sequence Analysis

Full sequences of (*rec*)DgHBP-1 and (*rec*)DgHBP-2, including their sequence information (molecular weight, amino acid number, isoelectric point, charged and aromatic residues count), as well as their hydropathy plots are shown in Figure 4.4. From the results, several sequence similarities and difference between DgHBP-1 and DgHBP-2 were noticed. First of all, both proteins carry more positive charge than negative charges in general, and the distribution of charges turns out to be bi-polar with most of the positive residues in the C-termini and the negative residues in the N-termini. Also, both proteins contain more than 10 % aromatic residues, spaced regularly by other residues like Gly, Ala, and Val. Last but not least, both proteins are highly hydrophobic. As shown in Figure 4.4d, about 70 % and 60 % of residues are more hydrophobic than Gly for both DgHBP-1 and DgHBP-2, respectively. In addition, a bi-polar distribution pattern was noticed, as the majority of hydrophobic residues are in the C-terminus, although DgHBP-2 fluctuates slightly more than DgHBP-1 across the hydrophilic/hydrophobic index.

4.3.4 Structural Characterization: Solid State FTIR

The lyophilized state (*rec*)DgHBP-1 and (*rec*)DgHBP-2 after expression and RPC purification were examined by FTIR spectroscopy to investigate their secondary structure at dehydrated state. As shown in Figure 4.5a and 4.5b, both proteins were dominated by β -sheets, with the amide I band centered at 1620 cm^{-1} corresponding to β -sheets for both proteins. Further semi-quantitative analysis conducted on the Amide I band (Figure 4.5c

Figure 4.4 Sequential information and hydropathy plots of DgHBP-1 and DgHBP-2 (a) and (b) Sequences of DgHBP-1 and DgHBP-2, respectively. Red: positively charged residues. Blue: negatively charged residues. Green: aromatic residues. (c) Molecular weight (Mw), amino acid number (A. A.), isoelectric point (pI), and acidic/basic/aromatic residues count ($R^{\text{acidic}}/R^{\text{basic}}/R^{\text{arom}}$). (d) Hydropathy plots of DgHBP-1 and DgHBP-2, using hydropathy values from Hopp T. P. et al.^[5]

and Figure 4.5d) also indicated the existence of β -turn and α -helical content for both proteins. The β -turn accounted for about 32 % of the conformation in DgHBP-2, a much higher content compared to ~16 % in DgHBP-1, which may be attributed to the higher Pro

content in DgHBP-2 (32 % vs. 16 % in DgHBP-1) since Pro is known to drive the formation of β -turns as exemplified in tropoelastin studies.^[7,8]

Figure 4.5 FTIR of lyophilized, solid-state (*rec*)DgHBP-1 and (*rec*)DgHBP-2. (a) and (b) FTIR spectra. (c) and (d) Deconvolution of the Amide I region spectra.

4.3.5 Structural Characterization: Solution State CD

Aqueous solutions of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 were prepared by solubilizing lyophilized proteins in 10 mM acetic acid, where both proteins were found to be well-dissolved. CD spectroscopy was used to characterize the conformation of proteins in aqueous solution state, where two major conformation transitions were noticed.

a. Concentration-dependent conformational transition

As shown in Figure 4.6a and b, both DgHBP-1 and DgHBP-2 exhibited a concentration-dependent conformational transition at room temperature. Interestingly, a similar transition was also reported for α -Elastin, a 70 kDa chemical fragmentation product from elastin often used as a model system to study elastin's secondary structure.^[9] More specifically, at low concentration (1 mg/ml), both DgHBP-1 and DgHBP-2 showed a negative band near 200 nm, which is typical of disordered or denatured proteins, implying a low order of correlation between peptide moieties.^[10] Upon increasing the protein concentration to 10 mg/ml, both HBPs showed distinct spectral changes characterized by a negative peak near 220 nm, indicating a shift towards the formation of β -structures (β -sheets and β -turn). A similar behaviour has been interpreted for α -elastin as a result of intermolecular hydrophobic association^[9], combined with the formation of coacervates from aqueous solutions of α -elastin gives via a concentration-dependent process at 37 °C.^[11] Therefore, a similar hypothesis could be made for DgHBP-1 and DgHBP-2, namely that hydrophobic interactions are mainly responsible for the formation of β -structures in DgHBP-1 and DgHBP-2, which may also drive the formation of coacervates, as discussed in *Chapter 5*.

It is noted that for DgHBP-1 at 1 mg/ml, there was also a negative shoulder at ~220 nm (Figure 4.6a), which suggests the possible presence of polyproline II (PPII) conformation associated with the high abundance of Ala and Leu.^[12,13] Due to the close location of PPII and β -sheets in the Ramachandran plot^[14], and the fact that GA-rich peptides can drive nanoconfined β -sheets formation in silk, a plausible option is that PPII transformed into more packed β -sheet (due to GAGFA/GFA motifs in DgHBP-1) at higher concentration.

On the other hand, the CD spectrum of DgHBP-2 was characterized by a negative shoulder peak at ~ 235 nm. This negative band has previously been reported for domain 26 of elastin-based poly-pentapeptides and attributed to Pro-Gly type of β -turns.^[7,8,15,16] Since the C-terminus of DgHBP-2 contains 3 repeats with Pro (GHGPY) that are absent in DgHBP-1, these data suggest the presence of Pro-Gly β -turns in DgHBP-2.

b. Temperature-dependent conformational transition

Since coacervation of elastin exhibits a LCST (lower critical temperature solution temperature) behaviour, whereby coacervation is induced upon heating the solution with concomitant changes in their CD signature^[17,18], it was investigated whether the same trend could be observed in HBPs and recorded the CD spectra of HBP solutions (1 mg/ml dissolved in 10 mM acetic acid) at increasing temperature. In contrast to elastin and ELPs, coacervation was not observed. However, both proteins exhibited significant shifts in their CD spectra (Figure 4.6c and Figure 4.6d).

For DgHBP-1, as temperature increased, the intense negative band at 200 nm decreased, indicating a transition into a more ordered state. The weaker negative band at 230 nm increased intensity and shifted to 225 nm (Figure 4.6c). These changes suggest that increasing temperature could facilitate intermolecular hydrophobic interactions, of proteins favoring more ordered structures (such as β -structures). The same pattern was reported for elastin-based poly-penta-peptides (PPPs)^[10], as well as α -elastin^[17]. It is noted however, that in contrast to elastin PPPs and α -elastin, self-coacervation for DgHBP-1 was not achieved by just thermally treating aqueous protein solutions, even with increased ionic strength (200-500 mM), which in Elastin's case, greatly widened the coacervation temperature range.^[19]

For DgHBP-2, a similar trend was observed with trivial differences. The major negative band shift for DgHBP-2 is 202 to 205 nm; for DgHBP-1 is 200 to 202 nm. The broad negative band at ~225 to 230 nm for DgHBP-2 shifted much more than DgHBP-1: at 10°C it hardly appeared (Figure 4.6d). This suggests obvious increased β -structure content, most probably coming from strong intermolecular hydrophobic interactions as temperature rose, as similarly reported for elastin Domain 26 and elastin PPPs.^[16,20,21]

Overall, the temperature-dependence spectra for DgHBPs suggest that increasing temperature facilitates intermolecular hydrophobic interactions, favoring more ordered conformation such as β -sheets and β -turns.

Figure 4.6 Conformational transitions of DgHBP-1 and DgHBP-2. (a) and (b) CD spectra of DgHBP-1 and DgHBP-2 of different concentrations in 10 mM acetic acid, showing concentration-dependent conformation. (c) and (d) CD spectra of DgHBP-1 and DgHBP-2 at different temperatures, showing temperature-dependent conformation.

4.4 Conclusions

In this chapter, several aspects have been established for the foundation of subsequent studies on DgHBPs self-coacervation. First, (*rec*)DgHBP-1 and (*rec*)DgHBP-2 were expressed in *E. coli* with adequate yields after purification. Secondly, based on the solubility test results of DgHBP-1 and DgHBP-2, acetic acid extraction has been used employed for both proteins, functioning as the first purification step. Next, RPC was conducted to obtain pure proteins. Subsequent characterization of protein secondary

structures was conducted first by FTIR in their freeze-dried solid state, suggesting high contents of ordered β -structures. Then, CD spectroscopy in their aqueous solution states revealed two conformational transitions: one triggered by increasing protein concentration, and a second one at increasing temperature. Rise in concentration and temperature both led to more ordered β -structures, while more β -turns for DgHBP-2. These transitions share intriguing homology to reported behaviors of α -elastin and elastin-based poly-penta-peptides, further confirming the hypotheses proposed in *Chapter 2* about DgHBPs' structural similarity to elastin. The modest differences between the secondary structures of DgHBP-1 and DgHBP-2 also provided insights about their sequence-structure relationships, specifically regarding how their different C-terminal repeating patterns might affect their conformations.

The main question of this thesis, whether DgHBPs are capable of self-coacervation or not, was still not answered in this chapter. However, the foundation have been laid down. Scaled-up production of recombinant proteins, purification of expressed DgHBP-1 and DgHBP-2, as well as a closer look at the protein sequences were all conducted with that purpose in mind. In addition, preliminary studies of DgHBPs' secondary structures using FTIR and CD have revealed features reminiscent to those of tropoelastin. Coacervation studies are detailed in the next Chapter.

References

- [1] Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez, J. H. Waite. *Nat. Chem. Biol.* **2015**, 11, 488–495.
- [2] A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite. *Science* **2008**, 319, 1816–1819.
- [3] D. J. Rubin, A. Miserez, J. H. Waite, C. Jérôme, J. S. Stephen. in *Adv. In Insect Phys.*, **2010**, pp. 75–133.
- [4] A. Miserez, D. Rubin, J. H. Waite. *J. Biol. Chem.* **2010**, 285, 38115–38124.
- [5] T. P. Hopp, K. R. Woods. *Mol. Immunol.* **1983**, 20, 483–489.
- [6] A. Rath, M. Glibowicka, V. G. Nadeau, G. Chen, C. M. Deber. *Proc. Natl. Acad. Sci.* **2009**, 106, 1760–1765.

- [7] S. Rauscher, S. Baud, M. Miao, F. W. Keeley, R. Pomès. *Structure* **2006**, 14, 1667–1676.
- [8] D. W. Urry, T. Hugel, M. Seitz, H. E. Gaub, L. Sheiba, J. Dea, J. Xu, T. Parker. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2002**, 357, 169–184.
- [9] A. M. Tamburro, V. Guantieri, D. Daga-Gordini, G. Abatangelo. *J. Biol. Chem.* **1978**, 253, 2893–2894.
- [10] D. W. Urry, B. Starcher, S. M. Partridge. *Nature* **1969**, 222, 795–796.
- [11] G. C. Yeo, F. W. Keeley, A. S. Weiss. *Adv. Colloid Interface Sci.* **2011**, 167, 94–103.
- [12] B. Boichicchio, A. Aït-Ali, A. M. Tamburro, A. J. P. Alix. *Biopolymers* **2004**, 73, 484–493.
- [13] A. M. Brown, N. J. Zondlo. *Biochemistry* **2012**, 51, 5041–5051.
- [14] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan. *J. Mol. Biol.* **1963**, 7, 95–99.
- [15] D. W. Urry. in *Methods Enzymol.*, **1982**, pp. 673–716.
- [16] D. W. Urry, T. L. Trapane, K. U. Prasad. *Biopolymers* **1985**, 24, 2345–2356.
- [17] A. M. Tamburro, V. Guantieri, D. Daga-Gordini, G. Abatangelo. *Biochim. Biophys. Acta - Protein Struct.* **1977**, 492, 370–376.
- [18] D. W. Urry, M. M. Long, B. A. Cox, T. Ohnishi, L. W. Mitchell, M. Jacobs. *Biochim. Biophys. Acta - Protein Struct.* **1974**, 371, 597–602.
- [19] B. Vrhovski, S. Jensen, a S. Weiss. *Eur. J. Biochem.* **1997**, 250, 92–98.
- [20] S. A. Jensen, B. Vrhovski, A. S. Weiss. *J. Biol. Chem.* **2000**, 275, 28449–28454.
- [21] D. W. Urry, R. G. Shaw, K. U. Prasad. *Biochem. Biophys. Res. Commun.* **1985**, 130, 50–57.

Chapter 5

Self-coacervation of DgHBP-1 and DgHBP-2

*Self-coacervation is the ability of a liquid solution of macromolecules to phase separate into two liquid phases: one containing a high concentration of the polymer (coacervate phase), and another containing much more diluted equilibrium solution. As mentioned in the previous Chapters, DgHBPs share a high degree of sequence and conformation homology with two other proteins with self-coacervation ability (mfp-3S and tropoelastin). This Chapter explores the self-coacervation behavior of DgHBP-1 and DgHBP-2, first by confirming the self-coacervation ability followed by thorough characterization of their coacervates' kinetics, the factors that control the coacervation behaviour, as well as their morphological, structural, and rheological properties. In addition, the thermodynamic aspects of DgHBPs self-coacervation was investigated using isothermal titration calorimetry, proposing a possible mechanism for the phase behavior. Finally, possible implications of self-coacervation on *D. gigas* beak formation in vivo are discussed.*

*This Chapter published partially as “Tan Y, Hoon S, Guerette PA, Wei W, Ghadban A, **Hao C**, Miserez A, Waite JH, *Infiltration of chitin by protein coacervates defines the squid beak mechanical gradient*, Nature chemical biology, 2015, 11(7), 488-495”.

*This Chapter submitted as “**Hao C**, Gabryelczyk B, Manimekalai M, Grueber G, Salentinig S, Miserez A, *Self-Coacervation of Modular Squid Beak Proteins – A Comparative Study*, Soft Matter, 2017”.

5.1 Introduction

As discussed in *Chapter 2*, the striking sequential similarities between DgHBPs and the two proteins in nature that are capable of self-coacervation sparked the author's interest to investigating the self-coacervation abilities of DgHBPs. This interest also derives from the fact that for both mfp-3S and tropoelastin, self-coacervation functions as a central way to insulate, transport and align large amount of proteins for further cross-linking, which is key to micro-managing the mechanical properties of each material. The question whether the same mechanism applies to *D. gigas* beak material was asked, therefore investigating the self-coacervation ability of *D. gigas* proteins is necessary. In *Chapter 4*, preliminary characterization results of DgHBP-1 and DgHBP-2's secondary structures have partially verified these hypotheses, for that they both showed similar transition into a more ordered conformation with increasing temperature or concentration, which is almost identical to tropoelastin's conformation under similar conditions.^[1-7] Yet, unlike tropoelastin, self-coacervation was not achieved for DgHBP-1 and DgHBP-2 by merely increasing the temperature.

In this Chapter, successful self-coacervation of DgHBP-1 and DgHBP-2 by tuning the pH and ionic strength of the system is demonstrated, which confirmed the major hypothesis in this dissertation. The morphological, structural, mechanical aspects of the coacervates were thoroughly studied via a range of methods. Kinetics, governing physico-chemical parameters, as well as thermodynamic mechanisms of DgHBP-1 and DgHBP-2 self-coacervation were explored, and a comparative study between the self-coacervation of DgHBP-1 and DgHBP-2 was conducted. Finally, a model for *D. gigas* beak formation is proposed, with self-coacervation being a central aspect of it.

5.2 Experimental Methods

A range of equipment and techniques were applied in order to study the morphology, extent, size, structure, thermodynamics, as well as the mechanics of DgHBPs' self-coacervation.

5.2.1 Self-coacervation of DgHBPs

Successful self-coacervation of DgHBP-1 and DgHBP-2 were achieved by adding a high concentration stock protein solution (0.65 mM, 10 mM acetic acid, pH 3.5) into an aqueous buffer with appropriate pH (6 to 8, optimum at 6.5 to 7) and ionic strength (100 to 1000 mM, maximum at 1000 mM). More details were described in *Sub-chapter 3.4*.

5.2.2 Optical Microscopy

Optical microscopy was used to visualize and capture the morphology of DgHBP-1 and DgHBP-2 coacervates, which appeared in the form of fluidic micro-droplets in equilibrium solution. For more details, please refer to *Section 3.6.1*.

5.2.3 Turbidity Measurements

Turbidity measurements at 600 nm wavelength is a common practice in the study of coacervates to quantify the extent of coacervation. Here it is used to monitor the effect of different pH and ionic strength on DgHBPs' self-coacervation. More details about measurement set-up and buffers used at different pH and ionic strength were described in *Sections 3.4 and 3.6.2*.

5.2.4 Dynamic Light Scattering

Dynamic light scattering (DLS) measurements were performed on DgHBP-1 and DgHBP-2 coacervates at different time after dispersion in order to monitor the dynamic change of their size distribution. The Z-average values were used as effective diameters of coacervates. For more details, please refer to *Section 3.6.3*.

5.2.5 Isothermal Titration Calorimetry

Isothermal titration calorimetry was used to characterize the thermodynamic parameters

(binding stoichiometry N , affinity constant K , coacervation enthalpy ΔH , entropy ΔS and Gibbs free energy ΔG) of DgHBP-1 and DgHBP-2 self-coacervation, and help determine the main mechanisms. More details are described in *Section 3.6.4*.

5.2.6 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was employed to study the secondary structures of DgHBP-1 and DgHBP-2 in the coacervates state. CD spectra were acquired for both proteins at different time after vigorous dispersion of coacervates phase in order to monitor the dynamic structural changes of the coacervates. More details about equipment measurement set-up were described in *Section 3.5.4*.

5.2.7 Small Angle X-ray Scattering

Small angle X-ray scattering was used to characterize the structural properties of DgHBP-1 and DgHBP-2 coacervates. Scattering measurements were performed at different time following vigorous dispersion of coacervates phase in order to monitor the dynamic structural transition of coacervates. Measurements were also performed for coacervates in buffers of different pH in order to unveil the influence of pH on the structure of coacervates. For details about SAXS measurement set-up and buffers used at different pH, please refer to *Sections 3.4* and *3.6.5*.

5.2.8 Rheology

Rheological measurements were conducted to characterize the rheological properties of DgHBP-1 and DgHBP-2 coacervates. Dynamic torsional shear mode and oscillatory frequency sweep were applied respectively, in order to obtain the viscosity and storage/loss modulus of coacervates. For more details, please refer to *Section 3.6.6*.

5.2.9 Contact Angle/Surface Tension Measurement

Contact angle and/or surface tension measurements were conducted to investigate the wetting and infiltration abilities of DgHBP-1 and DgHBP-2 coacervates. Measurements were done as described in *Section 3.6.7*.

5.2.10 Chitin Scaffold Infiltration Experiment of DgHBP-1 coacervates

In order to verify the infiltration ability of DgHBP-1 coacervates into the chitin scaffold in *D. gigas* beak, a, infiltration” experiment was designed as following: soft, untanned sections from native *D. gigas* beak were cut into $\sim 0.5 \text{ cm} \times 0.5 \text{ cm}$ slices, and then immersed overnight in a solution of 92.5 vol% H_2O , 5 vol% of H_2O_2 (30 wt%, Sigma-Aldrich), and 2.5 vol% NaOH to remove the native DgBPs existent in the beak. This procedure was repeated until full depigmentation of the slices was attained and then the slices (“scaffolds”) were lyophilized. Dried scaffolds were incubated with DgHBP-1 coacervates phases overnight, with the control being scaffolds immersed in DgHBP-1 solution phase (10 mM acetic acid, same concentration as the coacervates phase). Then the incubated chitin scaffolds were rinsed with MilliQ water to remove proteins absorbed onto the scaffolds’ surfaces, followed by drying and grinding in liquid nitrogen. The ground particles were then dissolved overnight with 50 μl of 5 % acetic acid with 8 M urea and centrifuged at 14,000 rpm for 3 min. The supernatants were analyzed by SDS-PAGE to verify the infiltration ability of coacervates. In parallel, the incubated, rinsed and ground scaffolds were analyzed by AAA to quantify the amount of proteins that have infiltrated the scaffolds. For details about SDS-PAGE and AAA, please refer to *Sections 3.5.1* and *3.6.8*.

5.3 Results and Discussion

5.3.1 Morphology of DgHBP-1 and DgHBP-2 Coacervates

After purification of (*rec*)DgHBP-1 and (*rec*)DgHBP-2 was achieved, various conditions

were tested to verify the hypothesis of their self-coacervation ability. At first, heating-up the protein solutions as done for tropoelastin was conducted. The author was able to monitor similar conformational transitions in aqueous protein solutions (10 mM acetic acid as a solvent), yet no visual turbidity was observed like in the case of tropoelastin. Other solvents (phosphate buffer and Tris buffer) did not induce coacervation due to the low solubility of DgHBP-1 and DgHBP-2 in them. Then, another method as following was devised: first stock solutions of high concentrations of DgHBP-1 (5 to 10 mg/ml) and DgHBP-2 (6 to 11 mg/ml) were prepared in 10 mM acetic acid, which were then pipetted into various other buffers to test the coacervation ability. It was found that coacervates phases were formed for both proteins when they were introduced from their stock solutions into 60 mM phosphate buffer at pH 6 to 7.5, with a volume ratio (buffer : protein solution) of 4 to 9.

This is illustrated in Figure 5.1a for DgHBP-1. Upon pipetting the soluble protein solution into the PBS buffer, the immediate and spontaneous formation of a cloudy phase distinct from the buffer solution phase occurred. If left undisturbed, this phase maintained its rigidity and structural stability, remaining a milky, semi-transparent phase that remained at the top of the vial. This phase is hereafter addressed as “condensed coacervates” phase. If left without stirring, the condensed coacervates phase existed for over 1 h without diffusing or sedimenting. However, vigorous pipetting this phase led to a homogeneous, turbid, and well-dispersed phase, hereafter referred to as “dispersed coacervates” phase. The dispersed coacervates phase were less stable than the condensed coacervates phase, as solid aggregates of proteins were observed after 20 to 40 min in the vial, suggesting a metastable equilibrium between the liquid coacervates and the solid aggregates phases. Under optical microscopy, the condensed coacervates appeared in the form of fluidic micro-droplets of 2 to 20 μm in diameter (Figure 5.1c) and the droplets merged into larger ones over time, as illustrated in Figure 5.2. Upon dispersion, the fluidic micro-droplets seemed to be disrupted through shearing action into smaller 2-5 μm droplets (Figure 5.1e). Some irregular-shaped aggregates were noticed as well, yet the dominant form was micro-droplets by far.

Figure 5.1 Macro-photos and optical microscopic images of DgHBP-1 and DgHBP-2 coacervates. (a) and (b) Self-coacervation process of DgHBP-1 and DgHBP-2, respectively. **i**: injection of stock protein solution into PBS buffer. **ii**: formation of a distinct liquid phase (coacervates phase) **iii**: dispersion of the prior coacervates phase within the buffer solution. **iv**: 20 min after dispersion of coacervates phase. Red circle in **ii**: condensed coacervates phase (for DgHBP-1 only). White dashed rectangle in **iii**: dispersed coacervate phase (for DgHBP-1 and DgHBP-2) (c) and (d) Optical micrographs of coacervate phase before dispersion. (e) and (f) Optical micrographs of coacervates phase after dispersion.

For DgHBP-2 (shown in Figure 5.1b) a cloudy phase spontaneously formed immediately after introduction of the soluble proteins into the buffer; however the coacervates phase of DgHBP-2 were less stable than DgHBP-1, with the coacervates sedimenting from the top

Figure 5.2 Micrographs of DgHBP-1 and DgHBP-2 coacervates after different time after initial formation, illustrating the “merging” behavior of DgHBP-1 coacervates. Red dashed ellipse indicates ongoing merging of two micro-droplets into a larger one.

of the vial to its bottom in a visible fashion. With a slight stirring action with the pipette tip, the coacervates phase mixed with the buffer phase almost immediately, resulting in a homogenous turbid solution, hereafter referred as “dispersed coacervates”. Precipitation in the dispersed coacervates phase was observed about 5 to 10 min after dispersion, as shown in Figure 5.1b, panel iv. Under the optical microscope, the coacervates of DgHBP-2 before dispersion also appeared in the form of micro-droplets, with an average diameter of 2 to 5 μm (Figure 5.1b), and no merging behavior was observed as shown in Figure 5.2. A large

amount of aggregates co-existed with the coacervates (Figure 5.1d). After dispersion, the micro-droplets are diluted in the solution, and no morphological changes were noticed, indicating a stable state of the dispersed coacervates, although an equivalent amount of aggregates were formed in this phase.

5.3.2 Kinetics of DgHBP-1 and DgHBP-2 self-coacervation: Size

In order to quantify the size growth of micro-droplets in DgHBP-1 and DgHBP-2 coacervation systems, DLS measurements were carried out.

As shown in Figure 5.3a, both protein coacervates system showed a rather homogeneous particle size distribution, as indicated by the smooth correlation function curves for certain time period (Figure 5.3c and Figure 5.3d). This behaviour indicates that immediately following dispersion of coacervates, the system remained relatively stable without premature precipitation. However, after 40 min for DgHBP-1 and 15 min for DgHBP-2, the size distribution became heterogeneous judging by the correlation functions in Figure 5.3c and Figure 5.3d, which usually comes from the presence of aggregation or precipitation in the system.^[8] The clear time difference in the appearance of precipitates suggests that DgHBP-1 coacervates are more stable than DgHBP-2, since the partial appearance of aggregates from DgHBP-1 coacervates is largely avoided as opposed to the behaviour of DgHBP-2.

During the time period over which particle size is homogeneous distributed, both protein coacervates exhibited a growing size of their micro-droplets. As also shown in Figure 5.3b, the average effective sizes of the particles in the system increased over time. However, for DgHBP-1 coacervates, the growing rate was relative slow, with an average droplet size of *ca.* 4000 nm after 50 min, whereas for DgHBP-2, the effective diameter grew to *ca.* 5000 nm after only 15 min.

Note that here only the “dispersed coacervate” phases were considered to generate the fitting curves as well as reliable size distribution diagrams. For “condensed coacervate”

phase in DgHBP-1 system, the correlation function curve was identical to aggregates. Therefore the size distribution of “condensed coacervate” could not be obtained.

Figure 5.3 DLS size measurements of DgHBP-1 and DgHBP-2 coacervates at different time after dispersion. (a) DLS profiles for DgHBP-1 (left panels) and DgHBP-2 (right panels) coacervates showing the intensity distributions of micro-droplet diameters in the coacervation systems. (b) Effective particle diameters (Z-average mean values) of DgHBP-1 and DgHBP-2 coacervates vs. dispersion time, respectively. (c) and (d) Correlation functions for the calculated distribution profiles shown in (a).

5.3.3 Kinetics of DgHBP-1 and DgHBP-2 Self-coacervation: Structure

As previously demonstrated in complex coacervation systems, the kinetic of droplets' growth is usually coupled with conformational shifts of the macromolecules concentrated

within coacervates' micro-droplets.^[9-11] Therefore, two methods to characterize the secondary structures of DgHBP-1 and DgHBP-2 coacervates, CD and SAXS, were employed to monitor possible dynamic conformational changes in the coacervates.

Based on CD spectra recorded at different times after dispersion of the coacervates, DgHBP-1 demonstrated a continuous conformational shift (Figure 5.4a). The CD spectra for the condensed coacervates indicated a similar conformation to DgHBP-1 (1 mg/ml) in 10 mM acetic acid, with two characteristic minima at near 208 and 220 nm, which suggest a possible PPII conformation, as described earlier in *Section 4.3.5*. On the other hand, the CD signature of the dispersed coacervate phase showed a dynamic, evolving pattern. The decrease in the absolute value of the 208 nm minimum, coupled with the rise of a broad minimum near 225 nm, indicates that the coacervates evolved towards β -sheets with time.^[12] Therefore, upon coacervation induction and dispersion, DgHBP-1 goes from a conformation of relative disorder to one of considerable order with large proportion of β -sheet conformation.

DgHBP-2 also exhibited a dynamic behaviour with CD spectral changes over time (Figure 5.4b) but these variations were distinct from DgHBP-1 and generally less pronounced. First of all, the condensed coacervate phase does not form. Upon contact with coacervation buffer, DgHBP-2 immediately disperses into the buffer within 3 to 5 min to form the dispersed coacervate phase. The minimum at 208 nm was slightly red shifted, its absolute intensity decreased but it remained the absolute minima. In addition, the broad minimum centred around 225 nm was only minor and only present after 120 min, suggesting that in contrast to DgHBP-1, β -sheets hardly formed in DgHBP-2 coacervates. Furthermore, for DgHBP-2 coacervates, the second minimum at 235 nm attributed to Pro-Gly β -turns for the concentrated DgHBP-2 solution was also found to appear over time. In summary, for DgHBP-2, drastic time-dependent conformation shift was not seen, as they mainly consisted of only one major minima at ca. 205 nm indicating a random coil structure with decreased intensity over time, likely due to coacervates sedimentation.

Figure 5.4 CD spectra of coacervates phase after different time after intentional dispersion in DgHBP-1 and DgHBP-2 self-coacervation system, respectively. (a) and (b) Dispersed coacervates phase (open symbols) at different time compared to protein solution phase (dash lines) and condensed coacervates phase (close symbols, only for DgHBP-1).

Another method employed to characterize the dynamic structural shift of dispersed coacervates was SAXS, which is very useful to investigate the conformation, shape and dimension of biopolymers in solution. As shown in Figure 5.5a, the scattering curves for DgHBP-1 coacervates at different times after dispersion all showed a similar pattern, characterized by the sharp turn/peak in the low q range. In order to better analyze the folding pattern, Kratky plot was plotted as shown in Figure 5.5c, which reveals that in the “condensed coacervates” phase, DgHBP-1 is unfolded/does not have a well-defined structure, as it showed a linear-rising curve, where no peak was found.^[13] Upon dispersion of the coacervates, the Kratky plot exhibited a characteristic maximum indicating conformational folding with enhanced intensity until 20 min, which then decreased again until 60 min, indicating an evolving degree of compactness during this period.

For DgHBP-2, on the other hand, the conformational transition was almost instantaneous, as shown in Figure 5.5b. All dispersed coacervates' scattering curves exhibited a similar pattern and were in stark contrast to the scattering profile of the solution phase. Kratky plots of DgHBP-2 coacervates (Figure 5.5d) indicated a conformational folding with a characteristic maximum at time 0, with a magnitude that decreases over time indicating a possible decreased scattering intensity due to precipitation, as previously discussed in CD studies.

Figure 5.5 SAXS for DgHBP-1 and DgHBP-2 dispersed coacervates at different dispersion time. (a) and (b) SAXS scattering curves for dispersed coacervates at different time as compared to solution phase and condensed coacervates phase (only for DgHBP-1) in DgHBP-1 and DgHBP-2 self-coacervation system, respectively. (c) and (d) Kratky plots calculating from scattering curves in (a) and (b), respectively, revealing the folding signature of coacervates.

5.3.4 Parameters affecting DgHBP-1 and DgHBP-2 Self-coacervation: Temperature

The role of temperature on self-coacervation has been exemplified in tropoelastin, where increasing the temperature of the initial aqueous solution induces coacervation, causing increased turbidity of the solution.^[14,15] For two-component complex coacervation systems, increasing temperature often results in the opposite behaviour, which may be attributed to the weakening of electrostatic interactions.^[16] For DgHBP-1 and DgHBP-2 self-coacervation systems (optimal buffer conditions, see *Section 3.4*), the impact of temperature was also tested. As shown in Figure 5.6, the turbidity of both coacervates remained relatively stable from 25 °C to 70 °C, with no significant changes detected. However, at 4 °C both coacervates demonstrated an obvious rise in turbidity, whereas at 80 °C DgHBP-1 showed a significant turbidity drop that was further corroborated by optical micrographs. This suggest that although details are still elusive, it is likely that the inter-molecular interactions leading to DgHBPs self-coacervation are multiple and include hydrophobic interactions as well as other intermolecular forces such as electrostatic interactions and hydrogen bonds, therefore leading their complex thermal behavior.

Figure 5.6 Turbidity measurement for DgHBP-1 and DgHBP-2 coacervates at different temperatures. Inset: optical image of DgHBP-1 and DgHBP-2 coacervates at 80 °C.

5.3.5 Parameters affecting DgHBP-1 and DgHBP-2 Self-coacervation: pH and Ionic Strength

As previously mentioned in *Chapter 2*, both complex coacervation and self-coacervation system are fundamentally influenced by pH and ionic strength, which could change the formation as well as the stability of coacervates. Therefore, for DgHBPs self-coacervation systems, studies on the influence of pH and ionic strength were also conducted, in two aspects: intensity of coacervation, and structure of coacervates.

The extent of coacervation was quantified by turbidity measurements, which is a common practice in the studies of complex coacervation. As shown in Figure 5.7, Turbidity measurement for both DgHBP-1 and DgHBP-2 indicate that in the range of pH and ionic strengths tested, both proteins form coacervates in a rather narrow “window”, and that in low salt concentration buffers, the proteins achieve maximum turbidity at pH (6-7) close to their their isoelectric points (6.03 and 6.15). This is because that at low or high pH, the molecules carries the same net charge, thus the long-range electric double-layer repulsion^[17] between the molecules would keep the proteins away from coalescence. As the pH approaches the pI of ~6.5 (the pH at which the positive and negative charges exposed on the molecule surface exactly neutralize each other) the net charge of the protein molecules decreases to zero, coupled with a decrease of long-range double-layer repulsion, thus the dominant interaction between any two molecules will be short range attractions (coulombic interaction and hydrophobic interaction), resulting in strong inter-molecular attraction. The results also showed that at high salt concentration (1 M), their coacervation happens even at pH 4. This is because the long-range electric double layer repulsion between the positively charged molecules, which are effective in preventing the coacervation, no longer functions due to salt-screening effect.^[18,19] In summary, both adjusting pH to ~pI and increasing ionic strength can decrease long-range repulsion and increase short-range attraction (coulombic interaction and hydrophobic interaction) that results in coacervation of proteins.

Regarding the structural properties at the nano-scale, SAXS measurements were carried

Figure 5.7 Turbidity measurement for DgHBP-1 and DgHBP-2 dispersed coacervates, respectively. (a) and (b) Turbidity as a function of pH at 100, 500 and 1000 mM ionic strength for DgHBP-1 and DgHBP-2 dispersed coacervates, respectively.

out on DgHBP-1 and DgHBP-2 dispersed coacervates in different buffers, with pH values from 5.5 to 10. As shown in Figure 5.8, the scattering curves for both coacervates vary at different pH value, with the highest turn/peak at low- q range appeared at pH 7.5 and pH 7 for DgHBP-1 and DgHBP-2, respectively. To better analyze their folding patterns, Kratky plots were obtained and plotted as shown in Figure 5.8c-d, which reveals that the maximum conformation folding for DgHBP-1 and DgHBP-2 coacervates happened at pH 7.5 and pH 7, respectively.

In addition, for DgHBPs under non-coacervation conditions (pH 4 and 10), SAXS scattering curves were fitted with simple power-law equation $I(q) \sim q^{-x}$, where the power-law coefficient x provide insights into the protein conformation (Figure 5.8e). At pH 4, this analysis indicated that DgHBP-2 was fully unfolded whereas DgHBP-1 appeared to be partially folded. A similar trend was obtained at pH 10 (Figure 5.8e), suggesting that DgHBPs present an elongated, cylindrical shape in these conditions. Although more accurate calculations will require measuring the scattering curves to lower q regimes, these results provide semi-quantitative insights on the nano-scale organization of DgHBPs. Thus, under non-coacervation conditions DgHBPs seem to arrange into a cylindrical shape that hinders compaction, in agreement with repulsive charge-charge interactions at these pHs. Under coacervation conditions, on the other hand, folding patterns are quite significant,

and more complex models are needed for scattering curve fitting, therefore will not be discussed here.

In summary, both turbidity measurement and SAXS unveiled that there are optimal physical conditions for the self-coacervation to happen, and that coupled with the optimization of conditions is the maximal folding pattern of coacervates.

Figure 5.8 SAXS for DgHBP-1 and DgHBP-2 dispersed coacervates at different dispersion time. (a) and (b) SAXS scattering curves for dispersed coacervates at different pH at controlled ionic strength (100 mM) for DgHBP-1 and DgHBP-2, respectively. (c) and (d) Kratky plots calculating from scattering curves in (a) and (b), respectively, revealing the folding signature of coacervates. (e) Power law fitting coefficient x values for DgHBP-1 and -2 at pH 4 and pH 10.

5.3.6 Mechanical Properties of DgHBP-1 and DgHBP-2 Coacervates: Viscosity and G'/G''

As mentioned in *Sub-chapter 1.5*, the significance of coacervation in nature material bioprocessing partially derives from coacervates' ability to maintain the mechanical properties of the relevant material. Therefore, the mechanical properties of DgHBP-1 and DgHBP-2 coacervates were characterized using rheology. In order to obtain the viscosity and storage/loss modulus of coacervates, two types of measurements were carried out: dynamic torsional shear viscosity measurement and oscillatory viscoelastic measurements

For viscosity measurements, DgHBP-1 condensed coacervates featured a shear-thinning behavior in dynamic torsional mode, characterized by a sudden 100-fold drop of viscosity above the critical shear rate of $\sim 15 \text{ s}^{-1}$ (Figure 5.9a, black curve). The nanoscale molecular structure of coacervates is often viewed as a bead-string model structure^[20,21], consisting of globular protein micelles arranged in a bead-string fashion. Applying shear induces elongation/relaxation of the beaded string structures, causing increased lateral inter-complex associations and coalescence of nano-complexes into denser microphases, which possibly translates into shear thinning behaviour.^[21] For DgHBP-2 coacervates, the viscosity at low shear rates ($< 1 \text{ s}^{-1}$) was more than two orders of magnitude lower than for DgHBP-1 and no obvious shear thinning was detected (Figure 5.9a, red curve), since the initial viscosity is already low. Similar to DgHBP-1, frequency sweep measurements (from 0.1 to 10 Hz) indicated a gel-like behaviour with the storage modulus G' dominating over the viscous modulus G'' ($G' > G''$) (Figure 5.9b). However, for a given protein concentration (11 mg/ml) G' of DgHBP-2 was about 2-order of magnitude lower than DgHBP-1, strongly indicating that DgHBP-1 coacervates are more densely cross-linked or more entangled. In fact, the absolute G' values of DgHBP-1 coacervates are close to some hydrogels^[22–24].

In terms of primary structure-property relationships, it is tempting to attribute the higher storage modulus of DgHBP-1 to the presence of the **GAGFA/GFA** repeats, since these are lacking in DgHBP-2. However, it cannot be ruled out that these variations may also be

related to differences of the of N-terminus domains between DgHBPs. In order to discern which of these domains dominate the rheological response, two consensus peptides derived from the repetitive C-terminus of DgHBP-2 and DgHBP-1, respectively, were designed. The first peptide (*pepDgHBP-2*) consisted of 5 repeats of the **GHGX₁Y** and a single C-terminus Trp residue, with the sequence: **GHGVY GHGVY GHGPY GHGPY GHGLY W**. The second peptide (*pepDgHBP-1*) corresponds to the sequence interval G113–Y135 of DgHBP-1: **GHGLY GAGFA GHGLH GFA GHGLY**. Thus, the main difference with (*pepDgHBP-2*) is the substitution of two **GHGX₁X₂** with the motifs **GAGFA** and **GFA**. Frequency sweep measurements for both peptides are presented in Figure 5.9c-d. Just like for full-length DgHBPs, *pepDgHBP-1* exhibited a much higher storage modulus, with values of *ca.* 20-40 Pa (0.1 to 1 Hz) compared to 0.1-0.4 Pa for *pepDgHBP-2* in the same frequency range. Furthermore, *pepDgHBP-2* exhibited a liquid-like behaviour with $G'/G'' < 1$ across all frequencies. These data demonstrate the occurrence of a tighter network in *pepDgHBP-1* coacervates since the storage modulus is well-known to scale with the cross-link density.^[25] Based on the CD spectra, it is suggested that these differences are due to the formation of short β -sheet domains formed by the **GAGFA/GFA** motifs that enhance the cross-link density in the coacervates; however further experimental evidence is needed to definitely confirm this proposition. It is also noted that the storage modulus of *pepDgHBP-1* is larger than that of the full-length DgHBP-2, which demonstrate that hydrophobic interactions from the pentapeptides dominate the rheological response over the chain length effect.

The unique shear-thinning of DgHBP-1 coacervates have intriguing biological implications.^[12] It implies that during the secretion and transportation of proteins, they pass from the beccublast cells to a chitin network with pores of ~10-50 nm. For a given volume flow rate within a porous network, the shear rate is known to scale as d^{-3} where d is the mean pore size.^[26] Thus for the densely cross-linked chitin/DgCBPs network, at a fixed flow rate, the shear rate on the wall of the inter-chitin porous network increases, and shear thinning of the coacervates is suggested to facilitate flow through the nano-porous network. The absence of shear thinning in DgHBP-2 coacervates suggest that DgHBP-1 and -2 play separate roles during their transport within the chitin networks during beak formation,

perhaps allowing the animal to finely tune the viscosity of DgHBP secretions during infiltration into the cross-linked network.

Furthermore, the unique viscosity and modulus properties of DgHBP-1 could also be exploited in applications such as drug delivery and tissue engineering. For example, coacervates of DgHBP-1 could be designed transport highly selective drug into target areas *in vivo* based on their “shear-thinning” ability.

Figure 5.9 Rheological studies on DgHBP-1 and DgHBP-2. (a) Viscosity of DgHBP-1 and DgHBP-2 coacervates as a function of shear rate, suggesting that only DgHBP-1 coacervates have shear-thinning behavior. (b) Frequency sweep measurements of storage moduli (G' , closed symbols) and loss moduli (G'' , open symbols) of DgHBP-1 and DgHBP-2 coacervates, suggesting that only DgHBP-1 coacervates have gel properties. (c) Viscosity of pepDgHBP-1 and pepDgHBP-2 coacervates as a function of the shear rate. (d) Frequency sweep measurements of storage (G' ; closed symbols) and loss (G'' ; open symbols) moduli of pepDgHBP-1 (blue boxes) and pepDgHBP-2 (pink circles) coacervates.

5.3.7 Thermodynamics of DgHBP-1 and DgHBP-2 self-coacervation

ITC was used to gain insights into the type of inter-molecular interactions occurring during the self-coacervation process. The measurements consisted in titrating DgHBP solution into the coacervation buffer, and the results are shown in Figure 5.10a and b for DgHBP-1 and -2, respectively. In both cases, positive heat rates were measured, denoting the endothermic nature of DgHBPs' self-coacervation. Fitting the isotherm curves with a single-binding model, thermodynamic parameters (Figure 5.10a-b, bottom insets) of self-coacervation were obtained. For both DgHBP-1 and -2 coacervation, the entropic contribution appeared to be significant and can be attributed to enhanced inter-protein hydrophobic interactions, protein conformational freedom, and loss of water during coacervation. The hydrophobic nature of the coacervation is consistent with the hydropathy plots results in Figure 4.4 since both proteins contain more than 60 % hydrophobic residues with a particular enrichment in their C-termini. The unfavourable enthalpy increase is possibly due to the breakage of hydrogen and ionic bonds arising from the release of ordered water molecules from the protein monomer surfaces during coacervation.^[27] However, the entropy gain compensates for the enthalpy increase ($T\Delta S > \Delta H$). In summary, ITC results show that self-coacervation of DgHBP-1 and -2 are predominantly entropy-driven through hydrophobic interactions, which is fundamentally different from complex coacervation governed by enthalpy-driven Coulombic attraction or specific ion pairing.^[28]

It should be mentioned that by far, nano-ITC measurements have not been previously reported for protein self-coacervation. Classical ITC measurements involve titrating solution A into solution B, as previously carried out for complex coacervates.^[24,29,30] In those cases, the binding site models are better established and readily interpretable, with the fitted stoichiometry (N) value reflecting the actual binding ratio of molecule A to molecule B.^[31] In this case, no suitable model for self-binding is currently available and this is no accepted theory about the details of inter-molecular interactions during self-coacervation. Thus, the physical meaning of the N value is to be taken cautiously. Various binding models currently available were applied to fit the data, of which the single-binding model provided the best results. With this model, the fitted number N for HBP-1 and -2 are

0.309 and 0.202 (Figure 5.10a-b, insets), respectively, which is lower than the assumed value of 1 in this model. According to Cooper et al.^[32], this abnormality could be explained in the following way. Firstly, the actual protein concentration involved in association is lower than its actual concentration, since proteins interact with themselves. Therefore, a significant fraction of the protein concentration in this model is sacrificed and attributed as a “ligand” concentration, which underestimates the N value. Second, the main driving force for self-coacervation appears to be hydrophobic interactions, therefore the simple assumption of non-cooperative binding would not be fully accurate, because cooperative interactions (hydrophobic forces) dominate in this case.^[32]

Figure 5.10 ITC measurement of DgHBP-1 and DgHBP-2 self-coacervation. (a) and (b) Thermogram (top) and isotherm (bottom) of DgHBP-1 and DgHBP-2 self-coacervation, respectively. Inset: Thermodynamic parameters of self-coacervation. The experiments were performed at pH 6.5 (phosphate buffer, ionic strength 100 mM) and 25 °C.

5.3.8 Wettability of DgHBP-1 and DgHBP-2 Coacervates: Contact Angle and Surface Tension

As mentioned in *Chapter 2*, one of the proposed functions for DgBPs in the formation of *D. gigas* beak is to be adsorbed by the chitin nanofibers network. In interface physics, this function usually requires good wettability of the liquid phase. Therefore, contact angle and surface tension measurements were measured for DgHBP-1 and DgHBP-2 coacervates.

As shown in Figure 5.11 both DgHBP-1 and DgHBP-2 coacervates demonstrated a lower contact angle than the control buffer on the same surface, indicating that the coacervates phase has a relatively low surface energy, which enables them to infiltrate the chitin scaffold in the early stage of beak maturation.

Surface tension measurement for DgHBP-1 and DgHBP-2 coacervates further confirmed their enhanced wettability as compared to the solution state. As shown in Figure 5.12, both coacervates showed substantial decrease of surface tension compared to their solution phases. These results, coupled with optical microscopic observation results, also indicate that DgHBPs coacervates have great potential to encapsulate therapeutics, due to their low surface tension, fluidic nature, and relative stable form in solutions.

Figure 5.11 Contact angle measurement for DgHBP-1 and DgHBP-2 coacervates. (a) Control buffer (10 mM acetic acid diluted into pH 6.5 phosphate buffer with 100 mM ionic strength at dilution ratio of 9). (b) and (c) DgHBP-1 and DgHBP-2 coacervates, respectively.

Figure 5.12 Surface tension measurement by pendant drop method for DgHBP-1 and DgHBP-2. (a) Surface tension diagram for phosphate buffer, solution and coacervates phases of DgHBP-1 and DgHBP-2, respectively. Insets are real-time captured images for each sample. (b) Summary chart of surface tension values for different samples. ($N = 3$, technical replicates, error bar represents mean values $\pm$ s.d.)

5.3.9 Infiltration Ability of DgHBP-1 Coacervates

As illustrated in *Section 5.3.6*, DgHBP-1 showed specific physico-chemical characteristics (fluidity and low surface tension). These characteristics make them suitable for infiltrating the initially soft chitin scaffold (made of hydrated chitin/chitin-binding protein nanofibers). To corroborate this hypothesis, infiltration experiment of DgHBP-1 coacervates was designed: soft sections of native beaks were sliced and incubated with *recDgHBP-1* coacervates or solution, and were then analyzed by SDS-PAGE to see whether the coacervates infiltrated into the chitin scaffold (Figure 5.13a).

Figure 5.13 Infiltration experiment of DgHBP-1 coacervates. (a) and (b) Schematic representation of the experimental and characterization procedures, respectively. (c) SDS-PAGE of infiltrated beak with coacervates phase and solution phase of DgHBP-1. L: ladder. P: purified DgHBP-1 for control. N: native *D. gigas* beak sections before de-proteinization. S: beak "scaffold" after de-proteinizing the native beak sections. cC: condensed coacervates. dC: dispersed coacervates. (d) Amino acid composition of (Ala + Gly) in infiltrated beak "scaffold" normalized by the "scaffold" dry weight before infiltration. $N = 3$; ** $P < 0.001$; * $P < 0.05$.

As residual proteins could still be present in the native beak which interfered with the assay, the cut wing slices were initially de-proteinized by alkali peroxidation prior to experiment.^[33] Protein after infiltration was extracted as shown in Figure 5.13b. SDS-PAGE results indicated the presence of DgHBP-1 after infiltration (Figure 5.13c). However this was only detected for scaffolds incubated with the condensed coacervates (S + cC) phase, yet not detected for scaffolds incubated with the protein solution (S + P), nor with the dispersed coacervates (S + dC). (Figure 5.13c) Given the fact that these outcome for (S

+ P) and (S + dC) could also be feasible if the total protein content was below detection limit of SDS-PAGE, therefore quantitative measurements of amino acid composition were also conducted for the scaffolds post-infiltration (Figure 5.13d). In those experiments, the total DgHBP-1 concentration in the solution (black bars) and the dispersed coacervates (green bars) were kept identical, which provided a strict control to evaluate the effect of coacervation during infiltration. The result showed that measured protein content for the beak scaffolds incubated with dispersed DgHBP-1 coacervates was about three times higher than with protein solution, after both 1 and 4 hours. This suggests that the coacervate phase was more efficient state to infiltrate the beak nano-porous chitin scaffold.

Protein content in the scaffolds infiltrated with the condensed coacervates for 4 hours (Figure 5.13d) was also measured. The results showed that protein content was significantly higher (25.8 ± 3.1 nmol/mg) in comparison with the dispersed coacervates (15.9 ± 5.2 nmol/mg) and the protein solution (6.2 ± 1.6 nmol/mg). This is consistent with the SDS-PAGE results demonstrates that DgHBP-1 is capable of infiltrating the native chitinous/DgCBPs network.

5.4 Conclusions

Summarizing the results from all above experiments, the following conclusions can be drawn.

a. Comparison between self-coacervation by DgHBP-1 and DgHBP-2

DgHBPs represent central biomolecular components of squid beaks, comprising about 80 wt% of their dry weight and largely responsible for the beak's impressive mechanical performance. DgHBPs have low complexity structure and contain hydrophobic pentapeptides repeats in their C-termini sharing homology with those from elastin, although these motifs are significantly different, notably the elevated His content of DgHBPs. Using turbidity and SAXS measurements, it has been shown that both DgHBPs are able to self-coacervate within a pH window near neutral pH, which can be broadened

by increasing the ionic strength, indicating that the hydrophobic interaction-driven self-coacervation process requires charge neutralization through salt-screening. ITC measurements confirmed the hydrophobic (endothermic) nature of the coacervation process. The results also indicated that DgHBPs coacervates are very dynamic, as evidenced by the evolution in particle size and conformational transitions detected by CD and SAXS. Another key finding of this study is that DgHBPs achieve maximum folding at or near their optimum coacervation conditions. Using contact angle and surface tension measurements, low surface tension for DgHBPs coacervates was verified, and chitin-network infiltration experiment for DgHBP-1 coacervates further confirmed their enhanced infiltration ability, which is vital during beak formation.

Comparative analysis also reveals significant differences between DgHBP-1 and -2. DgHBP-2 cannot coalesce into a condensed coacervate phase; however merging of coacervate droplets is faster than in DgHBP-1 and some aggregation is also observed in DgHBP-2. DgHBPs differ mostly in their rheological response. As opposed to DgHBP-1, DgHBP-2 does not exhibit shear thinning and its storage modulus is about 2-orders of magnitude lower. These differences are attributed to the penta- and tri-peptide repeats **GAGFA** and **GFA** that are only found in DgHBP-1, which are likely to act as physical cross-links that enhances both viscosity and storage modulus. (Details regarding this will be further explored in *Chapter 6*.) Rheological measurements on short consensus peptides support the view that these repeats are critical to enhance the visco-elastic properties.

b. Significance of DgHBPs self-coacervation in D. gigas beak formation

The second implication is about the role of DgHBPs play in the formation of *D. gigas* squid beak. Squid beak is made of one of the most wide-ranging stiffness gradient reported so far in natural materials. A structure with such a gradient must be produced with an accurate control of the molecules and processes involved. As the main protein component in the beak, DgHBPs demonstrate a series of unique phase behavior in aqueous solutions, featured by their “self-coacervation” ability. The coacervates of DgHBPs contain the following properties: high biopolymer concentration, low viscosity, low interfacial energy

and a shear-thinning behavior. These abilities would facilitate the flow of coacervate micro-droplets to their destinations without being limited by the high viscosity of a concentrated protein phase, therefore allowing DgHBPs to easily wet and diffuse through the chitin nanofiber. When DgHBPs arrive their target areas, owing to their hydrophobic nature they would repulse the water molecules present in the chitin scaffold, causing dehydration of the beak. They can also form inter- and intra-molecular histidine-catechol cross-links. Combining these two process (dehydration and cross-linking) leads to controlled sclerotization of *D. gigas* beak. In summary, self-coacervation of DgHBPs plays vital roles of transporting and assembling sclerotization proteins of *D. gigas* beak, and the entire process can be viewed as an efficient natural “biomanufacturing” process.

References

- [1] D. W. Urry, T. Hugel, M. Seitz, H. E. Gaub, L. Sheiba, J. Dea, J. Xu, T. Parker. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2002**, 357, 169–184.
- [2] D. W. Urry, B. Starcher, S. M. Partridge. *Nature* **1969**, 222, 795–796.
- [3] M. Mammi, L. Gotte, G. Pezzin. *Nature* **1968**, 220, 371–373.
- [4] D. W. Urry, T. L. Trapane, K. U. Prasad. *Biopolymers* **1985**, 24, 2345–2356.
- [5] D. W. Urry. in *Methods Enzymol.*, **1982**, pp. 673–716.
- [6] D. W. Urry, M. M. Long, B. A. Cox, T. Ohnishi, L. W. Mitchell, M. Jacobs. *Biochim. Biophys. Acta - Protein Struct.* **1974**, 371, 597–602.
- [7] A. M. Tamburro, V. Guantieri, D. Daga-Gordini, G. Abatangelo. *Biochim. Biophys. Acta - Protein Struct.* **1977**, 492, 370–376.
- [8] W. I. Goldburg. *Am. J. Phys.* **1999**, 67, 1152–1160.
- [9] W. C. Blocher, S. L. Perry. *Wiley Interdiscip. Rev. Nanomedicine Nanobiotechnology* **2016**, DOI 10.1002/wnan.1442.
- [10] C. Schmitt, C. Sanchez, S. Desobry-Banon, J. Hardy. *Crit. Rev. Food Sci. Nutr.* **1998**, 38, 689–753.
- [11] J. J. Water, M. M. Schack, A. Velazquez-Campoy, M. J. Maltesen, M. van de Weert, L. Jorgensen. *Eur J Pharm Biopharm* **2014**, 88, 325–331.
- [12] Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez, J. H. Waite. *Nat. Chem. Biol.* **2015**, 11, 488–495.

- [13] B. Zagrovic, J. Lipfert, E. J. Sorin, I. S. Millett, W. F. van Gunsteren, S. Doniach, V. S. Pande. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 11698–11703.
- [14] G. C. Yeo, F. W. Keeley, A. S. Weiss. *Adv. Colloid Interface Sci.* **2011**, 167, 94–103.
- [15] C. M. Bellingham, K. A. Woodhouse, P. Robson, S. J. Rothstein, F. W. Keeley. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **2001**, 1550, 6–19.
- [16] D. Priftis, M. Tirrell. *Soft Matter* **2012**, 8, 9396–9405.
- [17] A. W. Adamson, A. P. Gast, others. **1967**.
- [18] R. Perez-Jimenez, R. Godoy-Ruiz, B. Ibarra-Molero, J. M. Sanchez-Ruiz. *Biophys. J.* **2004**, 86, 2414–2429.
- [19] H. I. Petrache, T. Zemb, L. Belloni, V. A. Parsegian. *Proc. Natl. Acad. Sci.* **2006**, 103, 7982–7987.
- [20] W. Wei, Y. Tan, N. R. Martinez Rodriguez, J. Yu, J. N. Israelachvili, J. H. Waite. *Acta Biomater.* **2014**, 10, 1663–1670.
- [21] S. Kaur, G. M. Weerasekare, R. J. Stewart. *ACS Appl. Mater. Interfaces* **2011**, 3, 941–944.
- [22] S. Rammensee, D. Huemmerich, K. D. Hermanson, T. Scheibel, A. R. Bausch. *Appl. Phys. A Mater. Sci. Process.* **2006**, 82, 261–264.
- [23] R. Jin, C. Hiemstra, Z. Zhong, J. Feijen. *Biomaterials* **2007**, 28, 2791–2800.
- [24] D. Priftis, N. Laugel, M. Tirrell. *Langmuir* **2012**, 28, 15947–15957.
- [25] M. Rubinstein, R. H. Colby. *Polymer Physics*, Oxford University Press New York, **2003**.
- [26] R. L. Fournier. *Basic Transport Phenomena in Biomedical Engineering*, CRC Press, **2011**.
- [27] K. Luke, D. Apiyo, P. Wittung-Stafshede. *Biophys. J.* **2005**, 89, 3332–6.
- [28] J. Fu, J. B. Schlenoff. *J. Am. Chem. Soc.* **2016**, 138, 980–990.
- [29] C. G. De Kruif, F. Weinbreck, R. De Vries. *Curr. Opin. Colloid Interface Sci.* **2004**, 9, 340–349.
- [30] S. M. H. Hosseini, Z. Emam-Djomeh, S. H. Razavi, A. A. Moosavi-Movahedi, A. A. Saboury, M. A. Mohammadifar, A. Farahnaky, M. S. Atri, P. Van Der Meeren. *Food Chem.* **2013**, 141, 215–222.
- [31] E. Freire, O. L. Mayorga, M. Straume. *Anal. Chem.* **1990**, 62, 950A--959A.

- [32] A. Cooper. *Curr. Opin. Chem. Biol.* **1999**, 3, 557–563.
- [33] A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite. *Science* **2008**, 319, 1816–1819.

Chapter 6

DgHBPs-based Peptides and Variants

In Chapter 5, self-coacervation of DgHBP-1 and DgHBP-2 has been investigated in terms of morphology and properties, yet little light was shed on what specific sequence domains led to the self-coacervation. In this Chapter, the peptide domains of DgHBP-1 and DgHBP-2 that determine their self-coacervation ability have been established. Shorter peptide variants based on the full sequences of DgHBPs were synthesized, followed by self-coacervation testing. Biophysical characterization of the morphology and structure of DgHBPs-based peptides and variants are also presented in this chapter.

6.1 Introduction

In previous chapters, the sequence and structural similarities between DgHBPs and other proteins known to self-coacervate were firstly described. The hypothesis that DgHBPs are also able to form coacervates was then proposed and verified. However, a few critical questions remained unanswered. Notably, what specific sequence domains are responsible to the self-coacervation ability of DgHBP-1 and DgHBP-2? It was established to be likely due to the hydrophobic C-terminus repeating motifs, but which ones in particular and is there a minimum number of penta-repeats? And what roles do N-termini domains play?

Since self-coacervation appears to hold such a significant place in the bio-processing of squid beak, answering the above questions will help to design biomimetic functional materials based on the similar principles that the squid beak follows, namely assembling and delivering biomacromolecules at high concentration *via* self-coacervation, followed by controlled dehydration and inter-chain cross-linking. Therefore, further investigations on the specific domains of DgHBP-1 and DgHBP-2 is necessary to provide molecular-scale design guidelines that may be applied to mimic *D. gigas* beak. To achieve this goal, a series of short peptides (5 to 26 amino acids) and longer DgHBP variants (50 to 100 amino acids) based on the distinctive C-terminus repeating motifs of DgHBP-1 and DgHBP-2 was designed and prepared. In order to produce the selected variants, molecular methods were employed to mutate, express, and cleave DgHBP-1 and DgHBP-2 at pre-selected locations. The designs and sequences of those peptides and variants are shown in Table 6.1 and Figure 6.1. Finally, the self-coacervation abilities of all of the above DgHBPs-based peptides and variants were tested and their sequence patterns leading to self-coacervation were analyzed. This strategy allowed to gain a deeper understanding of the sequence design governing self-coacervation ability.

Table 6.1 Sequences of DgHBPs-based peptides and/or their positions on full protein sequences. “A. A.” are short for “amino acids”.

DgHBPs-based Peptides and Variants	Length or Description	Sequences and/or Positions on Full proteins	
DgHBP-1-based peptides and variants	5 A. A.	GAGFA	
		GHGLY	
	10 A. A.	GAGFA GHGLY	
		GAGFA GHGPY	
	23/25 A. A.	GAGFA GHGLH GAGFA GHGLY GAGFA (G98-A122, GA-25)	
		GHGLH GAGFA GHGLY GAGFA GHGLH (G103-H127, GH-25)	
		GHGLY GAGFA GHGLH GFA GHGLY (G113 – Y135, GY-23)	
	cleaved variants	(6×His-TEV-) G67-Y161 (DgHBP-1-V3-C)	
		(6×His-TEV-) G98-Y161 (DgHBP-1-V4-C)	
		G103-Y161 (DgHBP-1-V11-C)	M1-Y102-K (DgHBP-1-V11-N)
		G108-Y161 (DgHBP-1-V12-C)	M1-Y107-K (DgHBP-1-V12-N)
		G113-Y161 (DgHBP-1-V13-C)	M1-Y112-K (DgHBP-1-V13-N)
		G118-Y161 (DgHBP-1-V14-C)	M1-Y117-K (DgHBP-1-V14-N)
		G123-Y161 (DgHBP-1-V15-C)	M1-Y122-K (DgHBP-1-V15-N)
DgHBP-2-based peptides and variants	5 A. A.	GHGVY	
		GHGPY	
	26 A. A.	GHGVY GHGVY GHGPY GHGPY GHGLY W (GHGVY-G117-Y136-W, GW-26)	
	cleaved variants	A82-R172 (DgHBP-2-V1-C)	M1-R81 (DgHBP-2-V1-N)

Figure 6.1 Schematic representation of DgHBPs-based peptides and variants (25 amino acids and above). (a) DgHBP-1 full protein, GA-25, GH-25, GY-23, V3-C, V4-C, V11-C to V15-C, and V11-N to V15-N. (b) DgHBP-2 full protein, GW-26, V1-C and V1-N.

6.2 Experimental Methods

A range of equipment and techniques were applied in order to prepare DgHBPs-based peptides and variants and to characterize their self-coacervation properties.

6.2.1 Preparation of Short Peptides

Preparation of DgHBPs-based penta-peptides (GAGFA, GHGLY, GHGVY and GHGPY), deca-peptides (GAGFA GHGLY and GAGFA GHGPY), and polypeptides (GA-25, GH-25, GY-23 and GW-26) was done through contracted solid phase peptide synthesis (> 95 % purity) by GL Biochem (Shanghai) Ltd.

6.2.2 *E. coli* Recombinant Expression

Preparation of DgHBPs-based variants (DgHBP-1-V3-C, DgHBP-1-V4-C, DgHBP-1-V11-C/N to DgHBP-1-V15-C/N, and DgHBP-2-V1-C/N) was done using the following protocols:

For DgHBP-1-V3-C and DgHBP-1-V4-C, mutated genes were cloned to pNIC28-Bsa4 (containing cleavable N-terminal His-tag and kanamycin resistance gene) vector and transformed into *E. coli* cells, and then expressed according to previously described in *Subchapter 3.2*. The expressed protein variants were mainly found in the inclusion bodies of the cells. Cell lysis was conducted by microfluidization 6 times, followed by buffer washing steps and acetic acid purification methods. For more details regarding *E. coli* recombinant expression and primary protein purification, please refer to *Sub-chapters 3.2 and 3.3*.

For DgHBP-1-V11-C/N to DgHBP-1-V15-C/N, full protein genes of DgHBP-1 were first mutated by inserting a single Lys-encoding codon sequence in the genes, thereby allowing to introduce trypsin-cleavage sites spliced each protein into their C- and N- terminus variants at pre-determined sites. This strategy was possible because DgHBP-1 and DgHBP-2 do not contain any Lys. Thus, introducing a single Lys along the sequences allowed to control the trypsin-induced cleavage. Specifically, the mutated genes for each variant were expressed in *E. coli* cells as mutated proteins (DgHBP-1-V11 to DgHBP-1-V15) with an extra Lys after Y102, Y107, Y112, Y117 and Y122, respectively. After cell lysis by microfluidization, the expressed mutated proteins were found in both the inclusion bodies

of the cells as well as the soluble fraction. Subsequent purification then included buffer wash steps and acetic acid purification. More details regarding *E. coli* recombinant expression and primary protein purification are described in *Sub-chapters 3.3* and *3.4*. After purification, trypsin cleavage experiments (described in *Section 6.2.3*) were conducted for each variant to cut the full protein in two parts, the C-terminus and the N-terminus. Different cutting locations led to different variants (DgHBP-1-V11-C/N to DgHBP-1-V11-C/N).

For DgHBP-2-V1-C/N, two Arg (R81 and R172) within the full DgHBP-2 sequences were used as trypsin cleavage (described in *Section 6.2.3*) sites to obtain DgHBP-2-V1-N and DgHBP-2-V1-C, respectively. Details are described in *Sub-chapters 3.3* and *3.4*.

6.2.3 Trypsin Cleavage of DgHBP-1-V11 to DgHBP-1-V15 and DgHBP-2

After expression of mutated DgHBP-1 variants (V11 to V15) and DgHBP-2 full protein, trypsin cleavage experiments were conducted to cleave the proteins after the single Lys or Arg residue present in the full-length proteins and obtained the target variants. Details of trypsin cleavage experiment are as follow:

Following expression and acetic acid purification, the lyophilized DgHBP variants were dissolved in 50 mM Tris-HCl buffer (pH 8.8) at a concentration of 1 mg/ml. Then, Trypsin (“Trypsin Gold”, MS grade, Promega) dissolved in 50 mM acetic acid was added into the protein solutions at a volume dilution ratio of 1/1000. The mixed solution was incubated at 37°C for 1 to 3 h. 10 mg/ml PMSF of stock solution (isopropanol) was then added to terminate the cleavage at volume a dilution ratio of 1/100. SDS-PAGE was conducted on post-cleavage products to confirm successful cleavage. Reverse phase HPLC was subsequently employed on the cleaved proteins to separate the N-terminus C-terminus products.

6.2.4 Reverse Phase Chromatography

Following recombinant expression and an initial step of acetic acid purification, reverse phase chromatography (RPC) was applied to purify DgHBP-1-V3-C and DgHBP-1-V4-C. RPC was also employed to separate cleaved C-terminus and N-terminus of DgHBP-1-V11 to DgHBP-1-V15, and DgHBP-2-V1. For more details regarding RPC set up, please refer to *Section 3.3.4*.

6.2.5 SDS-PAGE and Western Blot

SDS-PAGE was used to confirm the molecular weight of expressed and purified DgHBPs-based variants, as well as to examine the results of trypsin cleavage experiments. For more details, please refer to *Section 3.5.1*.

During the expression of DgHBP-1-V3-C and DgHBP-1-V4-C, it was discovered that over-expression of impurity proteins obscured the target protein variant bands on SDS-PAGE gels. Therefore western blots with anti-bodies specific for the His-tag was used to confirm the expression of DgHBP-1-V3-C and DgHBP-1-V4-C. For more details, please refer to *Section 3.7.4*.

6.2.6 MALDI-ToF

MALDI-ToF was used to analyze the molecular weight of expressed and purified DgHBPs-based variants, as well as to examine the results of trypsin cleavage experiments. For more details, please refer to *Section 3.5.2*.

6.2.7 Optical Microscopy

Optical microscopy was used to visualize and capture the morphology of coacervates by DgHBPs-based peptides and variants, which appeared in the form of fluidic micro-droplets in equilibrium solution. For more details, please refer to *Section 3.6.1*.

6.2.8 Turbidity Measurements

Turbidity measurement at 600 nm wavelength is a common practice in the study of coacervates to quantify the extent of coacervation. Here it was used to monitor self-coacervation by DgHBPs-based peptides and variants. More details about measurement set-up and buffers used at different pHs and ionic strengths were described in *Section 3.6.2* and *Sub-chapter 3.4*.

6.2.9 Dynamic Light Scattering

Dynamic light scattering (DLS) measurements were performed to monitor the size distribution of coacervates by DgHBPs-based peptides and variants. The Z-average values were used as effective diameters of coacervates. For more details, please refer to *Section 3.6.3*.

6.2.10 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was used to study the secondary structures of DgHBPs-based peptides and variants. Measurements were done as described in *Chapter 3.5.4*

6.2.11 Rheology

Rheological measurements were conducted to characterize the mechanical properties of coacervates by DgHBPs-based peptides and variants. Dynamic torsional shear mode and oscillatory frequency sweep were applied in order to obtain the viscosity and storage/loss modulus of coacervates, respectively. For more details, please refer to *Section 3.6.6*.

6.3 Results and Discussion

6.3.1 Expression and Purification of DgHBP-1-V3-C and DgHBP-1-V4-C

Two variants, DgHBP-1-V3-C and DgHBP-1-V4-C, designed after the C-terminal sequence of DgHBP-1, were expressed in *E. coli* system. As shown in Figure 6.2a, both proteins were abundantly expressed. However the impurity proteins masked the target protein bands on the gel (rectangles in Figure 6.2a, left panel). Therefore, in order to further confirm the existence of the target proteins, Western blot of the protein bands after blotting the gels onto a PVDF membrane was carried out with a specific Anti-6X His tag® antibody.^[1] The results (Figure 6.2a, right panel) verified the expression of both protein variants.

To further purify the expressed DgHBP-1-V3-C and DgHBP-1-V4-C, a purification strategy similar to DgHBPs was applied, consisting of acetic acid extraction followed by reverse phase HPLC. The HPLC chromatograms shown in Figure 6.2c and Figure 6.2d indicated successful separation of DgHBP-1-V3-C and DgHBP-1-V4 from impurity proteins, as further verified by SDS-PAGE results of the target protein peak fractions (Figure 6.2b). Finally, MALDI-ToF experiments were conducted to confirm the molecular weight of DgHBP-1-V3-C and DgHBP-1-V4, which turned out to be 11,490 Da and 8,622 Da, respectively, and which correspond to their predicted values.

Figure 6.2 Expression and purification of DgHBP-1-V3-C and DgHBP-1-V4-C. (a) SDS-PAGE (left panel) and Western blot (right panel) of expressed DgHBP-1-V3-C and DgHBP-1-V4-C fractions after cell lysis. L: ladder. S: supernatants after cell lysis. P: pellets after cell lysis. (b) SDS-PAGE of DgHBP-1-V3-C and DgHBP-1-V4-C fractions after reverse phase HPLC purification (shown in (c) and (d)). (c-d) Reverse phase HPLC chromatograms of DgHBP-1-V3-C and DgHBP-1-V4-C, respectively. (e-f) MALDI-ToF spectra of purified DgHBP-1-V3-C and DgHBP-1-V4-C, respectively.

6.3.2 Expression, Cleavage and Purification of DgHBP-1-V11-C/N to DgHBP-1-V15-C/N

In order to study the self-coacervation ability of the N- and C- terminus of DgHBP-1 separately, as well as to determine the essential sequence repeating motifs required for coacervation, N- and C- variants (DgHBP-1-V11-C/N to DgHBP-1-V15-C/N) of different lengths and containing different number of repeats were expressed, cleaved, and purified.

The recombinant expression of DgHBP-1-V11 to DgHBP-1-V15 in *E. coli* system is shown in Figure 6.3a. Each protein variant was expressed abundantly and the gels also show a large amount of impurities. Subsequent acetic acid purification method identical to that of full the protein were conducted and the results are shown in Figure 6.3b. Reverse phase HPLC purification was also utilized to further purify the variants, as shown in Figure 6.3c and Figure 6.3d.

After DgHBP-1-V11 to DgHBP-1-V15 were successfully expressed and purified, trypsin cleavage was conducted. As shown in Figure 6.4a, successful cleavage for all variants was achieved, as demonstrated by the presence of two distinct bands of cleavage products on the gels. One band migrated at a slightly lower Mw than the non-cleaved variant, while the other band was lower than 10 kDa. Further separation of the C- and N- variants was conducted by reverse phase HPLC (Figure 6.4b). The purification was confirmed by SDS-PAGE (Figure 6.4c) and the molecular weights of cleaved C- and N- variants were ascertained by MALDI-ToF (Figure 6.4d), which matched to their respective predicted values in Figure 6.1a.

Figure 6.3 Expression and purification of DgHBP-1-V11 to DgHBP-1-V15. **(a)** SDS-PAGE of expressed DgHBP-1-V11 to DgHBP-1-V15 after cell lysis. **(b)** Acetic acid extraction of DgHBP-1-V11 to DgHBP-1-V15 after cell lysis. L: ladder. T: total fraction (before extraction). S1: supernatants after first acetic acid extraction. P1: pellets after first acetic acid extraction. S2: supernatants after second acetic acid extraction of P1. P2: pellets after second acetic acid extraction of P1. **(c)** Reverse phase chromatogram of DgHBP-1-V15 (representative of V11 to V15). **(d)** SDS-PAGE of DgHBP-1-V15 fractions after reverse phase purification (representative of V11 to V15).

Figure 6.4 Trypsin cleavage and purification of DgHBP-1-V11-C/N to DgHBP-1-V15-C/N. (a) SDS-PAGE of trypsin cleavage experiments. L: ladder. BC: protein variant before cleavage. AC: products after cleavage. (b) Reverse phase HPLC chromatograms of DgHBP-1-V11 cleavage products (representative of V11 to V15). (c) SDS-PAGE of cleaved V11 to V15 after RPC. (d) and (e) MALDI-ToF of DgHBP-1-V11-C and DgHBP-1-V11-N after reverse phase HPLC separation (representative of V11-C/N to V15-C/N), respectively.

6.3.3 Cleavage and Purification of DgHBP-2-V1-C/N

Studies on the C- and N- terminus of DgHBP-2 were also conducted by exploiting trypsin cleavage. Since there are two Arg residues (R81 and R172) in the original sequence of DgHBP-2, trypsin cleavage experiment was conducted with (*rec*)DgHBP-2. There is a tripeptide variant (GYL, at the end of C-terminus, after R172) existent in the cleavage products other than DgHBP-2-V1-C and DgHBP-2-V1-C, yet it was neglected in the discussion considering its trivial size and molecular weight compared to the two larger variants.

As shown in Figure 6.5c, the trypsin cleavage of DgHBP-2 full protein was successful, as suggested by the two bands of low molecular weight after cleavage compared to full length DgHBP-2 before cleavage. Subsequent reverse phase HPLC separation of C- and N-terminus did not work as expected, as indicated by the single peak in the chromatogram (Figure 6.5a) that still contained both variants (Figure 6.5c). Thus another separating method was developed.

It was previously proposed that the C-terminus parts of DgHBPs are mainly responsible for their self-coacervation behavior. Inspired by the fact that coacervation is often used as a purification method for proteins in food industry^[2], coacervation was employed as a method to separate the C-terminus of DgHBP-2 from N-terminus. After reverse phase HPLC purification of the cleavage products, the lyophilized peak fractions were dissolved in 10 mM acetic acid, then pipetted into pH 6.5 phosphate buffer (same as DgHBPs coacervation conditions) at a volume ratio of 9/1. Visual turbidity was immediately noticed (Figure 6.5b), indicating successful self-coacervation. After incubation of the coacervation system for 2 hours, the solution was centrifuged and both the pellet and supernatants were loaded onto an SDS-PAGE. The results (Figure 6.5d) showed that the method was successful: the pellet (which formed the coacervates) mostly contained the C-terminus, while the supernatant (which remained in the solution) contained the N-terminus. Further MALDI-ToF analysis (Figure 6.5e and Figure 6.5e) corroborated this result, confirming the respective molecular weights of DgHBP-2-C and DgHBP-2-N.

Figure 6.5 Trypsin cleavage and purification of DgHBP-2-V1-C/N. (a) Reverse phase HPLC chromatogram of DgHBP-2 cleavage products. (b) Formation of coacervates by adding the dissolved HPLC fractions into phosphate buffer. (c) SDS-PAGE of DgHBP-2 after cleavage, and fractions after reverse phase HPLC. L: ladder. BC: before cleavage. AC: after cleavage. (d) SDS-PAGE of DgHBP-2-V1-C/N separation by the coacervation method. L: ladder. P: pellets after coacervation. S: supernatants after coacervation. (e) and (f) MALDI-ToF spectrum of DgHBP-2-V1-C and DgHBP-2-V1-N, respectively, after separation by coacervation.

6.3.4 Self-coacervation Abilities of DgHBPs-based Peptides and Variants

Self-coacervation ability was tested for all DgHBPs-based peptides and variants in Table 6.1, using the conditions described in *Chapter 3*, and the test results are summarized in Table 6.2.

None of the DgHBPs-based penta-peptides or deca-peptides were able to coacervate, indicating that perhaps a minimum number of residues (likely more than 10) is needed to ensure inter-molecular interaction and, in turn, coacervation.^[3,4] For the 23/25/26 amino acid long peptides, Only GY-23 (based on DgHBP-1) and GW-26 (based on DgHBP-2) formed coacervates. This suggests that differences in the C-terminus repeat pattern between DgHBP-1 and DgHBP-2 strongly influence their self-coacervation abilities. Specifically, for DgHBP-1, the GAGFA/GFA-GHGXX' modular peptide (with a length of 23/25 amino acids) of DgHBP-1 did not necessary result in self-coacervation unless two GHGLYs are included. On the other hand, GW-26 is made of 5 GHGXY repeats but lacked the GAGFA motifs, and this construct was able to self-coacervate. These results point towards the critical role of GHGXY motifs to drive the self-coacervation ability of DgHBPs.

For DgHBPs-based variants, it is clear from the results that the C-termini are responsible for coacervation in both DgHBP-1 and DgHBP-2. Indeed most expressed or cleaved C-variants formed coacervates, while almost none of N-variants did. In addition, the results indicated that the minimum number of GHGXY to induce coacervation was 2. This is illustrated by DgHBP-1-V14-N and DgHBP-1-V13-C, where both contained 2 of such repeats, and all formed coacervates. Any number of GHGXY repeats above 2 led to coacervation, while any number below that did not, as confirmed by other variants in the series.

In summary, two important conclusions were reached: **(i)** the C-termini of DgHBPs are responsible for their coacervation abilities; and **(ii)** a minimum number of GHGXY repeats

is required for coacervation: at least 2 GHGXY repeats are needed in order to trigger coacervation.

Table 6.2 Coacervation test results summary of DgHBPs-based peptides and variants.

Red highlighted: no coacervation. Green highlighted: coacervation. At least 2 of the motifs “GHGXY” must be present to induce coacervation.

Type		DgHBP-1-based		DgHBP-2-based	
Peptides	penta-peptides	GAGFA		GHGVY	
		GHGLY		GHGPY	
	deca-peptides	GAGFA GHGLY		GHGLY GHGLY	
		GAGFA GHGPY			
	23/25/26-peptides	GA-25, GAGFA GHGLH GAGFA GHGLY GAGFA (G98-A122)		GW-26, GHGVY GHGVY GHGPY GHGPY GHGLY W (GHGVY-G117-Y136-W)	
		GH-25, GHGLH GAGFA GHGLY GAGFA GHGLH (G103-H127)			
GY-23, GHGLY GAGFA GHGLH GFA GHGLY (G113-Y135)					
Protein Variants	expressed and purified	DgHBP-1-V3-C, MHHHHHHSSGVDLG TENLYFQSMGHGG FA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA GHGLH GAGFA GHGLY GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G67-Y161)			
		DgHBP-1-V4-C, MHHHHHHSSGVDLG TENLYFQSM GAGFA GHGLH GAGFA GHGLY GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G98-Y161)			
	trypsin-cleaved and purified	DgHBP-1-V11-C, GHGLH GAGFA GHGLY GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G103-Y161)	DgHBP-1-V11-N, MQLYGAPAVG GVVENAVNAA ESGAAATHDA QGAYAEADTA GVLVDVNHA EH HDGVHDAS GYGFG GLAG HGGFA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA (M1-Y102-K)	DgHBP-2-V1-C, AAQGANAVH NDAAQSHSAL AAANTFGHGH APYAAY GHGVY GHGPY GHGPY GHGLY GHGLY GHGPY GHGLY GHGAF GHGLN AYAPL VGHGLR (A82-R172)	DgHBP-2-V1-N, MQFFGAGPFN TAHHS AVSDA AAAHHDAAGE YAQNAATG LL DTHHNNHDM THDLANGYGL HEHDEQHHGL ADGLHQEYAA R (M1-R81)
		DgHBP-1-V12-C, GAGFA GHGLY GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAHY (G108-Y161)	DgHBP-1-V12-N, MQLYGAPAVG GVVENAVNAA ESGAAATHDA QGAYAEADTA GVLVDVNHA EH HDGVHDAS		

			GYGFG GLAG HGGFA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA GHGLH (M1-Y107-K)		
		DgHBP-1-V13-C, GHGLY GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G113-Y161)	DgHBP-1-V13-N, MQLYGAPAVG GVVENAVNAA ESGAAATHDA QGAYAEADTA GVLDVNHAEH HDGVHDAS GYGFG GLAG HGGFA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA GHGLH GAGFA (M1-Y112-K)		
		DgHBP-1-V14-C, GAGFA GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G118-Y161)	DgHBP-1-V14-N, MQLYGAPAVG GVVENAVNAA ESGAAATHDA QGAYAEADTA GVLDVNHAEH HDGVHDAS GYGFG GLAG HGGFA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA GHGLH GAGFA GHGLY (M1-Y117-K)		
		DgHBP-1-V15-C, GHGLH GFA GHGLY GAGFA GHGLG LGGLH GALGH GALAH Y (G123-Y161)	DgHBP-1-V15-N, MQLYGAPAVG GVVENAVNAA ESGAAATHDA QGAYAEADTA GVLDVNHAEH HDGVHDAS GYGFG GLAG HGGFA GHGLY GPGFA GHGLL GAGYA GLGLH GAGFA GHGLH GAGFA GHGLY GHGLH (M1-Y122-K)		

6.3.5 Morphology of Coacervates by DgHBPs-based Peptides and Variants

Micrographs of coacervates by DgHBPs-based peptides and variants are shown in Figure 6.6. For DgHBP-1-based peptides/variants, the droplet size varies from 1-20 μm , indicating merging behavior of coacervates. This was especially obvious for GY-23 in Figure 6.6a. The merging of two droplets indicates a low surface energy of coacervates phase, and perhaps also strong inter-droplet attractions. For DgHBP-2 based peptides, the droplet size is about 1-5 μm , and hardly any merging behavior was observed. GW-26 coacervates remained fluidic and stable in terms of particle size, and almost no aggregation was seen after even 30 min of coacervation.

The morphology of GY-23 and GW-26 was similar to that of their “mother” protein, DgHBP-1 and DgHBP-2, respectively. This suggests that they represent the key consensus/sequence of respective protein to coacervation. The difference in their particle size, might come from the difference in their sequence. In other words, $(\text{GHGXY GAGFA})_n$ ($n > 2$) forms coacervates that merge, while $(\text{GHGXY})_n$ ($n > 4$) forms coacervates that do not merge.

6.3.6 Rheological Properties of Coacervates by DgHBPs-based Peptides and Variants

In *Chapter 5*, rheological studies for GY-23 and GW-26 coacervates were discussed. Here the figures were reproduced in Figure 6.7a-b. In the shear rate range 0.1-100 s^{-1} , GY-23 showed a much higher viscosity than GW-26. In the same frequency range (0.1 to 1 Hz), the storage modulus of GY-23 (*ca.* 20-40 Pa) was much higher than GW-26 (*ca.* 0.1-0.4 Pa). Also, for GW-26, the G'/G'' value was below 1 across all frequencies, which indicates a liquid-like behaviour. Since the storage modulus scales with the cross-linking density^[5], it can be concluded from these data that GY-23 has a tighter network of complexes. Since the main difference between these two peptides is the replacement of GHGXY motif in GW-26 by GAGFA/GFA motifs, one can surmise that GAGFA/GFA motifs enhances the cross-linking density in the coacervates, and also that the two types of repeating motifs

(GHGXY GAGFA *vs.* GHGXY) in DgHBP-1 and DgHBP-2 are responsible for their rheological differences. However further experimental evidence is needed to definitely confirm this proposition.

Figure 6.6 Optical microscopy images of coacervates formed by DgHBPs-based peptides and variants. (a) to (f) Optical micrographs of coacervates formed by GY-23, GW-26, DgHBP-1-V3-C, DgHBP-1-V4-C, DgHBP-1-V11-C and DgHBP-1-V15-N, respectively.

Figure 6.7 Rheological studies on coacervates of DgHBPs-based peptides GY-23 and GW-26. (a) Viscosity of coacervates formed by GY-23 and GW-26. (b) Frequency sweep measurements of storage moduli (G' , closed symbols) and loss moduli (G'' , open symbols) of coacervates formed by GY-23 and GW-26.

6.3.7 Structure of DgHBPs-based 23/26 Peptides

CD spectroscopy was only conducted for DgHBPs-based peptides. While pentapeptides and decapeptides only gave no reliable results due to their short chain lengths, self-coacervating DgHBP variants made of 23/26 amino acids provided ample information (Figure 6.8). Both GY-23 and GW-26 in their solution state displayed significant anti-parallel β -sheet conformation with the maximum peak centered around 220 nm.^[6] In their coacervate state, this conformation remained dominant with a rise of minima at around 195 nm, suggesting random coil content.^[7,8] Comparing these CD spectra to the native protein spectra in *Chapter 4* and *5*, consistency was observed, with a significant amount of β -structures. However, for proteins parallel β -sheets is the dominant form while for peptides, anti-parallel β -sheets accounted for most of the detected β -sheet structures. The different propensity towards parallel and anti-parallel β -sheets could be due to their relative energy in proteins and peptides.^[9–11] According to Chou et al^[9], for proteins with residues such as Gly, Ala and Leu, anti-parallel β -sheets is favored over parallel β -sheets; while for proteins with Val, Phe and Tyr, parallel β -sheets is more favorable. Since all above residues are

present in DgHBPs and DgHBPs-based peptides, it is possible that their relative abundance determined the dominant form of β -sheets in each biopolymer. However, more evidence is needed to definitely confirm this proposition.

Figure 6.8 CD spectra of DgHBPs-based peptides. (a) GY-23 at solution phase (1 mg/ml, 10 mM acetic acid) and coacervates phase (final concentration: 1 mg/ml). (b) GW-26 at solution phase (1 mg/ml, 10 mM acetic acid) and coacervates phase (final concentration: 1 mg/ml).

6.3.8 Variables Affecting Coacervation of GW-26 and DgHBP-1-V3-C: pH and Ionic Strength

Turbidity measurement for GW-26 self-coacervation at different pH and ionic strength (Figure 6.9a) suggested that, similar to DgHBP-2 full protein (Section 5.3.5), GW-26 also has a narrow window where coacervation occurs. The maximum turbidity was achieved around pH 7.5 with 1-2 unit's "window", which is around GW-26's isoelectric point (7.15). While increasing ionic strength of the solution, the coacervation window for GW-26 was widened, showing "screening effect" by salt.^[12,13] However, the salt screening effect for GW-26 is less obvious than DgHBP-2. When the ionic strength was raised to 1 M, the starting pH value for GW-26's coacervation was 6.5, whereas for DgHBP-2 coacervation could be initiated at pH as low as 4. This is likely ascribed to high content of His residues in the N-terminus of DgHBP-2. His carry a positive charge positive at pHs away below 6.05 and low ionic strength, and His-rich proteins/peptides are dominated by long-range

electric double-layer repulsions at those conditions, therefore could not coacervate easily.^[14] At high ionic strength however, the salt screening effect would neutralize these repulsions, and proteins/peptides became easier to coacervate even at low pHs, and the coacervation pH range for His-rich proteins/peptides is more obvious than His-poor proteins/peptides. Compared to DgHBP-2 (whose N-terminus are rich in His), GW-26 has a lower His content, and possibly resulted in a weaker salt screening effect at 1 M ionic strength.

Turbidity measurements for DgHBP-1-V3-C coacervates at different pHs and ionic strengths are shown in Figure 6.9b. The C-terminus of DgHBP-1 formed coacervates under different conditions than the full-length DgHBP-1. A wider pH range (pH > 6) could induce coacervation. Furthermore, increasing the ionic strength to 1 M could directly induce coacervation, regardless of pH. These results might be due to the fact that a 6 × His – Tev tag was present in DgHBP-1-V3-C variant (added for post-expression purification), which increased the His content within and improved its coacervation ability and salt-screening effects.^[15] To further study this variant's coacervation, His-tag need to be removed.

Figure 6.9 Turbidity measurement for (a) GW-26 and (b) DgHBP-1-V3-C coacervates, respectively. Turbidity as a function of pH at 100, 500, and 1000 mM ionic strength.

6.4 Conclusions

This chapter has looked closely into the sequences of DgHBPs and tested the possible domains of the sequences that are responsible for self-coacervation. Using a trypsin cleavage method, a number of C- and N- terminus variants of DgHBP-1 and DgHBP-2 were produced. After testing all peptides and variants for their self-coacervation behaviors, it was verified that the C-terminal domains harboring the highly repetitive motifs for both DgHBP-1 and DgHBP-2 are governing their self-coacervation behavior, and that a minimum of two of the penta-repeat motif **GHGXY** are required. Further characterization of these peptides and variants revealed that the morphology and rheological properties of coacervates are highly dependent on the repeating domains of the protein/variant/peptide, with **GHGXY GAGFA**-repeating polymers exhibiting merging droplets and higher storage moduli, in comparison to **GHGXY**-repeating polymers. Structural characterization on 23/26 peptides showed anti-parallel β -sheet conformation for both GY-23 (DgHBP-1 based) and GW-26 (DgHBP-2 based), whereas parallel β -sheet conformation is present in their full-length versions. Finally, turbidity measurements for GW-26 suggested that coacervation is affected by pH and ionic strength, although comparing to DgHBP-2, its coacervation pH range are less affected by ionic strength, because of the absence of His-rich N-terminus.

The investigation on DgHBPs-based peptides and variants helped deepening the understanding of DgHBPs coacervation mechanisms. It is now firmly established that the C-termini of DgHBPs are triggering the coacervation; and that the **GAGFA** repeats in DgHBP-1 are likely to form short β -sheet domains that enhance the cross-link density of coacervates, which is significant for their merging behavior and gel-like rheological properties.

References

- [1] P. N. Hengen. *Trends Biochem. Sci.* **1995**, 20, 285–286.
- [2] M. S. M. Wee, S. Nurhazwani, K. W. J. Tan, K. K. T. Goh, I. M. Sims, L. Matia-Merino. *Food Hydrocoll.* **2014**, 42, 130–138.

- [3] L. Ringstad, A. Schmidtchen, M. Malmsten, U. U. V, P. O. Box, S.- Uppsala. **2006**, 5042–5050.
- [4] B. Frohm, J. E. DeNizio, D. S. M. Lee, L. Gentile, U. Olsson, J. Malm, K. S. Akerfeldt, S. Linse. *Soft Matter* **2015**, 11, 414–21.
- [5] M. Rubinstein, R. H. Colby. *Polymer Physics*, Oxford University Press New York, **2003**.
- [6] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos. *Proc. Natl. Acad. Sci.* **2015**, 112, E3095–E3103.
- [7] R. K. Dukor, T. A. Keiderling. *Biopolymers* **1991**, 31, 1747–1761.
- [8] A. F. Drake, G. Siligardi, W. A. Gibbons. *Biophys. Chem.* **1988**, 31, 143–146.
- [9] K. C. Chou, G. Nemethy, H. A. Scheraga. *Biochemistry* **1983**, 22, 6213–6221.
- [10] J. S. Richardson. *Proc. Natl. Acad. Sci.* **1976**, 73, 2619–2623.
- [11] K.-C. Chou, H. A. Scheraga. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, 79, 7047–7051.
- [12] H. I. Petrache, T. Zemb, L. Belloni, V. A. Parsegian. *Proc. Natl. Acad. Sci.* **2006**, 103, 7982–7987.
- [13] R. Perez-Jimenez, R. Godoy-Ruiz, B. Ibarra-Molero, J. M. Sanchez-Ruiz. *Biophys. J.* **2004**, 86, 2414–2429.
- [14] A. W. Adamson, A. P. Gast, others. **1967**.
- [15] M. Carson, D. H. Johnson, H. McDonald, C. Brouillette, L. J. DeLucas. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2007**, 63, 295–301.

Chapter 7

Discussion and Future Work

This Chapter summarizes the findings of Chapter 4 to 6 about DgHBP-1 and DgHBP-2, their self-coacervation behaviors, and how it is controlled by their sequences. A model for D. gigas beak material production in vivo is proposed. Preliminary studies about chitin network, His-catechol cross-linking and DgHBPs' interplay are also presented in this Chapter and provide the foundation for further studies. Potential future work on D. gigas beak system and prospects of squid beak-inspired applications are proposed in this chapter.

7.1 General Discussion and Conclusion

As a biological material with unique structural properties, *D. gigas* beak has attracted the interest of material scientists, notably because of its fully organic composition (chitin, proteins, pigments and H₂O).^[1,2] While prior work had unveiled information about the relations between its compositional and mechanical gradients^[3], as well as about the proteins present within the beak^[4], a deeper understanding on the specific roles these proteins play and the physico-chemical transitions they undergo during formation of the beak remained unclear..

This thesis has provided in-depth understanding of DgHBPs and their self-coacervation behavior, a biophysical phase separation process proposed as an essential step during the maturation and sclerotization of *D. gigas* beak. From these studies, the two hypotheses and three objectives mentioned in *Chapter 1* have been addressed, and a proposal of beak formation model is presented in this Section.

7.1.1 Preparation of DgHBPs and DgHBPs-based Variants

In *Chapter 4*, a recombinant expression system for DgHBP-1 and DgHBP-2 has been established using *E. coli* cells that carry protein-encoding genes. Expressed proteins after cell lyses were initially solubilized with different solvents, the results of which led to subsequent development of their respective purification strategies involving solvent (acetic acid) extraction from lysed cells. After that, chromatograph techniques were utilized to obtain DgHBP-1 and DgHBP-2 in high purity for further studies.

The establishment of expression and purification methods of DgHBPs also provided the basis for systematic studies of truncated DgHBPs (*Chapter 6*), where N- or C-terminal parts of full-length DgHBPs proteins were produced. This was achieved using two strategies: (i) direct mutation and expression of the protein-encoding genes, followed by purification (DgHBP-1-V3-C and DgHBP-1-V4-C); (ii) insertion of lysine into the primary structure (DgHBP-1-V11-C/N to DgHBP-1-V15-C/N) or exploitation of the wild-type

arginine from the protein sequences, (DgHBP-2-V1-C/N) followed by expression, purification and cleavage experiment by trypsin, a lysine/arginine-specific digestive enzyme.

The above work has addressed the first objective proposed in this thesis, namely to set up an efficient recombinant expression and purification system for DgHBP-1 and DgHBP-2.

7.1.2 Characterization of DgHBPs and Self-coacervation

In *Chapters 4* and *5*, comprehensive characterization studies were conducted on the biophysical properties of DgHBP-1 and DgHBP-2, which established the self-coacervation behavior of DgHBPs. The sequence similarities of DgHBPs to the well-known self-coacervation protein, tropoelastin described in *Chapter 2*, translated into intriguing structural parallel, notably their dynamic nature, as revealed in *Chapter 4*. Further investigation into different variables (morphology, kinetics, structure, affecting parameters, etc.) of the self-coacervation behavior of DgHBP-1 and DgHBP-2 was presented later in *Chapter 5*, revealing its fundamental importance to the sclerotizing process of *D. gigas* beak: (i) delivering large amount of proteins to their destination; (ii) maintaining the micro-structure of the beak scaffold.

The studies on DgHBPs and their self-coacervation behavior substantiate the central hypothesis in this thesis, namely that a biopolymer permeation process might take place during the beak formation. During this process, large amount of DgHBPs are transported from the secretion sites located in the soft bucciblast gland tissues to the densely cross-linked rostrum region of the beak. Such a procedure would require a highly-concentrated and fluidic protein phase in order to infiltrate the biopolymer through the confined chitin nano-fibrous network, and self-coacervation meets all important criteria for this process to occur. This has also addressed the second objective of this thesis, namely to perform comprehensive studies on the biophysical features of DgHBP-1 and DgHBP-2, and to provide explanation as how they can be the dominant component of the beak hard rostrum.

7.1.3 Sequence Studies on DgHBPs Self-coacervation

In *Chapter 6*, the specific sequence motifs causing DgHBPs self-coacervation were identified, through preparation of a range of DgHBPs-based peptides and variants followed by respective coacervation studies. Based on the previously established expression system combined with simple protein engineering methods, C- and N-terminus variants of DgHBP-1 and DgHBP-2 were prepared. These investigations verified that was the C-termini are indeed responsible for the self-coacervation ability of DgHBPs. The importance of GHGXY pentapeptide repeats was highlighted, and it was revealed that at least two copies of the consensus repeat is needed for DgHBPs-based peptide/variants to exhibit self-coacervation ability.

Work on DgHBPs- peptides and variants have confirmed the second hypothesis of this thesis, namely the highly modular sequences of DgHBPs play a central physico-chemical role during beak formation. As it turned out, the highly conserved and repetitive domains of the C-terminus of DgHBPs are vital for the self-coacervation ability, which linked the molecular design of biopolymers to the overall properties of biological materials, as also revealed by other countless examples in nature, such as elastin^[5,6] and spider silk^[7,8]. This has also addressed the third objective of this thesis, namely to establish key primary structure-biophysical properties relationships of DgHBPs.

7.1.4 Proposed Model for *D. gigas* beak formation

Based on the findings of this this thesis, a bio-processing model for *D. gigas* beak is proposed.^[9] As shown in Figure 7.1, three steps involving different components of the beak are involved during beak maturation. First, chitin nano-fibers and chitin-binding DgCBPs are secreted together at the initial stage of beak formation, constituting the initial soft, hydrated scaffold. The function of DgCBPs, although still elusive, is believed to control hydrophobicity of chitin fibers, as natural chitin hydration level in *D. gigas* beak's proximal regions were found to be around 75 wt%, an abnormally high value for chitin is usually no more than 50 wt%.^[10,11] Secondly, DgHBPs are secreted and, after exposure to seawater

environment (around pH 7 and ionic strength 0.5 M), they form coacervate droplets (2 to 10 μm) characterized by a high fluidity. The coacervates phase infiltrates within the chitin network thanks to its “shear-thinning” behavior, namely its viscosity greatly decreases with increased flow rate (v) due to confined pore size in the network ($v \sim d^{-3}$, where d is the mean pore size).^[12] Finally, cross-linking of DgHBPs coacervates within the chitin network leads to the dehydration and strengthening of the beak scaffold, and involves the formation of His-catechol cross-links.^[4] It is worth mentioning that His residues in DgHBPs are almost always flanked by 2 Gly, which provides flexibility of the peptide backbone locally and minimizes the steric hindrance for His reactivity, both are critical for cross-link formation.^[9]

In summary, the demonstration that DgHBPs can self-coacervate, the identification of peptide motifs controlling this coacervation ability, as well as a deeper understanding about the natural material “bio-manufacturing” was developed. Novel insights for biomimetic engineering using peptides analogs and recombinant proteins were thus presented in this thesis.

Figure 7.1 Proposed model for *D. gigas* beak formation.^[9] Three steps in beak bioprocessing are (i) DgCBPs and chitin nano fibers are secreted to form the beak scaffold; (ii) coacervation of DgHBPs enables protein infiltration within the chitin network; (iii) aggregation and cross-linking by DgHBPs causes the dehydration and stiffening of the beak material.

7.2 Reconnaissance Work for Future Studies

Here, a few preliminary experiments on the interactions between DgHBP-1 and DgHBP-2, cross-linking of DgHBP-1 coacervates, and modification of chitin network is briefly presented and pave the way for future work described in *Sub-chapter 7.3*.

7.2.1 Mixed Coacervation by DgHBP-1 and DgHBP-2

To study the interactions between DgHBP-1 and DgHBP-2, a mixed coacervation experiment was conducted as following: DgHBP-1 stock solution (10 mg/ml, 10 mM acetic acid) and DgHBP-2 stock solution (11 mg/ml, 10 mM acetic acid) were mixed, then added in the phosphate buffer in the same way as for pure protein coacervation. The resulting coacervate phase was examined under the optical microscope (Figure 7.2), which showed micro-droplets similar to those in DgHBP-1 and DgHBP-2 coacervates, yet coalescence of micro-droplets was not observed as in DgHBP-1 coacervates. This suggests possible interference by DgHBP-2 (Figure 7.2a). Also, a dendritic microstructure was noticed in the system, unlikely to be precipitates, which indicates possible assembly of proteins in the coacervates form (Figure 7.2b). However, more comprehensive investigations on the interplay between DgHBPs are needed in order to confirm this hypothesis.

Figure 7.2 Optical microscopy of DgHBP-1 and DgHBP-2 mixed coacervates. (a) Presence of uniform micro-droplets in coacervates form. (b) Dendritic microstructures formed by proteins, possibly due to self-assembly.

7.2.2 His-catechol Cross-linking of DgHBP-1 Coacervates

In order to verify the cross-linking ability of DgHBPs coacervates, a His-catechol cross-linking experiment was conducted for DgHBP-1 coacervates. As shown in Figure 7.3a, stock solution of DgHBP-1 was pipetted into phosphate buffer that contains 4-methylcatechol, a compound found to be involved in the cross-links of *D. gigas* beak, the origin of which is still unclear.^[4] Then, the oxidizing reagent NaIO₄ was added to the coacervation system to trigger the histidine-Catechol cross-linking reaction. This led to the immediate formation of brownish complex that sedimented from the solution. Subsequent SDS-PAGE of the cross-linked product showed a faint smeared band on the gel (Figure 7.3b, lane 4), possibly due to the insoluble nature and high molecular mass of the cross-linked proteins.

Figure 7.3 DgHBP-1 cross-linking experiment. (a) Experimental design. Samples taken for subsequent SDS-PAGE are marked with number 1 to 4. (b) SDS-PAGE of samples taken at different experimental stages. L: ladder. 1-4: different samples taken during the experiments with each number annotated in (a).

7.2.3 Chitin Network Preparation: Histidine Functionalization of Chitosan

Preliminary studies to mimic *D. gigas* beak from a chitin precursor scaffold was also conducted. Since native chitin is highly insoluble in water, its soluble derivative, chitosan was used here. Chitosan is obtained by deacetylation of chitin under alkaline conditions.^[10] Due to the sparse research on DgCBPs so far, it is sensible to first study interactions between chitin (chitosan) network and DgHBPs, especially with regard to the histidine-catechol cross-linking reactions that form within the network. In order to mimic this DgHBPs-chitin mixture system in laboratory, the idea to create histidine-catechol cross-links between DgHBPs and functionalized chitin network came up. However, the sidechain of chitin is occupied by acetyl groups, therefore chitosan was considered for this experiment. An approach using amidation to modify the side chain of chitosan with an imidazole ring was conducted (Figure 7.4a). The resulting compound was characterized by 1D-NMR to confirm the degree of substitution (DS %). As shown in Figure 7.4b, by calculating the detected imidazole hydrogen (H_a and H_b) peaks from grafted Boc-Histidine, the DS % of imidazole on chitosan side chain was 15.5 %, which indicates a decent level of substitution. This approach has sparked other possibilities to functionalize the amine group of chitosan such as catechol groups, which may provide further insights for studies on *D. gigas* beak material, and biomimicry of the beak, as catechol could also form His-catechol cross-links with DgHBPs, which strengthens the chitosan network.

7.3 Outlook and Future Work

7.3.1 Further Studies on DgHBPs

Despite comprehensive studies on self-coacervation of DgHBP-1 and DgHBP-2, elucidating the structure of the coacervate phase is still incomplete, largely due to their highly dynamic nature and also due to the limitations of current structural characterization methods. More advanced techniques, for example solid-state NMR (ssNMR), could be exploited to monitor instantaneous conformational transitions during the self-coacervation process. Also, more characterization work (CD, SAXS, rheology, etc.) could be performed

(a)

Chitosan (92.4% deacetylated) Boc-Histidine Histidine-Chitosan (HIS-CS)

Solvent: H₂O (145ml, pH 5.5)1eq = n(-NH₂) = 1g × 92.4% ÷ Mw ( ) = 5.74 × 10⁻³ mol**(b)**

Figure 7.4 Histidine functionalization of chitosan. (a) Substituting reaction of the amide group of chitosan with Boc-Histidine. (b) NMR spectra of the reaction product indicating successful substitution. DS %: degree of substitution.

on DgHBPs-based peptides and variants in order to reveal the sequence-structure-mechanical property relationship of *D. gigas* beak system. More peptides could also be designed and synthesized based on DgHBPs' sequences, and especially in the repetitive C-terminus domain.

In addition, studies on the third member of the DgHBPs family, DgHBP-3, could be carried out, building on the current framework for DgHBP-1 and DgHBP-2. Although not undertaken in this thesis due to limitations in time and resources as well as due to its sequence homology to DgHBP-1 (Figure 2.10), studies on DgHBP-3 could provide additional information, for example to discern whether glycine-rich N-termini has an impact on the biophysical behavior of DgHBPs (Figure 2.10). After all, Nature is not known for redundancy, in that almost every species or protein has adapted for a unique reason during evolution.

Finally, hetero-protein interactions among DgHBPs could be looked into more closely. Whether these proteins perform interactive or cooperative roles during beak formation could be examined by structural and biophysical characterization of their mixed coacervates, as presented in *Section 7.2.1*, to grasp if any inter-protein interactions occur.

7.3.2 Studies on DgCBPs

As another family of proteins present in *D. gigas* beak, DgCBPs has so far drawn less attention, possibly due to their irregular modular structures and also because they are less abundant in the hard rostrum of the beak.^[9] However, it is of great interest to investigate their possible ability to modify the hydrophobicity of chitin, and to study how they fix chitin nano-fibers into a network with leaflet-like structures that functions as a scaffold prior to cross-linking.^[2] Recombinant DgCBPs could be produced in the similar way as DgHBPs, followed by primary characterization on their biophysical and structural properties, then chitin/chitosan binding experiments could be conducted on different DgCBPs with an emphasis on the hydrophobicity changes of the binding products. Furthermore, protein-protein interaction studies could be carried out for all DgBPs, which

would include both DgHBPs and DgCBPs, to determine any possible interactions between these two protein families.

7.3.3 Studies on Chitin Network

As mentioned in *Section 7.2.3*, functionalization of chitosan could provide a mimic system to study cross-linking reactions within the nature beak scaffold. Besides histidine-chitosan, chitosan could be functionalized with other moieties such as 4-methycatechol, tyrosine, DHCA (3, 4-Dihydroxyhydrocinnamic acid), and Dopa (3,4-dihydroxyphenylalanine)., Then “mock infiltration” experiments could be done using DgHBPs coacervates and modified chitosan, followed by inducing catechol-histidine cross-linking with control of the cross-linking reactions, with the ultimate goal to reproduce the stiffness gradient observed in native beaks. In order to confirm the nature of His-catechol cross-links formed in these experiments, MS/MS method could be conducted as done by Tan et al.^[9]

7.3.4 *D. gigas* Beak Bio-inspired Materials

Artificial composite materials could be developed according to the same design principles as found in *D. gigas* beak. For example, hydrogels with an inherent dehydration gradient could be achieved via controlled cross-linking among its components. Composite films containing the same components (chitin, water, DgBPs and pigments) as *D. gigas* beak could be synthesized *in vitro* by protein-engineering combined with polymeric materials processing, providing both bio-compatibility and mechanical strength, with a designed gradient of stiffness that could solve the current issue of interfacial stresses prosthetic implants.^[13,14]

Finally, coacervation of DgHBPs could inspire the design of synthetic polypeptides that could also form micro-droplets of coacervates under physiological conditions, which could act as vesicles for nutrient micro-encapsulation^[15,16], drug-delivery^[17,18] or protocell simulation^[19–22], all of which have attracted growing attention in recent years.

References

- [1] A. Miserez, Y. Li, J. H. Waite, F. Zok. *Acta Biomater.* **2007**, 3, 139–149.
- [2] D. J. Rubin, A. Miserez, J. H. Waite, C. Jérôme, J. S. Stephen. in *Adv. In Insect Phys.*, **2010**, pp. 75–133.
- [3] A. Miserez, T. Schneberk, C. Sun, F. W. Zok, J. H. Waite. *Science* **2008**, 319, 1816–1819.
- [4] A. Miserez, D. Rubin, J. H. Waite. *J. Biol. Chem.* **2010**, 285, 38115–38124.
- [5] L. D. Muiznieks, F. W. Keeley. *J. Biol. Chem.* **2010**, 285, 39779–39789.
- [6] L. Nivison-Smith, A. Weiss. *Regen. Med. Tissue Eng. - Cells Biomater.* **1997**, 323–340.
- [7] M. Xu, R. V Lewis. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87, 7120–7124.
- [8] R. V. Lewis. *Chem. Rev.* **2006**, 106, 3762–3774.
- [9] Y. Tan, S. Hoon, P. A. Guerette, W. Wei, A. Ghadban, C. Hao, A. Miserez, J. H. Waite. *Nat. Chem. Biol.* **2015**, 11, 488–495.
- [10] M. Rinaudo. *Prog. Polym. Sci.* **2006**, 31, 603–632.
- [11] R. A. A. Muzzarelli. *Chitin*, Elsevier, **2013**.
- [12] R. L. Fournier. *Basic Transport Phenomena in Biomedical Engineering*, CRC Press, **2011**.
- [13] P. Fleckman, J. E. Olerud. *Biomed. Mater.* **2008**, 3, 34006.
- [14] J. D. Fox, J. R. Capadona, P. D. Marasco, S. J. Rowan. *J. Am. Chem. Soc.* **2013**, 135, 5167–5174.
- [15] Z. Xiao, W. Liu, G. Zhu, R. Zhou, Y. Niu. *J. Sci. Food Agric.* **2014**, 94, 1482–1494.
- [16] K. A. Black, D. Priftis, S. L. Perry, J. Yip, W. Y. Byun, M. Tirrell. *ACS Macro Lett.* **2014**, 3, 1088–1091.
- [17] Y. Luo, Q. Wang. *Int. J. Biol. Macromol.* **2014**, 64, 353–367.
- [18] N. R. Johnson, Y. Wang. *Expert Opin. Drug Deliv.* **2014**, 11, 1829–1832.
- [19] T.-Y. Dora Tang, C. Rohaida Che Hak, A. J. Thompson, M. K. Kuimova, D. S. Williams, A. W. Perriman, S. Mann. *Nat. Chem.* **2014**, 6, 527–33.
- [20] S. Koga, D. S. Williams, A. W. Perriman, S. Mann. *Nat. Chem.* **2011**, 3, 720–724.
- [21] D. S. Williams, S. Koga, C. R. C. Hak, A. Majrekar, A. J. Patil, A. W. Perriman, S. Mann. *Soft Matter* **2012**, 8, 6004.
- [22] X. Huang, A. J. Patil, M. Li, S. Mann. *J. Am. Chem. Soc.* **2014**, 136, 9225–9234.