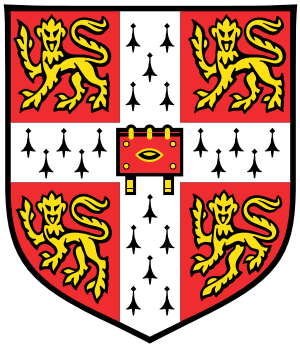
Biologically Inspired Engineering for Protein Stabilisation



Huseyin Burak Caliskan

Department of Engineering

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Homerton College

April 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Huseyin Burak Caliskan
April 2019

Biologically Inspired Engineering for Protein Stabilisation

Huseyin Burak Caliskan

Abstract

Every year millions of lives are saved by vaccination and millions more could be saved with more efficient vaccination coverage. Thermostability of vaccines is one of the major reasons why many children who need vaccines most do not get all the shots they need. Vaccines need to be kept below 8°C throughout manufacturing, delivering and shipping which requires a cold-chain. This involves carrying vaccines in a cool box for days even weeks to small villages or islands with no electricity or any other infrastructure. One of the grand challenges in global health is to produce vaccines that do not require refrigeration. Here, a biologically inspired vaccine stabilisation method is proposed. The method is based on the fact that proteins entrapped in the fossil avian eggshell crystals are well preserved up to thousands of years in Africa where high-temperature vaccine stability is needed. The persistence of the ancient intra-crystalline proteins suggests that protein incorporation into the inorganic host could retard protein degradation for long periods of time without refrigeration. In the present work, biomimetic protein incorporation in calcium carbonate was investigated to produce intra-crystalline heat-stable vaccines. Intra-crystalline protein persistence in the fossil record has been shown for ostrich eggshell proteins and the most durable proteins are known to belong to the C-type lectin-like protein family. It is thought that at least one C-type lectin-like protein is found in the eggshells of every species. The most well known C-type lectin-like protein is the OC-17 from chicken eggshell. OC-17 and other C-type lectins have the highest concentration in eggshells compared to other eggshell proteins which suggests that they could lead understanding of efficient protein incorporation in calcium carbonate crystals. Apart from ostrich and chicken, C-type lectins from other species have not been studied in detail. For this reason, eggshells from 14 species were studied in the present work. The aim of the investigation is to quantify the effect of the organic matter on the eggshells which could allow inferring protein incorporation efficiency of different species. The mechanical properties of the eggshells were studied with nanoindentation. This allows probing the differences in the elastic modulus and hardness of eggshells which are affected by the intra-crystalline protein content. It was found that the elastic modulus differs among species, which is lowest at around 10 GPa for Bali myna and highest at around 60 GPa for ostrich. Similarly, the hardness changes from around 1 GPa for Bali myna to around 3 GPa for rhea. The chemical analysis

was conducted with IR spectroscopy. The deviations in the absorption peaks of eggshells compared to pure calcium carbonate allows probing the extent of the amorphous structure of the eggshells. In addition, the comparison of the full-width at half maximum values of vibration modes in each spectrum provides information of the crystal order of the eggshells. Comparison of mechanical and chemical analyses of different species offers insights on the protein content of different eggshells which could lead identification of the most efficient C-type lectins in terms of protein incorporation ability. Because of the importance of eggshell C-type lectins for the vaccine preservation method proposed in the present work, the OC-17 was studied here in detail. First, a purification method was developed to extract OC-17 from chicken eggshell. Liquid chromatography was used to extract OC-17, OC-23 and lysozyme. Then, lysozyme was removed from solution using a lysozyme binding protein. In addition, OC-17 was cloned to synthesize a recombinant C-type lectin-like eggshell protein in bacteria for the first time. Growth conditions were optimized and OC-17 expression was verified with anti-Histag antibody. Lastly, the model protein incorporation into synthetic calcium carbonate was studied. Because the incorporated and reconstituted protein structure is of utmost importance, the secondary and tertiary protein structures of incorporated proteins were analyzed with circular dichroism and intrinsic tryptophan fluorescence spectroscopy. The effect of both the reconstitution and incorporation were studied independently as intra-crystalline proteins could denature during the interaction with the reconstitution buffer or during the incorporation process. It was shown that the model proteins BSA, lysozyme and diphtheria anti-toxin could be reconstituted using EDTA without structural change. The incorporation was found efficient for BSA only with an efficiency of %84 in terms of total amount of protein incorporated into the crystals. The secondary structure of BSA was shown to be stable during reconstitution and incorporation. A structural change in the tertiary structure was observed in BSA. Possible improvements to incorporate a target protein as a fusion construct using the OC-17 or using the ostrich intra-crystalline peptides as 'incorporation tags' were discussed in order to move towards a real-world application of biomimetic vaccine stabilisation.

Acknowledgements

I would like to thank to Prof Robin Langley for his supervision, continuous support and insightful comments on my work, Prof Sir Mark Welland for his supervision and for providing an excellent environment at the Nanoscience Centre, Prof Jim Huntington for accepting me in his lab at the Cambridge Institute for Medical Research, Prof Jim Woodhouse for lecturing me on Fourier Analysis (even if I was the only student), Prof Abir Al-Tabbaa for her help as my advisor at the Department of Engineering, Dr Melanie Keene for her help as my tutor at Homerton, Dr Alex Buell for his help on protein chemistry, Dr Janet Kumita for her continuous support, comments, suggestions and for our fruitful discussions on many topics, and for being extremely kind and generous. Dr Iris Batalha for help on chemistry, molecular biology and comments on the present work, Dr F. Isik Ustok for the cloning experiments, Dr David Labonte for the nanoindentation measurements and comments on the present work, Dr Gillian Fraser and Owain Bryant for kindly providing samples for the cloning experiments, Dr Richard Langford for his help on the EBSD, Dr Eric Rees for our discussion on FRET, Dr Caroline Trotter for our discussion on vaccines and especially for introducing me to the NIBSC, Dr. Paul Stickings, Dr Barbara Bolgiano and Dr Caroline Vipond from the NIBSC for inviting me to present my research on vaccine thermostability problem and for kindly providing the diphtheria antibody samples.

I would like to thank to the Lucasian Professor of Mathematics Mike Cates from Department of Applied Mathematics and Theoretical Physics for accepting me as a member of Edwards Centre for Soft Matter Research. I would like to thank to Prof Sir Christopher Dobson's group for accepting me to their team and to the Department of Chemistry. I would like to thank to the Cavendish Laboratory, Physics of Medicine for accepting me to their department where I did some of the most important experiments presented in this thesis. I would like to thank to Dr Emre Ozer from ARM and his wife Susan, Dr Julia Kenyon, Ian Willis and Stephen Reece from the Department of Medicine, Ewa Klimont from the Department of Chemistry, Dr Penny Hamlyn from GE Healthcare, Dr Jenny Molloy from the Department of Chemical Engineering and Biotechnology, Dr Chris Pearson and Dr Andrew Walsh from the Cambridge Enterprise, Prof Neel Joshi from the Wyss Institute for Biologically Inspired Engineering, Harvard University. I would like to thank to the Cambridge Trust for supporting

my PhD.

I would like to thank to my PhD examiners Dr Thierry Savin from the University of Cambridge and Dr Colin L. Freeman from the University of Sheffield for their comments and suggestions which improved the thesis in many aspects substantially.

Finally and particularly, I would like to thank to my parents and my sister for their continuous support.

I would like to dedicate this thesis to my parents ...

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

ACC Amorphous calcium carbonate

ATR-FTIR Attenuated total internal reflection Fourier-transfrom infrared spectroscopy

BCG Bacillus Calmette-Guérin

CD Circular dichroism

Diph.ab Diphtheria antibody (antitoxin)

DTaP Diphtheria, tetanus, acellular pertussis

DTwP Diphtheria, tetanus, whole-cell pertussis

EBSD Electron back scatter diffraction

EDTA Ethylenediaminetetraacetic acid

FLIM Fluorescence lifetime imaging

FRET-FLIM Forster resonance energy transfer-Fluorescence Lifetime Imaging

GFP Green fluorescent protein

HepA Hepatitis A

HepB Hepatitis B

Hexavalent DTaP-hepatitis B-Hib-IPV

Hib Haemophilus influenzae type b

HPLC High pressure liquid chromatography

HPV Human papillomavirus

IPTG Isopropyl {b}-D-1-thiogalactopyranoside

IPV Inactivated poliovirus vaccine

JE Japanese encephalitis

LC-MS Liquid chromatography mass spectrometry

Men A Meningitis A,

Men C Meningitis C

MMR Measles, mumps, rubella

NCBI BLAST National Center for Biotechnology Information-Basic Local Alignment Search Tool

OC-17 Ovocleidin-17

OC-23 Ovocleidin-23

OPV Oral polio vaccine

Pentavalent DTwP-hepatitis B-Hib

Pneumo Pneumococcal

PS Polysaccharide

PS-PCV PS-protein conjugate vaccine

SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEM Scanning electron microscopy

T, DT, dT Tetanus, diphtheria tetanus, diphtheria (low-dose) tetanus

TCSPC Time correlated single photon counting

Chapter 1

Introduction

1.1 A Challenge in Global Health: Temperature Sensitivity of Vaccines

One of the most demanding ventures Homo sapiens[1] are being tackled is the eradication of diseases. The fact that this has been accomplished only once in the entire human history lucidly explains the scope of the problem. The only disease that has ever been eradicated intentionally is the smallpox[2]. Smallpox was one of the most infectious and fatal diseases that had crippled people around the world for thousands of years. The mummy of pharaoh Ramses V had been found to carry the fingerprint of smallpox's rash of pustules still recognized since the death of the pharaoh in 1147 BC [3] which makes him the earliest known victim of the disease. Back then there was not much to be done against diseases but writing around 410 BC, in the History of Peloponnesian War, Thucydides describes the plague epidemic and notes a crucial detail: "No one caught the disease twice or if he did the second attack was never fatal [4]." Physicians of that time could have been intrigued on why people do not suffer from a contagious disease twice, however, discovering the intricacies of the adaptive immunity had to wait more than two millennia.

The very first scientific journal, the Philosophical Transactions of the Royal Society, had published a letter on smallpox in January 1714 [5]. It was sent to the Royal Society by an Italian physician, Emanuel Timonius in 1713 from Istanbul where Edward Wortley Montagu will travel to be the British ambassador to the Ottoman Empire [6]. It is thought that Lady Mary Montagu, the wife of the ambassador, had been aware of the letter published in the

[1] To be precise, the actual scientific term describing our species is not Homo sapiens ('wise man') but Homo sapiens sapiens, "... so wise we had to say twise [sic]" [1]

[2] In fact, there is one more disease, the rinderpest, which has also been eradicated but it was not a human disease rather it only infected some animal species. The eradication of rinderpest was declared in 2011 [2].

Philosophical Transactions because of her interest in the subject. Her brother had died of smallpox before she contracted the disease herself [6]. The anxiety for her children's health possibly motivated her to learn more about the inoculation practice (i.e. vaccination as we know it today) being applied in Istanbul. Through her connections to the Royal Society, Lady Mary had found the people who successfully operate the practice and inoculated her son in the presence of physicians of the embassy [6].

Although discovered independently in many different places including China, India and Africa, the practice of inoculation that sparked scientific interest probably originated from the Circassian women in the Ottoman Empire's capital [7]. The immunity against smallpox among Circassians had likely been discovered because of their unaffected faces from smallpox scars. The lack of immunity against Circassian beauty serendipitously led one of the most effective medical interventions in history, that is, vaccination.

A vaccine is any substance (prepared generally from the causative agent) that could trigger the body to produce antibodies for providing immunity against the agent. The first Nobel Prize in Physiology or Medicine in 1901 was awarded to Emil Adolf von Behring for his work on antibody therapy (back then known as serum therapy, that is, collecting blood plasma from an animal challenged with the diphtheria toxin and transferring the plasma to the people suffering from diphtheria[3]). The antibody response is the basis and the aim of vaccination but the development of new vaccines is known to be notoriously difficult (see [9] for a historical account on the subject).

Vaccination is one of the two crucial interventions that drastically improved global health[4]. Vaccination had been propelled into the scientific world by Edward Jenner [10] who had also coined the term 'vaccine'. He had published his findings in a book, Inquiry [11], after collecting the cowpox virus from a milkmaid and injecting it into his gardener's son, seemingly along with wishful thinking [7]. From the beginning of his efforts to spread vaccination around the world, he had realized that during the delivery of vaccines to different countries, high temperature deactivates the virus and decreases the efficiency of vaccination. Today, heat degradation of vaccines is still a major problem and the present work is aimed at developing a new method for vaccine stabilisation.

It is worth noting that the inoculation practice did not start with Jenner. Because of the efforts of Lady Montagu, including a book she wrote, inoculation did spread in England. Despite the success of inoculation, the procedure had met with skepticism. In 1760, Daniel Bernouilli worked on (probably the first) mathematical model in epidemiology [12] (also see

[3] https://www.nobelprize.org/prizes/medicine/1901/summary/. Extraction of antibodies from immunized animals is, interestingly, still in use today [8].

[4] The other major contributor to global health is sanitation, which is mostly related to providing clean drinking water and disposing sewage appropriately.

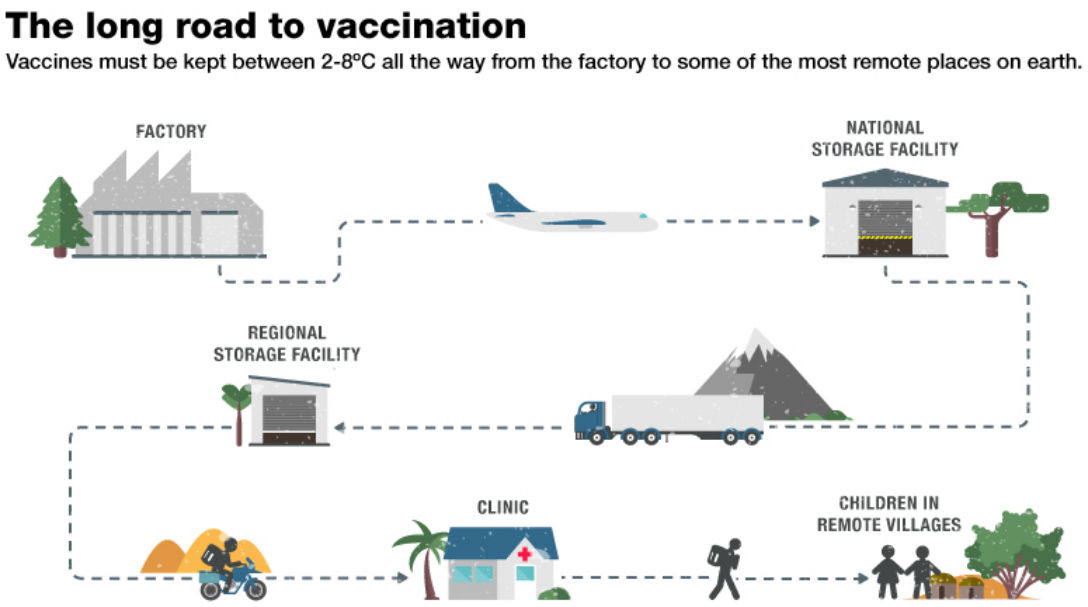


Figure 1.1: Vaccines must be stored around 2- 8 ℃ all the way from production to the vaccination of people in the most remote places. Image credit, World Economic Forum, weforum.org.

Prof Tom Körner's book 'The Pleasures of Counting' for a mathematical analysis of cholera by Dr John Snow in the 19^{th} century London [13]). Bernouilli analyzed the benefits of inoculation against smallpox and concluded that the inoculation is beneficial[5] (also [12]). Not everyone was convinced by Bernoulli's work though, one of whom was D'Alembert [12].

One of the challenges in global health is to produce vaccines that do not require refrigeration[6]. Most of the vaccines are made of proteins and are inherently heat-sensitive. Keeping vaccines in optimal storage conditions is a major problem in some of the developing countries, particularly, in Africa. The problem of vaccine delivery, unbeknownst to many in the developed world, is a tremendous obstacle to those[7] operating with poor settings to vaccinate people in the hard-to-reach areas in developing countries. Fig 1.1 depicts how vaccines are delivered to the most remote places around the world.

The World Health Organization (WHO) is the main institution responsible for the assessment of vaccines and orchestrate vaccination programs around the world. According to the

[5] 'Mathematics and Smallpox' a lecture on the history of inoculation by Prof Tom Körner from the University of Cambridge. https://www.gresham.ac.uk/lectures-and-events/mathematics-and-smallpox

[6] The production of heat-stable vaccines was listed as one of the grand challenges in global health by Bill & Melinda Gates Foundation (https://gcgh.grandchallenges.org/challenges)

[7] One such group is a non-profit organization Médecins Sans Frontières/Doctors without Borders. See 'Keeping vaccines cold in hot climates' at https://www.doctorswithoutborders.org/what-we-do/medical-issues/vaccination

'Temperature Sensitivity of Vaccines' report [14] by the WHO, all of the vaccines must be stored between 2 ℃ and 8 ℃ *at all times*. Keeping vaccines below 8 ℃ throughout their journey until they are used requires the 'cold-chain' system [15]. The cold-chain is crucial in Africa, where temperatures could reach above 45 ℃. Despite the importance of temperature sensitivity of vaccines, the stability data have many gaps, particularly for temperatures above 25 ℃. A schematic representation showing the heat and freeze sensitivity of vaccines is given in Fig 1.2. A more detailed vaccine stability data could be found in the 'Summary of Stability Data for Licensed Vaccines' report[8] by PATH, an international non-profit organization to improve the health of people around the world.

The fact that eradication of a disease is possible implies that through vaccination, more diseases could be eradicated in the future (see [16] for the eradication of polio). A disease-causing agent goes extinct when it fails in finding a host to infect. Vaccinations campaigns are being conducted in developing countries where these efforts decrease the burden of contagious diseases with the overarching goal of complete eradication [17]. However, the deployment of vaccines are still tackled with conventional methods including carrying them in cool boxes along with ice packs. The problems of current vaccine protection methods are twofold. First, keeping vaccines in a cool box on ice packs is non-trivial because of the distances between city centres and remote villages where vaccination needs to be conducted. Second, some of the vaccines degrade if the temperature gets too low (i.e. if the vaccines freeze, which is an issue for the vaccines that contain aluminum adjuvant because freezing causes aggregation of aluminum on which the vaccine is adsorbed [14, 18]). The latter problem is a subtle one and hitherto it is equally (if not more) problematic than the high temperature degradation.

1.2 Biologically Inspired Engineering for Vaccine Stabilization

The present work proposes a biologically inspired method to address temperature sensitivity of vaccines. The proposed method is based on the fact that some peptide sequences survive for thousands of years within biomineral crystals, such as $CaCO_3$ crystals of fossil avian eggshells [19],[20],[21],[22],[23],[24]. The presence of proteins in the crystal lattice of biological minerals has been known in archeology since 1950s [25] (also see[26] and [27]) but imitating intra-crystalline incorporation phenomenon *in vitro* for practical applications is yet to be realized. This natural protein preservation in avian eggshells might be reverse-engineered to incorporate vaccine proteins in $CaCO_3$ and incorporated vaccines might be stored at high

[8] https://www.path.org/resources/summary-of-vaccine-stability-data/

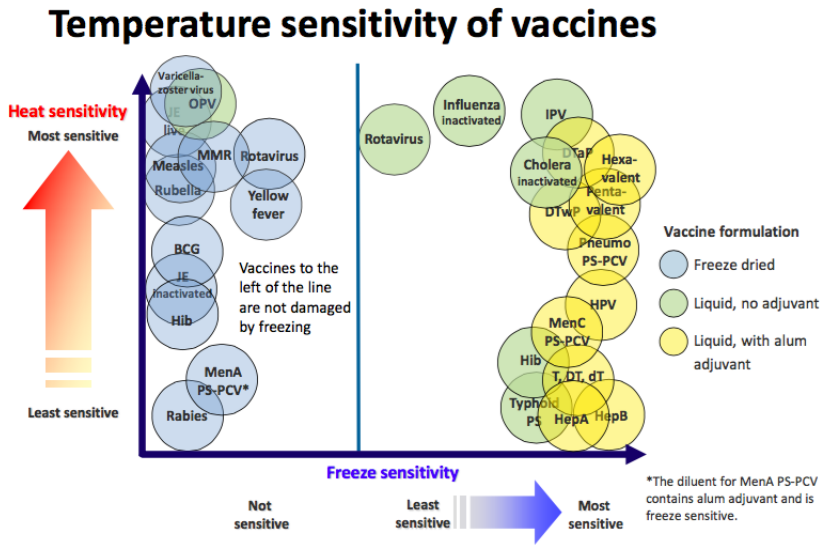


Figure 1.2: The heat and freeze sensitivity of freeze-dried and liquid vaccines. For abbreviations see the Nomenclature. Image credit, the PATH Vaccine Stability Report[7].

temperatures without refrigeration. Whenever needed, vaccine incorporated crystals could be reconstituted (i.e released from the inorganic host upon dissolution of $CaCO_3$) on-site as a ready-to-be injected vaccine solution. As explained in Section 1.1, degradation of vaccines at high temperatures is a major problem hampering vaccination programs in Africa. The fact that fossil eggshell samples excavated in Africa, where heat-stable vaccines are most needed, suggests that the proposed method of vaccine preservation could be able to keep vaccines safe in hot climates. For instance, it was shown that ostrich eggshell intra-crystalline proteins Struthiocalcins (SCA-1 and -2), extracted from eggshells excavated in Tanzania, are extremely well preserved up to 150,000 years at ambient temperature [19]. Inspiring from this phenomenon in nature, the overarching aim of the present work is to develop an intra-crystalline preservation method for heat sensitive vaccines.

It has been known for some time that upon dissolution of host $CaCO_3$ crystal, intra-crystalline proteins could be extracted and characterized even after millions of years [28]. Studying ancient proteins entrapped in fossil samples enables extraction of precious information [26]. For instance, entrapped proteins have been used to date fossil samples [29],[20] using amino acid geochronology (AAG) [30]. AAG is a method to calculate the age of an ancient sample using the racemic nature of amino acids, in particular, when the radiocarbon dating method (i.e. measuring half-time of ^{14}C) falls short.

Biological amino acids are left handed but these left handed amino acids spontaneously transform from the L-form (left handed) into the D-form (right handed) inside a crystal lattice over a span of thousands of years [30]. The analysis of the extent of transformation from L-

to-D form allows calculating the age of an ancient sample. It has been noted that the accuracy of AAG is higher when carbonate-based samples are used because carbonate minerals contain high concentration of proteins with which chemical analysis could be conducted [30]. The ostrich eggshell is an ideal source for AAG as it provides a thermodynamically closed system containing intra-crystalline proteins [29] . Once incorporated in the $CaCO_3$ lattice, proteins are isolated from the environment thus denaturation effects are largely avoided (i.e. unfolding is restricted possibly due to the immobilization of amino acid residues through strong binding towards $CaCO_3$).

There are two key issues to be addressed in order to harness intra-crystalline incorporation phenomenon for practical applications, such as increasing vaccine stability. Firstly, the effect of the incorporation process on the 3D structure of the incorporated protein needs analysis. In other words, the incorporation process itself should not cause structural changes on the protein. Secondly, the effect of high temperature on incorporated the proteins (i.e. while they are still in the $CaCO_3$ lattice) needs analysis to ensure that an increase in temperature does not cause denaturation, unfolding or any other structural change that might induce loss of function. It should be noted that, even if incorporation itself does not denature intra-crystalline proteins, dissolution of the host $CaCO_3$ to obtain proteins back could denature proteins because of the effect of dissolution buffer (e.g. low pH). Because there are two ways to dissolve $CaCO_3$, acid dissolution and metal chelation, both methods need to be tested to find the optimal reconstitution process.

Eggshell intra-crystalline proteins ovocleidin -17 (OC-17, from chicken) and Struthiocalcins (SCA-1 and SCA-2, from ostrich) have been identified and their X-ray 3D structures have been reported ([31] and [32] for OC-17 and SCAs, respectively). Furthermore, these 3D structures have been used in molecular dynamics simulations to probe the function of OC-17 [33] and to understand the survival mechanism of SCAs in ancient ostrich eggshell samples [19]. It should be added that the structural integrity of eggshell proteins could be unique to *in vivo* protein incorporation. More specifically, only the C-type lectin-like intra-crystalline proteins of eggshells might be preserved because of their strong binding energies to $CaCO_3$ (see [19] for SCAs). Still, the fact that eggshells offer exceptional protein preservation in hot climates in Africa could be used in a number of ways. One possibility is to produce recombinant OC-17-vaccine fusion protein to harness strong binding energy of OC-17 (or SCAs) to $CaCO_3$. This strategy might confer increased protein stability through retarding hydrolysis at the protein-mineral interface [19].

A schematic representation of the proposed method for intra-crystalline vaccine preservation is given in Fig 1.3. Incorporation of a target protein (e.g. a vaccine immunogen) in $CaCO_3$ lattice could be achieved by crystallization of $CaCO_3$ in the presence of the target protein or

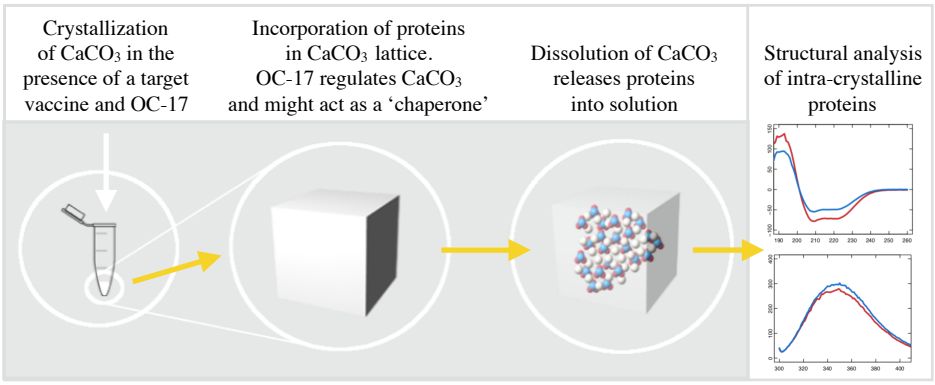


Figure 1.3: A schematic representation of the proposed method of vaccine incorporation into $CaCO_3$ mimicking intra-crystalline proteins in the avian eggshells. The OC-17 might act as a 'chaperone' to incorporate the target vaccine (or protein) that is not incorporated by itself. Alternatively, a genetic fusion of the target protein to OC-17 might also be used. Upon dissolution of the host $CaCO_3$ (i.e. reconstitution of the intra-crystalline protein), the 3D structure of the incorporated protein must be same with the native structure. The long term high temperature protein stability needs strong binding energy to $CaCO_3$ which might be achieved by genetically fusing the target protein to the OC-17 or to the best preserved SCA peptides in order to bind the protein complex selectively into $CaCO_3$ where hydrolysis is retarded.

more likely, the incorporation might require a 'chaperone' protein that makes it possible to incorporate the target protein into the crystal lattice. As will be discussed in Chapter 4, a viable candidate of being a chaperone protein is the OC-17. The scope of the present work including the experimental work to test the hypotheses expressed here is given at the end of this chapter (see Section 1.4).

1.3 Literature Review

1.3.1 Biomineralization Proteins in Natural Minerals

Although the importance of the proteins in biominerals as structural elements has been known for decades, the exact function of them is still largely unknown. The possible functions of proteins, such as sacrificial bonds or hidden lengths for enhanced fracture toughness, have been studied for a number of organisms. It is known, for instance, that the sacrificial bonds serve as an energy dissipation mechanism at the molecular scale [34]. The two of the well known examples of such reversible mechanisms include the mussel byssal thread and the gecko feet [35],[36].

The formation of a myriad of $CaCO_3$-based inorganic shells in both vertebrates and in-

vertebrates is controlled by specific proteins [37]. The biomineralization proteins could be divided into two major groups. The first group includes proteins that are found between the crystals of a biomineral assuming that the biomineral is not made of a single crystal itself, such as sea urchin skeletal elements [38]. This first group of proteins are called inter-crystalline proteins. The second group is the intra-crystalline proteins which are located inside the crystal lattices of biominerals. Contrary to the inter-crystalline proteins, intra-crystalline ones become entrapped into the growing crystal during biomineral formation [39]. The function of these proteins is still largely unknown and only a few of them have been studied in detail [40]. The question of whether or not they have a function to be entrapped into the crystal on purpose is open to inquiry. There is hitherto no direct evidence showing that these proteins have a specific function to become incorporated into the crystal lattice. It is probable that these proteins modify crystal growth or become entrapped into crystals during the process only because they are present in the crystallization medium. For instance, although its function is not modifying crystallization, lysozyme is present in the crystal lattice of the avian eggshells [41]. It has been known that the function of the lysozyme is to act as an anti-microbial defense system in the egg white [42]. The presence of the lysozyme in the crystal lattice suggests that the incorporation of lysozyme is a by-product of the crystal growth. In addition, an incorporated protein cannot function anymore as long as it is in the crystal lattice since the incorporation forms a closed system for a protein and isolates it from the environment.

There are a myriad of proteins in the avian eggshell crystal lattice and some of these proteins are only found in eggshells which makes the case intriguing. In addition, the concentration of intra-crystalline proteins differ substantially from each other [43]. The high concentration proteins might be structurally important for the eggshell formation while the low concentration ones might be considered as 'impurities'. It is worth noting that the incorporation is based on the chemical interactions between the amino acid residues of the proteins and the ions of the crystal [44],[45]. This means that the concentration of a specific intra-crystalline protein cannot be used as a direct evidence for assigning it a function to be an intra-crystalline protein. The fact that some of the intra-crystalline proteins are only found in one species and not in others suggests that at least some of these proteins should have an important function [43]. There are similar intra-crystalline proteins that have been conserved between species but there are also other proteins that have been evolved differently in different species to serve similar functions [46],[47]. It has been, for instance, proposed that some of the intra-crystalline proteins might have a structural function in the crystal lattice to inhibit crack propagation [38]. In other words, the intra-crystalline proteins could serve as a toughening mechanism in biominerals. The mechanism of toughening lacks concrete evidence except for the nacre which is the most studied and a well known example of organic-inorganic composite

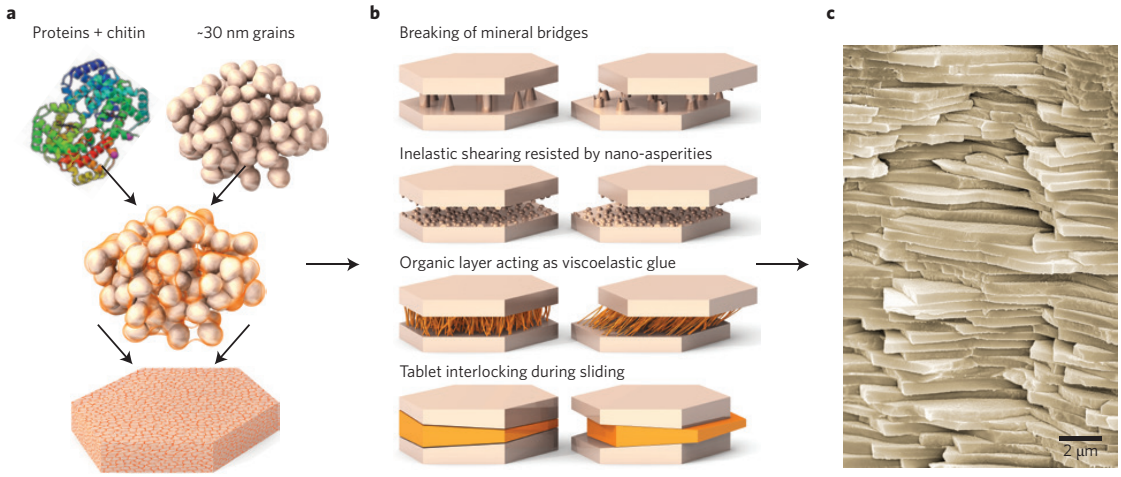


Figure 1.4: (a) The structure of the nacre. The inter-crystalline proteins (chitins) are located between $CaCO_3$ nanograins. There are several mechanisms (b) in the nacre structure for enhanced fracture toughness compared to the pure aragonite which is much more brittle. (c) The brick-and-mortar architecture is a main feature of the nacre where aragonite platelets are aligned on top of each other, between which thin protein layers are deposited [52].

material harnessing inter-crystalline proteins [48],[49],[50],[51]. Fig 1.4 shows the structure of the nacre where proteins are located outside the crystals of $CaCO_3$ instead of the inside of the crystal lattice. This distribution of proteins between crystals is the difference between the intra and inter-crystalline proteins.

A toughening mechanism has also been discussed for the eggshell [31]. It is reasonable that the intra-crystalline proteins might have evolved to bind onto specific crystal planes to avoid rupturing the crystal from its weakest point, that is, from the natural cleavage plane. The most common shell mineral among invertebrates is $CaCO_3$. Pure $CaCO_3$ crystal have a rhombohedral crystal lattice where calcium, carbon and oxygen atoms are located in a specific pattern as shown in Fig 1.5. The cleavage plane of $CaCO_3$ is the $\{10.4\}$ surface of the crystal. It has been shown *in vitro* that in the presence of some intra-crystalline proteins, $CaCO_3$ growth is affected in such a way that the protein binds onto the $\{10.4\}$ plane [31]. The evidence of an effect of molecular binding on the cleavage plane could be seen by cracking an eggshell and looking through the cross-section of shell. Contrary to any non-biogenic bulk $CaCO_3$ sample, the cross-section of a cracked eggshell would not yield an almost completely smooth surface (such as seen when a piece of chalk is broken) but is substantially rough in appearance. This fact suggests the presence of organic macromolecules bound to the $\{10.4\}$ plane. It is not unreasonable to assume that the crack propagation could be diverted, if not avoided completely, by harnessing the adsorption ability of proteins onto the cleavage planes of biominerals. This assumption is supported by the fact there are some specific proteins that

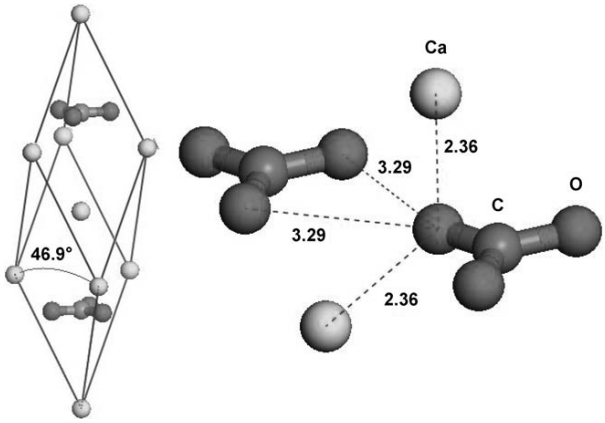


Figure 1.5: The rhombohedral crystal structure of pure $CaCO_3$ (left) and the position of atoms in a $CaCO_3$ crystal (right) with their inter-atomic distances in ångströms. C, O and Ca denote carbon, oxygen and calcium, respectively [57].

have the ability to recognize and interact with the growing crystal lattices [39]. One of the most intriguing protein family that could bind onto crystal planes is the antifreeze protein family found in fish species that inhabit cold waters close to the poles [53]. The function of the antifreeze proteins is to inhibit the formation of ice crystals using their specificity for the growing ice lattice [54]. It is crucial for the survival of the fish to avoid crystallization of ice in cold waters. There are some indications that the evolution of the proteins that have affinity towards specific crystal planes derived from a special family of proteins known as the 'Calcium dependent type lectin-like proteins' or shortly C-type lectin-like proteins [55],[56]. One of the members of this family is key for the eggshell formation and for this reason C-type lectin-like proteins will be explained later.

1.3.2 The Avian Eggshell: Function & Formation

A very brief history of the eggshell reveals the interest in the subject throughout the history. The very first record explaining the eggshell belongs, unsurprisingly, to Aristotle. He says, 'The eggshell is soft immediately after it is laid and hardened as it slowly cools down in contact with air' [58]. After vanishing for millennia, the subject was resurrected by a Cambridge scholar William Harvey saying, "...ascertained for certain by infallible experience that the egg in the uterus is almost always covered with a hard shell" [59]. Writing enthusiastically in 1625 Edward Topsell says: "Without the knowledge of fowles natural philosophie was very maymed [sic]"[60]. Thomas Wentworth Higginson recaps the subject in the 19^{th} century: "I think that, if required on pain of death to name instantly the most perfect thing in the universe, I should risk my fate on a bird's egg" [61].

The avian egg has the indisputable function of protecting and flourishing the growing offspring within [62]. The word egg evokes almost instantly a chicken image in the mind but long before the evolution of birds, dinosaurs had reproduced through eggs for millions of years. The closest relatives of birds were bipedal predatory dinosaurs, such as flightless Citipati and Troodon [63]. The fossilized eggs of these dinosaurs show bird-like features, such as the shell microstructure. These eggs were 15–18 cm long in size which is large compared to their body size. The eggs were found to be long and symmetrical unlike the sauropod eggs which were spherical in shape. Interestingly, gigantic 40 tonne sauropods like the Diplodocus and Brachiosaurus had laid eggs with a size close to an ostrich egg [64]. The shape of an egg, more specifically, the ellipticity and asymmetry of them, has been recently found to be correlated with the flight ability [65].

The avian eggshell is a paradoxical package. The main purpose of the eggshell is to withstand physical damage that could jeopardize the developing embryo and defend it against microbial attack [62]. Being robust and impervious comes with a cost but the eggshell is formed to resist a substantial amount of force while being fragile enough to allow the chick to crack open the shell. It is one of the most sterile environments in nature but at the same time it has surprisingly large number of pores ensuring that the developing chick does not suffocate [66]. Merging paradoxical features in an egg requires adaptations mastered for eons. The double-dome structure gives an inherent mechanical strength to the eggshell [67] and the material of choice for construction is important. The eggshell is made up of calcium carbonate. This inorganic compound is intrinsically fragile but the eggshell comes with an organic content that compensates the fragility. Around 5% of the eggshell content is organic matter, mostly proteins[68]. The exact function of these proteins are still largely unknown.

Current knowledge on the formation and structure of the eggshell is based mostly on one species, the domestic chicken, Gallus gallus. Although there are some significant differences in the shape, thickness and proteomics of the eggshells between species, the main architecture of the shell is remarkably similar [62, 65]. An important contribution to the knowledge of the eggshell formation was the sequencing of chicken genome which enabled conducting genetic and proteomic studies [69]. The comparative analysis of the eggshells based on their genomes is advancing with sequencing the genomes of different species. The second species whose genome was sequenced is the zebra finch, *T. guttata* [70].

The avian egg is the most complicated amniotic egg in oviparous vertebrates [68]. The organ where the egg is produced is known as the oviduct. The different parts of an avian egg are formed in different specialized compartments and the function of each of these compartments are known in some detail [66]. The formation of an egg starts when an ovum, which is a mature female reproductive cell, is discharged from ovary to a funnel like com-

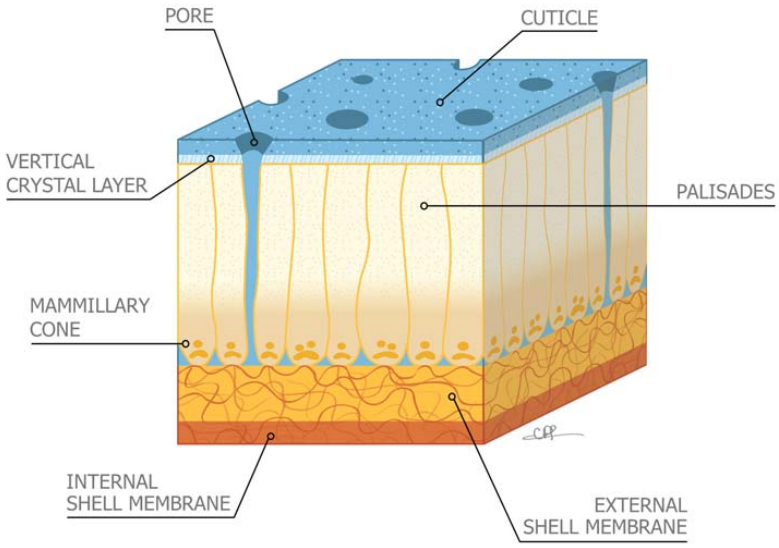


Figure 1.6: A schematic representation of the avian eggshell structure showing different layers through the cross-section of the shell [68].

partment called infundibulum [68]. The yolk is the first part of the egg deposited to surround the ovum. The yolk is surrounded in a membrane, called the vitelline membrane, during its passage through infundibulum [71]. After the infundibulum, the second compartment is the magnum where the albumen is deposited [43]. The yolk and albumen are passed through the next compartment, white isthmus [68]. This is the compartment where the shell membranes are synthesized. There are two membranes separated at the blunt end of the egg [72]. The shell is synthesized directly on the membranes starting from specific organic spots randomly distributed throughout the membrane [43]. It is intriguing that the shape of the egg is determined in this compartment as the shape is dictated by the membranes rather than the shell [65]. The next compartment is the red isthmus. Here, the shell formation is started with the deposition of an array of organic spots, called the mammillary spots [43]. This compartment is also known as the tubular shell gland. After the isthmus, the egg enters the uterus where the actual mineralization starts [68]. This part is known as the shell gland. Calcium carbonate is synthesized with the reaction of ionized calcium and bicarbonate ions in about 12 hours (in case of the domestic chicken [68]). The concentration of the calcium in the shell gland during formation is around 10 mM and much less than the bicarbonate concentration (70 mM) [73]. It should be noted that these concentration measurements might not reflect the actual values. After the mineralization of the shell is completed, a thin vertical layer of calcium phosphate is synthesized and on top of this layer, an antimicrobial coating is deposited which is known as the cuticle [66].

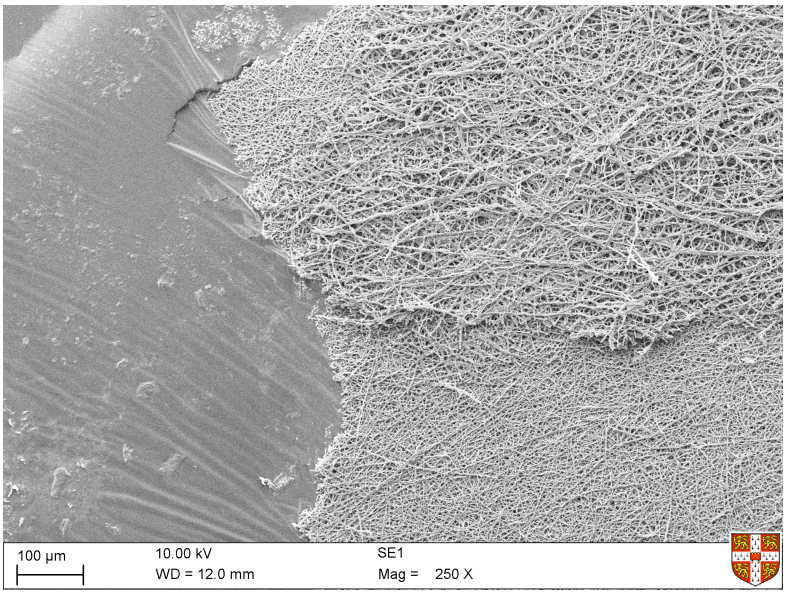


Figure 1.7: An SEM image of the chicken eggshell membranes separated from $CaCO_3$ layer with a sticky tape. The difference in the diameters of different types of collagen fibers is seen clearly.

The eggshell could be divided into 5 compartments. A schematic representation of these compartments is shown in Fig 1.6 as a cross-section of the eggshell. The first compartment is the membrane surrounding the albumen. There are three membranes first of which is in direct contact with the albumen and called the limiting membrane [72]. This membrane is very thin compared to the other membranes and difficult to differentiate from the inner membrane. The inner membrane is made of collagen fibres, proteins and glycoproteins with lysine-derived crosslinks[74]. The fibres of the inner membrane are made of type I, type V and type X collagen [74]. The diameter of the fibres in the inner membrane are between 0.1 to 3 µm and the thickness of the membrane is around 20 µm [72]. The outer membrane is composed of collagen fibres type I and type X but they are thicker compared to the inner membrane fibres as the diameters of the these fibres vary between 1 to 7 µm. The thickness of the outer membrane is around 60 µm [72]. Similar to the inner membrane, there are proteins and glycoproteins with lysine-derived crosslinks in the outer membrane. In total, more than 60 proteins have been observed in the membranes [75]. A scanning electron microscopy (SEM) image of the chicken eggshell membranes is shown in Fig 1.7.

These two membranes are separated from each other at the blunt end of the egg to form an air cell [76]. The air cell is gradually built up as the egg develops and before hatching, it reaches its maximum volume. The function of the air cell is to provide enough O_2 for the chick before cracking the shell. This is because the chick starts to use its lungs immediately after

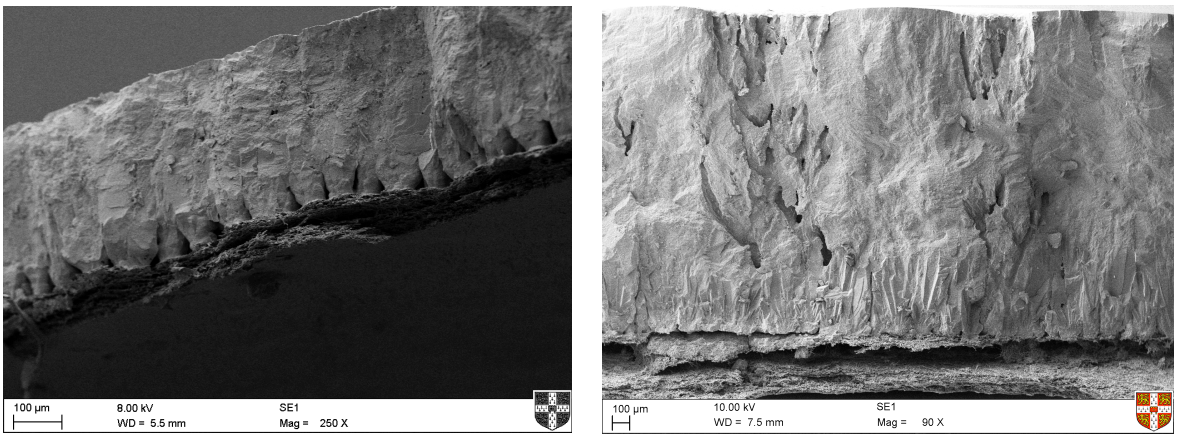


Figure 1.8: SEM images of the chicken (left) and ostrich (right) eggshell cross-sections. Conic structures where crystallization starts are seen on the membranes. Once the cones merge with their neigbours, columnar growth starts to construct the palisade layer perpendicular to the membranes. In the final step of the eggshell formation, a vertical crystal layer is added which is not visible in the images above.

its gill-like respiration through chorioallantoic membrane is terminated [77]. It is possible to separate the shell membranes form each other using the air cell.

The shell mineralization starts on organic spots, called the mammillary layer, distributed on the outer membrane [68]. This layer is thought to tailor mineralization by specific proteins inducing $CaCO_3$ growth [43]. One of the key proteins in this layer is OC-17 which will be explained in detail in Chapter 3.

The mineralization starts within the organic spots and expands while growing perpendicular to the membranes. This means that the crystals start to grow on a spot and the diameter of the crystal cluster expand to form a cone-like shape. Because the organic spots are distributed throughout the membrane these cone-shaped minerals merge with their neighbour once they contact. The initial growth in all directions is thus inhibited because of the space restriction and growth is oriented to the c-axis only, that is, perpendicular to the membrane [76]. SEM images of the chicken and ostrich eggshell cross-sections are given Fig 1.8 where conical growth pattern on the membranes can be seen.

This growth pattern is remarkably similar to the thermodynamically driven grain growth which has also been used to explain the self-similar mesostructure evolution of the growing mollusk shells [78]. This similarity is yet to be verified for eggshell growth and in doing so might shed light onto the thermodynamics of the eggshell formation.

The main mineral section in an eggshell is made of polycrystalline bulk $CaCO_3$. This 'palisade' layer is the thickest layer in the shell and it was called the palisade because of its columnar or railing like structure [68]. The growth of the palisade layer continues until the

vertical crystal layer, which is the next compartment, is deposited on the palisade. Both the mechanism of the inhibition of $CaCO_3$ growth (i.e. palisade layer) and the mechanism of the deposition of vertical crystal layer are not well understood. It might be possible that the inhibition is under genetic control. This means that inhibition and induction are regulated by proteins that are still unknown. Since phosphate ions inhibit $CaCO_3$ mineralization, it is possible that phosphate plays a dual role both in terminating shell growth and inducing the synthesis of the vertical layer of calcium phosphate [79]. A phosphoprotein, ovocalyxin-32 (OC-32) is found in the outer shell with a relatively high concentration. This suggests that OC-32 might also serve as a termination signal [80].

The last compartment after vertical crystal layer is the cuticle. The cuticle is a proteinaceous layer with a thickness of around 10 μm [66]. The cuticle serves as an anti-microbial defense system [81]. This is supported by the fact that it contains large amounts of glycoproteins [82]. The main function of this layer, however, is considered to avoid the excessive water loss through the shell pores [81]. The cuticle allows gases to diffuse into the egg for the growing chick. The diffusion depends on the humidity of the environment.

1.3.3 Mechanical Properties of Natural Materials and Avian Eggshells

Understanding of structure-function relationship of natural materials is an active field of research. An analysis of how biological organisms respond to mechanical failure offer new ways for the synthesis of light-weight, mechanically robust, nature-friendly materials. The trend in biomimetic materials is shifting from conventional mechanical testing to investigations of structure-property relationship from nanoscale to macroscale. With respect to the hierarchical biological materials, such as the cuttlefish bone [83], conducting traditional mechanical testing is not feasible as there are multiple layers in direct contact with each other at a scale of less than 1 mm. In submilimeter length-scales, there are substantial structural heterogeneity in biological materials due to their crystal orientation, shape or size and these differences contribute to different mechanical properties.

The main function of the avian eggshell is to protect the embryo. A hard shell providing a robust structure for embryonic growth is essential. The mechanical properties of the avian eggshell are fine-tuned to accomplish this feature. In order to achieve the level of sophistication needed for the eggshell, mechanical properties are adjusted to meet certain requirements for each species. The ostrich is challenged by harsh conditions in an open arid climate in Africa while the quail has to adapt to a different environment. There are evolutionary reasons for the differences of the physical and mechanical properties of different eggshells. For instance, the shape of an egg, that is, whether it is symmetrical or asymmetrical or whether it is elongated with a sharp pointed end or spherical, affects its overall mechanical response [84].

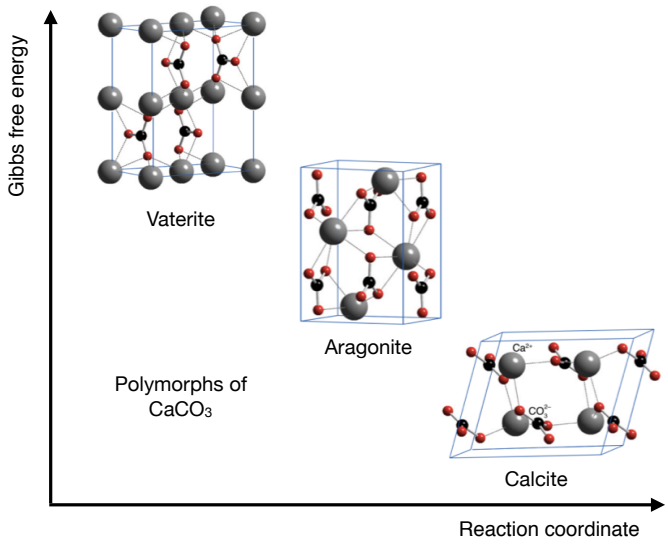


Figure 1.9: The structures of the polymorphs of $CaCO_3$ and their free energies of formation. Vaterite is the least stable form with the highest Gibbs free energy. Calcite is the most stable polymorph with the lowest free energy [85].

Although the main focus of the present work is to analyze and if possible imitate the intracrystalline incorporation of proteins for practical applications, the information that could be gained from the effect of proteins on eggshells' mechanical properties might be insightful. The reports on the effect of the overall structure on the mechanical properties had been reported decades ago before advanced methods developed for micro-scale measurements with accuracy. In terms of the eggshell cracking, there are two types of studies, one of which deals with the subject because of the commercial interest in cracking as it is a problem during the transportation of chicken eggs around the world. The second type of study focuses on the macro-scale mechanical properties to deal with hatching strategies and is mainly restricted to the evolutionary biology and zoology [86]. The eggshell coloration is also interpreted as a contribution to overall mechanical robustness of eggs [87]. Because of the scope of the present work, the literature that focus only on micro-scale measurements of the eggshell mechanical properties will be reviewed.

The mechanical properties of the eggshells from different species differ as expected. As mentioned earlier, the inherent robustness of the eggshell arises from its unique shape which allows distribution of pressure. This can be tested by applying a load to break an egg from either on its sharp end or on its side. Experimental and theoretical works have been reported to explain the mechanical response of eggs to an applied load. The physics of the geometry-induced rigidity of non-spherical shells has been studied in detail [88]. A thorough mathematical analysis to understand the effect of stiffness upon indentation of shell structures (also

explaining geometry-induced rigidity) has been published recently [67]. A numerical and experimental work on the localization of deformation on shells upon indentation has also been reported [89].

The inherent shape of the egg, however, is only one aspect of its mechanical properties. The eggshell is made of $CaCO_3$ which has stiffness and strength comparable to alloys and metals, such as aluminum. Pure $CaCO_3$ is a brittle material but the invertebrate shells of animals are made of polymorphs of $CaCO_3$ in unique architectures and in combination with various proteins to synthesize an organic-inorganic composite material. Fig 1.9 shows the three polymorphs of $CaCO_3$, vaterite, aragonite and calcite. The free energy of formation of $CaCO_3$ is highest for the vaterite and for this reason, thermodynamically the least stable form is the vaterite which can spontaneously transform into the more stable aragonite. Similarly, aragonite is less stable than calcite which makes calcite is the most thermodynamically stable polymorph of $CaCO_3$.

In Fig 1.10(a), an Ashby plot showing the mechanical properties of some natural and synthetic materials is given [52]. As can be seen, calcite and aragonite have stiffness and strength comparable to some metals and alloys. Intriguingly, as can be seen, once the inorganic aragonite is modified with organic components (e.g. nacre proteins), the fracture toughness of the composite product increases substantially compared to its components alone. This fact also explains the toughness of the eggshells as they are also products of organic-inorganic interplay.

It is known that there are some toughening mechanisms for biogenic calcites [49]. The hardness of the chicken egg is compromised with brittleness and similarly the quail eggshell is softer and more flexible which has evolutionary implications for hatching [86]. The crystal order of a crystalline material affects the resulting mechanical properties [90]. The geogenic is far superior to biogenic calcite because it is formed through natural pathways under extreme conditions, such as extreme pressures under the ground for thousands of years. This is because biogenic calcite lacks a perfect orientation of ions in the crystal lattice since the proteins interfere crystallization and affect resulting mechanical properties. This effect serves as a toughening mechanism to inhibit crack propagation and is achieved by binding proteins onto the natural cleavage plane of $CaCO_3$ [91].

Another the intriguing features of the avian eggshell is the role of the membranes in avoiding the collapse of the eggshell entirely when a part of the shell is damaged. Because the outer membrane is firmly bound to the shell, when an external force is applied, it does not fail all at once. The crack pattern after the breakage is rectangular fragments hold onto the membrane rather than a crack propagating along the shell and failing the whole structure. Although, there are reports in the literature on the indentation techniques for fracture toughness mea-

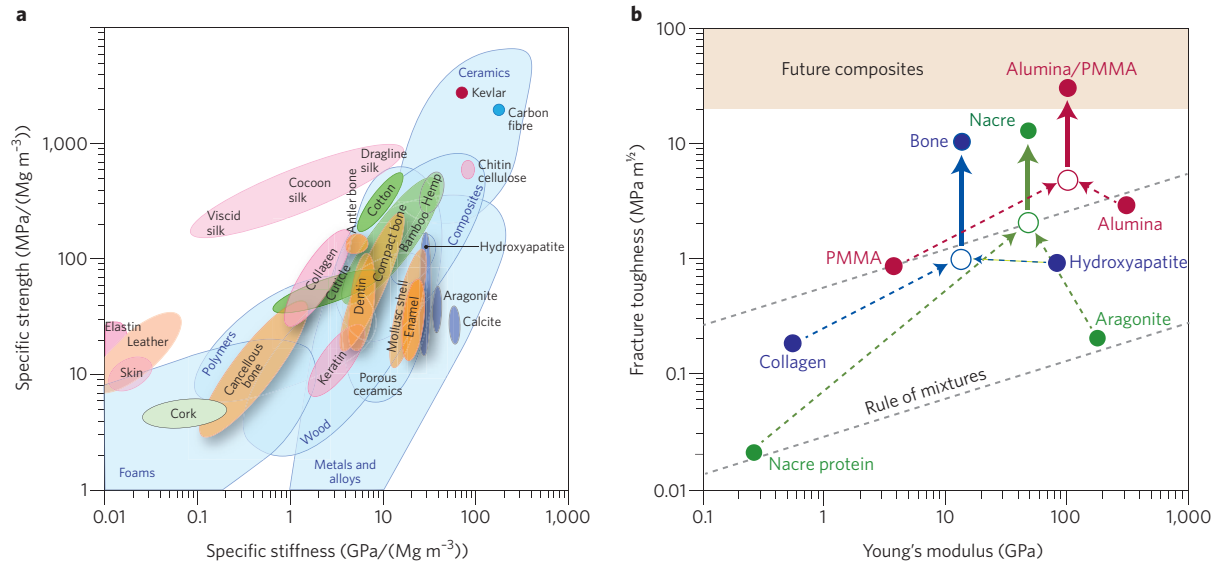


Figure 1.10: The mechanical properties of natural and synthetic materials (a) Ashby plot of specific values (normalized by density) of strength and stiffness (Young's modulus) (b) The resulting mechanical properties of organic-inorganic composites for natural materials and a synthetic one [52].

surements of natural materials [92], whether or not these measurements are well established is still controversial.

The vast majority of the literature uses the term 'mechanical properties of the eggshell' for the chicken egg only. The macro-scale mechanical properties of chicken eggs have been reported which employ conventional mechanical tests [93],[94],[95]. Being more relevant to the present work, regarding the micro-scale mechanical properties of the chicken eggshell, a number of studies have been reported, including the effect of microstructural changes [96],[97]. In terms of the micro-scale mechanical testing, there are limited number of studies that use nanoindentation for eggshells. Due to the reasons that will be explained in the Chapter 2, the nanoindentation testing protocol should be designed in such a way to collect reliable and reproducible data. The studies that report high variations in elastic modulus might have adopted non-ideal experimental protocols [98], such as using conventional load-hold-unload cycles, instead of partial unloading for cyclic indentation.

A recent work on the micro-scale mechanical properties of the chicken eggshell using nanoindentation has been reported [99]. This work focuses on the nanostructure of the chicken eggshell and shows that the shell has a nanogranular structure like other biominerals, such as mollusk shells, corals and fish otoliths [100, 101]. Although the quantitative data of nanostructure size distribution does not clearly show the increase of the nanostructure size, particularly for palisade layer because of the high standard deviation, it suggests that the nanostructure

size increases from the outer layers to the inner layers with the exception of the mammillary layer where the data are not statistically significant [99].

The eggshells are not single crystalline structures. Because of the organic phase controlling the crystal growth, misalignments in the crystal structures occur which have been shown using electron backscatter diffraction and two-dimensional X-ray diffraction (2D XRD) measurements [99]. The comparison of the chicken eggshell with geogenic calcite (e.g. Iceland spar as a control for calcite) using 2D XRD confirms the internal disorientation of the eggshell while geogenic calcite gives a sharp peak and rounded spots with a small angular spread [99]. These data imply the differences in the mechanical properties of single-crystalline calcite and $CaCO_3$-based biominerals because imperfections in crystallinity and an increase in amorphous structure changes mechanical properties. In the same work, the correlation of nanostructure with mechanical properties has been studied using an eggshell protein, osteopontin. It should be noted that the osteopontin is found in the eggshell [102] but it is not clear whether osteopontin controls the shell formation or not. As also showed in the same study, osteopontin is most abundant in the vertical crystal layer which is the outer part of the egg where $CaCO_3$ is inhibited to end the formation of the shell [79]. The vertical crystal layer contains hydroxyapatite crystals [103] and is known to contain proteins to inhibit calcification [80]. The fact that osteopontin plays a role in the formation of bone [104], which is made of calcium phosphate, implies that osteopontin might not be the actual protein for controlling the eggshell formation. It has been reported for the chicken eggshell that the hardness and elastic modulus both decrease from the outermost vertical crystal layer to the middle of the palisade layer and then increase again through mammillary layer [99]. In addition, it has been purported that there is a correlation between the nanostructure size and the mechanical properties. In other words, the formation of the inorganic structures involves the synthesis of nanosized crystals of the inorganic compound which are then combine to give the overall bulk mineral. This size-mechanical response relation has been explained by the Hall-Petch relationship which says that in the nanocrystalline materials, the hardness is inversely correlated with a decrease in the nanostructure size [105]. The correlation, however, does not hold from the middle palisade to mammillary layer [99]. An explanation for the lack of correlation has not been given except mentioning the possibility of different compositions or less structural homogeneity. An explanation of the lack of a correlation in the nanostructure sizes between the palisade and mammillary layers might be that the distribution of the size of nanostructures in the mammillary layer is statistically not significant which suggests that the average nanostructure size might not be lower than the palisade layer. The mechanical properties of the eggshell could be measured with nanoindentation but the experimental setup, measurements and interpretation of the data are all should be done accurately. The Chapter 2 of the present work aims to

address these issues systematically.

1.3.4 The Organic Matrix of the Avian Eggshells

Although the eggshell is made of around 95% inorganic material, there are surprisingly high number of proteins in the organic part which is around 5% by weight. The organic matrix of an eggshell consists of around 700 proteins. The concentrations of most of these proteins are too low compared to other intra-crystalline proteins [106],[107]. This fact confirms that some of the proteins are incorporated without no specific function for the eggshell crystallization because the eggshell specific proteins have concentrations comparable to any other protein in an organism. Among 700 proteins, nearly 40 of them have high enough concentrations to be structurally important [108]. In general, there are 3 different groups of proteins in the eggshell organic matrix. The first group belongs to the egg white proteins. The well known members of this group are the lysozyme, ovotransferrin and ovalbumin [109]. The second group belongs to the bone matrix proteins. The most well-known example of this group is the osteopontin [110]. The third group belongs to the eggshell specific proteins since these proteins are found exclusively in the uterus. The two classes of this group are called the ovocleidins and ovocalyxins [111],[112]. One of the most important proteins of the ovocleidin family is the OC-17 which is specific for the chicken eggshell and will be described in detail in the following sections. In this section, the chicken eggshell organic matrix proteins will be described as these proteins are studied in detail unlike other shell proteins found in different avian species. It should be noted that shell proteins are closely related between species but there are also some differences which will also be explained later.

A member of the ovocleidin group is the ovocleidin-116 (OC-116) which, unlike other ovocleidins, is not specific to the eggshell but found also in embryonic chicken osteoclasts and osteoblasts as well as young chick cortical bone [113]. Ovocleidin-116 is the first eggshell matrix protein to be cloned and although it is not specific to the shell, surprisingly it has the highest concentration among other shell proteins [114]. OC-116 has been found to be modified in many different ways. There are 22 phosphorylations sites on the serine or threonine residues [115]. In addition, at least two N-glycosylation sites are known [116]. It is also known to possess glycosaminoglycans [114]. The OC-116 gene has been shown to affect the eggshell thickness, elastic modulus and egg shape[117]. It should be noted that the effect of OC-116 on the shape of an egg is unlikely as it is known that the egg shape is dictated not by the eggshell but by the shell membranes [65]. Because the shell membranes are known to be synthesized in the isthmus, the role of OC-116 in shaping the egg becomes controversial as it is a protein found in the uterine fluid and not in the isthmus. Another protein that is found in high concentration in the eggshell organic matrix is the ovocalyxin-32 (OCX-32) [80]. The secretion of

OCX-32 has been found to correlate with the last stage of the shell formation[118]. This suggests that it is responsible for the termination of the mineral growth. The location of OCX-32 in the shell supports this hypothesis as it is found in the upper palisade layer, vertical crystal layer and cuticle [80]. In another study, though, OCX-32 has been found to be up-regulated in the strong eggshells compared to the weak ones [108]. This suggests that OCX-32 is active during the entire calcification process and might not specifically related with the end stage of shell formation. OCX-32 is also one of the phosphoproteins of the avian eggshell organic matrix [115]. Another member of the ovocalyxin class is the ovocalyxin-36 (OCX-36). This protein has been found to be active during the calcification period of the eggshell but its amino sequence has similarity with innate immune system proteins, such as bactericidal permeability protein [119],[112]. This similarity might indicate that the OCX-36 functions as a defense system to protect the egg and the oviduct rather than being a mineralization specific protein. Osteopontin is another eggshell organic matrix protein. It is a well-known protein because of its role in calcium mineralization in a myriad of different species [110]. Osteopontin is a 58 kDa protein found in the uterus yet it is also a common protein in mineralized tissues, such as bone [120]. It is found in the entire palisade layer and because it is related with mineralization in different tissues, the function of osteopontin is thought to be regulation of $CaCO_3$ growth [121]. A supporting experimental result for this hypothesis is that the osteopontin has been found to bind {104} crystal plane which is the natural cleavage plane of calcite [110]. This might indicate that the difference in the fracture behavior of the eggshell from other geogenic calcites arises from the proteins binding onto the specific crystal planes. Single nucleotide polymorphisms in the osteopontin gene has been shown to change the fracture toughness of the shell [117]. This supports the idea that osteopontin plays a role in the regulation of $CaCO_3$ crystals.

1.3.5 C-type Lectin-like Proteins

The terminology describing the C-type lectin-like proteins could be confusing. For this reason a short explanation is given here. There are different terms that are used interchangeably to describe the C-type lectin-like proteins. Lectins are a class of proteins originally from plants that are able to recognize and bind oligosaccharides [122]. Calcium dependent type lectins or shortly C-type lectins are a class of lectins that need calcium ions to be active [55]. They are called 'C-type' to distinguish them from calcium independent lectins. C-type lectin-like proteins are another class of proteins that have lectin-like activity but they do not belong to the lectin family [55]. These proteins are sometimes defined as proteins that have C-type lectin-like domains or C-type carbohydrate recognition domains [56]. The reason that lectin-like proteins do not have a well established terminology is that there are several proteins

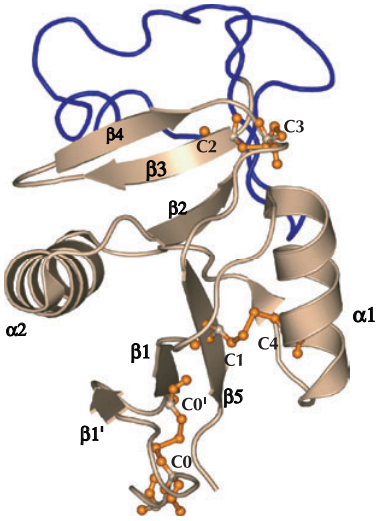


Figure 1.11: The common 3D structural motif of the C-type lectin-like proteins. Dark blue region represents the long loop domain. Cysteines (Cs) forming disulfide bridges are shown in orange. Secondary structure of alpha helices and beta strands are labelled as α_1, α_2 and, β_1-β_5, respectively [56].

in this family that have neither calcium dependency nor a lectin binding domain [56]. In short, the terms 'C-type lectin-like protein', 'C-type lectin-like domain', 'C-type carbohydrate recognition domain ' are used to describe protein families that have slight differences from each other. For instance, one of the domains in a protein could be a C-type lectin-like domain but the protein itself might not be a C-type lectin like protein. In the present work the 'C-type lectin-like protein' is used to describe the avian eggshell intra-crystalline proteins that are thought to transform amorphous $CaCO_3$ to calcite.

The common feature of the C-type lectin-like proteins is that they have a carbohydrate recognition domain [55]. It is assumed that they have evolved to recognize various carbohydrates and not only oligosaccharides [55]. The important point here is that this protein family has a highly conserved motif for carbohydrate recognition. Although, there are thousands of different C-type lectin-like proteins, the conserved motif is found throughout this family [123]. In terms of the amino acid sequence, each member of the C-type lectin-like protein family shows similarity with its conversed domain that is required for a unique 3D conformation [56]. The fold arising from the conformation is responsible for establishing the recognition domain [56]. Figure 1.11 shows the common 3D structural motif of the C-type lectin like proteins.

The recognition domain or domain fold is a double-loop structure consisting of an antiparallel β-sheet, which is the main domain loop and a long loop which is located in the core

domain [56]. An important motif of the C-type lectin-like proteins is the highly conserved disulfide bridges. In general terms, two of the four cysteines, C1 and C4, together connect one of the β strands to an α helix and this part of the protein forms the whole domain loop and the other two cysteines, C2 and C3, connect two of the five β strands (β3-β5) [56]. The rest of the protein has two α helices and another β-sheet with three β strands (β2,β3 and β4). The Ca^{2+}-dependent carbohydrate binding domain is the long loop region [56]. The most crucial motif that has been conserved in the C-type lectin-like proteins is the amino acid sequence 'WIGL' (tryptophan-isoleucine-glycine-leucine) which is found on the β2 strand and is used as a type of fingerprint in amino acid sequencing of the C-type lectin-like family [56].

The motif found in the C-type lectin-like proteins does not necessarily bind to oligosaccharides. On the contrary, most of the C-type lectin-like domains have affinity for some chemical groups other than saccharides [124]. One of the examples is some snake venoms. These protein-based toxins are known to agglutinate the blood using a recognition domain on their surfaces [125]. Another example is the antifreeze proteins that have been described previously. It should be clarified that the C-type lectin-like antifreeze proteins are known as Type II antifreeze proteins some of which do not depend on calcium ions [126]. Although, the sea raven type II glycoprotein has an ice-binding surface that is located at a different position than carbohydrate- or calcium binding domain, some other fish, such as the smelt and herring, do need calcium ions and shown to depress freeze point through binding with their calcium binding sites [127]. A change in the Ca^{2+} binding site has been shown to disrupt the binding of the protein to the growing ice crystal [127]. This shows that the antifreeze proteins specifically interact with ice crystals to inhibit the crystal growth.

The snake venoms and antifreeze proteins are examples of the evolution of the carbohydrate recognition domain from oligosaccharide binding to recognition of various chemical groups. Another example of the C-type lectin-like proteins is the calcium carbonate binding lithostathine [128]. Lithostathine is a pancreatic protein that has once been thought to play a role in the pancreatic stones [129]. It has been suggested that lithostathine might interact with calcium carbonate to inhibit stone formation [130] but the proposed interaction has been disproved in a follow-up study [131]. The Ca^{2+}binding ability of the C-type lectin-like proteins may serve as a modifier during biomineral formation. The evidence supporting this line of reasoning is that the C-type lectin-like proteins are found in different species of organisms having inorganic shells, such as mollusk and sea urchin [56]. The importance of the C-type lectin-like proteins for the formation of the avian eggshells arises from the fact that a protein, the OC-17, is a C-type lectin-like protein and is thought to be key for modification of $CaCO_3$.

1.3.6 An Avian C-type Lectin-like Protein: Ovocleidin-17

The OC-17 is an intra-crystalline C-type lectin-like protein found in chicken (Gallus gallus) eggshell and is one of the first identified eggshell specific proteins [111]. It should be noted that OC-17 may correspond to an earlier finding of a 15 kDa protein playing role in biomineralization [132]. The amino acid sequence and the X-ray crystal structure have been determined [133],[31]. OC-17 is composed of 142 amino acids with a molecular weight of 15,3 kDa based on its sequence. Unlike most of the biomineralization specific proteins, OC-17 is positively charged at pH 7 with a net charge of 7.3, an isoelectric point of 11.56 and a molecular extinction coefficient is 45880 $M^{-1}cm^{-1}$. The overall structural features of OC-17 (described in ref [31]) are as follows: It is a typical C-type lectin-like protein with a globular structure consisting of three long helix segments (α helices) and eight β-sheets. There are two groups of β-sheets oppositely arranged to each other which form two antiparallel β-sheet structure with β1-β2-β8-β3 in one group and β5-β4-β6-β7 in the other. The overall structure of OC-17 consists of two different orientations of its amino acids residues. The upper half contains second group of β-sheets with a short helix to connect β5-β6 and the lower half contains two of the α helices which surround the first group of β-sheets. There are six cysteine residues which form three disulfide bridges. The locations of the cysteine (Cys) residues that are connected with the disulfide bridges are Cys5-Cys16, Cys33-Cys138 and Cys113-Cys130. The disulfide bridges of OC-17 is one of the well preserved structural motifs of the C-type lectin-like protein family. Cys5-Cys16 has been shown to connect the loop structure, which is another common feature of the C-type lectin-like proteins, at the beginning of β1-β2 strand. The other Cys33-Cys138 connects α1 to β8. The second loop structure is built by the third disulfide bridge, that is, Cys113-Cys130 connecting β6 to β8 loop.

Similar to the other C-type lectin-like proteins, the disulfide bridges provide a large degree of stability to OC-17 [31]. OC-17 is the first biomineralization specific non-pathological protein whose three-dimensional structure has been determined at a resolution of 1.5 A° [31]. The tertiary structure of OC-17 is given in Fig 1.12. The amino acid sequence of OC-17 in one-letter notation is as follows: DPDGCGPGWV PTPGGCLGFF SRELSWSRAE SFCRRWGPGS HLAAVRSAAE LRLLAELLNA SRGGDGSGEG ADGRVWIGLH RPAGSRSWRW SDGTAPRFAS WHRTAKARRG GRCAALRDEE AFTSWAARPC TERNAFVCKA AA.

OC-17 has been thought to play a role in $CaCO_3$ crystal formation as it is found in the eggshell's intra-crystalline fraction. It is known to be secreted by special cells in the tubular shell gland [111] and found at high concentrations in the organic spots distributed randomly on the shell membrane [111]. On the one hand the actual location of OC-17 is thought to be the mammillary cores on the inner membrane attached to the bottom of the crystal layer. On

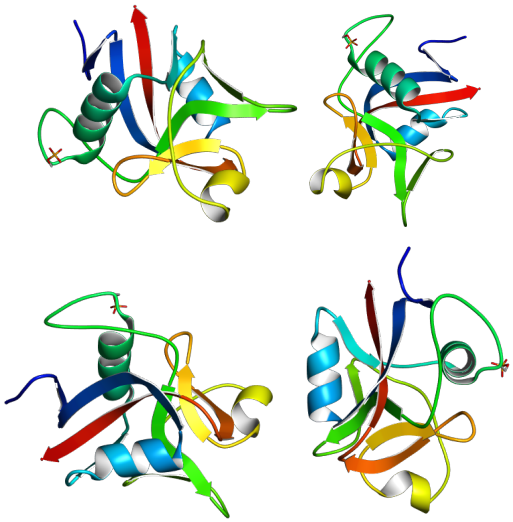


Figure 1.12: The crystal structure of OC-17 from four different angles. The 3D structures were taken from the Protein Data Bank Japan (https://pdbj.org/mine/summary/1gz2).

the other hand, although the concentration of OC-17 is high at the mammillary cores, it is also found throughout the palisade layer confirmed with immmunohistochemistry [111]. Bleach treatment of the eggshell has been shown to decrease the concentration of OC-17 once the purification of the whole shell extract has been done after the treatment [134]. This verifies that OC-17 is found in mammillary cores because only the organic content (i.e. membranes) is degraded by bleach but not the shell.

Unlike many C-type lectin-like proteins, OC-17 does not recognize or bind calcium ions but instead interacts with carbonate ions probably because of the charge at physiological pH [135]. Similar to the other C-type lectin-like proteins, OC-17 has the 'WIGL' motif in its sequence [56],[133]. An important feature of the OC-17 is that it is one of the phosphoproteins of the eggshell organic matrix [115]. OC-17 has two phosphorylated serines ($Ser^{61}$and Ser^{67}residues). The phosphate group at the $Ser^{61}$is located at the second α helix connecting $\alpha 2$ to $\beta 4$ in the lower half of the structure [31]. The exact location of the phosphate group at the atomic scale on the Ser^{67}, however, has been reported to be difficult to pinpoint because of the disorder in the vicinity of that serine residue [31].

OC-17 has a second glycosylated form in the eggshell organic matrix. This form, the OC-23, has the same amino acid sequence but with a higher molecular weight of 23 kDa because of a modified asparagine residue which is glycosylated [136]. The glycosylation site on Asn^{59} is a common motif in glycosylated proteins known as N-glycosylation consensus site, which should consist an amino acid sequence of asparagine (N)-alanine(A)-serine(S) in tandem [137]. The N-glycosylation consensus site has a general motif of asparagine(N)-

(X)-serine(S) or asparagine(N)-(X)-threonine(T) where X can be any one of the amino acids with the exception of proline [137]. Ser^{61}, which is a part of the consensus sequence, is not phosphorylated in OC-23 but phosphorylated in OC-17 [136]. The reason for the lack of phosphorylation of Ser^{61} is currently unknown. One possibility is the physical constraint of the residue 61 as the addition of a carbohydrate group to the Asn^{59} might restrict the addition of a phosphate to Ser^{61} separated by a single alanine residue from Asn^{59}. Another reason might be related to the function of OC-17 which may require the glycosylated form to be secreted to extracellular space. The deglycosylation of OC-23 enzymatically using N-glycosidase F has been shown to remove the carbohydrate group and reduce the molecular weight to 17 kDa [136].

1.3.7 The Function of the Ovocleidin-17

OC-17 is the only mineralization-specific protein whose function has been studied *in silico*. The molecular dynamics (MD) simulations are used to analyze the molecular motions of a sample of interest as a function of time [138]. Subsequent calculations harness equations of motions to understand how a molecule behaves in a defined environment [138]. The first MD simulation of a protein has been published on the mechanism of bovine pancreatic trypsin inhibitor in 1977 [139]. Since then, MD simulations have been improved substantially and applied on a variety of problems. An extension of MD simulations, known as metadynamics, uses the traditional MD approach with focusing on the free energy landscape sampling [140]. An example of such an analysis has been done for the eggshell specific OC-17 [33]. This study was the first analysis of a biomineralization protein to understand its function using simulations.

The onset of $CaCO_3$ crystallization was monitored through amorphous to calcite transition in the presence of OC-17. Using the UK national supercomputer, it was shown for two clusters of amorphous calcium carbonate (ACC) nanoparticles (192 and 300 formula units), that OC-17 binds to the ACC particles, mostly through arginine residues resembling a clamp [33]. The free-energy calculations suggest a mechanism for the function of OC-17 as follows: In the absence of the protein, there are three free-energy minima for the ACC nanoparticle. This corresponds to the polymorphs of $CaCO_3$. It is known that $CaCO_3$ has at least three polymorphs, vaterite, aragonite and calcite. Among them, vaterite is the least thermodynamically stable one. Aragonite is thermodynamically less stable than the calcite. The synthesis of aragonite requires magnesium ions for which reason the $CaCO_3$ shell of mollusks are composed of aragonite as a result of magnesium in sea water. Calcite is the most thermodynamically stable form. Geogenic $CaCO_3$ is calcite because the equilibrium conditions are met to satisfy thermodynamic requirements. The metadynamics simulations show that in the presence of

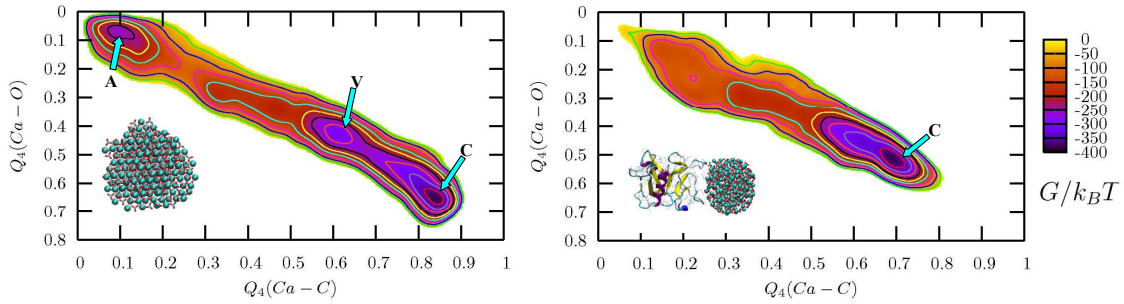


Figure 1.13: The Gibbs free energy maps of the effect of OC-17 on $CaCO_3$ growth (MD simulation of a 192 units of amorphous cluster. (Left) Amorphous cluster in water in the absence of OC-17. (Right) $CaCO_3$ cluster in the presence of OC-17 bound in water. The x and y axes show the symmetry in the arrangement of the carbon or oxygen atoms about calcium ions. The capital letters on the maps denote free energy minima where local atomic order indicates a specific polymorph: A is amorphous, C is calcite and V is vaterite-like. Colour code at the right refers to the free energy [33].

OC-17, thermodynamically less stable polymorphs do not form but rather the only polymorph observed was the calcite which has the minimum free energy [33]. In the absence of OC-17, the free energy barrier for calcite relative to ACC is 350 kj.mol^{-1}which enables the ACC to be long-lived [33]. In other words, ACC is thermodynamically metastable. The free energy landscape is completely different in the presence of OC-17 which hints that the function of the OC-17 is to transform an amorphous cluster to a crystal [33]. The free energy maps of $CaCO_3$ formation in the absence and presence of OC-17 are shown in Fig 1.13.

The mechanism of action of the eggshell protein OC-17 is pronounced in a larger ACC cluster in the same study. After the nucleation of calcite, the protein is detached from the crystals because of the change in the shape of the cluster. The proposed model for the function of OC-17 is that it acts as a 'catalyst' that transforms ACC to calcite nuclei and detaches to become available again for binding to another ACC cluster [33]. The studies on vertebrate and invertebrate $CaCO_3$-based minerals suggest that ACC to crystal transformation is a common theme in biomineralization [141],[142]. It can be said that this general principle is under genetic control (i.e. controlled by proteins). The follow-up studies of OC-17 using classical molecular simulations have also provided complementary results on the function of the protein [143],[135].

The follow-up study focuses on the interaction of OC-17 with calcite {10.4} surface at 310 K at pH 7 [135]. A number of conformations of OC-17 adsorbing onto calcite indicates that depending on the configuration of the protein, different amino acid residues control binding in different ways and the adsorption energies differ between configurations [135]. The simulations have been run both in vacuum and explicit solvent, and the difference between vacuum

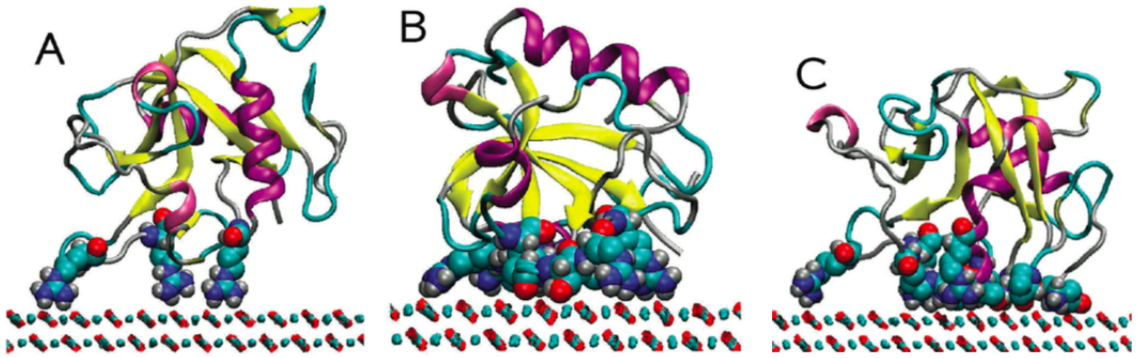


Figure 1.14: Three different configurations of OC-17 binding to the calcite {10.4} surface. (A) configuration 1, (B) configuration 2, and (C) configuration 3 [33].

and explicit solvent has been found to be very similar likely because of the stability of the OC-17 via the disulfide bridges.

The adsorption energy is an indication of the strength of the interaction between the protein and the mineral. A low adsorption energy shows that the interaction is stronger compared to higher energies. A negative adsorption energy is considered to be thermodynamically favorable. Among several possible configurations, three of them have been reported based on their adsorption energies [135]. These configurations are shown in Fig 1.14.

Configuration 1 has been shown to bind the oxygens of carbonates through the hydrogens of arginine. Configuration 2 has many residues interacting with the calcite surface whereby serines use the oxygen of the alcohol group and alanines use the carboxyl-oxygen to bind to calcium. Configuration 3 has also many residues interacting with the calcite including aspartic acid, proline and many arginine groups. It has been observed that the number of hydrogen bonds in Configuration 1 are substantially higher than the other two configurations [135]. In terms of the adsorption energies, Configuration 1 has the lowest energy which is intriguing given that it has the lowest number of residues interacting with the surface. The explanation for this lowest energy is that the arginine residues in Configuration 1 strongly bind to the mineral surface where they can dislodge water molecules effectively. On the contrary, weak binders other than arginine residues in other configurations are not able to detach water molecules easily. The molecular dynamics simulations provide detailed analysis of OC-17-calcite interaction and show that the surface water (water molecules around $CaCO_3$ clusters in solution) is an important part of the interaction. This is in agreement with other studies on the effect of water on $CaCO_3$ growth [144].

The possible interactions of OC-17 with different crystal faces of $CaCO_3$ have also been reported. Two stepped surfaces have been studied with molecular simulations [145]. Stepped crystal surfaces are likely to be observed in classical crystallization processes and because

biomineralization is likely to proceed through the non-classical pathway, the OC-17 is likely to interact with amorphous $CaCO_3$ rather than stepped surface or with {10.4} surface. It is worth noting that an experimental verification of the binding energies is challenging. The change in free energy upon protein-mineral binding might be calculated using an isothermal titration calorimetry (ITC) but the synthesis of ACC from $Ca^{2+}$and $CO_3{}^{2-}$ makes the measurements complicated. This is because the formation is of $CaCO_3$ is a highly endothermic process [146] and the heat change upon ionic interactions is difficult to measure with currently available ITC systems.

There are two aspects of the molecular dynamics simulations of OC-17 binding onto the calcite surface. On the one hand, the simulations provide detailed information about possible interactions at the protein-mineral interface and might lead further studies about protein-mineral interplay. On the other hand, it is widely accepted that the onset of biogenic mineralization includes the ACC phase which subsequently transforms into a crystal [142]. This means that the OC-17 should interact not with calcite surface but with ACC. Consequently, the interaction of OC-17 with ACC is needed which has also been addressed by the same authors.

The ACC could be found in two phases one of which is a metastable one found in biomineralization processes [147]. This phase is stable for a long period of time and contains water molecules within its structure. The other phase is thermodynamically unstable and transforms into a crystal in a short period of time [141]. The molecular dynamics simulation for ACC-OC-17 binding has been studied with anhydrous ACC as the proposed function of OC-17 is thought to be consistent with a rapid change from amorphous to crystal [143]. A clear difference between OC-17 on crystalline $CaCO_3$ and ACC is that the total number of residues interacting with the mineral surface is higher in the amorphous phase. In total, contrary to 3 arginine residues on the calcite surface, 15 residues are in contact with the ACC surface [143]. Moreover, some of the residues diffuse into the ACC slab [143]. In addition, some of the carbonate ions diffuse out the slab and interact with some residues [143]. A schematic representation of OC-17-amorphous $CaCO_3$ interaction is given in Fig 1.15.

The interacting groups have been found to be mostly positively and the only acidic residue binding to the ACC slab has been found to be the aspartate (92) [143]. The structural changes have been found to be negligible upon binding of OC-17 onto the ACC surface and the adsorption energy calculated as -977 $kj.mol^{-1}$ which is more than twice of the adsorption energy of OC-17 on the calcite surface [135]. This shows that the binding is stronger for OC-17 when it interacts with the ACC rather than crystal $CaCO_3$ surface. An important difference between the calcite and ACC interactions of OC-17 is the amount of order of water molecules on the mineral surface. While the order of water molecules is high on the calcite surface, the ACC surface has no organized layer [143]. This effect is reflected in the number of hydrogen bonds

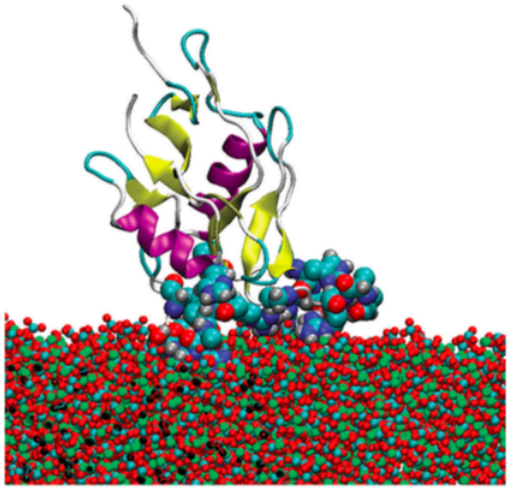


Figure 1.15: A schematic representation of the interaction of OC-17 with amorphous $CaCO_3$ [143].

between the water molecules and the amino acid residues. In addition, the number of water molecules are twice as much in calcite compared to the number of water molecules on the ACC surface [143].

As mentioned previously, calculating the actual free energies for large molecules, such as a protein, is financially challenging. A pseudo-free energy has been given in another study by the same authors for the {10.4} surface-water interaction [135]. Taking into account of the displacement of ordered water molecules from the surface gives a free energy of -188 kJ mol^{-1} for the {10.4} surface. The same method has been used for the amorphous case and the pseudo-free energy has been calculated to be -755 kj mol^{-1}which is much less compared to the calcite binding [143]. This means that the binding of OC-17 is thermodynamically more favorable to ACC than to calcite.

In summary, the molecular dynamics simulations of OC-17 suggests that this protein has a transformation function of amorphous $CaCO_3$ into a crystal calcite. It has been suggested that upon binding and transforming an amorphous cluster, the interaction of protein with the mineral surface becomes thermodynamically less favorable and it detaches from the crystal spontaneously. Similar to a catalyst, OC-17 becomes available again to bind another amorphous cluster to repeat the process. A schematic representation of the OC-17 catalytic cycle is shown in Fig 1.16.

It should be noted that the C-type lectin-like proteins have also been known for their role in the immunity [148]. It is possible that along with other anti-microbial proteins in the egg or in the shell, OC-17 might act as an immune response protein. This possibility, though, is unlikely because of the fact that OC-17 is a low molecular weight protein reminiscent of

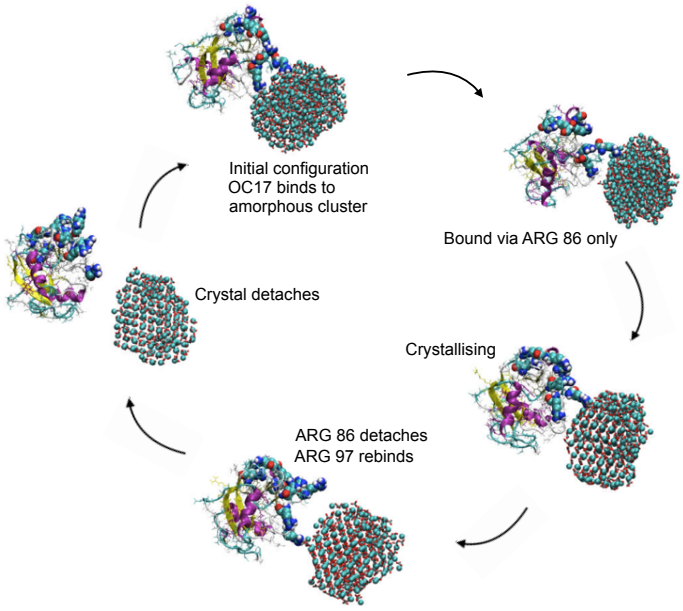


Figure 1.16: The catalytic cycle of OC-17 based on a proposed function for the protein using metadynamics simulations [33].

substrate specific binders akin to the fish antifreeze C-type lectins. In addition, an immune system protein is unlikely to be found specifically in the shell rather than in the egg where immunity is of utmost importance to protect the embryo.

1.3.8 The Experimental Studies on the Function of the Ovocleidin-17

1.3.8.1 The Purification of the OC-17

OC-17 is the first avian eggshell specific protein that has been purified and characterized [111],[31],[133]. There are, however, methodological problems in the first papers explaining the purification of OC-17. As described in the previous sections, OC-17 is positively charged at physiological pH. Based on the amino acid sequence, the isoelectric point of OC-17 is 11.56 which gives a net charge of 7.3 to the protein at pH 7. In the first papers dealing with the purification of OC-17, it has been reported that a two step purification was done with an ion exchange (DEAE Sephacel) and a hydroxyapatite resin [111]. The DEAE Sephacel is an anion exchange resin which means that it binds negatively charged proteins based on the pH of the buffer used. Because the reported buffer is Tris with a pKa value of 8, at this pH, OC-17 is positively charged, therefore, cannot bind to the column [111]. Subsequently a different method for purification has been described [133]. This method has been used to purify OC-17 for amino acid sequencing. High Pressure Liquid Chromatography (HPLC) has

been used with a reverse phase column (Vydac C4) with a linear gradient acetonitrile in 0.1% trifluoroacetic acid [133]. A polishing step was also done to discard impurities using two consecutively connected size exclusion chromatography columns at pH 5. In another study, a Jupiter C18 reversed-phase column has been reported [134] to purify OC-17 from a shell extract without further purification in contrast to the previous reports (e.g. [133]). Although HPLC is an efficient purification technique, reverse phase solvents are most likely to denature proteins. Denaturation might not be a problem for some applications, such as amino acid sequencing. Because the present work aims to use OC-17 for its function, it is important to obtain properly folded native structure of the protein and eliminate the risk of denaturation during the purification. For this reason, a new purification protocol was developed which is described in Chapter 3.

1.3.8.2 The Genomic Studies of the OC-17

One of the contributions that could lead further experimental studies on OC-17 is the publication of the full-length cDNA sequence [149]. Although there had been substantial efforts to obtain and clone the OC-17 gene, it was unavailable until recently. This lack of the identification of OC-17 was unexpected as the chicken was the first domestic animal whose genome has been sequenced [69]. In addition, contrary to the most of the genome sequencing data, Gallus gallus has the least gaps in its complete sequence [150].

The cDNA of the OC-17 has confirmed that the OC-17 gene is located in the gap region of the Gallus gallus genome [149]. The full-length cDNA of OC-17 has 626 base pairs and the open reading frame consists of 483 base pairs [149]. Another finding is that the full-length amino acid sequence not only contains the previously published sequence of OC-17 but also contains an extra sequence of 19 amino acids which has been found to be a signal peptide at the N-terminus of the mature protein [149]. The presence of a signal peptide further confirms that the OC-17 is synthesized to be used extracellularly. This is because the modification of proteins with a signal peptide is a process used during the post-translational modification to direct the proteins to the cell membrane where it is delivered to specific locations [151]. The signal peptides usually contain 15-30 amino acids and they are cleaved at the membrane by a special enzyme called a signal peptidase [151]. This enzyme is located at the inner membrane and recognizes N-terminal leader (signal) peptides as it is known that signal peptides contain a hydrophilic N-terminus and an adjacent hydrophobic core [151].

The expression levels of OC-17 in different tissues have also been studied along with the cDNA sequence. It has been shown that the expression level is highest in the uterus [149]. The isthmus also contains OC-17 but the other 13 tissues studied have shown that OC-17 is not produced elsewhere [149]. The presence of OC-17 in the isthmus is not unexpected as

this region is responsible for the synthesis of shell membranes. It is intriguing that the OC-17 expression is much higher in the uterus despite the fact that OC-17 is known to be located mainly in the mammillary cores on the shell membrane.

It has been noted that the expression level is highest in the 27^{th} week of the life span which corresponds to the development of uterine and egg laying stage [149]. The expression is significantly less in the 16^{th} and the amount is minimum in the 13^{th} . Starting from the 20^{th} week, the level starts to increase rapidly.

Recombinant OC-17 production in bacteria is conducted in the present work but other experimental studies on OC-17 are also limited. The function of OC-17 has been studied via *in vitro* crystal growth experiments [41, 134, 152] which are thought to be inconclusive. The change in the crystal morphology, which is given in most studies as the effect of proteins on biomineralization, is not indicative of the protein function because any protein whether biomineralization specific or not does affect crystal morphology. Probing the function of the biomineralization proteins needs novel approaches and molecular dynamics simulations have been shown to be one of the promising approaches leading experimental studies in a meaningful manner.

The proposed function of OC-17 through molecular dynamics simulations is yet to be confirmed experimentally. Although, the simulations indicate that the presence of the protein changes the free energy landscape, it is experimentally challenging to measure the change in the free energy with and without OC-17. It might be possible to probe the thermodynamic interactions of OC-17 with $CaCO_3$ using the advanced microfluidic calorimetry which has been shown to be highly sensitive even with volumes less than 5 nanoliter [153].

1.3.8.3 The Structure-Function Studies of the OC-17

The experimental studies of mineralization specific proteins are mostly based on *in vitro* crystallization and SEM imaging of crystals both of which are qualitative. A better approach could be probing the function of these proteins using ultra sensitive biophysical methods, such as FRET FLIM. Hitherto there are few studies on the structure-function relations of the avian eggshell proteins one of which suggest that major C-type lectin-like proteins from different species might have different functions. A comparison of ansocalcin, a C-type lectin-like protein from the goose eggshell has revealed a 36% of similarity of amino acid sequence with OC-17 [134]. An investigation of the structure-function relationship of the ansocalcin and OC-17 has indicated different functions despite their amino acid sequence and conformation similarity. *In vitro* crystallization experiments have shown that the growth patterns of $CaCO_3$ are different depending on the protein used. In the presence of OC-17 at low concentration between 10-100 $\mu g.ml^{-1}$ crystals have been observed to stick together and aggregate [134].

At higher concentrations aggregation is more pronounced but the corners of the crystals become smooth contrary to the sharp corners at low concentrations [134]. In the presence of the ansocalcin, the morphology has been found to be different. At low concentrations up to 10 $\mu g.ml^{-1}$ crystals have screw dislocations on {10.4} faces and at high concentrations over 50 $\mu g.ml^{-1}$ polycrystalline aggregates have been reported [134]. At higher concentrations from 100 to 500 $\mu g.ml^{-1}$ crystals exhibit spherical or more distorted shapes [134]. The difference between OC-17 and ansocalcin suggests that C-type lectin-like proteins might have distinct functions during mineralization as the locations, crystal growth patterns and net charges all differ between ansocalcin and OC-17. It has been noted that ansocalcin is unaffected by bleach treatment of the eggshells where membranes are dissolved leaving the inorganic matrix intact [134]. This means that ansocalcin is concentrated in the inorganic layer unlike OC-17 which is concentrated in the mammillary cores.

1.3.9 Protein-based Vaccines and Temperature Sensitivity Problem

Vaccines save millions of lives thanks not only to highly effective vaccines but also dedicated efforts to reach the remotest parts of the world (e.g. Médecins Sans Frontières). It could have been possible to prevent many more deaths if there would be heat stable vaccines at higher temperatures and hot climates. Vaccine instability at high temperatures had been recognized from the very early days of vaccination. Edward Jenner had realized that the smallpox vaccines he delivered to different countries lose their potency because of the high temperature degradation (see the Introduction for details). Today the temperature sensitivity of vaccines is still one of the major problems in developing countries. Vaccines being used across the world are temperature sensitive, and according to the WHO, all vaccines must be kept between 2-8 °C (or at -20 °C depending on the formulation of the vaccine)[14].

Historically, high temperature sensitivity of vaccines was the major problem for vaccine delivery [17]. The current situation is interestingly the opposite. Almost all of the current effort for vaccine stability is to prevent freeze degradation of vaccines rather than developing heat stable ones, as the latter one proved to be more challenging [154, 155]. The ideal solution for vaccine stability, however, is to produce vaccines that do not require refrigeration. The existence of freeze degradation problem arises only because of the lack of heat stable vaccines which makes it necessary to keep vaccines at low temperatures. Heat stable vaccines could eliminate the need for the cold-chain which is both difficult to establish and maintain [17]. Power shortages are known to occur in Africa and jeopardize the integrity and continuity of the cold-chain system. One of the other crucial advantages of heat stable vaccines would be the elimination of vaccine wastage as tens of thousands of degraded vaccines are discarded annually [17]. In addition, delivery of vaccines to hard-to-reach areas will be easier. Without

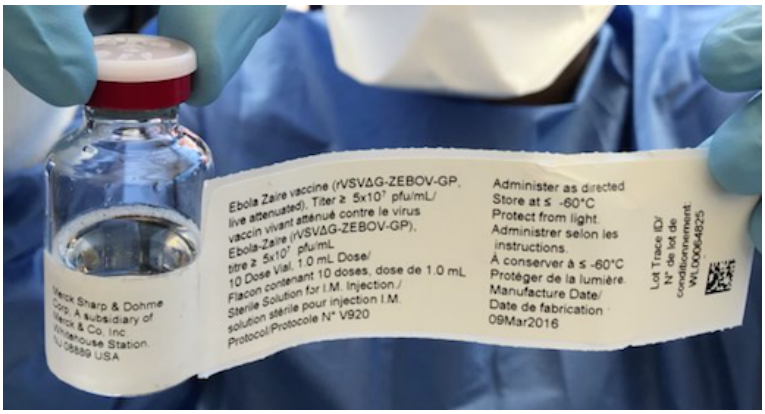


Figure 1.17: A recently developed attenuated virus-based Ebola Zaire vaccine by the Merck Sharp & Dohme to be used in the Democratic Republic of Congo. The storage temperature of the vaccine is -60 ℃ as is seen on the vial. Image credit, Reuters/Kenny Katombe, May 21, 2018.

the loss of activity of vaccines, vaccinations will be more effective in preventing diseases during emergency situations, such as earthquakes, after which it might not be possible to maintain a cold-chain system to store vaccines [156]. In the conflict zones, it is also hardly possible to keep vaccines under the recommended conditions [17].

One of the drawbacks of natural organisms is that they are inherently temperature-sensitive as they depend on proteins which function in a specific temperature range. The temperature range is narrow because almost all of the proteins are prone to degrade rapidly if there are fluctuations in the optimum temperature at which they evolved to survive [157]. Proteins need to be degraded and discarded to avoid the accumulation of aggregated debris in the cell [158]. The temperature range for most organisms is between 15-35 °C. An exception to this rule are extremophile organisms which, as their name indicates, could survive under extreme conditions [159]. Certain bacteria survive in the hydrothermal vents on ocean floors or in hot springs on the ground [160]. A well-known example of thermophile bacterium is the *T. aquaticus* [161]. Because of the high temperature adaptation to live in hot waters, the proteins from this bacterium display activity between 50-80 °C. For this reason, the DNA polymerase enzyme of *T. aquaticus*, known as Taq polymerase, is used in the Polymerase Chain Reaction (PCR) to unfold a DNA strand at around 50 °C, make a copy and repeat the process many times to synthesize DNA. The PCR is undoubtedly one of the breakthroughs in molecular biology, and this method is entirely based on the high temperature stability of Taq polymerase.

The temperature sensitivity of vaccines is closely related to protein stability as many vaccines are either antigenic proteins or any other type of biological entities, such as viruses. Although state-of-the-art techniques are used to develop highly effective vaccines, tempera-

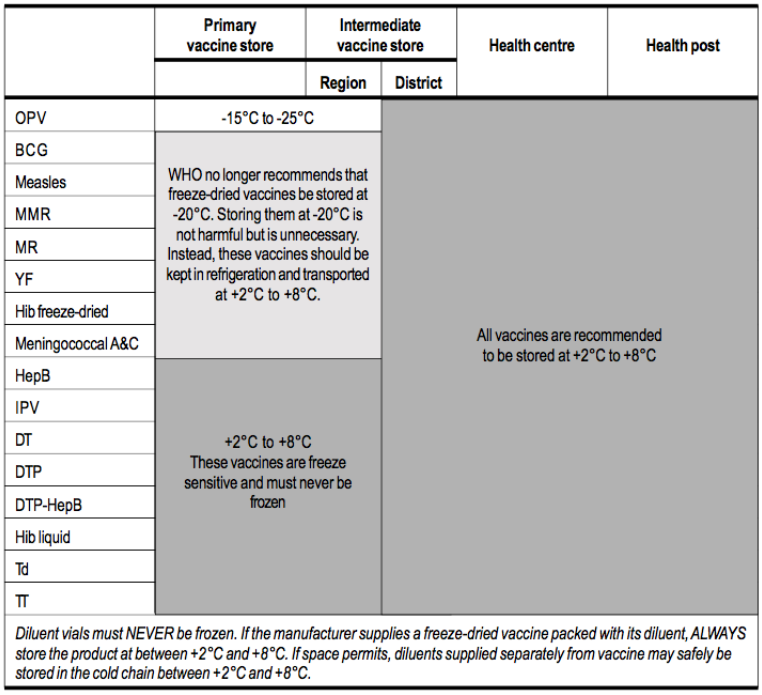


Figure 1.18: The storage temperatures for currently used vaccines recommended by the WHO [14].

ture sensitivity is becoming more problematic. An example is the Ebola Zaire vaccine. This recently developed vaccine is made of attenuated viral particles and requires a storage temperature of -60 ºC (Fig 1.17).

There are a number of reasons for the high temperature sensitivity of vaccines. One of the main reasons is the effect of temperature on the tertiary (or quaternary) structure of proteins. Proteins have complex 3D structures with which they interact with their substrates in a specific way. For instance, hemoglobin has a structure to bind and carry oxygen in the bloodstream [162]. A change in the tertiary structure of hemoglobin also changes the interaction of the protein with oxygen. This will cause either complete loss of function or a decrease in the efficiency of oxygen transport [162].

The overall tertiary structure of a protein is encoded in its amino acid sequence [163]. Because proteins contain many amino acids, they fold into an energetically stable form through the interactions of atoms. High temperatures might cause proteins to unfold either completely if the temperature is high or partially if temperature is below the protein unfolding temperature [164]. In the case of vaccine instability, environmental temperatures are not high enough to completely unfold proteins as the unfolding temperatures are generally higher compared to the environmental temperatures, for instance in Africa. Alterations in the tertiary structures, though, could be enough to distort the function of the protein. The main cause of the high tem-

Vaccine	Storage temperature, °C				
	2-8	20-25	37	>45	Freezing
Tetanus and diphtheria toxoids, monovalent or components of combined vaccines	Stable for >3 years	Stable for months	Stable for months	Unstable above 55°C	Unstable: do not freeze
Hepatitis B vaccine	Stable for >4 years	Stable for months	Stable for weeks	At 45°C, stable for days	Unstable; do not freeze
Measles, mumps, rubella vaccines	Stable for 2 years	Stable for at least one month	Stable for at least one week	Unstable	Stable
Yellow fever	Stable for >2 years	Stable for months	Stable for two weeks	Unstable	Stable
Pertussis vaccine	Stable for 18-24 months	Stable for 2 weeks	Stable for one week	10% or more loss of potency per day	Unstable; do not freeze
BCG vaccine	Stable for 1-2 years	Stable for months	Loss of no more than 20% after one month	Unstable	Stable
Oral poliovirus vaccine	Stable for up to 1 year	Stable for weeks	Stable for 2 days	Unstable	Stable
Inactivated poliovirus vaccine	Stable for 1-4 years	Stable for weeks	Stable for weeks	Little data available	Unstable; do not freeze
Polysaccharide vaccines (meningitis, pneumococcal)	Stable for 2 years	Stable for weeks to months	?	?	Unstable; Do not freeze
Conjugate polysaccharide vaccines (meningitis, Hib, pneumococcal)	Stable for > 2 years	Stable for > 2 years	May be unstable, depends on presentation	Unstable	If in combination with aluminum adjuvanted vaccine, do not freeze

Figure 1.19: The stability and lifetime data of vaccines commonly used around the world for immunization [14].

perature instability arises from these alterations [165]. The vaccine instability generally arises because of the aggregation of proteins due to the interactions of partially denatured proteins. Once the chemicals bonds are deformed because of the absorption of heat energy, new ones form as the formation of these bonds are now thermodynamically favourable.

Some of the vaccines are based on polysaccharide-protein conjugations, such as pneumococcal 7-valent vaccine. In this case, high temperature causes polysaccharide-protein dissociation [165]. The live-attenuated vaccines (weakened infectious viruses) are known to be less infective at high temperatures which means that the ability of the virus to cause a disease is low probably because the high temperature damages viral proteins that play roles in attacking the body [165]. Activity loss could not only happen at high temperatures but also during reconstitution of lyophilized vaccines.. In this case, the diluent may cause a heat shock if it is above a critical temperature that is too high for the vaccine [165].

High temperature instability has been a focus from the early days of vaccination but the current problem of instability also arises from freezing instead of high temperature [154].

Because the heat stability is still a major problem in delivery, vaccine transport out of the cold-chain system is considered unsafe and for this reason it is not being done. All the vaccines are kept in cold-chain system which also poses a number of problems. The first problem is the instability of aluminum adjuvants. In some vaccines, aluminum adjuvants are used to increase the efficiency of the vaccine but ice crystals distort the repulsion forces between aluminum particles and cause agglomeration or coagulation [166]. The degradation of aluminum adjuvants are easily recognizable using a shake test which makes it possible to discard the vaccine before using it for vaccination. Another reason for low temperature instability is cold denaturation [18]. It is known that proteins have stability curves as a function of temperature [164]. At a specific low temperature, the free energy of unfolding becomes negative and denaturation occurs as it is now thermodynamically favourable.

Although novel tools have been developed for the delivery, storage and stability assessment of vaccines, there are still many unknowns about temperature stability data and it is still challenging to store and deliver vaccines where they could make a difference [165]. A more detailed knowledge of the stability of vaccines outside the cold-chain for longer periods could enhance the effect of immunization campaigns, particularly, in Africa. The short term aim in vaccine delivery is to harness the inherent stability of currently used vaccines to possible maximum extend [167, 168]. The current practice, however, is still keeping vaccines at low temperatures although this poses the risk of freeze damage.

The storage conditions [14] of commonly used vaccines are given in Fig 1.18. Except from the OPV, all other vaccines do not need to be frozen or kept below -15 °C. The freeze-dried vaccines can be kept at -20 C yet it is neither needed nor recommended by the WHO. According to the WHO, all vaccines must be stored between 2 °C and 8 °C at all times. This requirement summarizes the fact that all vaccines are hitherto unstable without refrigeration [14].

During the eradication of the smallpox, the need for a reliable cold chain system has been realized [15]. Despite all of the scientific and logistic efforts, in developing countries, cold chain requirement and keeping vaccines at low temperature still accounts up to 80% of vaccination costs [169]. The stability and lifetime data of common vaccines can be seen in Figs 1.19 and 1.20. A detailed stability data for vaccines produced by different manufacturers around the world could be found in the 'Summary of Stability Data for Licensed Vaccines' report [170] by PATH, an international non-profit organization to to improve the health of people around the world by advancing technologies, strengthening systems, and encouraging healthy behaviors.

There are a number of strategies to increase the protein stability. One of these strategies is to modify proteins chemically by changing one or more of their amino acids. Genetic engi-

Vaccine	Storage temperature, °C				
	2-8	20-25	37	>45	Freezing
Hepatitis A vaccine	Stable for 2 years	?	Stable for 1-3 weeks or more	?	Unstable: do not freeze
Human diploid cell rabies vaccine	Stable for 3-5 years	Stable for 18 months	Stable for 4 weeks	Stable for several weeks	Stable
Japanese encephalitis B vaccine, inactivated	Stable for 1 year	Stable for 28 weeks	Stable for 4 weeks	Unstable	Stable
Japanese encephalitis B vaccine, live	Stable for 1.5 years	Stable for 4 months	Stable for 7-10 days	Unstable	Stable
Inactivated cholera and typhoid vaccines	Stable for > 2 years	Stable for years	Stable for 6 months	No data available	Unstable; do not freeze
Live attenuated cholera and typhoid vaccines	Stable for 1 year	Stable for 7 days	Stable for 12 hours	Unstable	Stable
Influenza, inactivated	Stable for up to 1 year	?	?	?	?
Influenza, live	Stable for 60 hours	Unstable	Unstable	Unstable	Store frozen, do not refreeze
Varicella vaccine	Stable for 1.5 years	?	?	?	May store frozen, do not refreeze
Rotavirus vaccine	Stable for > 2 years	Stable for 2 years	?	?	Stable

Figure 1.20: The stability and lifetime data of vaccines commonly used around the world for immunization [14].

neering allowed protein modification with which new proteins with increased stability could be synthesized. Although protein engineering techniques, such as site directed mutagenesis [171], have been promising, rational design of proteins is still a nascent field of research. This is because proteins are highly complex molecules and the structure prediction is still in its infancy. De novo protein design [172] will undoubtedly have a major impact in the near future yet hitherto protein stability studies require many cycles of trial-and-error.

Another method of increasing protein stability is fine-tuning the solution conditions, such as pH and ionic strength. The addition of stabilisers is a major path followed to increase the stability. One of the most well-known and highly efficient stabiliser is sucrose or any other non-reducing sugars [173]. Similarly, other different polyols, such as glycerol, xylitol and sorbitol are also effective stabilisers. The mechanism of protein stabilisation with sucrose is explained in two different ways [174]. One of the mechanisms is the vitrification which is explained by kinetics and is based on the immobilization of proteins in a glassy amorphous matrix [175]. The other mechanism is water replacement which is explained by thermodynamics and is based on the replacement of the hydrogen bonds between water and the protein with hydroxyl groups of the sugar [176].

Non-reducing sugars, such as sucrose and trehalose, are effective stabilisers as they form a glassy matrix and drastically reduce molecular mobility. The term glassy state (and rubbery state which will appear below) needs explanation. If a liquid is cooled slowly, it is possible that

a transition occurs to form a crystalline material at a temperature that can be defined through a liquid-solid equilibrium curve [177]. If the same liquid is cooled rapidly, it can be transformed into an amorphous solid which is commonly termed as glass [177]. The glassy state of an amorphous solid is hard and brittle, and at a temperature below its melting point the material undergoes another transition, which is called the glass transition, from a glassy state into a rubbery state [177]. Above the glass transition temperature, the viscosity of the amorphous solid increases to such an extent that it exhibits properties of a rubber rather than of a viscous liquid [178]. This is known as the rubbery state. Some sugars have glassy states which can be used to stabilise certain molecules [179]. Glassy matrix could surround the target substances and immobilize them in a matrix in which molecular mobility is substantially reduced. In other words, this matrix provides an 'infinitely' viscous medium in which no chemistry occurs [180]. It is worth noting that because of the lack of water in the medium, this phenomenon is akin to the extremophile organisms that go through anhydrobiosis to remove the water from their body to survive in a desiccated state [181].

The formation of a glassy state in sugars depends on the glass transition temperature of the sugar. Above this temperature glassy state transforms into a rubbery state [179]. The disaccharides trehalose and sucrose are being used as excipients to stabilise pharmaceutically active substances. Trehalose in particular is used widely because of its high glass transition temperature [182]. It has been reported that molecular dynamics simulations of trehalose, sucrose and glucose (a non-reducing sugar) have glass transition temperatures of 392, 347, and 325 K, respectively [179]. This confirms experimentally measured values and shows that among these carbohydrates, trehalose could be the primary choice to be used as a stabiliser yet its high cost might be a problem in which case using sucrose might be an option.

Although, the amorphous carbohydrates with optimal glass transition properties are being used in the pharmaceutical industry as stabilisers, disaccharide sugars need to be added at high concentrations which might be problematic in terms of viscosity. Because the average volume of a single dose of vaccine injection is 0.5 ml, dissolving sugars may be a problem in terms of volume of the vaccine. The stability enhancing properties of trehalose and sucrose are nonetheless of interest for vaccine stability at high temperatures. Although glass forming excipients are not the focus of the present work, different approaches for increasing vaccine stability using additives will be reviewed here. Sugar stabilisation methods have been reported for some of the vaccines [183]. Recent examples include additives, such as disaccharides, but also nanoparticles, which is an an example of mimicking sucrose stability but at much lower concentrations compared to sucrose.

Trehalose and sucrose mixture has been reported to preserve live poxviral and adenoviral vaccine vectors at 45 °C for several months without refrigeration [180]. In the study, the

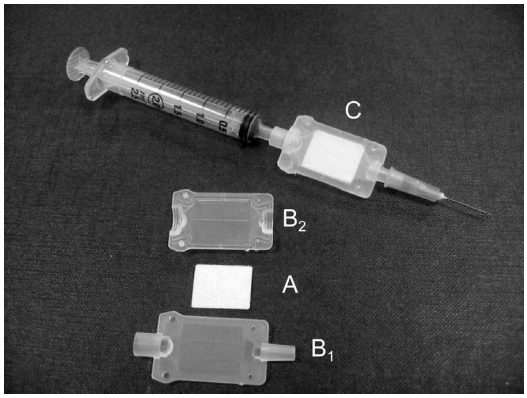


Figure 1.21: Image of the commercial HydRIS system which preserves some viral vaccines at around 45 °C embedded on a dissolvable membrane using an optimized sugar glass [180].

viral vectors immobilized on two different membranes in the presence of 0.5 M trehalose-sucrose mixture and diluted overnight at room temperature. Although both of the vectors are unstable without sugar on the membranes, addition of sugar has been reported to increase stability substantially [180]. The ability of sugars and in particular of trehalose, to confer stability for up to 6 months have been explained by the ability of trehalose in substituting water the in hydrogen bonded hydration shell of the proteins. The chemistry of trehalose to substitute water has been known [182]. Temperature fluctuations during vaccine shipment, however, pose different challenges compared to heat stability tests in proof-of-concept studies. Moreover, the sucrose-induced stabilisation does not always confer stability to viruses [184]. An advantage of the membrane-coated vaccines is to use a syringe-driven membrane holder to dissolve and reconstitute the vaccine in one step to be ready for injection [180]. Fig 1.21 shows the membrane (A) on which the viral vaccine particles in sugar glass embedded via drying. The membrane is placed between two medical-grade polymer housing (B_1 and B_2). The device is attached to a syringe-needle to be used as a conventional injection system because the buffer in the syringe dissolves the sugar glass of the vaccine rapidly. This study shows that it might be possible to immobilize vaccines in solid materials and prepare for injection whenever needed without complex procedures.

An alternative approach for vaccine stability is to mimic sugar stabilisation with using different types of additives or nanoparticles. Similar to the hydration effects observed for sugars, polyethylene glycol (PEG) has been studied as an alternative additive to replace sugars [185]. Although PEG has been shown to increase lifetime of adenovirus serotype 5 both at room temperature and at 37 °C, it has been noted that PEG also elicits an immune response and anti-PEG antibodies, and hypersensitivity up to 50% of people exposed to it has been noted [185]. In the same study, chemically modified anionic gold nanoparticles have been studied

for their electrostatic interactions with the virus particles. No increase in stability *in vivo* has been reported for anionic nanoparticles [185]. It should be noted that a study purporting stabilisation of vaccines and antibiotics in silk up to 6 months at 60 °C has been retracted 4 years after its publication [186].

An example from the literature that should be mentioned here is the paper claiming the production of heat-stable vaccines without refrigeration [187]. The title of the paper taken from the Gates Foundation's Grand Challenges in Global Health indicates the development of a new chemical method to preserve vaccines at high temperatures (T > 45 °C) as required in the real world. The stabilization of Japanese encephalitis virus vaccine has been reported via an eggshell inspired biomineralization. It has been reported that a $CaCO_3$ shell around the virus particles might protect the viral vaccine at high temperatures. However, the result of the work has been summarized as the stabilisation of the viral vaccine at room temperature for 7 days. It should be noted that, according to the WHO report on Temperature Sensitivity of Vaccines, the Japanese encephalitis virus vaccine is already stable for 37 °C for 3 months without any chemical intervention.

It is worth noting that $CaCO_3$ encapsulation, which is loading target molecules (e.g. small molecule drugs) into the pores of a porous $CaCO_3$ powder or adsorbed onto $CaCO_3$ surface, has been applied for some biopharmaceuticals but there is a fundamental difference between encapsulation and intra-crystalline incorporation. Encapsulation does not load the target molecules into the crystal lattice of $CaCO_3$. In other words, the strategy is not intra-crystalline incorporation but instead $CaCO_3$ is used as an inert 'vehicle' to carry molecules (for different examples of encapsulation see [188, 189] for small molecules, [190] for proteins, also see the examples below). Encapsulation methods use synthetic or natural polymers to assist for loading target molecules into microcapsules of $CaCO_3$ (e.g dextran sulfate/poly-L-arginine [190]). Another common method is to use polyelectrolytes for layer by layer assembly to obtain core/shell protein/$CaCO_3$ structure [191–193]. These encapsulation methods that use porous $CaCO_3$ is not relevant to the present work for several reasons. The main reason is that the intra-crystalline incorporation is a much more robust protection mechanism for proteins because the proteins are completely isolated from the environment, whereas in encapsulation proteins are exposed to the environment (i.e. in the pores of $CaCO_3$ where, for instance, solutions can diffuse). Another reason is that an ideal vaccine preservation method should contain vaccine immunogens ready to be used whenever needed. This requires reconstitution of proteins incorporated into $CaCO_3$ in a buffer that is safe and prepared easily on-site to be injected into human body. Encapsulation does not involve dissolving $CaCO_3$ and injection of $CaCO_3$ is harmful to the human body. In addition, encapsulated proteins need to be stable for a long time at high temperatures which could be achieved if they are immobilized into the lattice to

Figure 1.22: The Zipline drone used for delivering a medical cargo upon an order. The drone flies in a 80 km service radius from Zipline distribution centres in Rwanda and Ghana. Image credit, Zipline, flyzipline.com

avoid denaturation. Encapsulated proteins are located in the pores or adsorbed on the surface of $CaCO_3$ along with other chemicals which cannot be injected into human body. Encapsulation do not confer immobilization-based stability to the proteins. Furthermore, ensuring protein entrapment into $CaCO_3$ lattice could be tested by treating $CaCO_3$ particles in NaClO solution which would dissolve surface-adsorbed proteins as well as proteins loaded into the pores. Intra-crystalline proteins would not be affected by this process but the encapsulated proteins are exposed to the effects of solutions. In addition, heat treatment of protein encapsulated $CaCO_3$ and subsequent protein structure analysis of the encapsulated proteins are yet to be reported in the literature.

Recently, intra-crystalline incorporation of viruses in $CaCO_3$ has been proposed [194]. Two viruses have been studied, one of which has been reported to be incorporated in $CaCO_3$ [194]. Although this method has been proposed to preserve vaccines at high temperatures, neither the infectivity nor the heat stability of the incorporated virus has been reported. This might be because of the reconstitution process used in the study. It has been reported that reconstitution was done with 0.5 M acetic acid which drastically reduced the number of viral particles to an extent that the particle concentration was too low to be detected even with gel electrophoresis or transmission electron microscopy [194]. It will be shown in Chapter 4 (see Section 4.3.3) that the reconstitution buffer is a critical factor for the stability of incorporated proteins and EDTA should be used instead of acetic acid.

Apart from chemical protection, alternative vaccine delivery methods are being investigated. In late 2016, a drone-based system, Zipline, was launched to deliver essential medicines with a battery powered drone that is capable of carrying 1,8 kg of medical cargo. The Zipline

drones are launched from distribution centres currently operating in Rwanda (since October 2019) and Ghana (April 2019). These drones can fly in a range of 80 km to deliver medicals ordered through an app by a doctor. Fig 1.22 shows an image of the Zipline drone delivering a cargo with its parachute landing system.

1.4 Scope of the Present Work

The overarching goal of the present work is to develop an *in vitro* intra-crystalline incorporation method to preserve vaccines at high temperatures outside the cold-chain. Production of vaccines that do not require refrigeration is one of the major problems in global health and developing methods to confer high temperature stability to vaccines is of particular importance in particular to the developing countries. As explained in Section 1.2, the rationale of this work is based on the intra-crystalline incorporation of proteins in avian eggshells, particularly, the proteins that survived in the ancient eggshell samples. The experimental work will be presented in three parts: characterization of eggshells from different species, extraction and cloning of an intra-crystalline protein, OC-17, and incorporation of the model proteins in $CaCO_3$ *in vitro*.

Chapter 2 presents the first part of the experimental work which covers characterization of eggshells from 14 species using nanoindentation and Fourier-transform infrared (FTIR) spectroscopy to determine mechanical and chemical properties, respectively. The rationale for the eggshell characterization is to understand to what extent organic matrix affects the properties of a given eggshell. For instance, organic content was probed to interpret the change in the IR spectrum of an eggshell as an indication of protein incorporation into its $CaCO_3$ lattice. The species that show crystal 'disorder' in their spectra (i.e. deviations in the spectrum of the eggshell compared to the spectrum of pure $CaCO_3$) could offer insights about the amount of intra-crystalline proteins. Protein incorporation in crystal lattice increases the disorder which in turn suggests high protein concentration. It might be possible to indirectly infer which species contain high amount of intra-crystalline proteins using fingerprint IR peaks of $CaCO_3$. For instance, comparison of the ostrich eggshell, which is known to contain strongly bound intra-crystalline proteins (known as Struthiocalcin-1 and -2 from the C-type lectin-like family), it could be possible to reveal other species that also contain high-affinity intra-crystalline proteins that have not been identified yet. It is worth noting that at least one C-type lectin-like protein is likely to be found in each of the avian eggshell species but most of these proteins are yet to be identified.

Chapter 3 presents the second part of the experimental work which covers the extraction and (partial) purification of the OC-17 from chicken eggshell. It also covers cloning

and expression of recombinant OC-17 and cloning the Ivy protein (an inhibitor of vertebrate lysozyme). A method was developed to purify OC-17 which requires the Ivy protein to separate lysozyme from the eggshell protein solution because, although the rest of the eggshell proteins could be removed, lysozyme has highly similar chemical properties with OC-17 which makes it difficult to remove lysozyme from OC-17. Cloning of the OC-17 was done for several reasons. The main reason was to obtain high amount of protein without relying on eggshell extraction. Moreover, the expression of OC-17 in bacteria avoids the risk of denaturing the protein during acid dissolution of the eggshell. Expressing recombinant OC-17 could also be useful for functional studies, that is, probing the interaction of OC-17 with nascent $CaCO_3$ lattice. As will be explained in Chapter 5, Förster Resonance Energy Transfer-Fluorescence Lifetime Imaging (FRET-FLIM) could be used to study OC-17-$CaCO_3$ interaction. A plasmid encoding for the OC-17 would allow introducing point mutations to analyze the effect of single amino acid substitution on the function of OC-17. For instance, changing the arginine-86 could cause a loss-of-function mutation because of disrupting OC-17-amorphous $CaCO_3$ binding which could be probed via FRET-FLIM. Lastly, cloning would also allow expressing fusions of the OC-17 (e.g. other proteins genetically fused to OC-17, such as GFP-OC-17) to monitor protein stability upon incorporation through GFP fluorescence (see Chapter 5, Section 5.2.2).

Chapter 4 presents the third part of the experimental work which covers the protein incorporation experiments using the model proteins BSA, lysozyme and diphtheria antibody. In order to achieve intra-crystalline protein preservation, incorporation of the target protein needs to be done without denaturing or unfolding the target protein. However, the structural changes in intra-crystalline proteins extracted from $CaCO_3$ are not known. The lack of information on the change of the 3D structures of proteins during incorporation is partly because these proteins are found only in eggshells which means there are no 'native' proteins with which to compare the intra-crystalline ones. For instance, cloning of the OC-17 would allow obtaining a 'native' version (i.e. one that has not been involved in incorporation) of the protein whose structure could be compared to the intra-crystalline one to probe its structural changes (if any happens) during incorporation into the eggshell. Furthermore, the reconstitution of proteins from $CaCO_3$ is also a critical step during which structural changes could be triggered because of the interaction of $CaCO_3$ dissolution buffer with the incorporated protein. For this reason, the effect of the reconstitution buffer was studied. Because there are two ways to dissolve $CaCO_3$, acid dissolution and metal chelation, both methods were tested using acetic acid dissolution and EDTA chelation. The structural analyses were conducted by circular dichroism (CD) and intrinsic tryptophan fluorescence spectroscopy. CD data were further analyzed by a web server-based CD data analysis software CAPITO.

Chapter 5 draws conclusions from the experimental work and explains the future work that needs to be conducted for OC-17 and for any target protein incorporation. Potential improvements that could be done to increase the efficiency of vaccine/protein incorporation and strategies to move towards real-world applications are presented.

Chapter 2

Characterization of Avian Eggshells and Implications of the Effect of the Organic Matrix

2.1 Introduction

This chapter presents the results of the mechanical and chemical characterization of the eggshells from 14 different species. The overall aim of the characterization experiments is to study the effect of the organic matrix on the mechanical and chemical properties of the eggshells. The organic matrix is thought to be key both for the formation and for the overall material properties of the eggshells. Because the overarching aim of the present work is to understand and emulate intra-crystalline protein incorporation, expanding the knowledge on the different intra-crystalline content could allow mimicking organic-inorganic interactions better. In other words, in order to harness protein incorporation in $CaCO_3$, various species need to be studied as they might contain unknown C-type lectin-like proteins that are more efficiently incorporated into $CaCO_3$. The strategy followed here is to probe the effect of the intra-crystalline proteins on the mechanical and chemical properties of the eggshells using two highly sensitive techniques. The differences between material properties among different species could allow inferring -albeit indirectly- which species is affected more from the intra-crystalline content because these proteins have an effect on the crystal lattice unlike the inter-crystalline proteins. The effect of the inter-crystalline proteins is not likely to interfere because their concentration is both too low to be detected by FTIR and also cannot affect the lattice parameters as they are located between the crystals. Because the C-type lectin-like family is the major intra-crystalline content, it is expected that the high local concentration of these proteins could

have distinctive mechanical and chemical 'signatures' reflected on the eggshell properties. It is worth noting that the high local concentration might imply high binding affinity of these proteins to $CaCO_3$.

The mechanical properties were studied by nanoindentation for which the strategy of conducting the measurements accurately was explained in the Materials & Methods. The chemical characterization was conducted by spectroscopy. As will be explained below, FTIR spectroscopy allows probing the organic content based on the 'response' of lattice distortions upon changes in the crystal order of $CaCO_3$. The organic content (i.e. proteins) in the eggshells affect the lattice parameters of $CaCO_3$, which provides a way to infer the extent of change that the proteins impose on the eggshell. The results are presented for each species separately for both the nanoindentation and FTIR spectroscopy. The quantitative rankings of the key properties are presented in tables to compare species better. The last part of the chapter discusses the results with a focus on the differences between species where the effect of the organic content might be the reason for the observed differences. The overall aim is to 'predict' species that might potentially contain high affinity C-type lectins. As the struthiocalcins from the ostrich eggshell is known for their high affinity towards $CaCO_3$, the comparison of other species with ostrich could lead further studies to discover new C-type lectins from different species. It should be noted that there is a myriad of different avian species whose C-type lectin proteins have not been studied. It could be possible to discover new proteins that have evolved to exhibit high binding affinity to $CaCO_3$ thus could be more efficient compared to OC-17 or struthiocalcins.

2.2 Materials & Methods

2.2.1 Materials

The avian eggshells were donated by the Zoologischer Garten Wuppertal (Zoo Wuppertal), Dusseldorf, Germany, with their certificate of origins. It should be noted that collecting the eggs of wild birds is prohibited in the UK by the Protection of Birds Act 1954. All of the eggshells studied here were obtained for scientific purposes only. In total, eggs from 14 species were studied (up to 4 eggs per species where possible). Table 2.1 shows the Latin and English names of the species studied in the present work.

The mechanical properties of the eggshells were analyzed using nanoindentation. Nanoindentation tests were performed with a Hysitron TI-700 Nanoindenter (Minneapolis, MN) with a Berkovich indenter tip. The chemical properties of the eggshells were studied using infrared spectroscopy. A Perkin-Elmer FTIR Spectrometer with an ATR Imaging Spotlight 400 Mi-

Latin name	Common name
S. camelus	Ostrich
R. pennata	Rhea
D. novaeguineae	Laughing kookaburra
A. cinerea	Grey heron
N. auritus	African pygmy goose
B. sandvicensis	Hawaiian goose
B. scandiacus	Snowy owl
S. mollissima	Common eider
P. sordida	Hooded pitta
H. leucocephalus	Black bulbul
E. elegans	Elegant crested tinamou
S. fischeri	Spectacled eider
P. pulchellus	Beautiful fruit dove
L. rothschildi	Bali myna

Table 2.1: The Latin and common names of the avian species whose eggshells studied in the present work.

croscope (Waltham, MA) was used to collect IR maps and the corresponding spectra. The peak heights and full-width at half maximum values were calculated using the Perkin-Elmer Frontier Spectrum software.

For both methods eggshell samples were embedded in epoxy and polished. The Acree-Kleer Could Mounting System, plastic clamps and molds were purchased from Met-Prep (Coventry, UK). The Omega Pol 200 bench-top orbital polishing machine was used with four silicon carbide abrasive papers in tandem, P240 (58.5 µm), P1200 (15.3 µm), P2500 (8.4 µm) and P4000 (5 µm). The abrasive papers were purchased from Spectrographic Ltd (Leeds, UK).

2.2.2 Methods

2.2.2.1 Preparation of the Eggshell Samples For Mechanical and Chemical Tests

The eggshell samples for both nanoindentation and ATR-FTIR Imaging were prepared as follows: A piece of shell was fixed within a polymer clamp which holds the sample so that the cross-section is exposed to the surface. The clamp holding the sample was placed in a plastic mold in which epoxy resin was loaded. Acree-Kleer epoxy powder was mixed with its solution with a volumetric ratio of 2:1. Epoxy curing was left unperturbed for 30 min. After polishing, each sample was monitored using the optical microscope of FTIR Imaging Spotlight 400 to ensure the exposure of the cross-section. The optical microscopy images of two of the samples are shown in Fig 2.1.

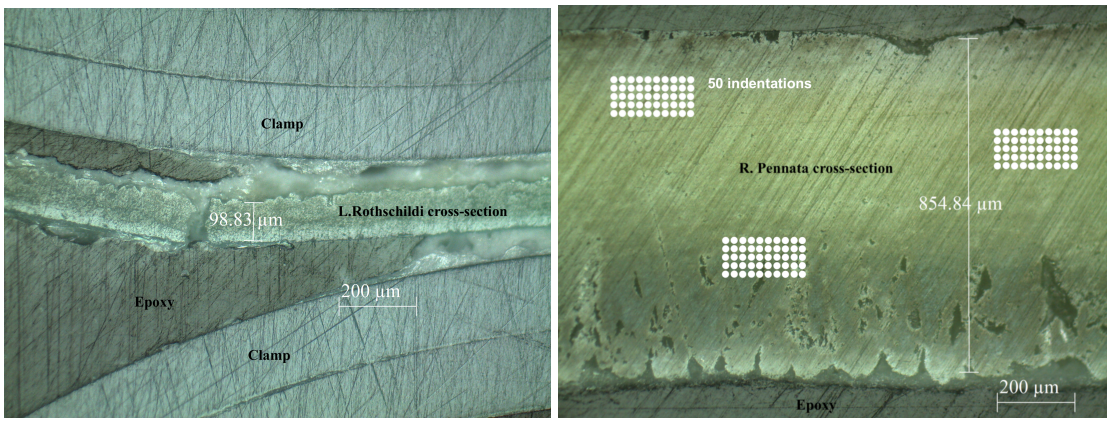


Figure 2.1: The optical microscopy images of two of the eggshell samples after the polishing step. The images belong to one of the thinnest and thickest samples to show the differences among the sample areas. (Left) Bali myna eggshell. (Right) Rhea eggshell. Indentations were performed on 50 points at three locations as schematically shown with white dots. The epoxy resin and plastic clamp holding the samples during epoxy curing are labelled on the image at the left. Scale bars show 200 µm. The cross-sections of Bali myna and rhea are 98.83 µm and 854.84 µm, respectively.

A possibly important point for the analysis of biological materials in general (and here for nanoindentation and FTIR mapping) is the surface roughness of the specimen. Although polishing with abrasive papers in tandem finishing with P4000 (5 µm) provides a smooth surface, indentation could be affected from roughness. An improvement strategy is to use ion-milling (see the Discussion for further details and explanation of a preliminary test done on the ostrich sample).

2.2.2.2 Nanoindentation and the Experimental Procedure for Measuring the Mechanical Properties of the Eggshells

Nanoindentation testing differs from conventional mechanical tests. In a conventional test, such as tensile or compression tests, a sample with an approximately constant cross-sectional area is gripped and pulled (in tensile test) or pushed (in compression test) in a direction perpendicular to the cross-section [195]. In contact mechanical testing, such as instrumented indentation, a probe is brought into contact with the surface of a sample. Because the indenter tip has a diameter between 50-100 nm (for a new indenter) [196], the area on which measurement is done is much less than the total volume of the sample. This allows probing the local microscale mechanical properties contrary to the traditional tensile test which gives an average of the mechanical response over the whole sample.

Two mechanical properties were studied with nanoindentation: elastic modulus and hard-

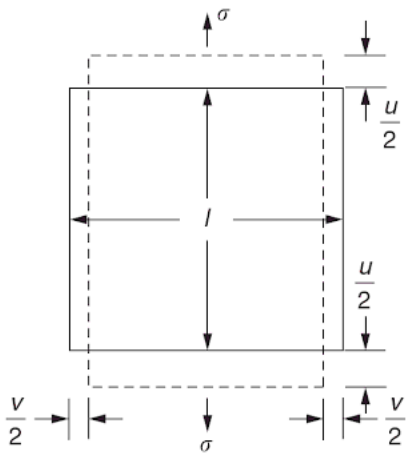


Figure 2.2: Strain as a response of stress on a square block with a face length, l. The total amount of shrinking inwards is shown as v (v/2 on both sides) and the total increase in length parallel to the stress (σ) is shown as u (u/2 on both sides) [197].

ness. Elastic modulus is a measure of the resistance of a material to elastic deformation [197]. It is calculated from the relation between stress and strain [197]. Stress, σ, is the force divided by the area on which the force is applied [197]. Because forces are measured in newtons, stress is measured in newtons per area, in SI units, N/m^2 [197].

$$\sigma = \frac{F}{A} \tag{2.1}$$

Strain, ε_n, is the response of materials to stress. A material shrinks inwards under a tensile stress, that is, being pulled perpendicular to stress direction. In this case, tensile strain is defined by the ratio of the amount of shrinking inwards to the original length of the face on which tensile stress is acting (Fig 2.2). Strain has no units as it is the ratio of two lengths.

$$\varepsilon_n = \frac{u}{l} \tag{2.2}$$

Elastic modulus is connected to the Hooke's law which describes stress-strain relation with the empirical observation that for small strains, the strain is very nearly proportional to the stress [197]. In other words, solid behaviour is linear elastic so that tensile stress is proportional to tensile strain [197] as follows:

$$\sigma = E.\varepsilon_n \tag{2.3}$$

In the equation above, E is the Young's modulus and determining E requires measuring the values of stress and strain accurately. In practice, elastic modulus can be calculated from a stress-strain curve [195]. In this case, elastic modulus is the slope of the initial, linear-

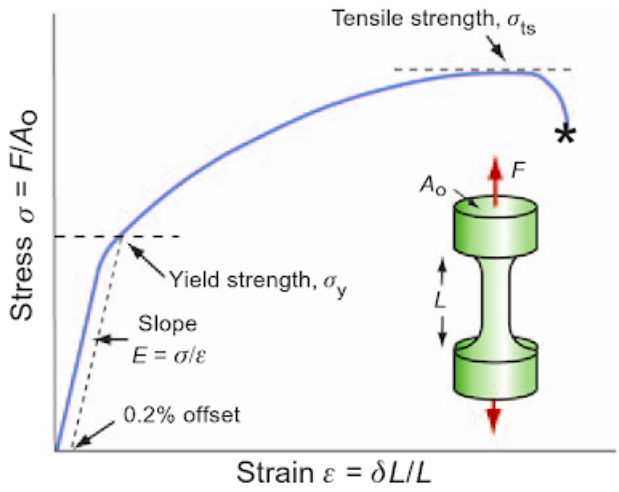


Figure 2.3: A stress-strain curve depicting the response of a material under loading. The initial part of the slope (linear-elastic) is used to calculate the Young's modulus, E (in case of tensile or compression test) [195].

elastic part of the curve [195] as is seen in Fig 2.3. Nanoindentation testing allows to extract elastic modulus based on these fundamental principles of mechanics [198]. The slope of the unloading part of the load-displacement curve is used to calculate the modulus [198]. The method of extracting modulus will be explained after describing the hardness.

Hardness (H in SI units) is a measure of strength [195]. It is calculated by indenting a material by a pointed (e.g. conical) or a spherical tip (e.g. hardened steel ball) and dividing the force by the area of the indent projected on the surface of the material [195]. Hardness is generally defined as a measure of a material's resistance against penetration of an indenter [52]. In other words, hardness is the resistance to plastic deformation [199]. Conventionally in indentation testing, hardness is calculated from the mean contact pressure at maximum load (P_{max}) divided by the projected area of contact (A) [196]:

$$H = \frac{P_{max}}{A} \tag{2.4}$$

An accurate value of the projected area of contact is key for nanoindentation. For this reason, the hardness calculation should be done with care, particularly, for conditions at less than full plasticity [196] (also see the alternative definition of the hardness below). In addition, the exact functional meaning of hardness -in particular for biological materials- is still being established [200]. Although it is still possible to calculate the hardness in the absence of plastic deformation [198], hardness is described as the resistance to plastic deformation because the sharp tip (e.g. a three-sided pyramid) causes near-immediate onset of plastic behaviour which

might not be the case with spherical tips [198].

The basic principle of nanoindentation is as follows: The probe of a nanoindenter is pushed into the sample in a load-controlled setup, allowed to stay in the sample for a period of time, retracted and during the process three variables are studied: load (P), displacement (h), and time (t) [201]. Using the P–h–t behaviour, it is possible to study various mechanical properties, including elastic modulus, hardness, viscoelasticity and poroelasticity [201]. An automated stage allows mapping the mechanical properties of an area of interest on the sample by repeating the same process automatically at different points separated by a few μm apart.

During nanoindentation, the sample is loaded with a peak force, generally with a hold period after reaching the peak load, and then unloaded. During this single cycle, mechanical properties are measured. A typical load-unload curve is shown in Fig 2.5. Then the data is analyzed with mathematical models developed for indentation analysis. Among these methods, the most widely used (even in-built in the commercial nanoindenters) is the Oliver-Pharr method [202] which is based on the work of Sneddon who derived the relationships between load, displacement and contact area for many simple punch geometries [203].

Depending on the material being tested, the response of the material to an applied load might be elastic or plastic as is seen in Fig 2.4. Glass remains elastic while being indented with a spherical tip following the relationship of P~$h^{3/2}$. Elastic to plastic transition can be traced from the load-displacement curve for aluminum as a transition from P ~ $h^{3/2}$ to P ~ h, again using a spherical tip. The separation of elastic-plastic response with sharp tips, such as a three-sided pyramid Berkovich tip, is not straightforward because the immediate induction of plasticity. In this case, the relationship between load and displacement is described as P ~ h^2 as is seen in Fig 2.4.

It is possible to calculate the modulus and hardness under the assumption that the unloading curve is elastic [198, 202]. As explained above, slope of the linear-elastic part of a stress-strain curve is used to calculate the Young's modulus which defines the stiffness of the material. Because the unloading curve in nanoindentation is assumed to be elastic, the reduced modulus (E_r) is calculated using the slope of the initial part of the unloading curve, dP/dh, which is defined as the unloading curve stiffness, S. Fig 2.5 shows a typical nanoindentation load-displacement curve. The maximum displacement is defined as the displacement at the peak load. The contact displacement h_c is a key quantity to calculate the modulus and the hardness accurately, and will be explained in detail below. The final displacement, h_f, is expected to be less than the h_{max} based on the assumption that the unloading curve is elastic (otherwise h_f would be equal to h_{max}). The constant ε is a property of the indenter geometry.

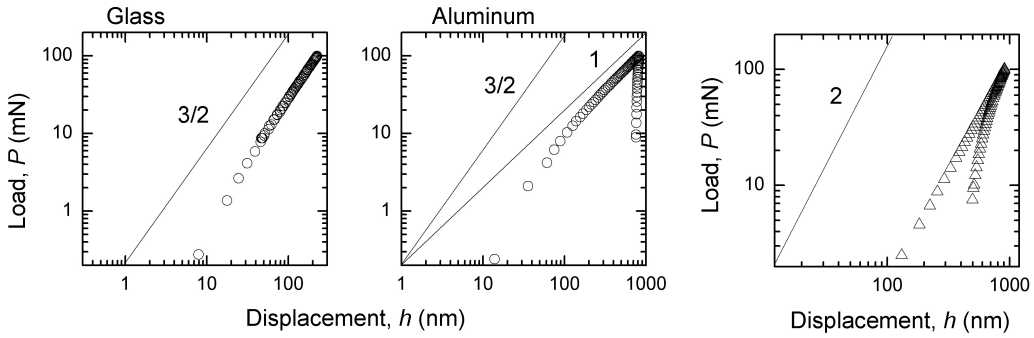


Figure 2.4: (Left) Elastic and elastic-plastic behaviour of fused silica glass and aluminum under a spherical indenter tip. While the response of fused silica response is elastic, aluminum exhibits plastic deformation above 100 nm displacement. Relationship for elastic behaviour changes from P ~ $h^{3/2}$ to P ~ h. (Right) Fused silica glass indentation response with a sharp tip. Behaviour follows P ~ h^2 relationship but the unloading curve has changing, functional response [198].

The unloading curve stiffness is related to the reduced elastic modulus as follows:

$$E_r = \frac{S.\pi}{2.\sqrt{A_c}} \tag{2.5}$$

As mentioned above, this relationship, based on the work of Sneddon [203], relates the reduced elastic modulus (E_r) to the contact area of the nanoindenter tip (A_c) which is related to h_c. The reduced modulus (E_r) is used in the equation above because it accounts for the modulus of the indenter. The modulus is calculated using the reduced modulus as follows [196]:

$$\frac{1}{E_r} = \frac{(1-\nu)^2}{E} + \frac{(1-\nu_i)^2}{E_i} \tag{2.6}$$

where subscript i applies to the indenter properties and ν is the Poisson ratio. A key quantity for calculating the reduced modulus is A_c, the contact area, which is the projected area calculated from the contact displacement h_c (Fig 2.6) via a calibration function of the form:

$$A_c = A_c(h_c) \tag{2.7}$$

The conventional hardness calculation assumes a perfectly sharp indenter tip but in reality the indenter tips approximate a sphere. The projected area of the indenter tip should be determined accurately to be used for subsequent calculations of the modulus and hardness. The

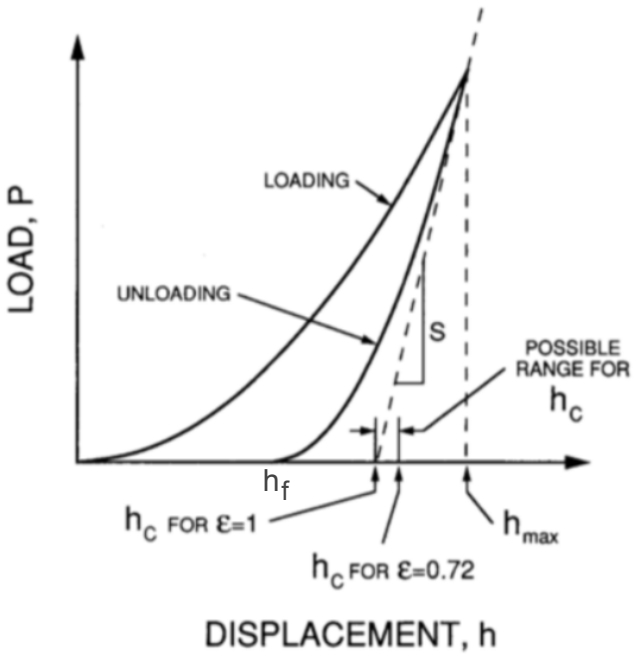


Figure 2.5: A typical load-displacement curve of a single load-unload cycle during nanoindentation testing. Displacement at the peak load is shown as h_{max}, and the final displacement after unloading is h_f. A key quantity for the modulus and hardness calculations is the contact depth, h_c, which is determined with a set of calibration measurements on a well-known material. The constant ε differs for each indenter geometry [202].

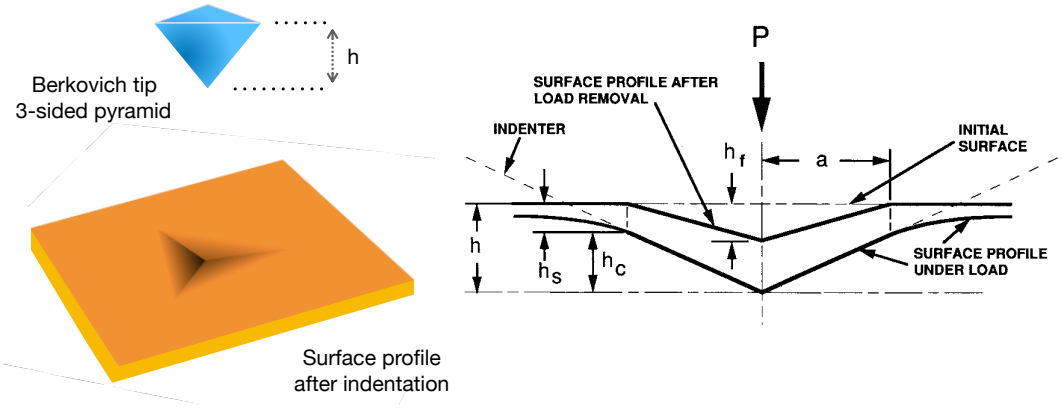


Figure 2.6: A schematic representation of an indent showing various quantities used for the modulus and hardness calculations. The maximum depth is denoted with h, final depth, h_f, contact depth, h_c, and surface displacement, h_s [202].

hardness can be redefined using $A_c(h_c)$ as follows [198]:

$$H_c = \frac{P_{max}}{A_c(h_c)} \tag{2.8}$$

The functional relationship between the contact area (A_c) and contact displacement (h_c) is given by [198]:

$$A_c(h_c) = C_0.h_c^2 + \sum_{k=1} C_k.h_c^{1/2^{\cdot(k-1)}} \tag{2.9}$$

C_i terms are coefficients related to the indenter and should be determined by calibration using materials with well-established modulus and Poisson ratio. In the present work, calibration was done with a fused silica glass (i.e. fused quartz) whose modulus (E) is 69.9 GPa and Poisson ratio (ν) is 0.17. The area function can be determined after calculating h_c, and then A_c using the equations below [198]:

$$h_c = h_{max} - 0.75\frac{P_{max}}{S} \tag{2.10}$$

$$A_c = \frac{S^2.\pi}{4.E_r^2} \tag{2.11}$$

The contact area for each indent on the fused quartz standard at different peak loads (e.g. 0.1, 0.3, 1, 3, 10, 30, 100, 300 mN) is calculated and $A_c - h_c$ data are fitted to the equation below:

$$A_c(h_c) = C_0.h_c^2 + C_1.h_c + C_2.h_c^{1/2} + ... \tag{2.12}$$

It has been shown that the first three terms are dominant and in most cases the first two terms are sufficient [204]. Once the area contact is known, E_r and H_c can be calculated from the equations above.

A cyclic indentation test was used for testing and the reasons of using cyclic mode will be explained in the Result & Discussion section. Shortly, the main reason for not using conventional single load-hold-unload cycle is because eggshells are highly porous non-isotropic materials contrary to the commercial isotropic single crystals or well-known polymers. This means that the deeper the indentation, the more likely to hit a pore which would result in 'free' displacement and deflate the estimated hardness. Cyclic loading at different depths allows monitoring the change in hardness at each depth thus provide a more reliable measurement.

The targeted peak loads were 2, 4, 6, 8 and 10 mN to measure the local material properties. In each loading cycle, 50 indentations were performed on a square grid. The replications of

the measurements were done in triplicate whenever possible. The number of eggs from some species was limited to only one, and when this is the case a single eggshell was indented on at least three different locations on the cross-section. Moreover, in order to increase accuracy, more than one location was indented on all of the samples. For instance, three indentation tests were done (each with 50 indentations) on different parts on the cross-section of a sample.

The loading was set to include hold periods so that after contacting the surface with the indenter tip, the load was applied and then held for 2 s at each load. Before increasing the load for the second one, a partial unloading was applied. The partial unloading was set to be 50% of the previous peak load. The indenter tip was returned to zero after the last peak load.

2.2.2.3 Fourier-Transform Infrared Spectroscopy and the Experimental Procedure for Mapping the Chemical Fingerprints of the Eggshells

Spectroscopy is a type of chemical analysis based on the interaction between light and matter. IR spectroscopy is a chemical fingerprinting technique based on the interaction of a sample with light in the infra-red region of the electromagnetic spectrum. The absorption (or transmission) of IR radiation excites molecular vibrations in the sample. This process can be thought of the vibration of an elastic string between two point masses. The extension of the spring is opposed by a restoring force which causes the system to vibrate upon release. The force constant of a string is related to the elasticity of the string. In a molecule, the force constant is related to the strength of the chemical bond and the energy of the vibration can be calculated using the harmonic oscillator model [205]. The vibrations of individual chemical bonds can be used to identify the functional groups in a sample which allows identifying the molecule if various vibrational modes are detected in a range of frequencies [206].

In an IR spectroscopy measurement the amount of energy absorbed is not the primary objective but the frequency of the vibration provides information about the chemical structure of the sample. In IR spectroscopy, wavenumber ($\tilde{\nu}$) is used instead of frequency (ν). Wavenumber is the reciprocal of wavelength, λ, and is measured in cm^{-1}. This indicates the number of waves in the infrared radiation in one centimeter. There is a simple relationship between wavenumber, frequency and the speed of light (c):

$$\nu = \frac{c}{\lambda} = c.\tilde{\nu} \tag{2.13}$$

FTIR spectrometers differ from the conventional ones which had been used to scan through all different frequencies in tandem which is a time consuming process. FTIR spectrometers have a fundamentally different signal processing setup which drastically reduces the time of analysis. FTIR has the advantage of collecting data at all frequencies simultaneously [207].

An interferometer allows processing the multi-frequency IR data as an interferogram, not in frequency domain but in of length [207]. The working principle of the FTIR spectrometer is depicted in Fig 2.7.

An IR beam is passed through a beamsplitter to produce two identical beams with the same intensity. One of the beams is directed to a fixed mirror and the other is to a movable mirror. Because of the mirrors is moved, the two beams pass through different paths with a difference of δ and they interfere constructively (if $\delta = 0, \lambda, 2\lambda..$) or destructively (if $\delta = \lambda/2, 3/2\lambda...$) [205]. In an IR absorption measurement, a large number of wavenumbers is scanned and where $\delta \neq 0$, waves would be out of phase and cancel each other. Once the interfered beams pass through the sample, absorption may occur at a specific wavelength where interference pattern is no longer the same as a result of absorption of the IR at specific wavenumbers by the sample [205]. An intensity signal as an interferogram is produced (I(δ) versus δ) which is then processed using Fourier transform to convert the length domain data into the frequency domain data as an intensity versus wavenumber spectrum [207]. Fig 2.8shows an interferogram which is converted to an IR spectrum via Fourier transform.

IR mapping is an enhanced version of IR spectroscopy to produce an IR map of a given area on the sample by scanning the area to collect copious spectra successively. IR mapping uses an attenuated total internal reflection (ATR) Objective (Fig 2.7, right). The ATR Objective has a probe that scans a specified area of interest [208]. The reason for using IR mapping technique is to analyze each eggshell sample in its 'native' state. In other words, obtaining a map of an eggshell cross-section shows the 'actual' spectrum of the cross-section. This is not possible with the conventional IR spectroscopy because in a conventional setup a powder sample needs to be prepared to collect its spectrum. Preparing the powder changes the order of the crystals in $CaCO_3$ [90]. This is in fact what was observed using the conventional IR analysis upon grinding the eggshell samples to a powder as will be shown in the Results section. Another advantage of mapping is the fact that biological samples are not homogenous [208]. Mapping of an area could reveal chemical differences, particularly, if there are local variations in micron scale [208].

The size of the scanned area varies between species as the thickness of the cross-section differs from one species to the other. The background scans were collected before sample mapping. The spectra were collected from wave numbers 4000 to 720 cm^{-1} unless otherwise stated. The resolution was set to 12 cm^{-1} which means that the maximum retardation (path difference, δ) in the interferometer is 1/12 cm. The pixel size was set to 1.56 μm. In each of the IR map, an area from the highest intensity section was chosen to collect an average spectrum. Wherever the absorption of a sample is not high enough to be reliable, a different section of the sample was mapped. During the entire period of mapping, the imaging detector

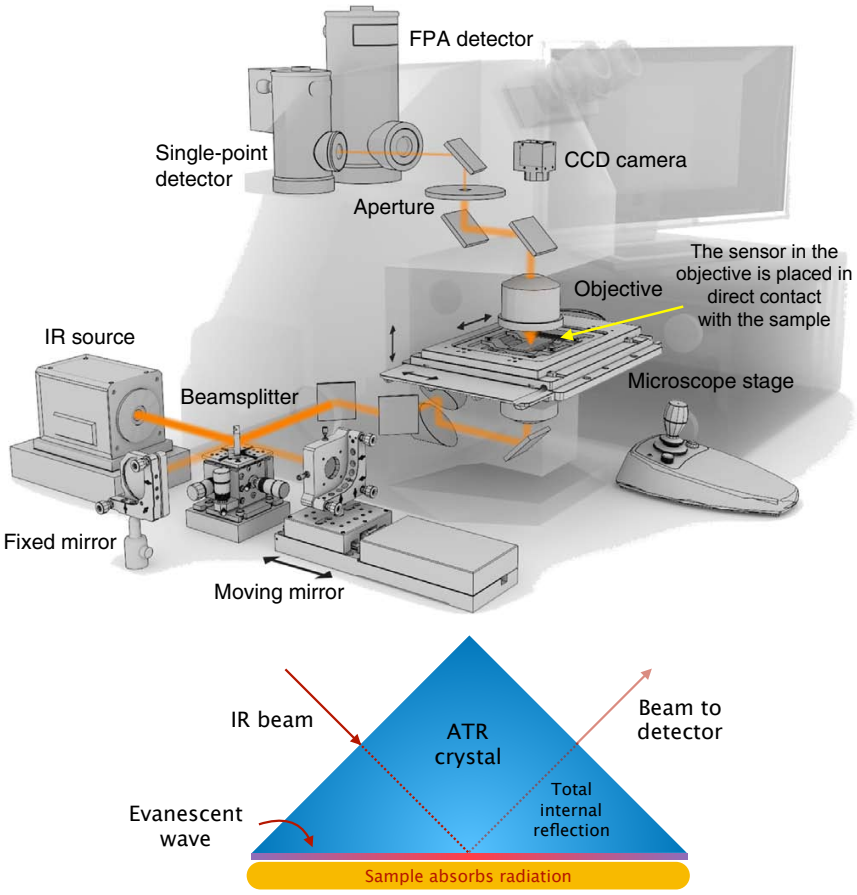


Figure 2.7: (Top) A schematic of an FTIR Imaging spectrometer. Bottom left shows the basic setup of an interferometer which is key in FTIR spectroscopy. IR beam is splitted in two parts with equal intensity. One of the beams is directed to a fixed mirror while the other is to a movable one. The different paths of the two beams because of different paths they take is reflected in an interferogram. Fourier transformation converts the interferogram in length domain (moving distance of the mirror) into frequency domain (wavenumber at which the absorption occurs). The schematic at the bottom depicts the IR light-sample interaction principle with a sensing mechanism based on an evanescent wave and attenuated total internal reflection (ATR) of IR beam [208].

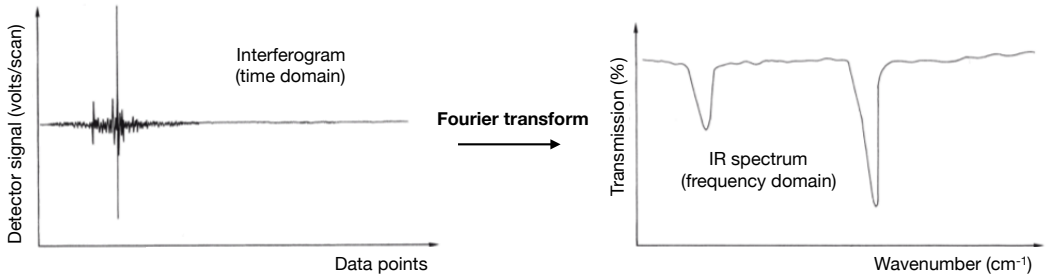


Figure 2.8: Fourier transform is used to convert the detector signal in an interferogram into a frequency domain IR spectrum. In an FTIR spectrometer, interferogram is produced in length domain as the detector signal is captured during the movement of the mirror producing interference patterns. IR spectra are given either as transmission or absorption intensity. Plots were modified from [205, 207].

was kept in liquid N_2 and the level was controlled to avoid going below a threshold.

The spectra of the eggshell cross-sections were plotted from wavenumbers 2000 to 720 cm^{-1}. The average spectra from a given high absorption area were averaged automatically by the software. The maximum limit of the wavenumber, which is 4000 cm^{-1}, was set up for data collection, however, $CaCO_3$ has no vibration modes above 1420 cm^{-1}. For this reason, the results were shown from 2000 to 720 cm^{-1}. Figure 2.9 depicts how the IR map of a sample corresponds to the resulting spectrum.

2.2.2.4 Data Analysis for Nanoindentation and FTIR Mapping

The reduced elastic modulus for each indentation test was determined from the unloading portion of the load displacement curve. An example of the load-partial unload-total unload curve from the present work is shown in Fig 2.10 (left). The plot belongs to the owl sample. Time versus load plot is shown in Fig 2.10 (right). As can be seen, a hold period of 2 s was set for each peak load. At the end of the testing, the sample unloaded completely (at around 32s). The fitting of the raw data from the load-partial unload-total unload testing was done using the Hysitron software. Data plotting and statistical tests were done in the R software.

With respect to the chemical analysis, assignment of peaks of IR spectra was done using the reported wavenumbers of $CaCO_3$ in the literature (e.g. [57, 90, 209]). A commercial $CaCO_3$ powder from the Sigma-Aldrich was used for comparison. The theoretical absorption spectrum of high order crystal $CaCO_3$ is seen in Fig 2.11 with three vibration modes, and the change in spectra if crystal order is changed.

The method for analyzing the crystal order in the present work is based on the relative intensity of absorption among the samples and the change of the overall absorption (i.e. full-width at half-maximum (FWHM) values of vibration modes [57, 90, 209]). The ratios of v_3/v_2

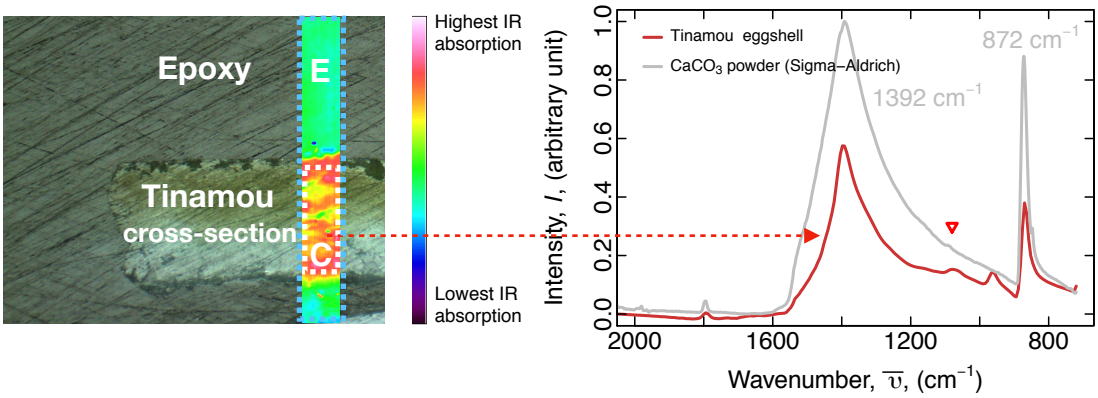


Figure 2.9: The optical microscope image of tinamou eggshell sample is shown to explain how the FTIR imaging corresponds to the resulting spectrum. The cross-section of tinamou eggshell is seen embedded in epoxy. FTIR (ATR Objective) sensor is brought into contact with the sample and an area is chosen for scanning (blue box in the microscope image). High absorption of IR is reflected as a red-orange area on the map and blue-purple colour indicates low absorption. In order to take an average spectrum from the highest intensity part on the cross-section, the FTIR software allows defining an area (white box in the microscope image) from which an average spectrum is calculated. The spectrum will be explained in more detail in the Results section.

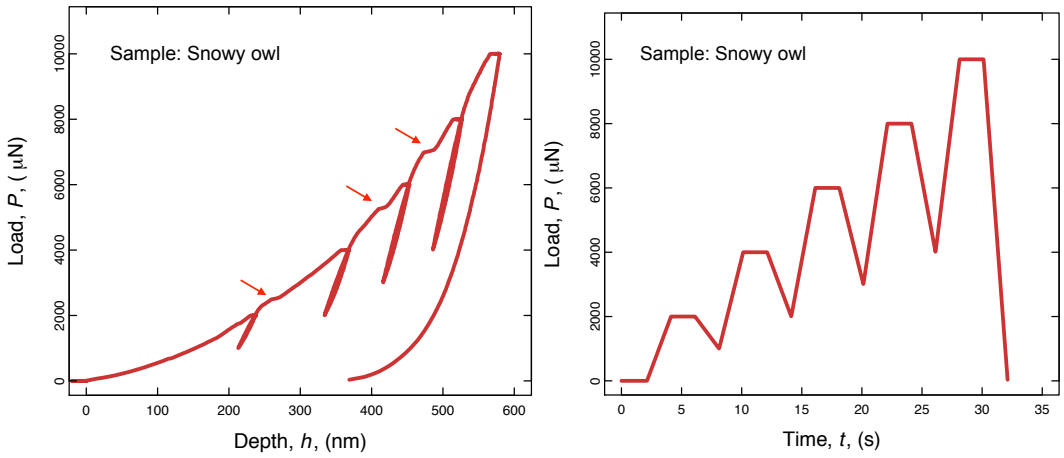


Figure 2.10: (Left) Load-unload curve of cyclic indentation of owl eggshell sample. Red arrows show 'pop-in' effects which cause a shift from expected curve during loading. Note the hysteresis formations during partial unloading. (Right) Time versus load curve of the same sample showing holding time (2s) at each peak load and complete unloading after maximum load (10mN).

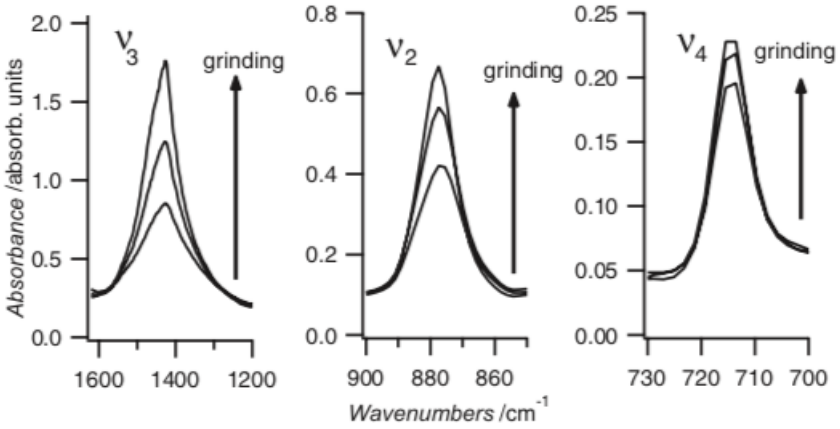


Figure 2.11: The IR spectrum of $CaCO_3$ for each of its three vibration modes. An increase in crystal order could be monitored as an increase in peak height/sharpness for a sample that is processed to change its crystal order, for instance, by grinding it to a fine powder [90].

were calculated for each sample to find a correlation between the order and peak intensity. In addition, as a quantitative measure of the sharpness and symmetry (e.g see [90]), FWHM values were also calculated. The calculations were done using the height and intensity of each peak after baseline correction in the Perkin-Elmer Frontier Spectrum software. The increase or decrease of the crystal order of the sample affects the resulting IR spectrum which could be used to probe a sample's crystal/amorphous content by its spectrum [90, 209, 210]. The implications of differences of IR spectra between samples are explained in the Results & Discussion sections.

2.3 Results

2.3.1 Results of Nanoindentation Analysis of Eggshells

Mechanical properties were tested with instrumented nanoindentation. Figures 2.12, 2.13, and 2.14 show the indentation modulus and Figures 2.15, 2.16 and 2.17 show the indentation hardness results of the eggshells from 14 species of birds.

Elastic moduli were found to vary between 20 to 60 GPa depending on the species studied. Hardness values were found to vary between 1 to 3 GPa. Increasing the indentation load after each partial unloading was observed to decrease both elastic modulus and hardness in most species. With increasing load, it was observed that the change in both modulus and hardness tend to be more 'stable' as the shape of the curves suggests approaching a final value. The implication of this observation will be explained in the Discussion.

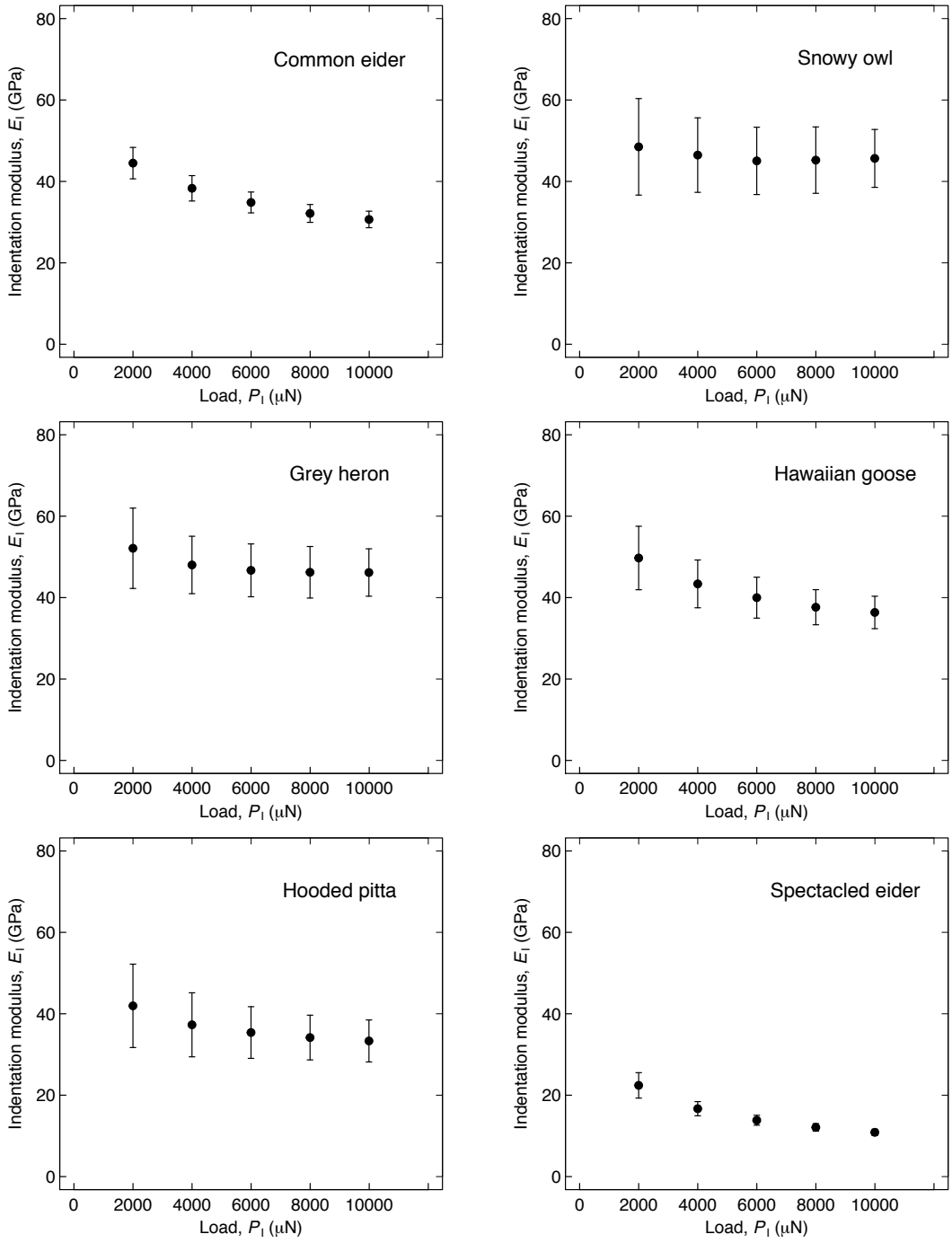


Figure 2.12: The indentation modulus results of common eider, owl, heron, Hawaiian goose, hooded pitta and spectacled eider for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

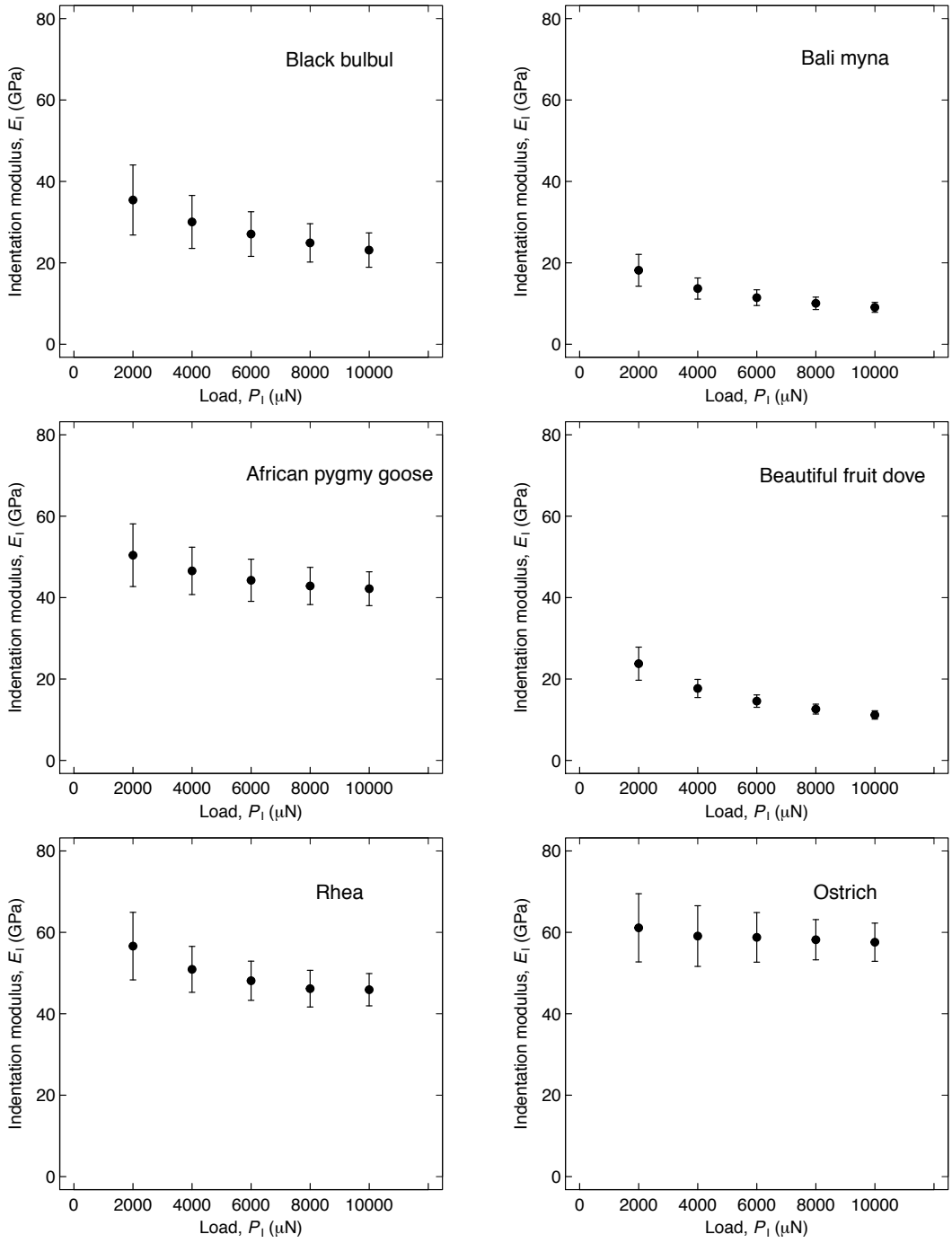


Figure 2.13: The indentation modulus results of black bulbul, Bali myna, pygmy goose, fruit dove, rhea and ostrich for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

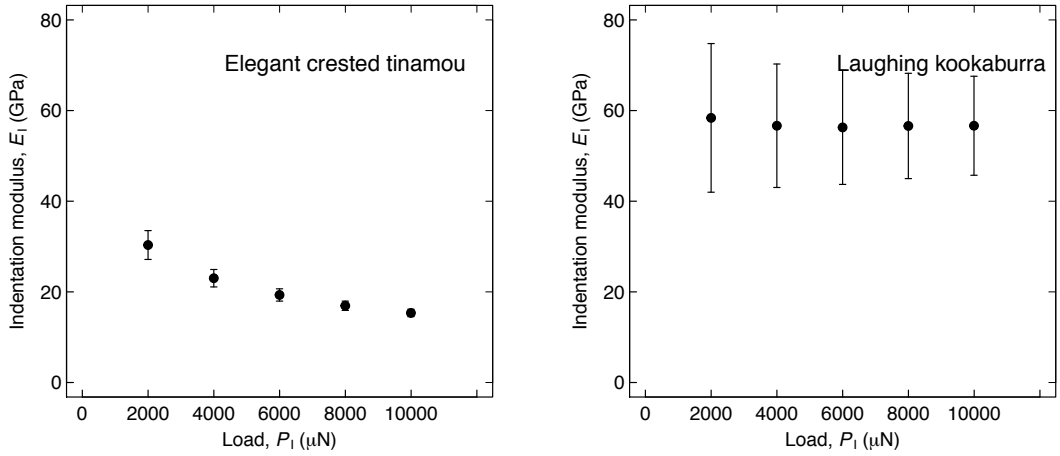


Figure 2.14: The indentation modulus results of tinamou and laughing kookaburra for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

For 2 mN, the highest elastic modulus, 61.11 ± 8.37 GPa, was observed for ostrich. The lowest elastic modulus was observed for Bali myna, 18.17 ± 3.90 GPa. For 10 mN, the highest elastic modulus, 57.57 ± 4.69 GPa, was again observed for ostrich and the lowest elastic modulus was again observed for Bali myna, 9.07 ± 1.21 GPa. The ranking of elastic modulus of all species from highest to lowest values both for 2 and 10 mN are shown in Table 2.2.

Regarding the hardness, for 2 mN, rhea has the highest value, 3.04 ± 0.86 GPa. The lowest hardness belongs to Bali myna, 1.69 ± 0.71 GPa. The rest of the hardness values were found not to diverge substantially from the modulus trend. Although heron and owl exhibit much less hardness compared to their modulus ranking, all other samples have comparable hardness to their rankings in modulus. For 10 mN, rhea has again the highest value, 2.68 ± 0.41 GPa. Similarly, the lowest hardness belongs to Bali myna, 1.15 ± 0.34 GPa. The ranking of hardness of all samples from highest to lowest values both for 2 and 10 mN are shown in Table 2.3.

Table 2.4 shows the degree of decrease in modulus (and hardness) as the slope of the modulus (and hardness) with increasing load. For instance, because the decrease is most pronounced for tinamou, the slope is the lowest, -18.10x10^{-4}. In most cases the decrease in modulus and hardness follows the same trend. For instance, the decrease in both modulus and hardness for tinamou are the highest. In some cases, the decrease in modulus and hardness follows a different trend from each other, that is, the degree of decrease in modulus is less (or more) pronounced compared to hardness. For instance, the decrease in modulus in Bali myna is more pronounced compared to its hardness (see individual plots in Figs 2.13,2.16 and the ranking in Table 2.4).

Table 2.5 shows modulus to hardness ratios at 2mN and 10 mN in a descending order. The highest ratio belongs to kookaburra for both 2 and 10 mN. This is because kookaburra has

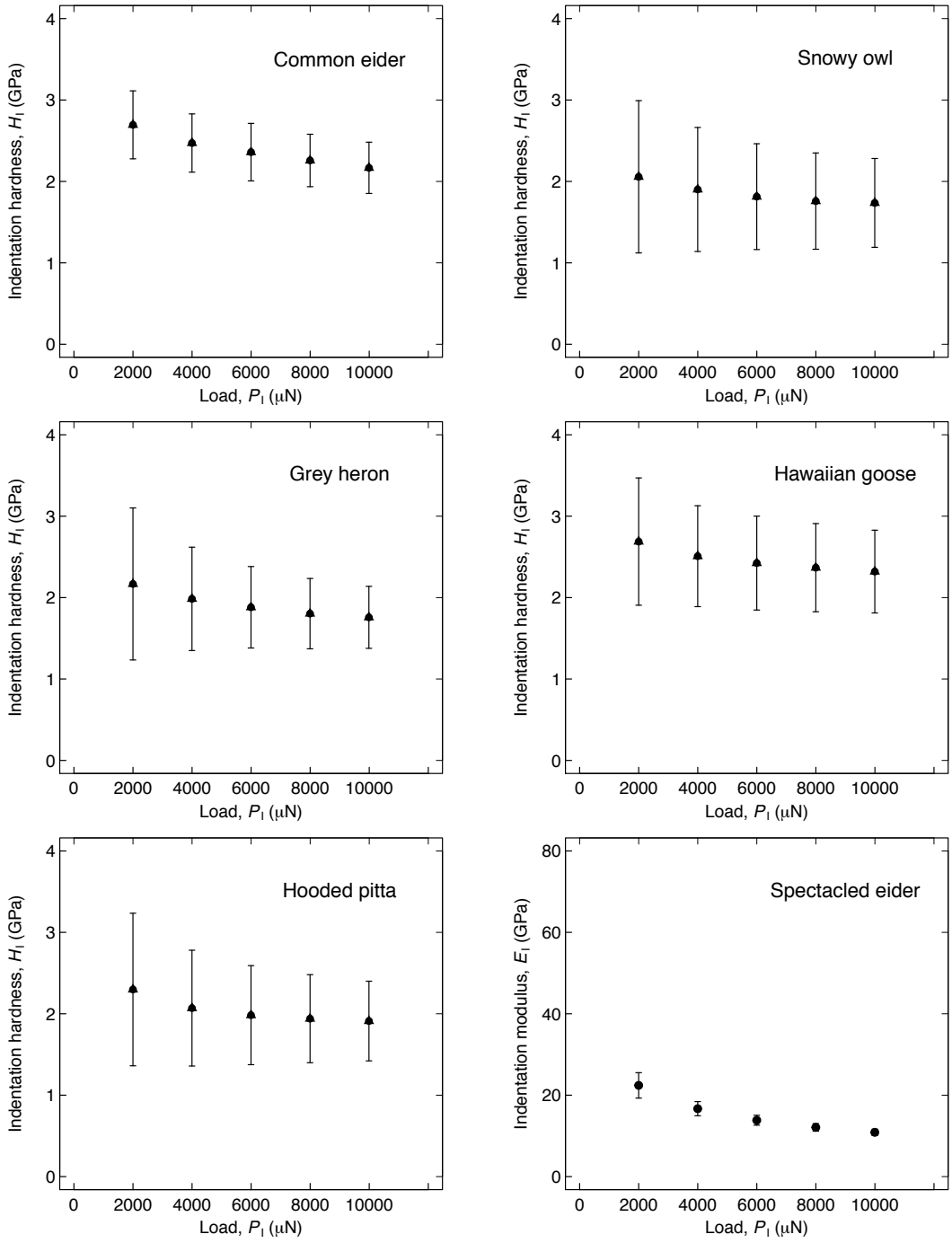


Figure 2.15: The indentation hardness results of common eider, owl, heron, Hawaiian goose, hooded pitta and spectacled eider for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

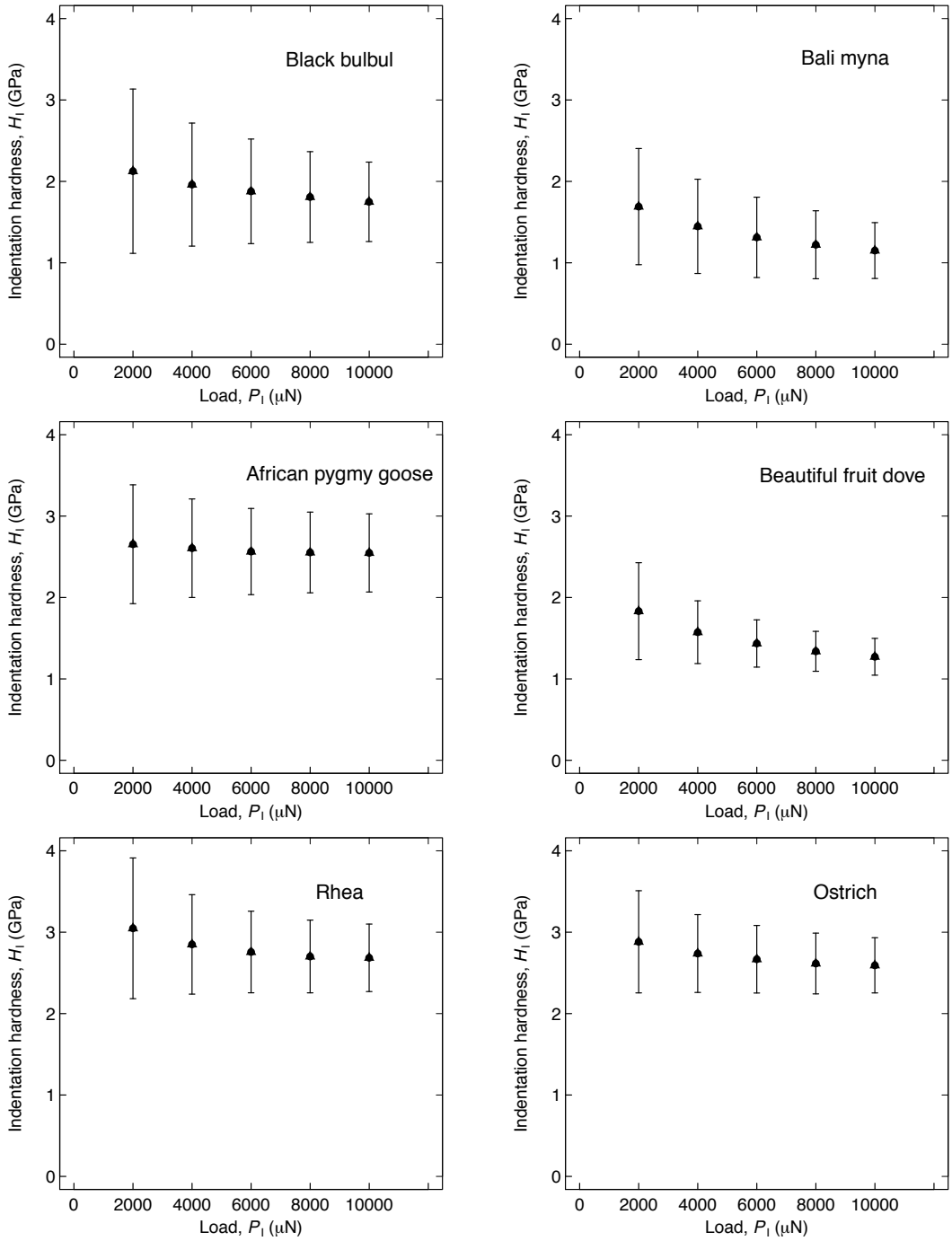


Figure 2.16: The indentation hardness results of black bulbul, Bali myna, pygmy goose, fruit dove, rhea and ostrich for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

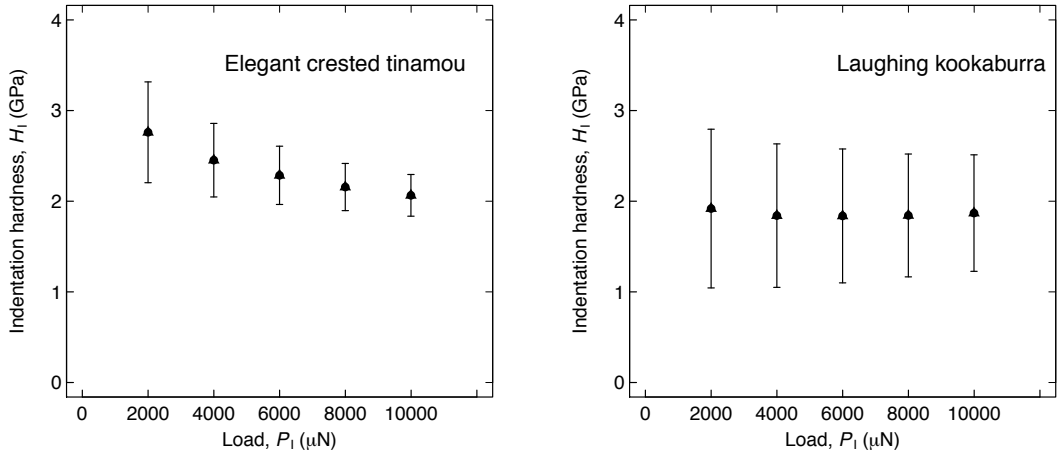


Figure 2.17: The indentation hardness results of tinamou and laughing kookaburra for five peak loads in cyclic mode nanoindentation. The error bars show standard deviation.

Species	Modulus at 2 mN	Species	Modulus at 10 mN
Ostrich	61.11 ± 8.37	Ostrich	57.57 ± 4.69
Kookaburra	58.38 ± 16.39	Kookaburra	56.65 ± 10.92
Rhea	56.62 ± 8.29	Heron	46.16 ± 5.82
Heron	52.11 ± 9.86	Rhea	45.91 ± 3.96
Pygmy goose	50.41 ± 7.69	Owl	45.65 ± 7.10
Hawaiian goose	49.72 ± 7.79	Pygmy goose	42.18 ± 4.15
Owl	48.49 ± 11.85	Hawaiian goose	36.35 ± 3.98
Common eider	44.49 ± 3.85	Hooded pitta	33.32 ± 5.15
Hooded pitta	41.94 ± 10.23	Common eider	30.66 ± 2.02
Black bulbul	35.43 ± 8.60	Black bulbul	23.13 ± 4.22
Tinamou	30.33 ± 3.18	Tinamou	15.35 ± 0.81
Spectacled eider	25.17 ± 3.12	Spectacled eider	12.76 ± 0.77
Fruit dove	23.76 ± 4.06	Fruit dove	11.17 ± 1.03
Bali myna	18.17 ± 3.90	Bali myna	9.07 ± 1.21

Table 2.2: The modulus results of the eggshell samples from highest to lowest for 2 and 10 mN with standard deviations.

Species	Hardness at 2 mN	Species	Hardness at 10 mN
Rhea	3.04 ± 0.86	Rhea	2.68 ± 0.41
Ostrich	2.88 ± 0.62	Ostrich	2.59 ± 0.33
Tinamou	2.76 ± 0.55	Pygmy goose	2.54 ± 0.47
Common eider	2.69 ± 0.41	Hawaiian goose	2.31 ± 0.50
Hawaiian goose	2.68 ± 0.78	Common eider	2.16 ± 0.31
Pygmy goose	2.65 ± 0.73	Tinamou	2.06 ± 0.23
Hooded pitta	2.29 ± 0.93	Hooded pitta	1.91 ± 0.48
Spectacled eider	2.26 ± 0.66	Kookaburra	1.86 ± 0.64
Heron	2.16 ± 0.93	Spectacled eider	1.85 ± 0.25
Black bulbul	2.12 ± 1.00	Heron	1.75 ± 0.38
Owl	2.05 ± 0.93	Black bulbul	1.74 ± 0.48
Kookaburra	1.91 ± 0.87	Owl	1.73 ± 0.54
Fruit dove	1.83 ± 0.59	Fruit dove	1.27 ± 0.22
Bali myna	1.69 ± 0.71	Bali myna	1.15 ± 0.34

Table 2.3: The hardness results of the eggshell samples from highest to lowest for 2 and 10 mN with standard deviations.

Species	Slope of Modulus (x10^4)	Species	Slope of Hardness (x10^5)
Tinamou	-18 (R^2= 0.8828)	Tinamou	-8.441 (R^2= 0.9251)
Common eider	-16 (R^2= 0.9159)	Fruit dove	-6.779 (R^2= 0.9073)
Hawaiian goose	-16 (R^2= 0.8916)	Bali myna	-6.531 (R^2= 0.9159)
Black bulbul	-15 (R^2= 0.9020)	Common eider	-6.350 (R^2= 0.9480)
Fruit dove	-15 (R^2= 0.8882)	Black bulbul	-6.179 (R^2= 0.8926)
Spectacled eider	-14 (R^2= 0.8843)	Spectacled eider	-5.090 (R^2= 0.9577)
Rhea	-13 (R^2= 0.8282)	Heron	-5.013 (R^2= 0.9168)
Hooded pitta	-10 (R^2= 0.8433)	Hooded pitta	-4.523 (R^2= 0.7848)
Pygmy goose	-10 (R^2= 0.8852)	Hawaiian goose	-4.390 (R^2= 0.8912)
Bali myna	-10 (R^2= 0.8792)	Rhea	-4.367 (R^2= 0.8194)
Heron	-6.8 (R^2=0.6661)	Owl	-3.923 (R^2= 0.8715)
Ostrich	-3.9 (R^2=0.8441)	Ostrich	-3.502 (R^2=0.8738)
Owl	-3 (R^2= 0.4750)	Pygmy goose	-1.325 (R^2= 0.8332)
Kookaburra	-1 (R^2=0.2566)*	Kookaburra	-0.487 (R^2= -0.0618)*

Table 2.4: The slopes (indicating the decrease in modulus and hardness with increasing load) and their fitting parameter R^2 values of modulus (left) and hardness (right) of each species to quantify the degree of decrease in modulus/hardness with increasing load. Data ranking indicates from highest to lowest decrease in modulus/hardness. *Note that because there is no decrease (there is even a slight increase) in modulus/hardness for kookaburra, linear fitting becomes obsolete as is seen from the R^2 values.

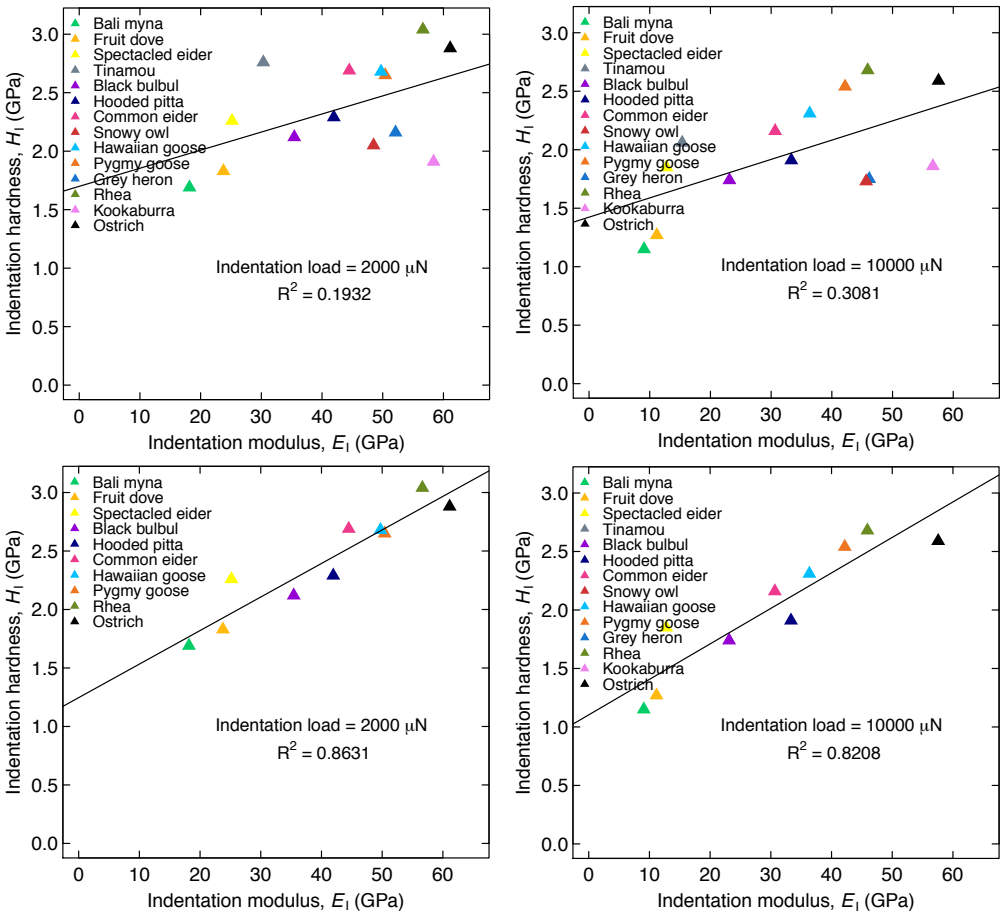


Figure 2.18: (Top) Modulus versus hardness results for all species at 2 mN and 10 mN where no linear correlation is found. (Bottom) If the outliers are omitted, then a linear correlation emerges.

high modulus but low hardness, compared to the general trend among other species (e.g. Bali myna has both low modulus and low hardness). It should be noted that although the hardness of Bali myna decreases more compared to its modulus from 2 mN to 10 mN, the ratio of modulus to hardness is affected minimally. This is because the main factor for the differences between species is affected mainly by the modulus and not hardness (e.g. differences between the highest and lowest modulus between species is around 6 orders of magnitude at 10 mN while the hardness is less than 3). It is worth noting that the species that have unexpectedly high or low hardness values can be inferred from Table 2.5. This feature of some species affects the modulus-hardness correlation as explained below.

It might be insightful to compare the modulus and hardness values both at 2 and 10 mN. It is known for various biological materials that there is a linear correlation between modulus and hardness which means that if a biological material has a high modulus it is also expected

Species	E / H at 2 mN	Species	E / H at 10 mN
Kookaburra	30.56	Kookaburra	21.87
Heron	24.12	Ostrich	21.48
Owl	23.65	Pygmy goose	21.13
Ostrich	21.21	Hawaiian goose	20.47
Pygmy goose	19.02	Heron	19.87
Rhea	18.62	Owl	19.03
Hawaiian goose	18.55	Rhea	18.17
Hooded pitta	18.31	Common eider	17.91
Black bulbul	16.71	Hooded pitta	16.57
Common eider	16.53	Black bulbul	13.21
Fruit dove	12.98	Tinamou	8.82
Spectacled eider	11.13	Fruit dove	8.79
Tinamou	10.98	Bali myna	7.88
Bali myna	10.75	Spectacled eider	7.37

Table 2.5: The modulus to hardness ratios for 2 mN and 10 mN in a descending order.

to have a high hardness [200]. It is clear that at both 2 and 10 mN there is no linear correlation between modulus and hardness in the eggshell samples ($R^2 = 0.1932$ and 0.3081, respectively, Fig 2.18). If the outliers in terms of unexpectedly high/low hardness values (kookaburra, tinamou, heron and owl) are omitted, though, a linear correlation seems to emerge ($R^2 = 0.8631$ and 0.8208, respectively, Fig 2.18). The implication of the emergence of a linear correlation once the outliers are omitted is explained in the Discussion.

2.3.2 Results of FTIR Mapping of Eggshells

As mentioned in the Methods section, IR spectroscopy analysis of the eggshells is non-trivial in a conventional setup because grinding samples into powder changes specimens' spectra. This is a result of a change in the crystal order of the sample as explained in detail in the literature [57, 90, 209]. This is in fact what was also observed in the present study. Fig 2.19 shows the IR spectra of two of the eggshell samples upon grinding them into powder. Although in terms of their spectra, ostrich and common eider are different from each other (see the ATR Objective maps below), upon grinding, differences cannot be observed anymore. This is the case even for ν_2 and ν_4 modes which are known to be most sensitive to the crystal order [90]. For this reason, the ATR Objective IR mapping was conducted for an accurate analysis. In Fig 2.19 the spectrum of $CaCO_3$ is given for comparison with the vibration modes and their corresponding wavenumbers. The spectrum was collected with a commercial powder sample using the Perkin-Elmer Frontier 400 system.

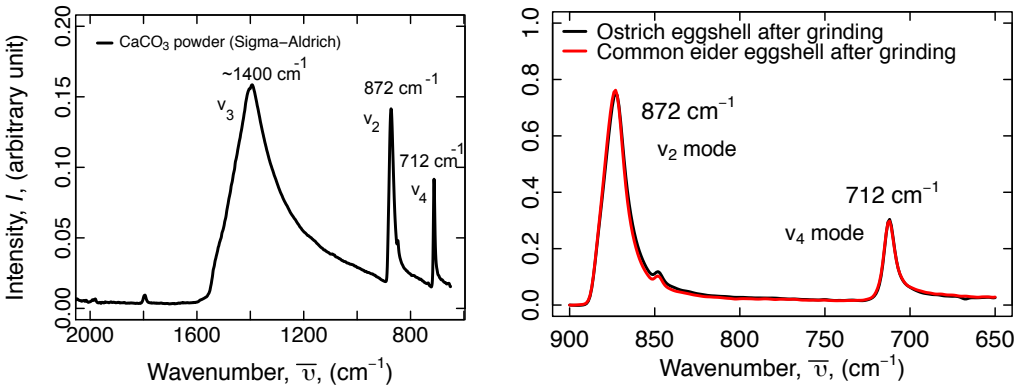


Figure 2.19: (Left) The spectrum of $CaCO_3$ shows the vibration modes and corresponding wavenumbers. (Right) The spectra of ostrich and common eider taken from a conventional IR spectrometer setup using powder samples. Differences between samples are not observable contrary to the IR mapping (see below).

The Chemical characterization of the eggshells of all species were done using ATR-FTIR mapping using the same samples that were used for nanoindentation. Figures 2.20, 2.21 and 2.22, show the results of the ATR-FTIR mapping.

The resulting spectra show characteristic calcite peaks. The absence of the peaks specific for aragonite and vaterite can be verified from the spectra. Aragonite has a sharp peak at 1083 $cm^{-1}$and a ν_2 peak shifted to 855 cm^{-1}along with a shoulder at 844 cm^{-1}[141]. There is also no peak at 744 cm^{-1} which is characteristic for vaterite. All samples have absorption peaks at the specific vibration modes of $CaCO_3$ which, in some cases, shift slightly compared to the geogenic calcite. There are four vibration modes of carbonate minerals and their corresponding wavenumbers are as follows: 1080 cm^{-1} (ν_1, symmetric stretch), 872 cm^{-1} (ν_2, out-of-plane bending), 712 cm^{-1} (ν_4, asymmetric stretch) and 1420 cm^{-1} (ν_3, symmetric stretch) .

In the resulting spectra, two of the vibrations, 876 cm^{-1} (ν_2) and 1420 cm^{-1} (ν_3) show the most pronounced peaks. The 1080 cm^{-1} (ν_1) peak is much less pronounced. For half of the samples (common eider, heron, Bali myna, kookaburra, rhea, ostrich, hooded pitta) ν_1 peaks are not even pronounced enough to show any significant difference from each other.

The intensity of 1080 cm^{-1} (ν_1) peak provides information about the existence of the amorphous content (i.e. disorder) in the sample [211–214]. This means, although being qualitative (based on the overall shape of the broad peak), ν_1 peak is used to infer the crystal degree through the amorphous content. There is no data for 713 cm^{-1}(ν_4) mode because of a shortcoming of the ATR- FTIR Imaging microscope used in this work, that is, the mapping accessory has the lowest limit at 720 cm^{-1}. Another note for the spectra is that the 1420 cm^{-1} (ν_3) mode of $CaCO_3$ was observed to shift to a lower wavenumber around 1390 cm^{-1} which is still consistent with $CaCO_3$ because the spectrum of the commercial $CaCO_3$ powder was shown

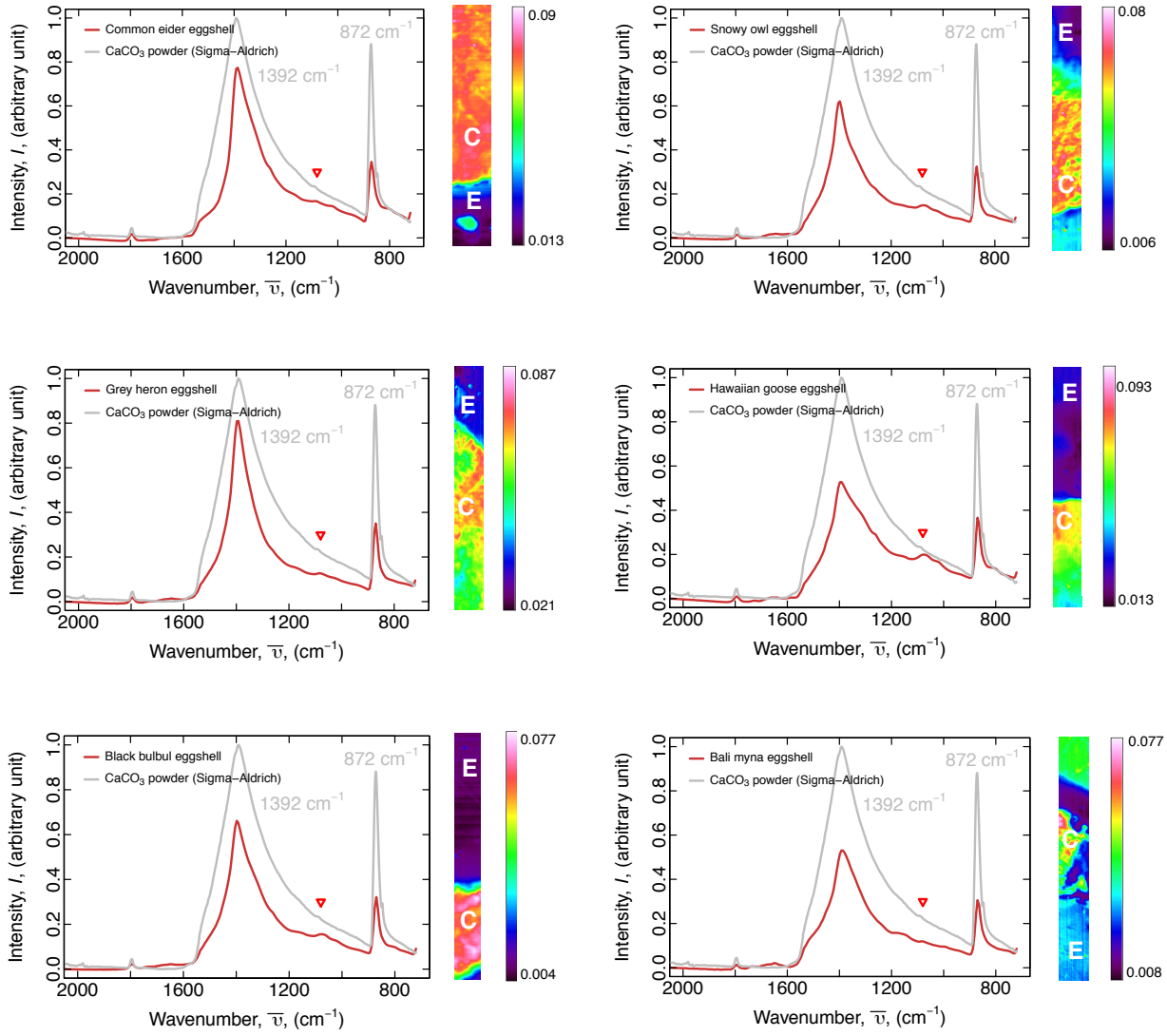


Figure 2.20: The ATR-FTIR spectra and IR maps of common eider, owl, heron, Hawaiian goose, black bulbul and Bali myna eggshell cross-sections. In each graph, the spectrum of commercial $CaCO_3$ is given for comparison with the eggshell sample. 1392 cm^{-1} and 872 cm^{-1}peaks correspond to v_3 symmetric stretch and v_2 out-of-plane bending modes of $CaCO_3$, respectively. The red triangle shows 1080 cm^{-1} v_1 symmetric stretch mode which indicates amorphous structure. The colour scale shows absorbance intensity. C, cross-section, E, epoxy.

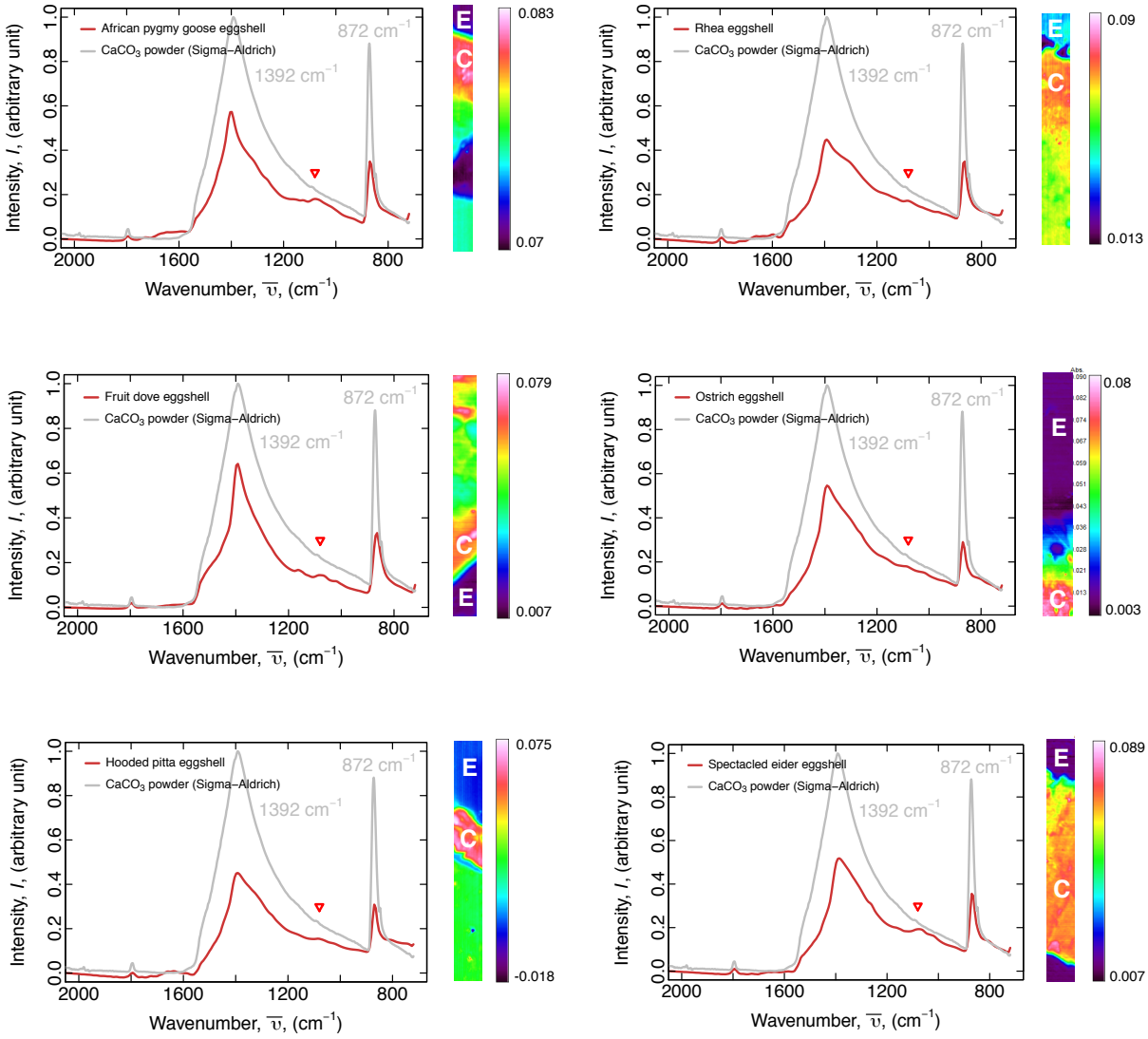


Figure 2.21: The ATR-FTIR spectra and IR maps of *N auritus,* rhea, fruit dove, ostrich, hooded pitta and spectacled eider eggshell cross-sections. In each graph, the spectrum of commercial $CaCO_3$ is given for comparison with the eggshell sample. 1392 cm^{-1} and 872 cm^{-1}peaks correspond to v_3 symmetric stretch and v_2 out-of-plane bending modes of $CaCO_3$, respectively. The red triangle shows 1080 cm^{-1} v_1 symmetric stretch mode which indicates amorphous structure. The colour scale shows absorbance intensity. C, cross-section, E, epoxy.

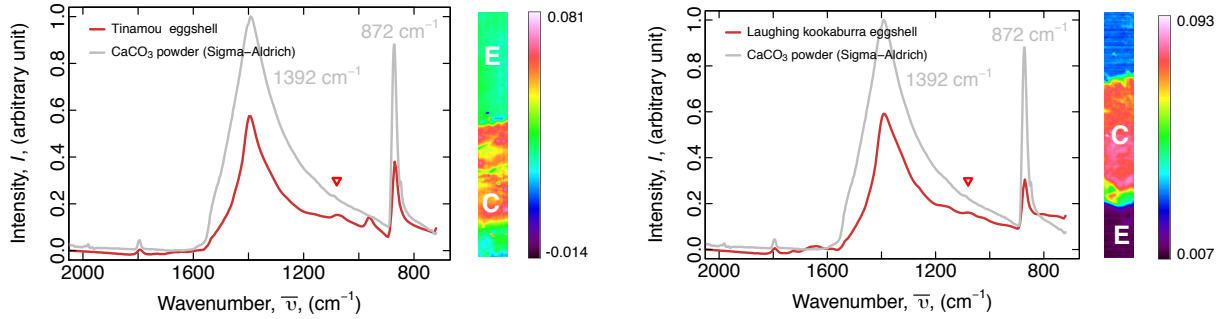


Figure 2.22: The ATR-FTIR spectra and IR maps of tinamou and kookaburra eggshell cross-sections. In each graph, the spectrum of commercial $CaCO_3$ is given for comparison with the eggshell sample. 1392 cm^{-1} and 872 cm^{-1}peaks correspond to v_3 symmetric stretch and v_2 out-of-plane bending modes of $CaCO_3$, respectively. The red triangle shows 1080 cm^{-1} v_1 symmetric stretch mode which indicates amorphous structure. The colour scale shows absorbance intensity. C, cross-section, E, epoxy.

Sample	FWHM of v_3 (cm^{-1})
Heron	99
Owl	121
Common eider	122
Black bulbul	130
Fruit dove	131
Tinamou	132
Kookaburra	144
Bali myna	149
$CaCO_3$ powder (Sigma-Aldrich)	165
Pygmy goose	173
Hawaiian goose	186
Spectacled eider	188
Ostrich	193
Hooded pitta	202
Rhea	212

Table 2.6: The full width at half maximum values of the eggshell samples for v_3 peak.

Sample	v_3 / v_2
Calcite (Ideal crystal [theoretical])	3.50
Common eider	2.78
Kookaburra	2.74
Heron	2.69
Ostrich	2.59
Black bulbul	2.50
Owl	2.35
Fruit dove	2.29
Bali myna	2.06
Hooded pitta	2.05
Pygmy goose	1.97
Spectacled eider	1.81
Hawaiian goose	1.81
Tinamou	1.75
Rhea	1.60
$CaCO_3$ powder (Sigma-Aldrich)	1.33

Table 2.7: The v_3 / v_2 values of the eggshell samples compared to the theoretical value for an ideal calcite crystal.

to have a v_1 peak at 1392 cm^{-1}. Because IR spectroscopy provides the chemical fingerprint of a sample, it is not surprising that all species have characteristic $CaCO_3$ spectra and they do not reveal any high intensity organic content, probably because of the low concentration of the protein content. The most significant difference in the IR spectra between species is the intensity and symmetry of v_3 peaks around 1393 cm^{-1} based on the FWHM values. This is because a lower FWHM value indicates a higher intensity which in turn results in sharper peaks with a high amount of asymmetry. It can be seen from the spectra that the higher the intensity of v_3 peak, the sharper it becomes (e.g. comparison of rhea versus heron).

The intensity of v_2 peaks around 876 cm^{-1} show differences yet the symmetry does not reveal any substantial variability. In $CaCO_3$ samples from any origin, v_1 peak at around 1080 cm^{-1} does not show a high intensity. Regarding this mode, Hawaiian goose and pygmy goose show two of the most pronounced peaks compared to the other samples. Another sample with a different v_1 peak is the spectacled eider. An outlier among all samples is the tinamou, which shows both the v_1 peak and another peak at around 970 cm^{-1}.

The results of v_3/v_2 ratios and the FWHM values are presented in Table 2.6, Table 2.8 and Table 2.7. For ease of comparison, FWHM of v_3, FWHM of v_2, and v_3/v_2 ratios were given. In case of the FWHM of v_3, the ranking was done in an ascending order (Table 2.6) because the lower the FWHM, the higher the crystal order [90] (also see [57, 209, 211, 214]). The

Sample	FWHM of v_2 (cm $^{-1}$)
$CaCO_3$ powder (Sigma-Aldrich)	20
Owl	22
Hawaiian goose	23
Heron	23
Kookaburra	24
Rhea	25
Ostrich	25
Spectacled eider	25
Hooded pitta	25
Tinamou	25
Bali myna	26
Common eider	27.5
Pygmy goose	28
Black bulbul	29
Fruit dove	45

Table 2.8: The full width at half maximum values of the eggshell samples for v_2 peak.

reason for correlation is that the increasing order of crystal gives rise to sharper peaks with increasing intensity thus FWHM decreases rapidly [90] (also see [210]).

In the case of FWHM of v_2, the ranking was also done in an ascending order (Table 2.8) but the difference between most of the samples is not high enough to extract crystal order information. Unlike the FWHM values, the v_3/v_2 ratios was given in a descending order in Table 2.7.

2.3.3 Correlation Between the Mechanical and Chemical Properties

Quantitatively, the comparison of the modulus, hardness, modulus/hardness ratio, FWHM of v_3, FWHM of v_2 or v_3/v_2 ratio do not reveal any correlation between the mechanical and chemical properties. The lack of the v_4 mode in spectra is a shortcoming because if it was available, the ranking of the crystal order of the samples would be more accurate without the need to use other vibration modes (i.e. v_3, v_2). As explained above, although the conventional IR spectroscopy allows collecting the v_4 peak at 714 cm^{-1}, upon grinding, it does not reveal differences anymore.

One of the ways to determine the extent of a correlation between the mechanical and chemical properties of the eggshells of different species is to visualize the measured quantities (i.e. modulus, hardness etc.) in a way that makes it easy to compare these quantities. For this reason, a correlation map is produced for the species studied in the present work. The species

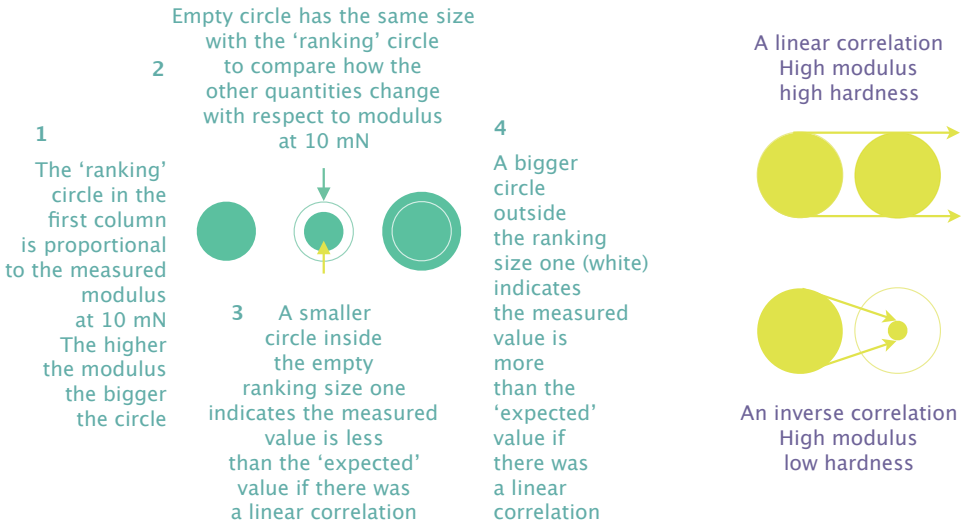


Figure 2.23: An explanation of how the correlation between the mechanical and chemical properties of the eggshells is visualized in the correlation map in Fig 2.24.

are first ranked from the highest modulus to the lowest as measured at 10 mN. The other quantities were then ranked compared to the modulus at 10 mN. This allows the visualization of data to follow how a specific quantity changes between species. For instance, if there was a linear correlation between mechanical and chemical properties, then a species with a high modulus would also exhibit a high crystal order as inferred from the IR spectra. In the correlation map, each colour represents one of the species. The size of the circles are drawn to scale which means that the bigger the circle the higher the quantity being represented. A detailed explanation of how the symbols represent the experimental data in the correlation is explained in Fig 2.23. The correlation map is seen in Fig 2.24.

Although the chemical properties ought to affect mechanical properties, nanoindentation and FTIR spectroscopy did not offer a 'global' correlation that fits all the species. Still, it may be possible to extract some 'local' correlation which is reflected in the correlation map as a pattern of circles growing in size from top left to bottom right. This 'shift' might indicate a higher crystal order for the species that have relatively low modulus and hardness but this trend does not hold for all species. It is not unreasonable to expect a linear correlation between high modulus and high crystal order. This is because, in general, amorphous minerals tend to be 'weaker' in terms of their mechanical properties. Whether or not the 'shift' in the map implies an inverse correlation between low modulus-hardness and high crystal order is not clear. This is because an important indication of the crystal order that could be extracted from an IR spectrum is the ratio of the v_2 and v_4 vibrations of $CaCO_3$. The lack of this ratio for the eggshells is a shortcoming which should be avoided by either collecting the spectra with an IR

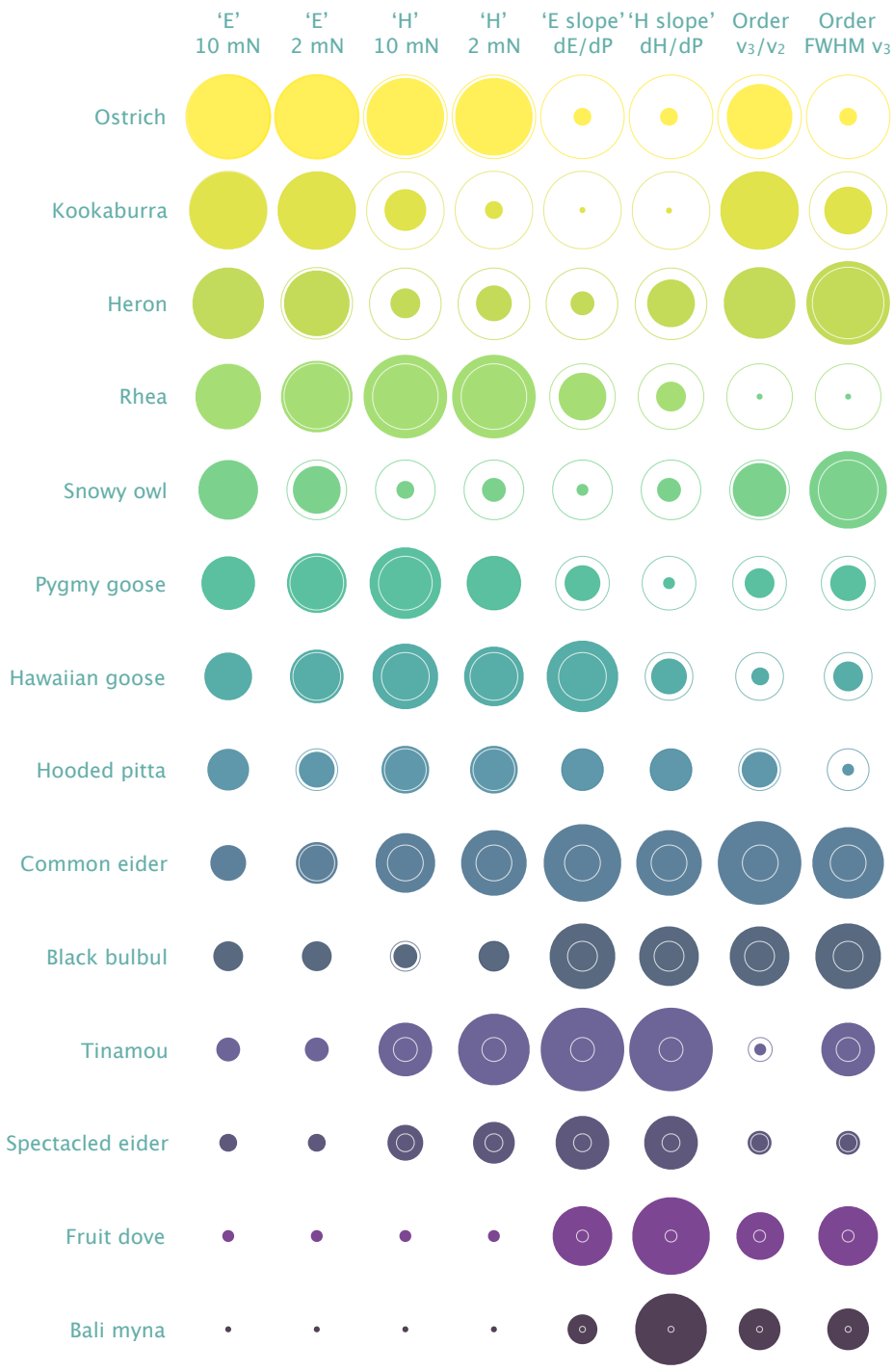


Figure 2.24: A correlation map of the mechanical and chemical properties of the eggshells. The elastic modulus and hardness are denoted with 'E' and 'H'. The 'Slope(s)' correspond to the degree of change in the modulus and hardness as the load increases (see Table 2.4). The 'Order(s)' denoted with v_3/v_2 and FWHM of v_3 represent the IR measurements to infer the crystal order of each of the eggshells. For the details of how the circles visualize the correlation, see Figure 2.23.

mapping sensor that could scan below 710 cm^{-1} or collecting data on the crystal properties (e.g. crystal size) of each species using an EBSD in an SEM. This would allow comparing the crystal structures of different species more accurately.

2.4 Discussion

The eggshells from 14 species were studied to investigate their mechanical and chemical properties. This investigation was expected to provide information on the organic-inorganic interplay, such as the effect of the organic matrix. Nanoindentation results show some differences between species in terms of modulus and hardness. FTIR mapping suggests that there are differences in the crystal orders between species either as an effect of the organic matrix or as a result of the amorphous content. Indentation and FTIR mapping data do not reveal a correlation between mechanical and chemical properties. The implications of the results and possible improvement strategies for both measurements are discussed below.

2.4.1 Nanoindentation Measurements of the Eggshells

A common theme of all the indentation results of the eggshells is a decrease in modulus and hardness with increasing load. The decrease follows the same trend for modulus and hardness (i.e. degree of the decrease is same for both properties). In some samples, such as rhea or hooded pitta, the decrease seems to approach a 'final' value through the maximum load. In some other samples, such as owl or kookaburra, there is a slight decrease after the first load and then there is no further decrease (there is even a slight increase for kookaburra). The important point is the reason of the decrease. The most likely reason is the pop-in effect. The pop-in effect arises here from the fact that eggshells are highly porous structures so that during an indentation, sooner or later the probe hits one of the pores which causes a shift in the load-displacement curve. Once the indenter hits a pore, the measurement is affected, as if, there is a displacement while in fact there is free space thus resulting a misleading decrease in the hardness (or modulus or both). Another reason which is likely to happen is cracking. This means that because of the loading, brittle samples crack and then a pop-in effect is observed. Cracking decreases modulus because it is driven by a release of strain energy [215].

Because some samples are more difficult to test mechanically, such as biological ones, different strategies have been developed to increase the accuracy of indentation measurements. One of these strategies is to conduct cyclic loading at different peak loads [198]. Before conducting cyclic indentation (i.e. partial unloading method), conventional single load-unload measurements were done on the eggshells, however, it became clear that the single load-unload

tests do not provide consistent, reproducible results. The reason is that the loading-unloading curves were indeed found to include pop-ins [216] which most probably arise because of the pores but might also arise from cracking [217].

Another possibility for the decrease in hardness (and modulus) is the 'indentation size-effect'. It is described as a decrease in hardness with increasing indentation depth. This decrease is followed by an increase and a final decrease before 'saturation' [218]. The size-effect is well known for metals [218] but it is not entirely clear to what extent it also holds for biological samples. Another point about mechanical measurements is the slightly higher standard errors for hardness measurements compared to elastic moduli. This might also arise from the combination of pop-in and size effects. It is important to note that the pop-in effect and surface roughness might compromise the mechanical and (to a lesser extent) chemical measurements. Possible improvements/refinements for data collection and analysis could prove beneficial (see below 'Strategies for the Refinement of Nanoindentation and FTIR Mapping Analyses').

The organic content of biominerals affects material properties depending on the function of the biomineral. The mechanical properties of eggshells are no exception and exhibit differences compared to the geogenic $CaCO_3$. The indentation modulus of pure calcite is around 77 GPa [219]. It is known that biogenic calcite has different moduli in different organisms, for example, although both eggshells and mollusks are biogenic $CaCO_3$, mollusk *A. rigida* has a modulus around 70 GPa [220]. The chicken eggshell has been reported to have a modulus around 50 GPa at the palisade layer [99]. As explained in the Results section, moduli of eggshells vary between 10 to 60 GPa at 10 mN. Considering the difference, it could be said that the different species have adopted strategies to optimize modulus of shells which might be related to the evolutionary pressures.

The hardness results vary between 1 to 3 GPa at 10 mN. Pure calcite has a hardness of 2.5 GPa [220]. More than half of the species studied here exhibit less hardness below 2 GPa. This mechanical adaptation might be related to hatching strategies. An eggshell should be robust enough to withstand physical damage (e.g. against predator attack) but also should be fragile enough for the chick to hatch. An example from the parasitic birds (those that lay eggs in other nests) relates the hardness of the eggshells to a surviving strategy. It has been reported that the common cuckoo, *C. canorus*, has a higher hardness compared to their respective hosts (cuckoos lay eggs in the nests of other 'host' birds) to increase the chance of survival [221]. Further studies on other species could shed light on differences and their implications. The contribution of the organic matrix to the mechanical outcomes could help understand eggshell evolution better but most species -including many of the species studied here- have not been studied or reported in the literature yet.

There are some unexpected results regarding the elastic moduli of eggshells. The first one is kookaburra as an outlier. With respect to the phylogeny, all living birds are in the subclass Neornithes [222]. After this subclass, birds divide into two major category, Neognathae and Paleognathae [222]. Ostrich, rheas, emus, cassowarries, kiwis, extinct moas and elephant birds, and tinamous belong to paleognathae [222]. All other birds belong to neognathae [222]. Although there is a long divergence in terms of evolution between kookaburra and ostrich (and rhea), moduli of these species are the highest. It is interesting that kookaburra has a high modulus in the first place considering that it follows a different trend with a low hardness (For the modulus-hardness relationship see below). The decrease in modulus and hardness with increasing load is lowest for kookaburra which might be related with high modulus/low hardness property. The comparison of the organic matrix of ostrich and kookaburra could reveal key insights but the extraction of the organic matrix and characterization of the proteins is limited for ostrich and has not been done for kookaburra. Another unexpected result was found for common eider and spectacled eider. Although they are closely related in the phylogenetic tree, the former has a modulus of 45 GPa whereas the latter has almost half of it around 24 GPa. A detailed comparison between kookaburra and common eider could reveal the reasons behind high modulus and low hardness relationship.

The hardness and modulus results of tinamou make it another outlier. Although it belongs to the paleognathae [222], tinamou exhibits surprisingly low modulus around 30 GPa but its hardness is close to the ostrich. Contrary to the kookaburra, tinamou has the opposite trend, having low modulus and high hardness. The decrease in the modulus and hardness with increasing load is the highest for tinamou. As mentioned above, it is the lowest for kookaburra. Because the high modulus-high hardness trend is not followed both by kookaburra and tinamou, extraction and comparison of the organic matrix of kookaburra and tinamou may help to unveil a pattern where the intra-crystalline proteins affect modulus or hardness. As mentioned above, the possible bias from pop-in effects or surface roughness should be eliminated to show that kookaburra or tinamou does exhibit the same mechanical properties that were shown here. On the other hand, it is thought that multiple indentations at different sections for each sample increased the accuracy of measurements conducted here.

A correlation between modulus and hardness might be useful to see if there is a linear relationship between these parameters. It is known that for a myriad of materials, including the biological ones, the modulus and hardness are linearly correlated [200] (also see Fig 1-10 in [198]). Here, both at 2 and 10 mN, there is no linear relationship between the modulus and hardness. Once four samples with unexpectedly high/low hardness are removed, the rest of the samples seem to follow a linear pattern as would be expected. This assumes that the outliers have unexpectedly high/low hardness which might not be the case and it might be that

they have low/high modulus. Possible strategies to increase the accuracy of the mechanical measurements will be explained at the end of this section.

2.4.2 FTIR Mapping of the Eggshells

The chemical characterization of the eggshells might provide further insights on the effect of the organic matrix. IR absorption of proteins themselves are unlikely to be revealed in the spectra because the concentration of them is too low to be detectable. The organic content plays an indirect role because the incorporation of proteins changes the lattice parameters of $CaCO_3$, thus giving rise to differences in the resulting spectra [223],[57].

A strategy to study the effect of the organic matrix is to determine to what extent an eggshell contains amorphous phase or more generally disordered mineral. The ATR-FTIR Imaging allows IR mapping of the samples without any sample preparation that can cause a change in the crystallinity (see [209] for the effect of grinding) hence the comparison.

The amount of amorphous content could be approximated from the v_1 peak as it is known that a broad and diffuse peak around 1080 cm $^{-1}$ is an indication of disorder [211, 212, 214]. The 1080 cm $^{-1}$ peak does appear in some species in a more pronounced way compared to the others. For instance, the amorphous content of Hawaiian goose seems to be higher than other eggshells. Although, 1080 cm^{-1} is an indication of disorder, and even the area under it could be calculated, it is more reliable to do a quantitative comparison using the peak heights and the FWHM values [57, 90, 209, 210]. A comparison between the species for determining the crystal order could be done with using the v_3/v_2 ratio and FWHM values (see in particular [90]). Because the samples were mapped in their 'native' states (i.e. without grinding which if done does change crystal order), spectra were not normalized for analysis which would change the FWHM values. The FWHM values are in some cases less revealing, because, in particular for FWHM of v_2, values are too close to each other for a reliable comparison. For instance, for 5 samples the v_2 values are equal. It is reasonable to rely more on the v_3/v_2 ratio and compare it to the FWHM of v_3 mode. As a consistent example, rhea has the lowest v_3/v_2 ratio which means it has high disorder and it is verified by the fact that it has the highest v_3 FWHM. The results of the comparisons (Tables 2.6,2.7) could provide insights about the extent of the organic matrix. For example, samples with the highest disorder such as rhea, suggest that the amount of the intra-crystalline proteins could be high enough to increase the disorder hence the resulting spectra. This might also mean that the rhea C-type lectins bind to calcite with high affinity thereby increasing local protein concentration. Because it is known that struthiocalcins from ostrich have strong binding energies [19], extracting the rhea C-type lectin(s) might offer high affinity proteins for $CaCO_3$ binding. The implication of a high affinity C-type lectin will be discussed below.

2.4.3 Implications of the Characterization Data on Selection of New C-type Lectins

The incorporation of proteins and their possible long-term preservation would only be possible if the concentration of the intra-crystalline protein is high enough to be feasible for a real world application. While the chemistry of incorporation could be tuned to load the target chemicals into crystals [224], proteins and more importantly vaccine antigens are challenging to manipulate since they should retain their strict 3D structures. The C-type lectins could be used to facilitate incorporation of a given target protein by either adding it in the solution as a 'chaperone' or as a genetically-fusing it to the target protein. The presence of lysozyme in the eggshell suggests that it is incorporated along with another 'chaperone' protein because lysozyme was shown to be not incorporated alone (see Chapter 4). The most likely protein to act as a 'chaperone' is OC-17. In addition, because the C-type lectins tend to be incorporated efficiently, they could allow incorporation of the target proteins at higher concentrations.

Synthetically, highly efficient proteins can be selected using directed evolution [225, 226]. Analyzing various eggshells may provide a directed evolution 'experiment' that has been on-going for millions of years between different species. The amino acid sequence differences among the avian C-type lectins likely make some of them better suited for incorporation than others, for instance, struthiocalcins from ostrich.

Based on the v_3/v_2 ratios and FWHM of v_3, rhea C-type lectins -if extracted- might be promising strong binders of calcite. It might also be interesting to find out why some species (e.g. ostrich) have not one but two C-type lectins. It is worth noting that OC-23 [136], suggests that the second C-type lectins in avian eggshells are being eliminated during evolution and a 'ghost' protein might be left with the same amino acid sequence but post-translationally modified for a yet unknown function (e.g. glycosylation in OC-23).

Tinamou might also be a promising candidate to study the intra-crystalline proteins because the mechanical and chemical characterizations suggest that it could provide new C-type lectin(s) that have high affinity towards calcite. It should be noted that this interpretation is based on the disorder of the IR spectra assuming a high local protein concentration causing the higher disorder. Lastly, determining the gene encoding these C-type lectins and sequencing corresponding cDNAs to enable molecular cloning would be important contributions. Hitherto the only cloned avian C-type lectin is OC-17, cloning of which will be explained in Chapter 3.

2.4.4 Strategies for the Refinement of the Nanoindentation and FTIR Mapping Analyses

A possible improvement for the nanoindentation measurements could be done by preparing samples with ion milling after polishing with silicon carbide abrasive papers. Although P4000 (5 μm) abrasive paper provides a smooth surface, ion milling is substantially better as it was understood that electron diffraction pattern could only be obtained after ion milling (Gatan Ilion^{+}II system was used to prepare ostrich eggshell for collecting its diffraction pattern using electron backscatter diffraction in a FEI Helios NanoLab FIB/SEM. Abrasive paper polishing did not provide enough smoothness to map the surface for diffraction. After ion milling, Kikuchi lines were observed as a result of proper diffraction [227]).

Another improvement might be tested if the pop-in effects are 'corrected' from the raw data manually. The horizontal shifts could be discarded from the data and the new data could be compared to the original one. This procedure might allow testing if the pop-ins do affect the measurements substantially. The raw data can be processed automatically to discard the pop-ins if they cause substantial errors. With respect to the indentation size-effect, it could be probed by testing the eggshells with increasing indentation depths systematically and monitoring the change in hardness. To some extent this is what was done in the present work yet in order to monitor if there is a size-effect, the depth should be increased further because the size-effect causes a decrease following an increase and then a final decrease before 'saturation' [218]. It is possible that at the depths studied here, only the first decreasing phase was observed.

The ion milling could improve FTIR mapping as well. Samples after ion milling could be probed with the ATR Objective to see if there is significant differences with smoother surfaces but if the diffraction pattern of each of the eggshell sample could be obtained, then it would be possible to compare the crystal orders and crystal sizes. A more accurate quantitative data on the crystal order is expected to be more insightful than the FTIR mapping alone.

Chapter 3

Cloning of OC-17 and Ivy, and Extraction of Wild-type OC-17

3.1 Introduction

This chapter presents the results of the recombinant protein expression and purification along with the wild-type OC-17 extraction from the eggshell. The overall aim of the cloning experiments is to express OC-17 in bacteria both to obtain high concentration of protein without relying on eggshells and to optimize the recombinant OC-17 expression to be able to modify OC-17 for functional studies. For instance, FRET-FLIM could be used to probe the function of OC-17 and point mutations in OC-17 where key amino acids in the active site are changed systematically one at a time could provide insights and might verify molecular simulations regarding the function of OC-17 (see Chapter 5 Section 5.2.1). Furthermore, the recombinant OC-17 would allow synthesizing fusion proteins/vaccines genetically attached to OC-17 in order to be incorporated into $CaCO_3$. The genetic fusion of proteins or peptides to the vaccine antibodies could be done through the persistent peptides in the fossil avian eggshells, such as struthiocalcin's well preserved peptide sequence YSALDDDDYPKG [19]. In the present work, only OC-17 and Ivy expressions are presented. The same strategy could be used to fuse the well-preserved peptides to the N or C terminus of the target proteins to immobilize them into $CaCO_3$.

As will be explained below, the purification of the wild-type OC-17 from the chicken eggshell is non-trivial. More importantly, the effects of the eggshell incorporation and purification on the 3D structure of the protein are unknown. In addition, there is no other source from which OC-17 could be purified and compared to the eggshell purified protein. For these reasons, cloning of OC-17 into bacteria and subsequent expressions and purifica-

tions are needed. Still, for the comparison of the wild type to the recombinant OC-17, the wild type OC-17 needs to be isolated from the eggshell. The isolation of OC-17 from the shell extract is not straightforward as lysozyme is also present in the eggshell. The similarity of the chemical properties of lysozyme and OC-17 makes it difficult to separate these proteins. The methods described in the literature require high concentration acid treatment and HPLC purification using organic solvents, which are likely to denature the protein (see Chapter 1 Subsection 1.3.8.1). The strategy followed here is to use a lysozyme inhibitor protein to bind to lysozyme and separate the new lysozyme-inhibitor complex from OC-17 using the affinity tag in the inhibitor. The lysozyme inhibitor, Ivy protein, is encoded by the *ivy* gene which is found in some strains of the *E. coli* where the protein had evolved to defend the cell against lysozyme's ability to disrupt the cell wall of the bacteria (see Subsection 3.2.2.2).

The details of the cloning, expression and purification are presented in the Materials & Methods section. Then, in the Results section, the extraction of OC-17 from the eggshell, cloning and expression of the Ivy protein and cloning and optimization of expression of the recombinant OC-17 were explained in detail. Lysozyme binding ability of the Ivy protein was also tested. The verification of OC-17 expression was conducted by anti-Histag antibody as the recombinant OC-17 contains 6xHistag to assist purification. OC-17 was cloned to be secreted into the periplasmic space yet the expression of the protein in the growth medium was also tested. The protein expression was induced by IPTG which is added to the growth medium and both the periplasmic space and the medium fractions were tested for OC-17 expression at different growth temperatures. The last part of the chapter discusses the results of the cloning and purification steps. Strategies for improvement of the expression and cloning strategies and implications of the C-type lectin-like protein/peptide-assisted incorporation of proteins/vaccines into $CaCO_3$ are discussed.

3.2 Materials & Methods

3.2.1 Materials

The *E. coli* cells (strain K-12 MG 1655) were donated by Dr Gillian Fraser from the Department of Pathology, University of Cambridge. The primers and the plasmid pET26b were purchased from Sigma-Aldrich. The XL1-Blue Competent cells were purchased from Stratagene. The Q5 High-Fidelity 2X Master Mix was purchased from New England Biolabs. The 2X ReddyMix PCR Master Mix and FastDigest DpnI were purchased from the Thermo Scientific. PCR purification, gel extraction, plasmid purification kits were purchased from the QIAGen. OC-17 gene was synthesized by Invitrogen Thermo Scientific. LB growth medium,

agar plates, buffer solutions were obtained from Department of Hemaetology, University of Cambridge. All purification columns were purchased from the GE Healthcare unless otherwise stated. The SDS-PAGE gels were purchased from Life Technologies. The western blot membranes and ultrafiltration centrifugal filters were purchased from Merck Millipore.

3.2.2 Methods

3.2.2.1 Extraction and Purification of the Wild-type OC-17

OC-17 was extracted from the eggshell by dissolving shell pieces in 5% (v/v) acetic acid overnight at room temperature on an orbital shaker. The extract solution was centrifuged at 4500 rpm for 10 min and first filtered with Whatman Grade 1 filter paper to remove shell pieces and undissolved membranes. Then the solution was filtered with 0.22 µm filter paper using a vacuum filtration system. The filtered solution was concentrated with 10 kDa cut-off ultrafiltration centrifugal filters. CM Sepharose cation exchange resin was packed in a 50 ml syringe and equilibrated with 50 mM acetic acid, pH 5. Concentrated extract was loaded to the column to remove the acetic acid from protein solution. Unbound fraction was discarded. Elution was done with 0.9 M $(NH_4)_2SO_4$- 0.1 M NaH_2PO_4 using two column volumes of solutions. Phenyl Sepharose hydrophobic interaction resin was packed in a 50 ml syringe and equilibrated with 0.9 M $(NH_4)_2SO_4$- 0.1 M NaH_2PO_4 . Stepwise gradient was done in two steps with 0.7 M and 0.5 M $(NH_4)_2SO_4$- 0.1 M NaH_2PO_4 at pH 7.4. Unbound fractions were discarded. Elution of the rest of the proteins was done with 0.1 M NaH_2PO_4- 0.1 M NaCl pH 7.4. Elution was concentrated, diluted with the elution buffer and concentrated again to ensure complete buffer exchange. CM Sepharose was equilibrated with 0.1 M NaH_2PO_4- 0.1 M NaCl pH 7.4. Elution from hydrophobic interaction column was loaded to the equilibrated CM Sepharose and unbound fraction was collected. Then the column was washed with 0.1 M boric acid pH 9.2 harnessing chromatofocusing to remove negatively charged proteins at pH 9.2. Final elution was done with 0.1 M NaH_2PO_4- 1 M NaCl pH 7.4. Buffer exchange was done into 0.1 M NaH_2PO_4- 0.1 M NaCl pH 7.4. Fractions were analyzed with SDS-PAGE.

Because lysozyme is also present along with OC-17 (and OC-23), it could be separated from the protein solution using the ivy protein. Ivy is an inhibitor of lysozyme and its interaction with lysozyme forms a protein complex that can be removed from the OC-17 solution. The cloning, synthesis, purification and function of Ivy protein is described below.

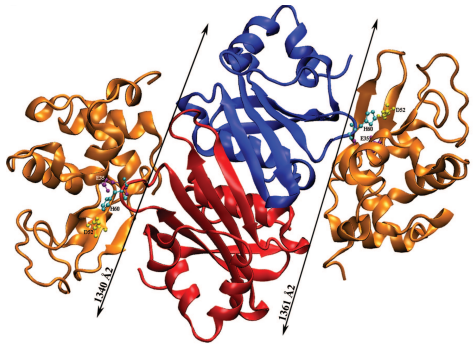


Figure 3.1: The 3D structure of Ivy-lysozyme complex. Ivy protein is encoded by ivy gene in *E. coli* K-12 strains and is a potent inhibitor of lysozyme. Ivy structure is shown in red-blue and two lysozymes are shown in orange [229].

3.2.2.2 Cloning of the Ivy Gene for the Synthesis of Inhibitor of Vertebrate Lysozyme Protein (Ivy)

Ivy (Inhibitor of vertebrate lysozyme, formerly known as ykfE) is a potent inhibitor of the type C lysozyme (UniProt P0AD59). Lysozyme-C is found in the chicken egg and the eggshell. Ivy is encoded in bacteria including some strains of *E. coli*. *Ivy* gene has been found in the K-12 MG1655 strain [228]. It is encoded by a 474-base open reading frame to be translated as a homodimer protein with 157 amino acid [228]. The 3D structure of Ivy protein interacting with two lysozymes is shown in Fig 3.1. Ivy protein has been studied for its function after initial purification attempts indicated an interaction with lysozyme including co-purification and co-crystallization. Subsequently, Ivy protein has been shown to inhibit lysozyme through a strong interaction [228]. It is worth noting that the purification of OC-17 from eggshell extract results (in almost all cases) co-elution of lysozyme even when different purification techniques are used. It is, however, unlikely that OC-17 has an interaction with lysozyme. In order to circumvent the problem of co-elution of lysozyme with OC-17, Ivy protein could be used to remove the lysozyme from the eggshell protein solution. This could be useful for the eggshell extraction of wild-type OC-17 rather than expression of the recombinant one in bacteria. In the present work, *Ivy* gene was cloned from the *E. coli* K-12 1655 and purified via 6XHistag fragment. Ivy protein was analyzed for its interaction with lysozyme by testing Ivy protein's ability to bind lysozyme through capturing Ivy-lysozyme complex in a nickel affinity chromatography. Ivy protein was then used for OC-17 purification from the eggshell protein extract. *Ivy* gene sequence (474-nucleotide) is presented in Appendix A.

The genomic DNA of *Ivy* was extracted from the *E. coli* K-12 MG 1655 and purified with the QIAprep Spin commercial kit using QIAgen Miniprep handbook. The cloning of *Ivy* was done using the protocol in the literature [228]. The only difference was that instead of the

plasmid pQE60 used in the reported protocol [228], another plasmid, pQE30, was used in the present work. In the reported protocol [228], 6XHistag is located at the C-terminus. In the pQE30 plasmid, 6XHistag is located at the N-terminus. In order to replicate the reported protocol the *Ivy* gene (*ykfE* gene in Fig 3.5) was inserted upstream 6XHistag so that it would be read at the C-terminus.

The forward primer for pQE30 is as follows:

AGAGGATCGCATCACCATCACCATCAC

The reverse primer for pQE30 is as follows:

CATAGTTAATTTCTCCTCTTTAATGAATTCT.

The forward are reverse primers for *Ivy* gene were designed with overlapping regions to plasmid primers. The primers used for cloning are given below:

The forward primer for ivy is GAGGAGAAATTAACTATGggcaggataagctcggg

The reverse primer is GTGATGGTGATGCGATCCTCTtttaaaattaaagccatccgg

The uppercase letters in *Ivy* primers show the overlapping regions with plasmid primers. The plasmids were dissolved in water after centrifugation. The plasmid and the *Ivy* DNA fragments were amplified with PCR using Thermo 2X ReddyMix PCR Master Mix. The DNA fragments were analyzed with 1% agarose gel. A high concentration band was excised and the DNA was extracted from the band using QIAquick Gel Extraction Kit. The construction of plasmid was done using Gibson assembly with PCR using Q5 High-Fidelity 2X Master Mix.

Transformation of plasmids to the XL1-Blue Competent cells was done using the manufacturer's transformation guidelines. In order to select the cells that contain the plasmid, transformed cells were grown overnight at 37 ºC on an LB agar containing ampicillin. Six colonies were taken into a 5 mL medium and grown overnight at 37 ºC. Plasmid DNA extraction was done with QIAprep Spin Miniprep Kit based on the protocol in QIAprep Miniprep handbook. The plasmid was sequenced at the Genewhiz, Cambridge, UK. The sequencing result is presented in the Appendix A.

After sequence verification, the cells were transferred into the expression system (*E. coli*) and a 100 mL overnight culture was grown at 37 ºC with constant shaking at 200 rpm. The cells then transferred to 1 L growth medium and protein expression was conducted at 37 ºC until desired the optical density is reached. The medium was centrifuged, cells were harvested and resuspended in a sodium phosphate buffer pH 8 and then cell lysis was done with a French press system. The solution was then centrifuged and the supernatant was dialyzed against a sodium phosphate buffer pH 8 containing 25 mM imidazole and purified with a nickel-affinity column in an AKTA Fast Protein Liquid Chromatography (FPLC) system with an 0.5 M imidazole gradient. Then further purification was done using a size-exclusion chromatography column and the collected fractions were analyzed using SDS-PAGE.

3.2.2.3 Cloning of *OC-17* Gene for the Synthesis of OC-17 Protein

As explained in the Chapter 1, the cDNA encoding for OC-17 has been reported in the literature. In order to clone *OC-17* for protein expression in bacteria, the cDNA sequence was obtained from the GenBank (KF835610). The primers used for the cloning are given below:

OC-17 forward primer is caccctcagttcgaaaaaagcgcAGATCCGGATGGTTGTGGTCC

OC-17 reverse primer is gtcattaatggtgatggtggtgatgGGCTGCTGCTTTACAAACAAATGC

The forward primer for pET26b is catcaccaccatcaccattaatgac and the reverse primer is tgcgctttttttcgaactgagggtg. The lowercase letters in the OC-17 primers show the complementary regions with the plasmid primers. The plasmid construct is seen in Fig 3.2.

The plasmids were dissolved in water after centrifugation. The plasmid and *OC-17* DNA fragments were amplified in PCR using Thermo 2X ReddyMix PCR Master Mix. The DNA fragments were analyzed with 1% agarose gel. The band that contains the PCR product was excised to obtain the fragment by dissolving agarose in 50 ºC water. The DNA was extracted from the dissolved gel using QIAquick Gel Extraction Kit.

The construction of plasmid to clone *OC-17* into pET26b was done using Gibson assembly with PCR using Q5 High-Fidelity 2X Master Mix. Transformation of the plasmids into the competent cells was done using the manufacturer's transformation guidelines. In order to select the cells that contain the plasmid, transformed cells were grown overnight at 37 ºC on an LB agar containing kanamycin. After the colony selection, cells were taken into 5 mL medium and grown overnight at 37 ºC. The plasmid DNA extraction was done with QIAprep Spin Miniprep Kit based on the protocol in QIAprep Miniprep handbook. The plasmid was sequenced at Genewhiz, Cambridge, UK. The sequencing result can be found in the Appendix A.

After the sequence verification, plasmids were transferred to Tuner DE3 competent cells for expression and two 10 mL overnight cultures (LB and TY medium containing 1% glucose) with 50 µg/mL kanamycin were grown at 30 ºC with constant shaking at 200 rpm. The following day, 50 mL media were inoculated with 0.5 mL overnight culture. The new cultures were grown at 37 ºC until optical density at 600 nm (OD600) reaches 0.5-0.6. Different IPTG concentrations (0, 0.1, 0.25, 0.5 and 1 mM) were tested to find the optimum growth condition. Half of the each LB and TY medium cultures incubated either at 30 ºC or at 25 ºC overnight. The following day, 2 mL of each of four test cultures were centrifuged at 8000 rpm and one of the test cultures from each medium were directly analyzed using the dot blot method against anti-Histag antibody.

The pellets of the remaining two test cultures resuspended in B-PER (Bacterial Protein Extraction Reagent) containing 0.2mg/ml lysozyme, 5UI DNAase I, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, and incubated at room temperature for 30 min followed by centrifugation. Because the

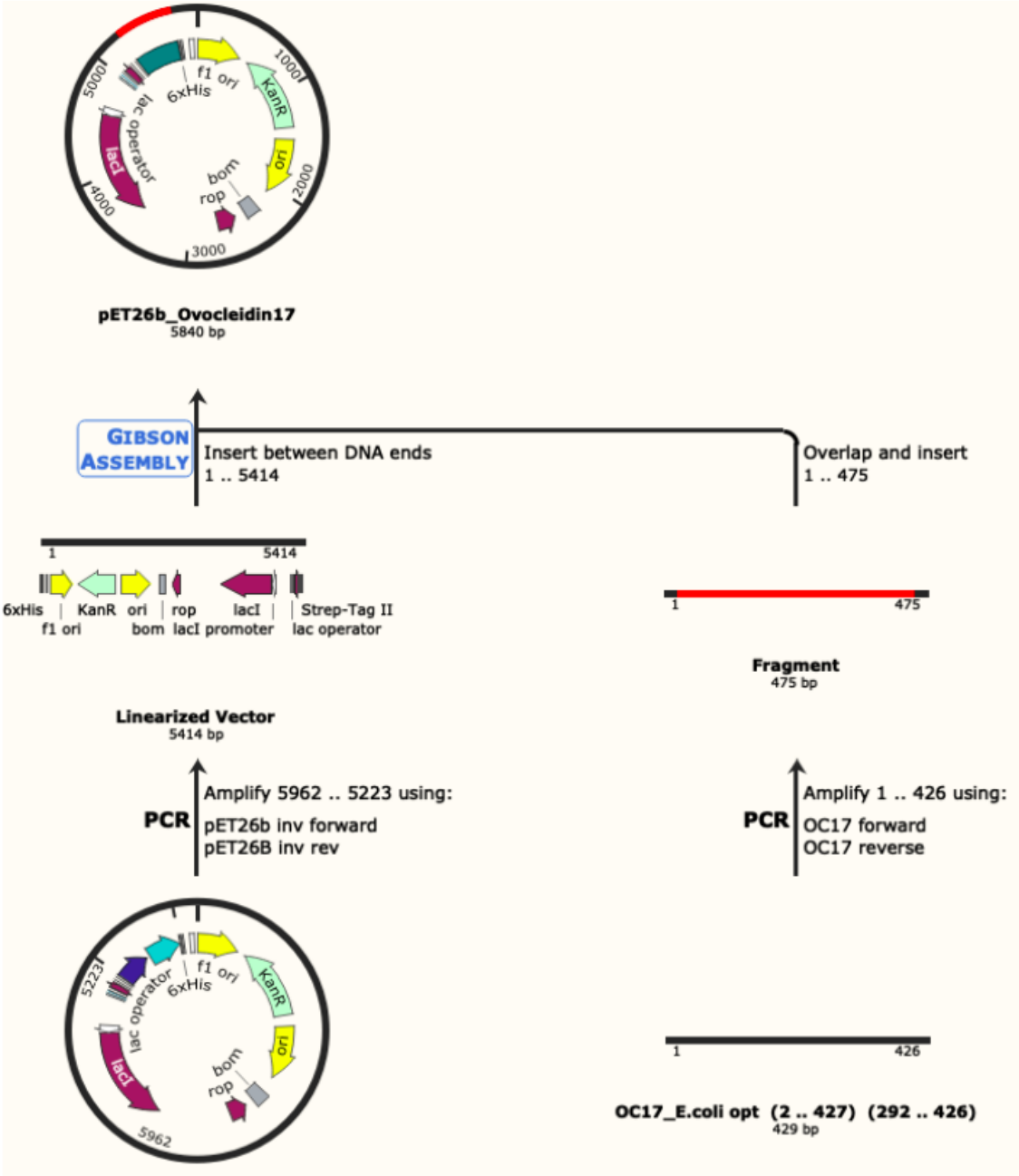


Figure 3.2: The plasmid construct for cloning of *OC-17* gene into pET26b. The SnapGene image shows PCR and Gibson assembly steps.

B-PER reagent is a cell lysis medium that solubilise proteins, the whole cell protein extract was used for the dot blot analysis of OC-17. After medium selection for optimum growth, IPTG concentration optimization was done again for LB medium in order to find whether lower concentrations of IPTG (below 0.1 mM) would result in higher expression level. The same procedure was repeated as above except that overnight culture was inoculated with 1 mL of medium and incubated at 30 ºC only until optical density at 600 nm (OD600) reaches 0.5-0.6. After IPTG addition, temperature was decreased to 25 ºC and the cultures were grown overnight. The cells were then harvested by centrifugation at 4000xg for 20 min and the medium was collected. In order to extract the periplasmic proteins only, pellets were resuspended in 30 mM Tris-HCl containing 20% sucrose, pH 8.0. The cells were kept on ice and ice-cold 500mM EDTA was added with a final concentration of 1 mM. The solution was then centrifuged at 4000xg for 20 min at 4 ºC. The supernatant was discarded and pellets were resuspended in 5 mM $MgSO_4$ and incubated on ice for 10 min. The solution containing periplasmic proteins was centrifuged at 4000xg for 20 min at 4 ºC. After the periplasmic proteins were extracted, in order to verify OC-17 synthesis, fractions were tested with SDS-PAGE and the resulting gel was analyzed with the western blot imumunodetection against anti-Histag antibody.

3.3 Results

3.3.1 Purification of the Wild-type OC-17 from Eggshell Extract

A new purification protocol for the OC-17 form the eggshell (here called the 'wild-type') was developed. Figure 3.3 shows the SDS-PAGE images of the eggshell protein extract after a 3 step purification of OC-17 by a cation exchange (pH 5), a hydrophobic interaction (pH 7.4) and another cation exchange using the chromatofocusing method. It is possible to identify three proteins in the purified eggshell extract (Well 10, Figure 3.3): lysozyme, OC-17 and the glycosylated form of OC-17 which is OC-23. The presence of OC-17 and lysozyme was further confirmed by LC-MS peptide mass fingerprinting (see Chapter 4, Section 4.3.1).

The presence of OC-23 is not problematic as it is the glycosylated form of OC-17 and both proteins exists in their natural environment in the shell gland. For this reason, it is reasonable to assume that the OC-23 might as well play a role in the amorphous to crystal transition of $CaCO_3$. For this reason, OC-23 was not further separated from OC-17. It is worth noting that the separation of OC-23 from OC-17 using Concanavalin A Sepharose, a glycoprotein purification resin, was not successful.

The presence of lysozyme has both negative and positive implications. The negative implication is that in order to probe the function of OC-17 in transforming amorphous $CaCO_3$ to

crystal, using, for instance FRET, pure OC-17 is needed (this purification can be done using the ivy protein described below). The positive implication is that the presence of lysozyme offers a way to test the ability of OC-17 to act as a 'chaperone' protein that facilitates the incorporation of other proteins. The rationale of this hypothesis will be discussed in detail in Chapter 4.

There are two key points that were shown experimentally. First, lysozyme is naturally present in the eggshell (i.e. it is incorporated *in vivo*) and, second, although lysozyme is present in the eggshell it is not incorporated *in vitro* (i.e. crystallization in the presence of lysozyme does not incorporate lysozyme into the synthetic $CaCO_3$, see details in Chapter 4). This means that there might be a 'mechanism' in the eggshell formation that allows incorporation of lysozyme. It can be hypothesized that the key factor in this mechanism might be the presence of OC-17. OC-17 might modify $CaCO_3$ crystal growth in such a way that protein incorporation becomes favourable. If OC-17 acts as a chaperone, then it is plausible that it could play a similar role *in vitro*, and facilitate the incorporation of other proteins, such as lysozyme, into $CaCO_3$.

In order to study the ability of OC-17 to entrap other proteins in $CaCO_3$ and its role on amorphous to crystal transition, the protein(s) need to be stable in water, as other buffers interfere with the $CaCO_3$ formation. The solution of OC-17 and lysozyme is stable in 0.1 M NaH_2PO_4- 0.1 M NaCl, but precipitates when the buffer is exchanged to water (and also to TRIS or to PBS). When the precipitate is filtered out, the OC-17 bands clearly show a decrease in concentration, but not lysozyme, indicating that the stability problem might be related to OC-17 (Fig 3.3. Wells 11 and 12).

3.3.2 Cloning and Purification of the Ivy Protein and Ivy-Lysozyme Interaction

The pQE30 plasmid vector was used to clone the Ivy gene. The simulation and experimental agarose gel images of the Ivy gene PCR amplification are shown in Fig 3.4. Six replications were prepared for PCR amplification to avoid experimental errors (i.e. pipetting very low amounts of solution) and to obtain high enough concentration of the target gene.

The three of the six replications (wells 6-8) were selected for further cloning experiments because of their high concentrations. The reason that the rest of the replications have low concentrations is likely because of pipetting errors as the volumes of the DNA solutions are generally very low in PCR (~5 μL).

The plasmid construct that was prepared for the ivy expression is shown in Fig 3.5. As explained in the Materials & Methods (see Section 3.2.2.2), *Ivy* gene was inserted in the C-

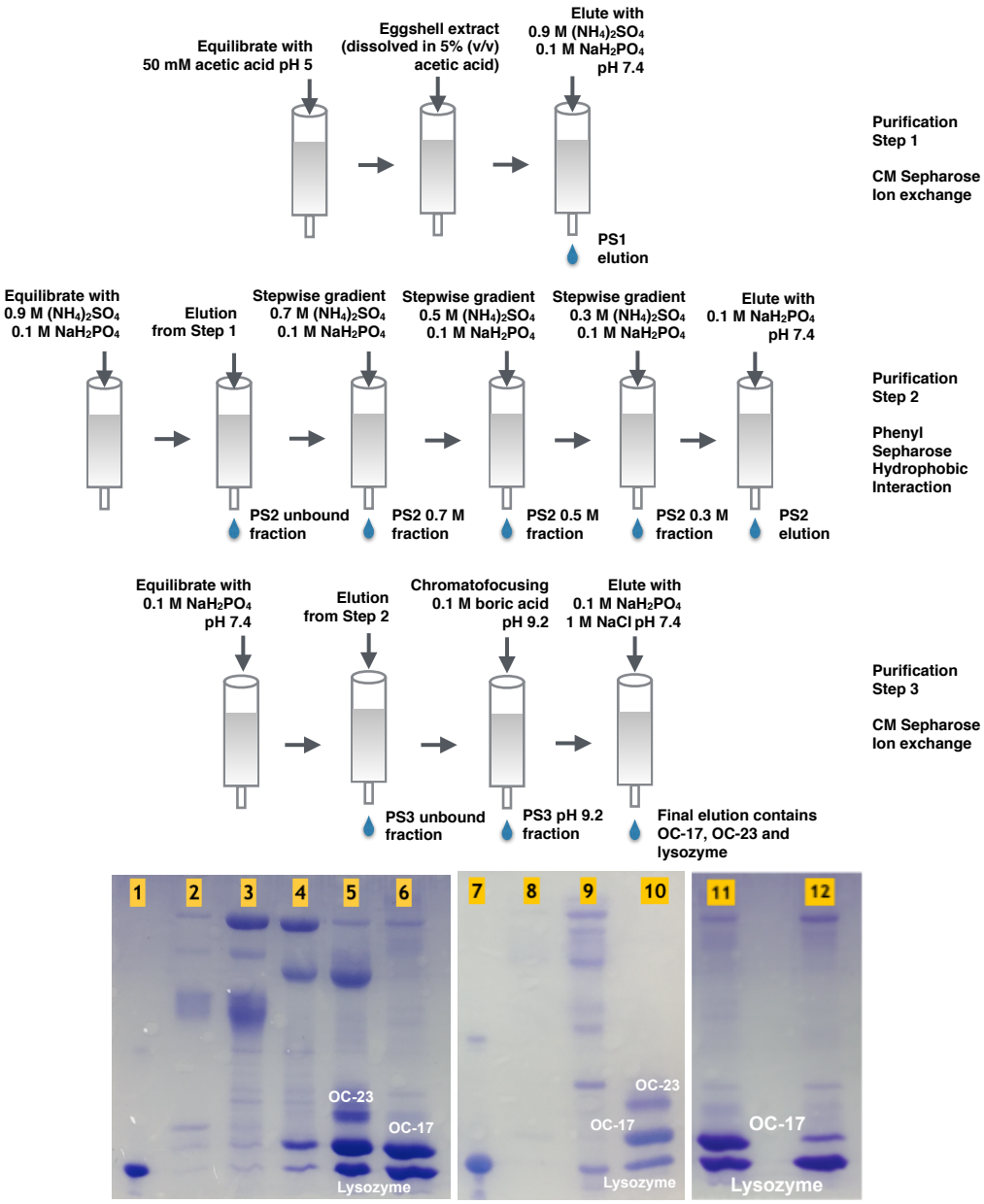


Figure 3.3: A schematic representation of OC-17 purification in 3 steps and the SDS-PAGE images of the collected fractions. PS1, PS2 and PS3 denote purification steps 1, 2 and 3. The step 1 is used to change the buffer of the protein extract from 5% (v/v) acetic acid to $(NH_4)_2SO_4$-NaH_2PO_4. The SDS-PAGE images of the step 2 samples (1-6) and the step 3 samples (7-10) are as follows: Well 1- commercial lysozyme (used as a marker); well 2- PS2 unbound fraction; well 3- PS2 0.7 M fraction; well 4- PS2 0.5 M fraction; well 5- PS2 0.3 M fraction; well 6- PS2 elution; well 7- commercial lysozyme; well 8- PS3 unbound fraction; well 9- PS3 pH 9.2 fractions; and well 10- final elution. The SDS-PAGE image at the right (samples 11 and 12) shows the effect of the buffer exchange on OC-17. Well 11- Protein solution containing lysozyme and OC17 after partial purification (the sample in well 6 was tested for buffer exchange); well 12- Sample 11 after the buffer exchange. Once the buffer is changed, precipitation occurs and filtering out the precipitate using 0.22 µm filter results in a significant loss of the OC-17 band concentration.

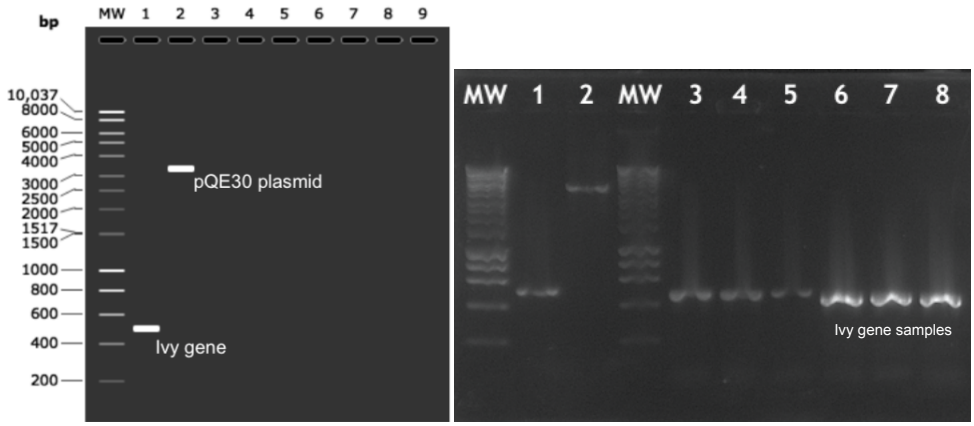


Figure 3.4: The simulation (left) and experimental (right) agarose gel images of the Ivy gene PCR amplification. MW denotes molecular weight marker. In the simulation, well 1 shows the Ivy gene and well 2 shows the pQE30 plasmid. In the experiment, well 1 shows the extracted Ivy gene from the *E. coli* K-12 MG1655 and well 2 shows the pQE30 plasmid. Well 3-8 show the replications of Ivy gene PCR amplification. (Samples 6-8 were used in the experiments because of their higher concentrations. Samples 3-5 have low concentrations probably because of pipetting errors.

terminus and not into the multiple cloning site (MCS). This protocol by-passes the N-terminus 6xHistag originally located in the N-terminus in the pQE30 plasmid. The sequencing result of the plasmid is presented in the Appendix A.

After protein expression, extraction and dialysis, the samples were loaded to a nickel affinity column for ivy purification. Although the Ivy protein was cloned with 6XHistag, another purification step was needed. An additional step of size exclusion was used to obtain pure Ivy protein. Figure 3.6 shows the SDS-PAGE images of the fractions after the nickel affinity and size exclusion purifications. Wells 5-9 show two characteristic, high concentration bands of Ivy protein after the nickel affinity purification, since the protein exists both as a monomer and a (homo)dimer and SDS-PAGE was run under non-reducing conditions (i.e. homodimer was not reduced). The monomer band appears at 15 kDa. After the size exclusion column, Ivy protein was pooled from the fractions that do not contain impurities (see Fig 3.6).

In order to confirm that the band appearing at 30 kDa corresponds to the Ivy dimer rather than a contamination, another SDS-PAGE was run under reducing conditions. In addition, the effect of the growth medium on the protein expression was tested by growing the bacteria in two different media: Tryptone-yeast (TY) and lysogeny broth (LB). The resulting SDS-PAGE image is seen in Fig 3.7. Under reducing conditions, a single and higher concentration band corresponding to the monomeric Ivy protein was observed at 15 kDa, as expected.

In order to verify the ability of the Ivy protein to bind lysozyme, purified Ivy was incubated

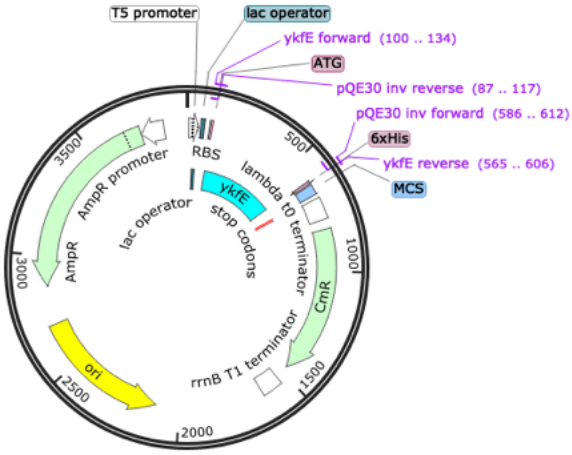


Figure 3.5: The pQE30 plasmid construct for cloning the *Ivy* gene. *Ivy* is shown as *ykfE* (its previous name) in light blue. Note that the pQE30 6XHistag is normally located at the N-terminus in the pQE30 but *Ivy* was inserted to the C-terminus to replicate the protocol in the literature (see Methods 3.2.2.2 for details).

with lysozyme at different protein ratios and loaded into the nickel affinity chromatography. The proteins were eluted with imidazole, which disrupts Histag interactions with the nickel in the column. The unbound and eluted protein fractions were analyzed in SDS-PAGE (Fig 3.8).

The ivy-lysozyme interaction can be observed from the resulting SDS-PAGE image. Because the Ivy protein has a 6XHistag for nickel affinity, once lysozyme binds to Ivy protein, the new ivy-lysozyme complex is bound through nickel affinity. The elution of the ivy-lysozyme complex was done using imidazole which disrupts Histag interactions (wells 10-12). The details of the each step is explained in Fig 3.8. Note that after the elution of the proteins, there are three proteins bands instead of one Ivy-lysozyme complex. This is because of the reducing conditions of SDS-PAGE which dissociates the Ivy protein from lysozyme.

After the confirmation of the ability of Ivy protein to bind lysozyme, it is possible to remove the lysozyme from the eggshell extract to obtain pure OC-17. In order to confirm that lysozyme could be removed from the eggshell protein extract, a partially purified protein solution from the eggshell was incubated with the Ivy protein. After incubation, the protein solution was loaded into the nickel affinity chromatography again using the same conditions as for previous lysozyme-Ivy protein experiments. It was expected that the lysozyme would be immobilized in the nickel affinity column along with the Ivy protein while OC-17 (and other proteins) would not be captured as it does not contain a Histag. The SDS-PAGE image of the nickel affinity chromatography of partially purified eggshell proteins in the presence Ivy protein is shown in Fig 3.9. As expected, the lysozyme was separated from the eggshell proteins in a single step using the Ivy protein (see details in Fig 3.9).

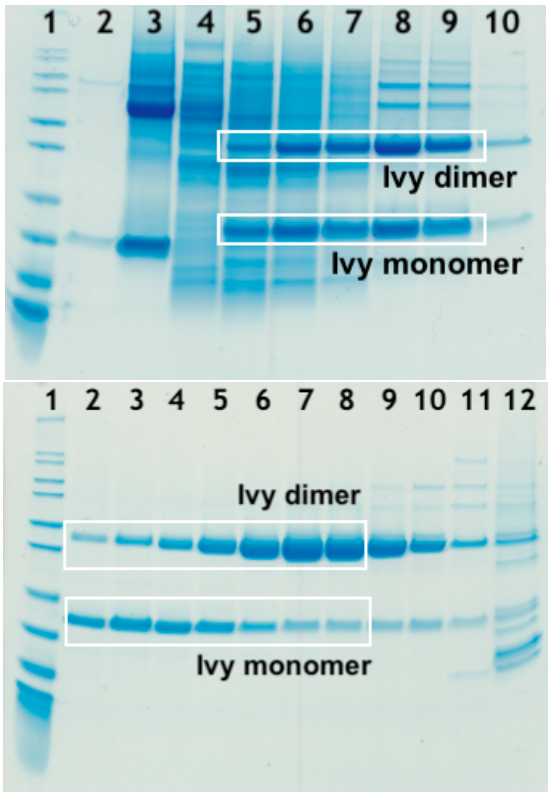


Figure 3.6: The SDS-PAGE images of the nickel affinity and size exclusion purifications of the Ivy protein. (Top) The Nickel affinity fractions. Two high concentration bands in wells 5-9 show the Ivy protein in white boxes as a monomer and a dimer. (Bottom) The size exclusion fractions of the Ivy protein again as a monomer and a dimer. The fractions that contain impurities, wells 9-12, were not pooled for the subsequent experiments.

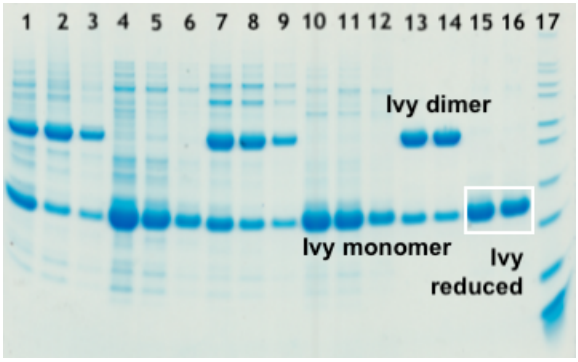


Figure 3.7: The SDS-PAGE image showing Ivy protein's monomer and dimer bands under reducing and non-reducing conditions. Samples are as follows: 1,2,3- TY medium under reducing condition; 4,5,6- TY medium under non-reducing condition; 7,8,9- LB medium under non-reducing condition; 10,11,12- LB medium under reducing condition; 13,14- purified Ivy protein under non-reducing condition; 15,16- purified Ivy under reducing condition (white box); 17- protein marker.

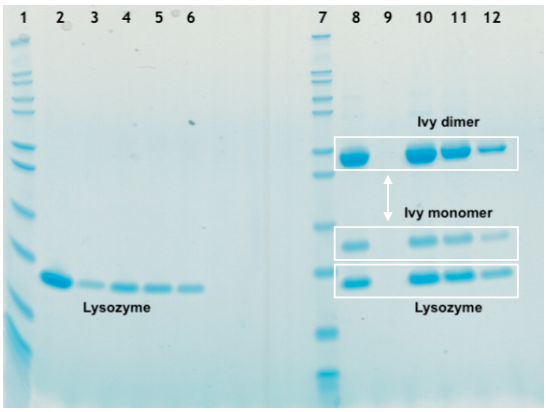


Figure 3.8: The SDS-PAGE image of the nickel affinity chromatography to test lysozyme binding ability of Ivy protein. Wells 1 and 7- protein marker; wells 2,3,4,5,6- lysozyme at different concentrations before mixing with Ivy protein; well 8- lysozyme and Ivy protein mixture before loading them into nickel affinity chromatography (Ivy protein is seen as a monomer and a dimer above lysozyme); well 9- unbound fraction after protein loading into nickel affinity column (Note the lack of the protein bands shown with the white arrow which confirms the Ivy-lysozyme interaction because, although lysozyme has no Histag, it is captured through the nickel affinity along with Ivy protein); wells 10,11,12- elution of the proteins upon imidazole gradient.

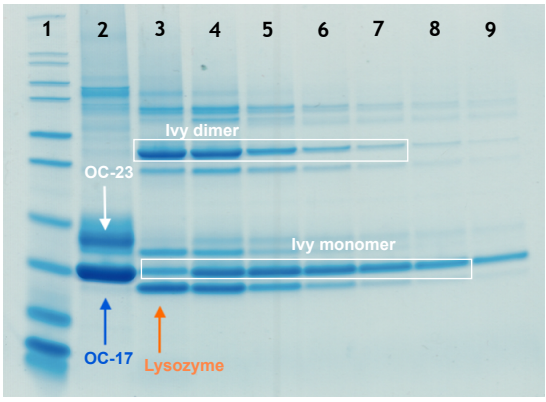


Figure 3.9: The SDS-PAGE image of the nickel affinity chromatography to remove the lysozyme from the eggshell extract using the Ivy protein. Well 1- protein marker; well 2- unbound fraction (i.e. because eggshell proteins do not contain Histag, unbound fraction contains all the proteins loaded into the column except the lysozyme. The lack of the lysozyme (in well 2) confirms that the Ivy protein can be used to separate the lysozyme from the eggshell extract hence from OC-17); wells 3,4,5,6,7,8,9; protein elution upon imidazole gradient.

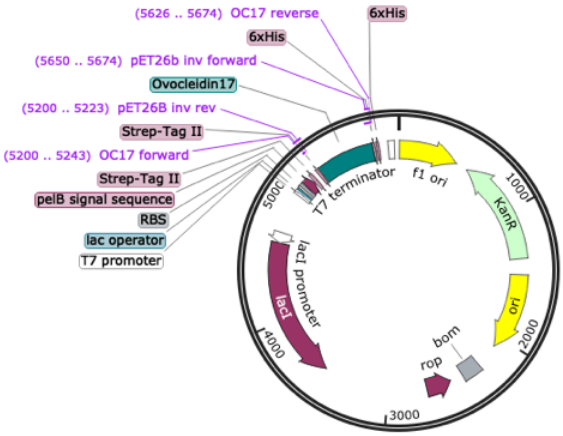


Figure 3.10: Plasmid pET26b construct for OC-17. *OC-17* gene was cloned into pET26b using Gibson assembly. 6XHistag is located at C-terminal since *pelB* signal sequence needs to be located at N-terminal (see the text for details).

3.3.3 Cloning of OC-17 and Optimization of Growth Conditions for High Expression Levels

The plasmid pET26b was constructed for the OC-17 expression as shown in Fig 3.10. The resulting plasmid contains the *OC-17* gene with two tags to facilitate purification: a Strep-Tag II (downstream to *OC-17*) and a 6XHistag (upstream to *OC-17*). The other Strep-Tag II and a 6XHistag seen in the Fig 3.10 belong to a previous construct and were not removed from the plasmid.

A *pelB* sequence was also inserted upstream of *OC-17*. The *pelB* signal sequence allows translation product to be secreted to the periplasmic space located between the outer and inner membrane of *E. coli*. The *pelB* signal sequence needs to be located at the N-terminal and following translocation, it is cleaved from the synthesized section so that it does not become a part of the polypeptide chain. Because the N-terminal region is processed in the presence of *pelB* sequence, 6xHistag should be located at C-terminal, otherwise it is also cleaved with *pelB*.

The plasmid also contains the T7 promoter followed by a *lac* operon. This is a control level added to the plasmid to avoid the risk of toxicity of the target protein to the host cell. A chemical inducer, such as IPTG, is added to induce translation; in the absence of such inducer, translation is well repressed. However, even in the absence of an inducer, there is a basal expression of T7 RNA polymerase, which might be problematic if the target protein is toxic to the host. *T7/lac* promoter provides a control level for keeping basal expression at a minimum level in some specific hosts such as one used in this work (λDE3 lysogen). The addition of glucose provides this control level because in the presence of glucose, cyclic adenosine

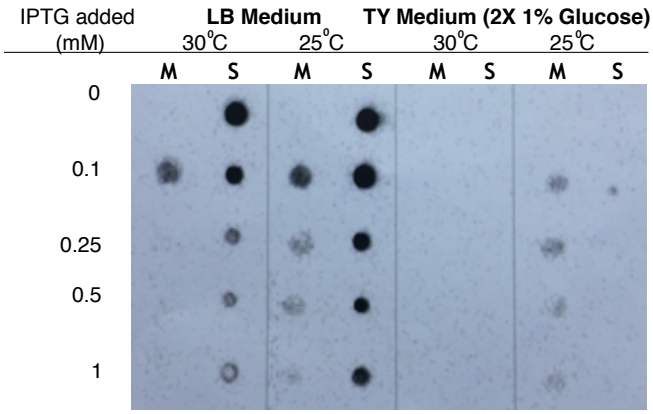


Figure 3.11: The dot blot image of anti-Histag antibody labeled proteins secreted into the growth medium (M) and cell extract soluble fraction (S) for LB and TY 1% glucose media.

monophosphate (cAMP) level is low. The cAMP is necessary for the translation of T7 RNA polymerase, the addition of glucose decreases cAMP and thus decreases the expression of T7 RNA polymerase. The impact of the different growth media and IPTG concentrations on the OC-17 expression were further assessed.

The dot blot analysis for media selection is shown in Fig 3.11. There is almost no expression in the TY medium even when supplemented with 1% glucose. In addition, although the plasmid contains *pelB* sequence to secrete the target protein into the periplasmic space, low concentrations of protein were detected in TY medium after overnight incubation at 25 ºC. In the case of the LB medium, at 30 ºC, no protein was observed in the medium at IPTG concentrations higher than 0.1 mM. Although the expression increases after overnight incubation at 25 ºC, concentration of the protein in the medium is much lower as expected. Therefore, another set of IPTG concentrations below 0.1 mM were tested: 0, 0.01, 0.025, 0.05, 0.1, 0.25 and 0.5 mM. The extraction of proteins was done as for previous set of experiments as explained above (see Methods, Section 3.2.2.3) and the samples were analyzed by SDS-PAGE. The resulting gel was transferred to a blotting membrane to run a western blot immunodetection against anti-Histag antibody. The western blot result of OC-17 fractions tested for the optimum IPTG concentration is shown in Fig 3.12.

The optimal IPTG concentration was found to be 0.05 mM, since OC-17 was present at a high concentration both in the periplasm and the medium. Although the plasmid expressing OC-17 contains pelB sequence, which allows the protein secreted to periplasmic space, because of the small size of OC-17 (15.3 kDa), it is also secreted to the medium similar to other small sized proteins. The same protocol was used for the large scale expression (in 2L) and both medium and periplasmic proteins were collected. The medium was directly concentrated using the VivaFlow peristaltic pump-driven concentrator. The periplasmic space proteins were

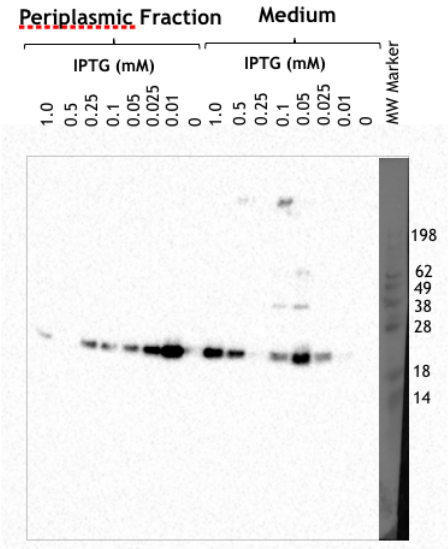


Figure 3.12: The western blot image of the OC-17 fractions tested for optimal IPTG concentrations both in the medium and periplasmic space.

extracted as explained above (see Methods, Section 3.2.2.3) and loaded into an immobilized nickel affinity column. The elution was done with increasing imidazole concentration as usual for elution procedure. However, no protein was detected either in the medium or the periplasmic space. The SDS-PAGE image of nickel affinity chromatography fractions is shown in Fig 3.13.

3.4 Discussion

The purification methods of OC-17 described in the literature are laborious (e.g [111]) and involve several chromatographic steps (e.g [230],[133]) leading to low concentrations of the purified protein. Here, an alternative method was developed to purify OC-17 from the eggshell using a cation exchange, a hydrophobic interaction and the Ivy-lysozyme interaction. In addition, in the present work, it was shown that the OC-17 can be expressed in bacteria at high concentrations and, to our knowledge, it is the first time that the cloning and expression of this protein has been reported.

Although the OC-17 was not purified further to remove the OC-23, high performance size-exclusion resins are likely to remove it because of the molecular weight difference (8 kDa). Another option is to deglycosylate OC-23 using a deglycosylation enzyme, such as the N-glycosidase F. It has been shown that this method is able to remove the carbohydrate group from OC-23 [136]. It is also possible to remove the carbohydrate group using a chemical deglycosylation method [231]. While there are options to remove OC-23 from OC-17, both

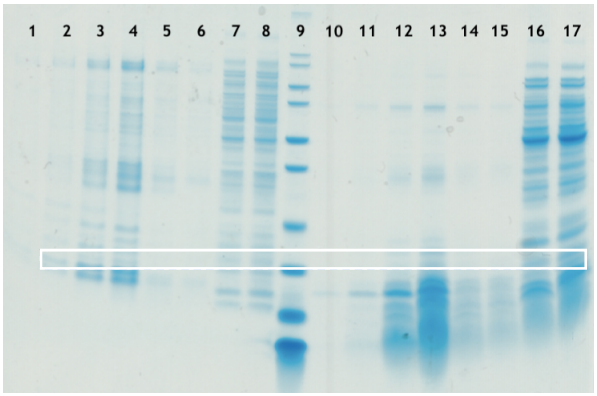


Figure 3.13: The SDS-PAGE image of the OC-17 large scale expression. The image shows protein fractions from two purifications: First, proteins in the medium (wells 1-8) and second, proteins in the periplasmic space (wells 10-17). Samples are as follows: 1- Proteins in the medium before loading the whole medium into the nickel affinity column; 2- The unbound proteins eluted from the nickel affinity column; 3,4,5,6,7,8- The binding proteins eluted by an imidazole gradient; 9- protein marker; 10-Periplasmic space proteins before loading into nickel affinity column; 11-The unbound periplasmic proteins eluted from the nickel affinity column; 12,13,14,15,16,17- The binding periplasmic proteins eluted by an imidazole gradient.

proteins are naturally found in the shell gland which suggests that the glycosylation of the protein might be important for its function.

The presence of lysozyme in the eggshell [232], a protein which have similar molecular weight and isoelectric point to OC-17, makes the purification of OC-17 from the eggshell challenging. As shown here, an efficient, one-step separation of lysozyme from OC-17 is to use the Ivy protein. The cloning and purification of Ivy are straightforward because *Ivy* gene could be cloned from a range of *E. coli* strains and the construction of plasmid with 6xHistag allows purifying the Ivy using a nickel affinity column [228]. The separation of lysozyme from the eggshell extract is known to be non-trivial as shown here and as reported in the literature (e.g. see [136]). The strong interaction of Ivy protein with lysozyme [228, 229] offers a single-step separation of lysozyme to purify OC-17 from the eggshell extract.

One of the problems encountered during the present work is the instability of OC-17 either before or after purification. More specifically, it was observed that upon the dissolution of eggshells in acetic acid (5% v/v) the protein solution tends to precipitate irreversibly which is also the case reported in the literature (e.g [133]). After a partial purification, even if the protein solution contains three proteins only (OC-17, OC-23 and lysozyme), precipitation occurs if the buffer of the solution is exchanged to H_2O, TRIS or PBS (see Results Section

3.3.1). The precipitate was shown to be OC-17 confirming the instability of the protein. In order to probe the function of OC-17 and to test the ability of it to act as a chaperone protein, the OC-17 should be stable, preferably, in H_2O for the incorporation experiments as $CaCO_3$ crystallization is affected by the buffer used. However, many proteins are expected to be unstable in H_2O because of the lack of ions stabilizing the protein structure. A potential strategy to circumvent OC-17 stability problem is to change the buffer of the protein into NaCl at relatively high concentrations (e.g. 0.25 M). This strategy might be beneficial because it was observed that the precipitate partly recovers after adding NaCl, although precipitation is triggered again upon further concentration by ultrafiltration. In addition, the extraction of OC-17 at high enough concentrations requires the dissolution of many shells, since the concentration of the protein in each shell is relatively low. This, in turn, complicates the extraction process by increasing the volume of the extract which should be concentrated by ultrafiltration. A possible way to solve this problem is to use peristaltic pump-driven Viva Flow Cross Flow Cassette[1] to concentrate large volumes of the eggshell extract in a short period of time.

The recombinant protein production would be an attractive alternative, considering the problems of extracting and purifying OC-17 from the eggshell. The recombinant OC-17 production would allow expression of higher amounts of protein needed for probing the function of the protein. One of the methods that could be used to analyze the function of OC-17 is FRET [233, 234]. Fluorescent labelling of OC-17 could offer a way to monitor $CaCO_3$- OC-17 interactions as a function of fluorescence lifetime. When coupled with the time-correlated single photon counting (TCSPC), FRET-FLIM experiments might provide quantitative data on the interaction of OC-17 with growing $CaCO_3$ (see Chapter 5, Section 5.2.1).

Cloning allows introducing point mutations to a protein. The coupling simulation results [33, 143] with precise single amino acid changes could lead FRET experiments which have the potential to follow loss-of-function in the protein thus might confirm OC-17-$CaCO_3$ interaction through specific amino acid substitutions. It would also be relevant to run molecular dynamics simulations of OC-17 with one point mutation at a time so that the effect of substitution of an amino acid from the active site could be analyzed *in silico*.

Cloning also makes it possible to compare the 'wild-type' and recombinant OC-17 because a protein could be denatured during incorporation into $CaCO_3$ and it is not unreasonable that OC-17 could be denatured during incorporation into the eggshell. However, the X-ray structure of OC-17 has been reported [31] using the eggshell extracted protein, which suggests that OC-17 is not denatured upon incorporation in the eggshell. It might be the case that

[1] https://www.sartorius.com/shop/ww/en/USD///vivaflow-200-laboratory-cross-flow-cassette/c/M_Vivaflow_200

OC-17 is able to refold upon extracted from the eggshell (e.g. during the subsequent steps of purifications).

As mentioned above, one potential useful application of OC-17 could be to use it as a chaperone protein to facilitate the incorporation of other proteins or even protein-based vaccines into $CaCO_3$ for long term protein preservation. One way to investigate this hypothesis is to genetically fuse a fluorescent marker protein, such as GFP to OC-17 and test the incorporation behaviour of GFP-OC-17 into $CaCO_3$.

Chapter 4

Analysis of Secondary and Tertiary Structures of the Intra-crystalline Proteins

4.1 Introduction

This chapter presents the results of the structural analysis of the intra-crystalline proteins incorporated into $CaCO_3$ *in vitro.* The overall aim of the structural analysis is to understand the effect of incorporation on the secondary and tertiary structures of the target incorporated proteins after dissolution of the host $CaCO_3$. The 3D structure of a protein is of utmost importance for its function. Because the overarching goal of the present work is to preserve vaccine proteins in $CaCO_3$ lattice to avoid heat degradation through immobilization, stability of the 3D structure of the incorporated protein is crucial. In addition, the effect of dissolution of the host $CaCO_3$ is also important as the release of the protein to the solution is achieved by a dissolution buffer which itself could affect the protein structure. Because there are two ways to dissolve $CaCO_3$, acid dissolution and metal chelation, both ways were tested compared to the positive control protein samples. After optimization of the dissolution step, model protein incorporation into $CaCO_3$ *in vitro* was conducted using three proteins with different properties, such as molecular size and electric charge. The three model proteins tested are BSA, lysozyme and diphtheria anti-toxin. The structural analysis was conducted by two spectroscopic methods: circular dichroism and intrinsic tryptophan fluorescence spectroscopy. These two methods provide complementary structural information. Circular dichroism allows monitoring the secondary structure of the protein while intrinsic tryptophan fluorescence provides information on the tertiary structure.

The persistence of the intra-crystalline proteins in the avian eggshell implies that it could also be possible to preserve a given target protein in synthetic $CaCO_3$ but the effect of high (and if needed low) temperature on the incorporated protein needs analysis. For this reason, eggshells were heat-treated to emulate the high ambient temperature that could be observed in Africa where vaccine thermostability is most needed. After the heat-treatment of the eggshells, OC-17 was purified at small scale for preliminary experiments. The structure of the OC-17 from the heat-treated eggshells was then compared to OC-17 purified from the eggshells stored at room temperature. As there are a myriad of proteins in the eggshell, verification of OC-17 is needed which was done by LC-MS using the protein band excised from the SDS-PAGE of protein purification. The stability of the protein structure in the eggshells might be unique to the eggshell incorporation or to the few specific proteins. For instance, OC-17 might preserve its structure in the eggshell but not in the synthetic $CaCO_3$. Similarly, only OC-17 but not other eggshell proteins might be intact in $CaCO_3$. For this reason, model proteins need to be incorporated and their structures should be analyzed compared to the positive controls. Lastly, the presence of lysozyme in the eggshell suggests that some of the proteins in the eggshell are incorporated because of the 'chaperone' effect of the other proteins. More specifically, because lysozyme is an anti-microbial protein, it is not expected to function in the $CaCO_3$ crystals. In addition, as will be shown in the Results section, lysozyme is not able to be incorporated alone. This means that lysozyme is incorporated by other proteins that act as 'chaperones'. This effect was tested using BSA as a potential chaperone as it is incorporated efficiently. The quantification of the protein incorporation for model proteins is also presented.

4.2 Materials & Methods

4.2.1 Materials

The diphtheria antibody was kindly provided by the National Institute for Biological Standards and Control (Potters Bar, UK), a vaccine pre-qualification laboratory of the World Health Organization. All chemicals and proteins (chicken egg white lysozyme [CAS 12650-88-3, lyophilized powder] and bovine serum albumin [CAS Number 9048-46-8, lyophilized powder]) were purchased from Sigma-Aldrich. All solutions were prepared with deionized water and pH adjustment was done with concentrated HCl or NaOH if needed. The solutions were filtered either through vacuum filtration with 0.45 or 0.2 μm filter papers or using syringe-driven filters. Buffer exchange of the protein solutions was done using Amicon Ultra centrifugal filters (Merck, Darmstadt, Germany) using Sorvall ST 8 bench-top centrifuge (Thermo Scientific, Massachusetts, USA). Intrinsic tryptophan fluorescence and CD measurements were

done in quartz cuvettes from Agilent (Santa Clara, USA) and Starna Scientific (Essex, UK), respectively. The concentration of the proteins was measured with a microplate reader from BMG Labtech (Ortenberg, Germany) at 280 nm. The molecular extinction coefficients were calculated theoretically using amino acid sequences of the proteins with ExPASy ProtParam or Innovagen PepCalc softwares. The heat treatment of $CaCO_3$ and the eggshell samples was conducted in a Heratherm table-top laboratory oven from Thermo Scientific (Massachusetts, USA).

4.2.2 Methods

4.2.2.1 Partial Purification of OC-17 for Preliminary Structural Analysis upon Heat Treatment

As explained in Chapter 3, Section 3.2.1, a new protocol was developed for the purification of OC-17. However, before the development the protocol, preliminary tests of the effect heat treatment on the intra-crystalline proteins were conducted using partially purified OC-17. Because the results of the structural studies upon heat treatment are presented in this chapter, the partial purification method will be explained, although it was not continued further. The partial purification was done after dissolving the eggshells in 5%(v/v) acetic acid solution and loading the whole protein extract in a CM Sepharose ion exchange column in 50 mM acetic acid pH 5. Then, a chromatofocusing protocol was applied to eliminate the contaminant proteins starting from pH 5 through to pH 11 using a stepwise gradient at pH 6 citrate buffer, pH 7.5 PBS, pH 9 TRIS and pH 11 boric acid. Although OC-17 was obtained as the primary protein after partial purification (in terms of its concentration compared to low concentration contaminants), the protocol was found to be neither repeatable nor efficient (i.e. yielding low amounts of protein) thus this protocol was not further used and another protocol was developed (Chapter 3, Section 3.2.1).

4.2.2.2 Circular Dichroism Spectroscopy

Dichroism is an optical property of some materials where a linearly polarized light in a particular plane is absorbed more than the light absorbed in the plane polarized perpendicularly to the first. Circular dichroism (CD) is a special case of dichroism in which the light is not linearly but circularly polarized. In fact, a linearly polarized light can be shown to be a combination of right circularly polarized light and left circularly polarized. Fig 4.1 shows how a combination of right and left circularly polarized light gives rise to a linearly polarized one. Fig 4.1 (A) and (B) show two linearly polarized light beams perpendicular to each other, and their electric fields are also shown at the right for clarity. Note that light is an electromagnetic

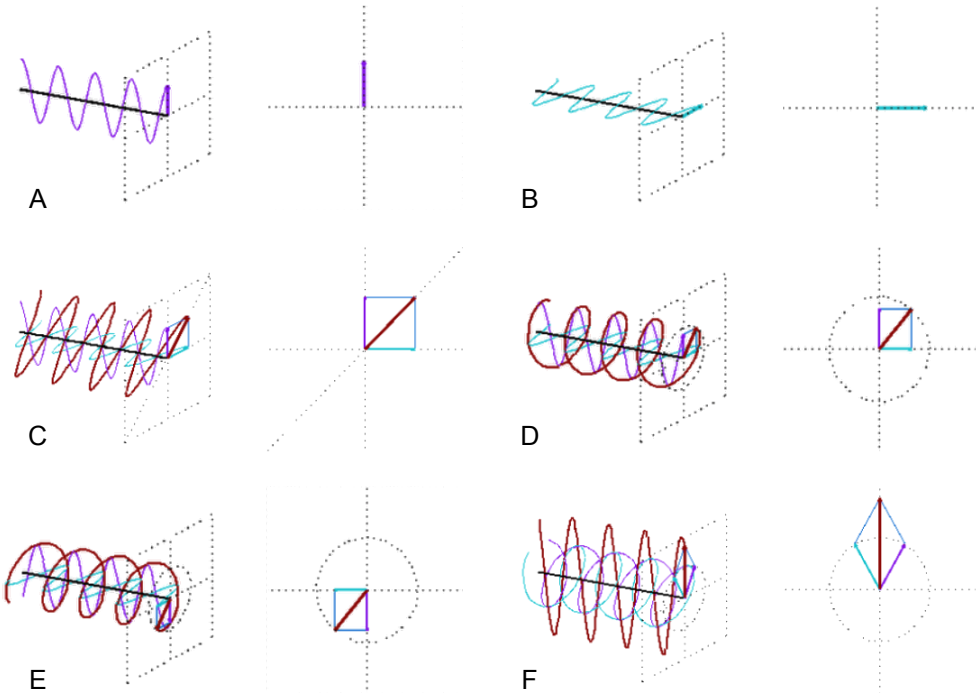


Figure 4.1: A schematic representation of how two linearly polarized lights are added to produce a circularly polarized one. It is also shown that the addition of two circularly polarized lights rotating in opposite directions produces a linearly polarized one. A more detailed explanation is given in the text. Image credit, https://cddemo.szialab.org

wave which consists of an oscillating electric field and an oscillating magnetic field perpendicular to it. In the figures below, only the electric fields are shown which have the same wavelength and amplitude. If these beams are in-phase as seen in Fig 4.1 (C), then they are added together to form a new linearly polarized light. If same beams are out-of-phase with a difference of 90 °C, then the resulting beam is circularly polarized light as seen in Fig 4.1 (D). When the phase difference between two linearly polarized beam is -90 °C, then the rotation of the circularly polarized light changes as seen in Fig 4.1 (E). Importantly, if a right and a left circularly polarized light have the same wavelength and amplitude, then the addition of these produces a linearly polarized light. This means that a linearly polarized light is in fact a combination of two circularly polarized lights rotating in opposite directions.

Circular dichroism property of materials can be used to probe the structure of a given sample. Using CD for the structural analysis of a sample is a type of spectroscopy because the analysis is solely based on the interaction of light with matter. As is the case in all spectroscopic methods, when a sample interacts with light, it can absorb the light which can be detected as a decrease in the intensity. In other words, the amplitude of the electromagnetic wave decreases. In addition to absorption, the velocity of the light can also change once it tra-

verses a sample. This change in velocity depends on the optical properties of the sample. The change in the velocity of light is related to the refractive index of the sample. The refraction index is defined as the velocity of light in vacuum divided by the velocity of light in the sample. Fig 4.2 shows a linearly (A) and a circularly (B) polarized light absorbed by a medium. The intensity decreases as a result of absorption. The refractive index of the medium is assumed to be 1 which means that the light is not refracted. In contrast, Fig 4.2 shows a linearly (C) and a circularly (D) polarized light not absorbed by the medium but refracted while the beams traverse the sample. The refractive index of the medium is greater than 1 which decreases the wavelength of the light while it is in the medium. Because there is no absorption, the intensity does not change.

CD spectroscopy is based on the CD effect observed in some materials. In this case, right circularly polarized light is absorbed to a different extent than left circularly polarized or vice versa. Fig 4.2 (E) depicts the CD effect for a medium that absorbs right circularly polarized light only. Because a linearly polarized light is a combination of right and left circularly polarized lights, the resulting beam after absorption no longer oscillates linearly (as shown at the left in F) or circularly but it follows an ellipsoidal path (at the right in F).

The absorption of circularly polarized light above assumes that the medium (i.e. sample) only absorbs light and not refracts it. As explained previously, even the materials that do not absorb light have a refractive index. Some materials have an optical property of refracting light depending on the polarization of light. This is called birefringence (i.e. double refraction). In case of the circularly polarized light, circular birefringence allows a material to refract right and left circularly polarized lights differently. Fig 4.3 (A) and (B) show the effect of circular birefringence in a medium that refracts right circularly polarized light. The result is a linearly polarized light but its oscillation is no longer vertical, in other words, its polarization angle is changed. Materials do not only possess dichroism or birefringence but they possess both. Fig 4.3 (C) and (D) depict the basis of circular dichroism spectroscopy. CD spectroscopy provides structural information about a given sample if the sample both absorbs and refracts right and left circularly polarized light in a different extent. The resulting light is not only rotates in an elliptical path but also changes its polarization angle as in seen in 4.3 (D, right).

It is possible to extract structural information of biological molecules using CD spectroscopy because circular dichroism and circular birefringence arise from the asymmetric secondary structures of proteins, such as α-helices or β-sheets. The far UV range in CD spectroscopy provides information on the secondary structures. The signals of various secondary structures are shown in Fig 4.4 (left). The near UV range could be used to collect data on the tertiary structure as shown in Fig 4.4 (right). In the present work, intrinsic tryptophan fluorescence spectroscopy was used to probe the tertiary structures of the intra-crystalline proteins. It

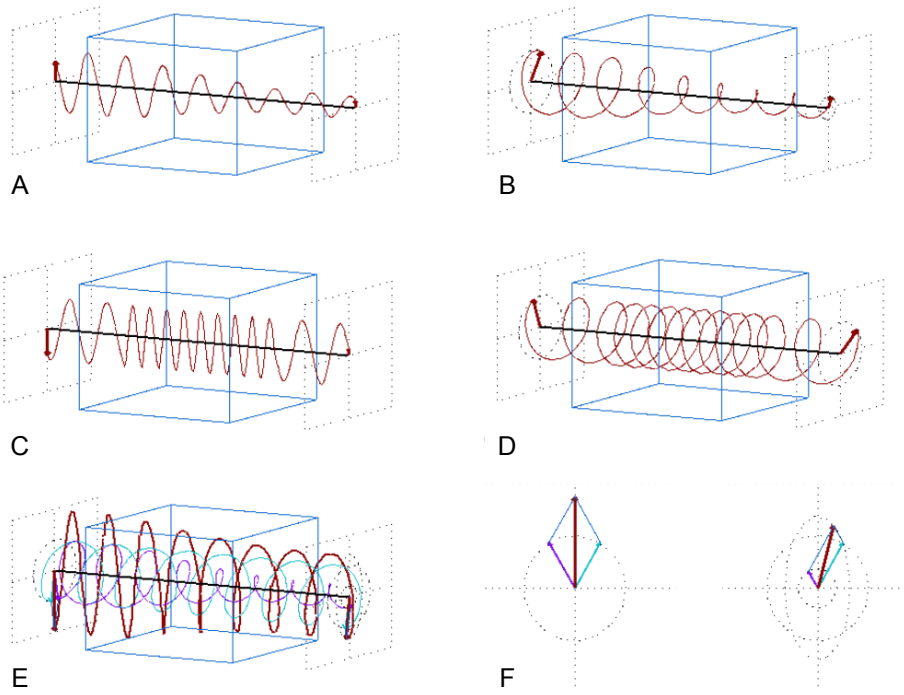


Figure 4.2: A schematic representation of absorption (A and B) and refraction (C and D) of light during interaction with matter. Circular dichroism occurs when right circularly polarized light is absorbed differently than left one (E) or vice versa. (F) shows the beam before (left) and after (right) it traverses the sample. The light after traversing the sample is not linearly polarized but follows an ellipsoidal path. Image credit, https://cddemo.szialab.org

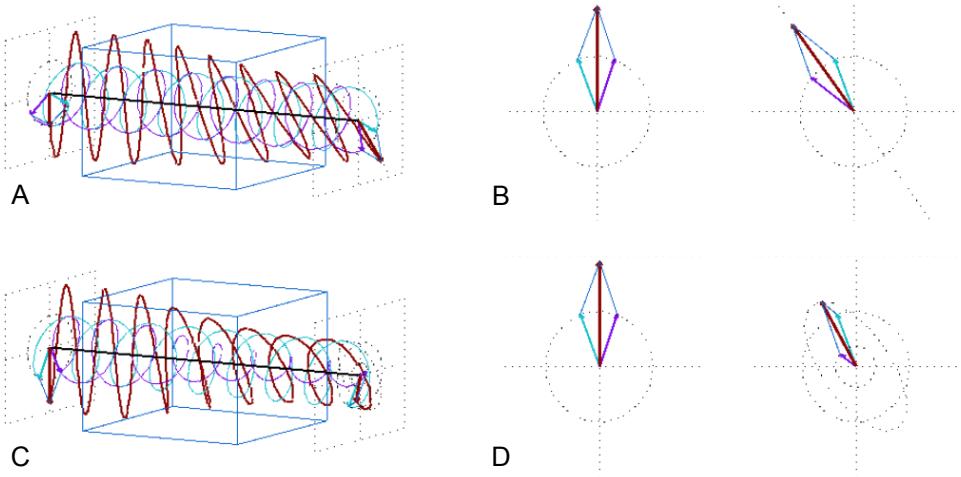


Figure 4.3: A schematic representation of circular birefringence (A and B) and the combination of dichroism and birefringence as the basis of the CD spectroscopy. Image credit, https://cddemo.szialab.org

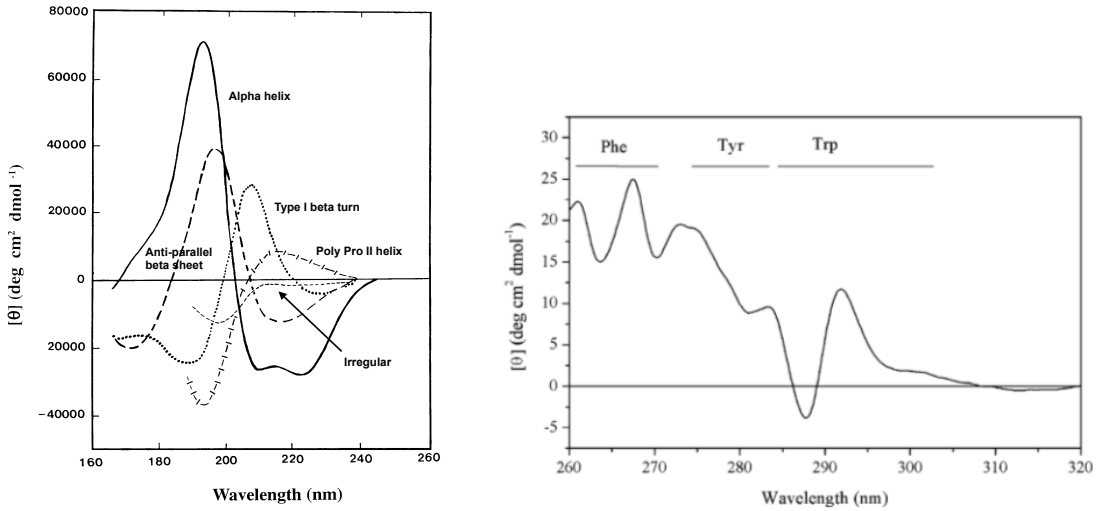


Figure 4.4: (Left) The far UV CD spectra of protein secondary structures. (Right) The near UV CD spectrum showing phenylalanine, tyrosine and tryptophan signals [235].

should also be noted that the CD spectrum of an active and inactive form of an enzyme could have the same CD spectrum which means that some proteins (e.g. genetically modified ones) can have the same secondary or tertiary structure but can be non-functional [235].

The CD measurements were carried out on a Chirascan CD Spectrometer (Applied Photophysics, Surrey, UK). A 400 µL, 1 mm path length quartz cuvette was used for the measurements. The samples were scanned from 260 to 190 nm with a bandwidth of 1 nm. The background measurements were collected using pure H_2O. Experiments were done at room temperature. The data smoothing was done to collected spectra using the Chirascan data software of Chirascan CD Spectrometer. Subsequent CD data analysis to interpret the spectra in terms of protein folding was done using the Capito [236], a web server-based analysis and plotting tool for CD data.

4.2.2.3 Intrinsic Tryptophan Fluorescence Spectroscopy

Fluorescence is the emission of a photon by an electron in a molecule that is excited. More specifically, fluorescence occurs when an electron absorbs enough energy to move into an excited state from where it is allowed to return to the ground state orbital. Emission of a photon is a result of returning of the electron to the ground state where it is paired to the ground state electron with an opposite spin [237]. If the spin orientation of the excited electron is the same with the second electron in the ground state, then the transition from the excited to the ground state is forbidden which results in phosphorescence with much slower emissions (milliseconds to seconds compared to nanoseconds in fluorescence) [237].

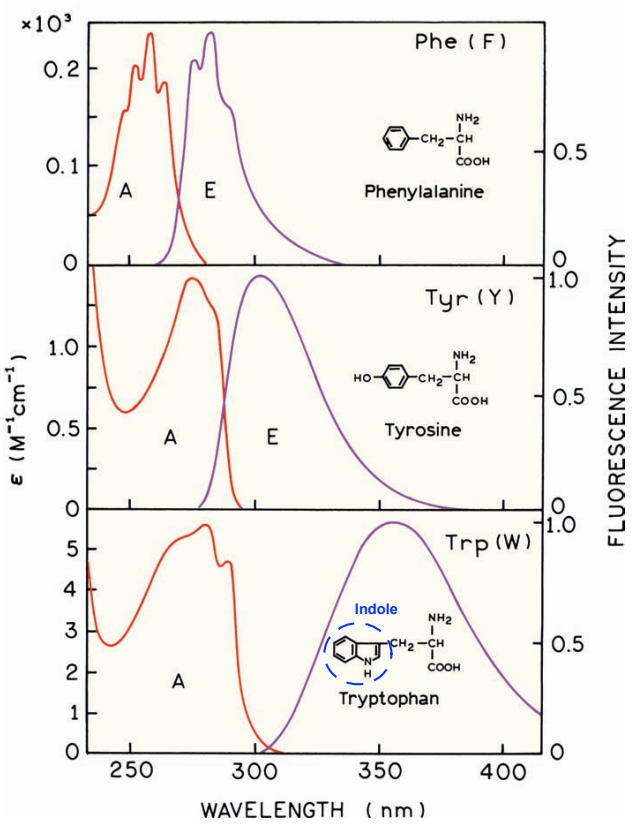


Figure 4.5: The absorption and emission spectra of phenylalanine, tyrosine and tryptophan and their chemical structures [237]. The indole group of the tryptophan is shown in blue dotted circle.

Three of the amino acids in proteins, tryptophan, tyrosine and phenylalanine possess fluorescence because of their aromatic groups that absorb energy in the UV range of the electromagnetic spectrum. Among these amino acids, tryptophan has an indole group which allows it to absorb UV light around 295 nm. The emission of phenylalanine is negligible and could only be used in case of the lack of both tryptophan and tyrosine in a protein which is very rare. Tyrosine emission is known to quench as a result of its interactions with peptide chains or transferring the energy to the nearby tryptophan(s). The absorption and emission spectra of phenylalanine, tyrosine and tryptophan and their chemical structures are shown in Fig 4.5. Intrinsic tryptophan fluorescence is used to probe the structural change in a protein because the emission of tryptophan is highly sensitive to its local environment. The change in intrinsic tryptophan fluorescence can arise from various processes, such as ligand binding to the protein, protein-protein interactions and protein unfolding. The fluorescence of tryptophan arises from its indole group which is shown in Fig 4.5.

The excitation wavelength of tryptophan is 295 nm and the reason for using this wave-

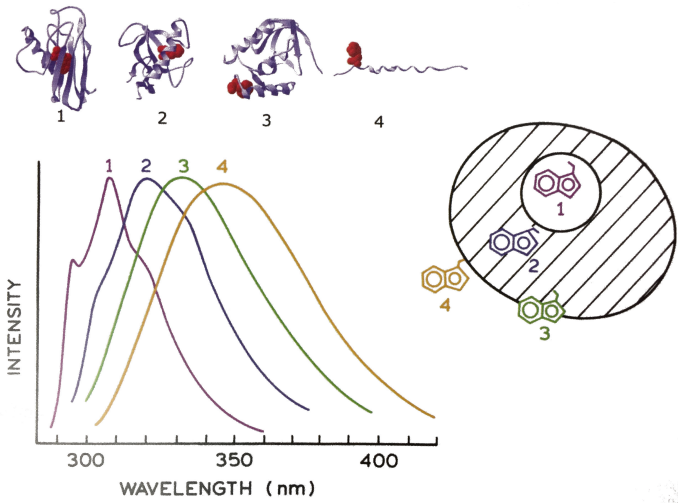


Figure 4.6: Redshift in the emission of tryptophan in different proteins with different surface exposure of the amino acid [237]. (1) apoazurin Pfl, (2) ribonuclease T, (3) Staphylococcal nuclease, (4) glucagon.

length is to avoid excitation of tyrosine. The emission of tryptophan occurs at 350 nm in H_2O and the fluorescence lifetime changes from 1 to 6 ns [237]. Depending on the position of a tryptophan residue in a protein, the emission maximum wavelength can change. For instance, buried tryptophan residues show blueshift compared to the surface exposed ones. Fig 4.6 shows an example of how tryptophan location is reflected in the emission wavelength. In apoazurin Pfl, the tryptophan residue is buried in the protein whereas in other proteins tryptophan exposure to the environment increases. The result of the surface exposure is a redshift in the emission spectra. This change of tryptophan emission can be used to probe the structural deviations of the tertiary structure of proteins. For instance, if a protein undergoes partial or complete unfolding, then the emission spectrum of native protein would be different compared to the final protein.

In addition, it is also possible to analyze if a given protein has surface exposed tryptophan(s) using fluorescence spectroscopy. Because the indole group is responsible for the fluorescence in tryptophan, a buffer solution containing a quencher of indole would result a redshift in the spectrum compared to the spectrum taken in a buffer without the quencher. In this case, the properties of the buffer such as pH, and the effect of buffer on the structure of the protein tertiary structure should be taken into account. If the quenching buffer does not impose structural changes on the protein, then it can be used to probe the surface exposed tryptophan(s). The examples of indole quenchers are iodide, acrylamide and disulfide bonds [237]. Disulfide bond triggered quenching can be observed when a histidine is relocated close to a tryptophan residue upon structural reorientation.

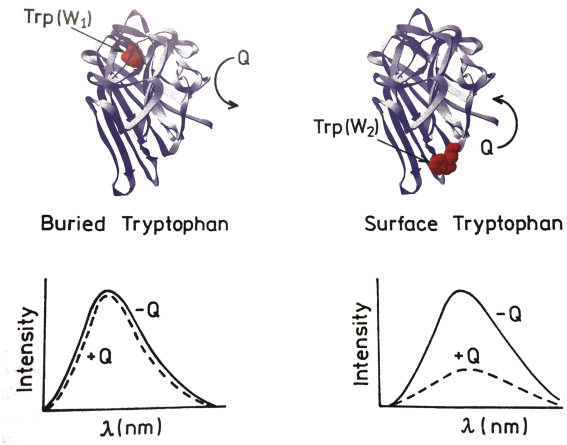


Figure 4.7: The difference between the spectra of buried a tryptophan and a surface exposed one. A surface exposed tryptophan can be probed in a quenching buffer (Q) in which the change in the emission is attributed to the quenching of tryptophan on the surface compared to no quenching in the buried one as it is not accessible to the buffer solution [237].

The high local sensitivity of tryptophan against environmental or structural changes is used in the intrinsic tryptophan fluorescence measurements where probing the tertiary structure compared to the native state of the protein reveal whether or not a change occurred in the overall 3D structure. Determining the effect of intra-crystalline incorporation including the effect of reconstitution of a protein is of utmost importance in the present work. In this respect, intrinsic tryptophan fluorescence is particularly useful because the comparison of a model protein before and after intra-crystalline incorporation is needed to determine possible changes in the 3D structure of the incorporated protein. Because the only difference between the native state of the model protein and the incorporated one is the incorporation process, a difference in the emission spectra can be attributed to the structural changes during crystallization.

In the present work, the model protein BSA was analyzed both for the effect of the reconstitution buffer (EDTA) and for the effect of the intra-crystalline incorporation. There are two tryptophans in the BSA and five in the diphtheria anti-toxin. It should be noted that multiple tryptophans give rise to more complex fluorescence responses compared to a single tryptophan in a protein. In the present work, multiple tryptophans do not complicate the analysis because the key part of the analysis is the comparison of the emission spectra of the same protein before and after the incorporation without a change in the experimental conditions.

Although OC-17 was not analyzed for its tryptophan fluorescence, it has eight tryptophan residues which would allow intrinsic fluorescence measurements in different ways for various purposes. For instance, the interaction of OC-17 with $CaCO_3$ might be studied using recombinant OC-17 in which the tryptophan located close to the reaction centre is changed to

a different amino acid via site-directed mutagenesis and analyze the change in the emission spectrum. Changing a tryptophan to a different amino acid or adding one in case there is no tryptophan in a given protein is an alternative method that was used to study proteins for various purposes [237]. As will be explained in Chapter 5, fluorescence resonance energy transfer may also be used to probe OC-17 during $CaCO_3$ crystallization.

Intrinsic tryptophan fluorescence measurements were performed with Cary UV/VIS Spectrophotometer (Agilent, Santa Clara, USA). All samples were analyzed under the same conditions (in H_2O at room temperature). A 3 mL, 10 mm path length quartz cuvette was used. The excitation wavelength was set to 295 nm and the emission was recorded from 300 nm to 450 nm. The photomultiplier tube (PMT) detector voltage was set to 700 V and the step size was set to 5 nm. Because the intra-crystalline protein samples are reconstituted using EDTA solution, buffer exchange was conducted prior to the fluorescence and CD measurements. Buffer exchange was done using centrifugal ultrafiltration tubes in a bench-top centrifuge at 4500 rpm. The samples were centrifuged to 500 µL, diluted to 15 mL in H_2O and concentrated again. This dilution/concentration cycle was repeated three times to ensure complete exchange of the buffer.

4.2.2.4 LC-MS Analysis of OC-17 and Lysozyme from the Eggshell for Protein Identification

LC-MS was used to identify the proteins upon extraction from the eggshell. An SDS-PAGE was run to obtain separate protein bands for the subsequent LC-MS analysis. The gel was run under denaturing conditions in a 12% resolving gel in Bio-Rad MiniProtean Electrophoresis System. Commercial lysozyme (lyophilized powder) was prepared in H_2O and loaded into the SDS-PAGE as a marker to identify lysozyme location in the resulting gel because lysozyme in also present in the eggshell extract. After running the gel at 200 V, the gel was stained using Brilliant Blue R solution for 15 min at room temperature on an orbital shaker at 100 rpm. Then the gel was destained in 5% acetic acid, % 40 methanol solution on an orbital shaker overnight. The target protein bands (i.e. OC-17 and lysozyme) were excised from the gel and kept in 10% methanol solution until the LC-MS analysis. All other subsequent sample preparations were done by the Cambridge Centre for Proteomics (CCP) staff. LC-MS analysis was conducted using Thermo Scientific Lumos and Dionex 3000 RSLCnano (Massachusetts, USA) at the Department of Biochemistry. MASCOT LC-MS data analysis software (Matrix Sciences, Boston, USA) was used to match the amino acid sequences of the target proteins from the National Center for Biotechnology Information, Basic Local Alignment Search Tool (NCBI BLAST) search (Protein IDs or OC-17 and lysozyme Q9PRS8 and P00698, respectively).

4.2.2.5 Incorporation of Model Proteins in $CaCO_3$ and Reconstitution of Intra-Crystalline Proteins

The protein incorporation experiments were conducted in deionized H_2O. This is because the salts in a buffer solution do modify $CaCO_3$ formation which could lead to different phases of the mineral. The final $CaCO_3$ and protein concentrations are 100 mM and 0.05 mg/mL, respectively (for model protein entrapment experiments, BSA and lysozyme were prepared with a concentration of 0.5 mg/mL for a 10 mL crystallization experiment). The co-crystallization solutions were prepared in 10 mL tubes. After mixing 4.5 mL of 222 mM $CaCl_2$ and $NaHCO_3$ solutions in a 10 mL tube at room temperature, 1 mL of 0.5 mg/mL protein was added immediately, mixed and the solution was kept at 4 °C overnight. In order to ensure that only intra-crystalline proteins were extracted, overnight samples were centrifuged, supernatants were discarded and crystals were incubated with %5 NaClO solution for 3 h. Then, the crystal samples were washed extensively three times and each time the supernatant water was discarded after centrifugation. The dissolution of the crystals was done with 200 mM 5 mL EDTA solution at room temperature overnight. The resulting solution (i.e. EDTA dissolved $CaCO_3$ containing intra-crystalline proteins) was transferred to a centrifugal ultrafiltration tube and centrifuged first to 500 µL, diluted to 15 mL in H_2O and concentrated again. This dilution/concentration cycle was repeated three times to remove EDTA completely and change buffer to H_2O.

Four replicates were prepared for each experiment. In order to analyze samples after the co-entrapment of BSA and lysozyme, proteins were separated using chromatography. CM Sepharose ion exchange column was first washed with deionized H_2O and equilibrated with 100 mM PBS pH 7.4. The concentrated protein solution (0.5 mL) was diluted in 1 mL PBS and loaded to the column. The elution was collected to obtain BSA as it is negatively charged at pH 7.4. Lysozyme (positively charged at pH 7.4) was washed with 1 M NaCl and buffers of the proteins were exchanged to deionized H_2O. The concentration of each protein was measured using their absorbance at 280 nm and their molecular extinction coefficients which were calculated using their sequence.

4.3 Results

4.3.1 LC-MS Identification of OC-17 and Lysozyme in the Eggshell

In order to confirm that the target protein studied in the eggshell is OC-17, LC-MS was conducted to the protein band upon running an SDS-PAGE of the eggshell extract. Lysozyme was also identified in the eggshell extract to confirm the presence of lysozyme in the eggshell as

an intra-crystalline protein. The molecular weights of lysozyme and OC-17 are very close to each other, determining both lysozyme and OC-17 via LC-MS further confirms the location of the protein bands in the SDS-PAGE since the location of the bands are used to identify and assess which protein is retained upon purification experiments.

LC-MS amino acid sequences of lysozyme (80% coverage) and OC-17 (78% coverage) from the eggshell extract are given below. The sequences that match the UniProt entries (Q9PRS8 and P00698 for OC-17 and lysozyme, respectively) are given in bold letters.

The amino acid coverage of the lysozyme sample:
MRSLLILVLC FLPLAALGKV FGR**CELAAAM KRHGLDNYRG**
YSLGNWVCAA KFESNFNTQA TNRNTDGSTD YGILQINSRW
WCNDGRTPGS RNLCNIPCSA LLSSDITASV NCAKKIVSDG
NGMNAWVAWR NR**CKGTDVQA WIR**GCRL

The amino acid coverage of the OC-17 sample:
DPDGCGPGWV PTPGGCLGFF SRELSWSRAE SFCRRWGPGS
HLAAVRSAAE LR**LLAELLNA SR**GGDGSGEG ADGR**VWIGLH**
RPAGSRSWR**W SDGTAPRFAS WHR**TAKARRG GR**CAALRDEE**
AFTSWAARPC TERNAFVCKA AA

The matching amino acid coverage for lysozyme and OC-17 were calculated to be 80% and 78%, respectively. In a typical LC-MS protein identification experiment, a coverage of 80% upon comparison of target sequences from NCBI BLAST is accepted as overwhelmingly conclusive. It should be noted that, the LC-MS results are presented in this chapter because the heat treatment experiments (Section 4.2.2) and the presence of lysozyme as an intra-crystalline (Section 4.3.5 and Discussion) have key implications for protein incorporation in $CaCO_3$ as a long term preservation method during high temperature storage.

4.3.2 Preliminary Structural Analysis of the Partially Purified OC-17 Before and After Heat Treatment

As explained in detail in Chapter 1 (see Section 1.2), intra-crystalline protein incorporation might be used for preservation of protein-based vaccines at high temperatures without refrigeration as long as the 3D structure of the incorporated protein is not affected after exposed to high temperatures (~45 °C). In order to test whether or not the protein structure is affected at high temperature, OC-17 could be tested as a model intra-crystalline protein.

As a preliminary investigation for the effect of heat on intra-crystalline proteins, OC-17 was partially purified and CD spectra were collected before and after the heat treatment of the protein at 50 °C for 60 h. Figure 4.8 shows the CD spectra of three samples of partially purified OC-17. The CD spectrum of the partially purified OC-17 was used as a positive control (red curve in Fig 4.8) for heat treatment experiments because OC-17 was extracted from eggshell at room temperature. Then the positive control sample was heat treated (in a sealed eppendorf tube to avoid evaporation) at 50 °C for 60 h in a laboratory oven. After the heat treatment the CD spectrum of the sample was collected (dark blue curve in Fig 4.8) and found to be structurally different compared to the positive control, possibly due to denaturation at high temperature. In order to analyze the effect of heat on $CaCO_3$ incorporated protein (i.e. incorporated OC-17), an eggshell was heat treated at 50 °C for 60 h before the protein extraction was done. OC-17 was then extracted from the heat treated eggshells (using the same extraction protocol as for the positive control) and analyzed with CD spectroscopy at the same concentration with the other samples. The spectrum of OC-17 from the heat treated shells (light blue curve in Fig 4.8) suggests that as long as the OC-17 is in the $CaCO_3$ lattice in the eggshell, it could preserve its structure even at 50 °C at least for 60 h.

The preliminary result of structural integrity of the intra-crystalline protein (i.e. OC-17) after heat treatment was the rationale to further investigate the possibility of high temperature preservation of other proteins in $CaCO_3$ lattice. It should be noted that the stability of OC-17 does not confirm the potential of heat stability of any given protein upon intra-crystalline $CaCO_3$ incorporation. The stability of OC-17 might be specific either to OC-17 itself or to the eggshell mineralization-specific proteins, such as SCA-1 and SCA-2 fro*m* ostrich [19, 32, 46].

4.3.3 The Effect of the Reconstitution Buffers on Protein Structure

As explained in the previous section, the heat treatment tests of OC-17 show that the intra-crystalline incorporation might be a promising method to preserve proteins at high temperatures. However, other proteins apart from OC-17 should also be shown for their heat stability in the $CaCO_3$ lattice. For this reason, three model proteins were studied for their efficiency to be incorporated in $CaCO_3$: BSA, lysozyme and diphtheria antibody.

In order to analyze the effect of the heat treatment only, other potential denaturation processes should be eliminated. One such potential denaturation might happen during the reconstitution of intra-crystalline proteins (i.e. dissolution of the protein incorporated $CaCO_3$). In other words, the proteins in the $CaCO_3$ lattice might be denatured during dissolution of $CaCO_3$ and not because of the heat treatment. For this reason, it should be shown that the dissolution process does not affect the 3D structure of the model intra-crystalline protein.

Because there are only two ways to dissolve $CaCO_3$ (i.e. metal chelation and acid dis-

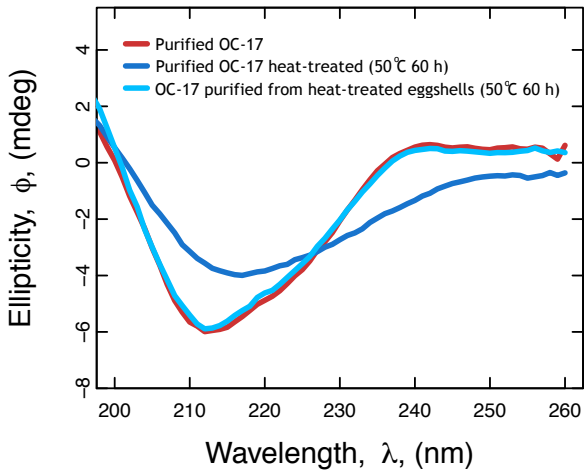


Figure 4.8: The CD spectra of partially purified OC-17 from the eggshell, before and after heat treatment as a preliminary test for the effect of the heat treatment on the protein. The red curve shows the partially purified OC-17 without any heat treatment as a positive control. The dark blue curve shows OC-17 after the heat treatment of the positive control sample at 50 °C for 60 h. The light blue curve shows OC-17 extracted from the eggshells that were heat treated at 50 °C for 60 h before OC-17 extraction was done.

solution), both methods were tested to find the optimal method to minimize (or preferably eliminate) the risk of any structural change during the reconstitution step. First, the acid reconstitution was tested. In order to analyze the effect of 5% (v/v) acetic acid solution on a model protein, BSA in H_2O was compared to BSA in 5% (v/v) acetic acid solution. A commercial BSA (lyophilized powder) was dissolved in 5% (v/v) acetic acid and, after an incubation period, analyzed using CD spectroscopy. The incubation time was fixed as for the time required to dissolve protein incorporated $CaCO_3$ samples (i.e. if the dissolution of the incorporated samples conducted for 30 min in 5% (v/v) acetic acid, then the incubation of BSA in 5% acetic acid (v/v) was also conducted for 30 min).

The top row in Figure 4.9 shows the CD spectra of the three replicates of acid incubated BSA samples compared to the positive control BSA in H_2O. The CD spectra suggest that there is no significant effect of acid reconstitution on the protein structure. Yet, as an alternative method, the effect of EDTA on BSA was also tested. Similar to the acid incubation, BSA in H_2O was compared to BSA incubated in 0.2 M EDTA solution. The middle row in Figure 4.9 shows the CD spectra of the three replicates of EDTA incubated BSA compared to the positive control BSA in H_2O. The difference in the amount of noise between acetic acid and EDTA reconstitution arises from the effect of far-UV on the acetic acid. Far-UV range was not included in the spectra because of the significant amount of noise and the wavelength range was set to 220-260 nm. (Buffer exchange could be done prior to CD analysis but this route

was not pursued as will be explained below. It should be added that EDTA causes significant noise throughout the spectrum range. For this reason, the buffers of EDTA incubated samples were changed to H_2O prior to CD analysis.)

Although the CD spectra of the acetic acid and EDTA incubation tests suggest that the reconstitution buffer does not affect protein structure significantly (in particular, in the case of EDTA), acetic acid was not used and EDTA was chosen for the subsequent experiments because the acid reconstitution of intra-crystalline proteins was found to be inconsistence. This can be seen in the bottom row in Fig 4.9. The three replicates of the intra-crystalline BSA were prepared and reconstituted in 5% (v/v) acetic acid. However, the CD analysis upon reconstitution shows significant fluctuations in the protein concentration between the replicates. On the other hand, EDTA reconstitution was found to be consistent as will be shown below. (Another important reason for using EDTA is because the acetic acid reconstitution provides much lower concentrations of the intra-crystalline proteins. When the replicates were reconstituted in acetic acid an average concentration of 0.13 mg/mL was measured whereas the replicates dissolved in EDTA gives an average of 0.42 mg/mL.)

EDTA reconstitution was further tested for its effect on the tertiary structure. The same samples that were used for the CD spectra were analyzed using intrinsic tryptophan fluorescence. Fig 4.10 shows the CD and intrinsic tryptophan fluorescence spectra next to each other for the three replicates. The CD spectra of the EDTA incubated samples show slight differences compared to the positive control. In order to further analyze the CD data, a web server-based CD data analysis software CAPITO [236] was used. CAPITO analyzes CD data by plotting ellipticity values at 200 and 220 nm (see [236] for details) and compares the spectrum with the CD spectra of proteins whose 3D structures previously determined by X-ray crystallography. The grey squares in the CAPITO analysis in Fig 4.10 represent proteins in their native and molten globule states taken from the Protein Circular Dichroism Data Bank (PCDDB) and white boxes represent proteins in their pre-molten globule and unfolded states taken from the literature [238]. The molten state of a globular protein is a compact yet partially denatured state which preserves the secondary structure of the native protein but lacks a significant amount of the tertiary structure [239]. The pre-molten globule state is similar to the molten state in the sense that it is also an intermediate between the folded and unfolded state of a protein. The pre-molten state differs from molten state as it is less compact, and unlike the molten state, the secondary structure is much less ordered [240].

Both the positive control and EDTA incubated samples were within the range of the native conformation which further suggest that EDTA could be used for as a reconstitution buffer without affecting the protein structure. In terms of intrinsic tryptophan fluorescence data, the positive control and EDTA incubated samples show no difference. This suggests that the

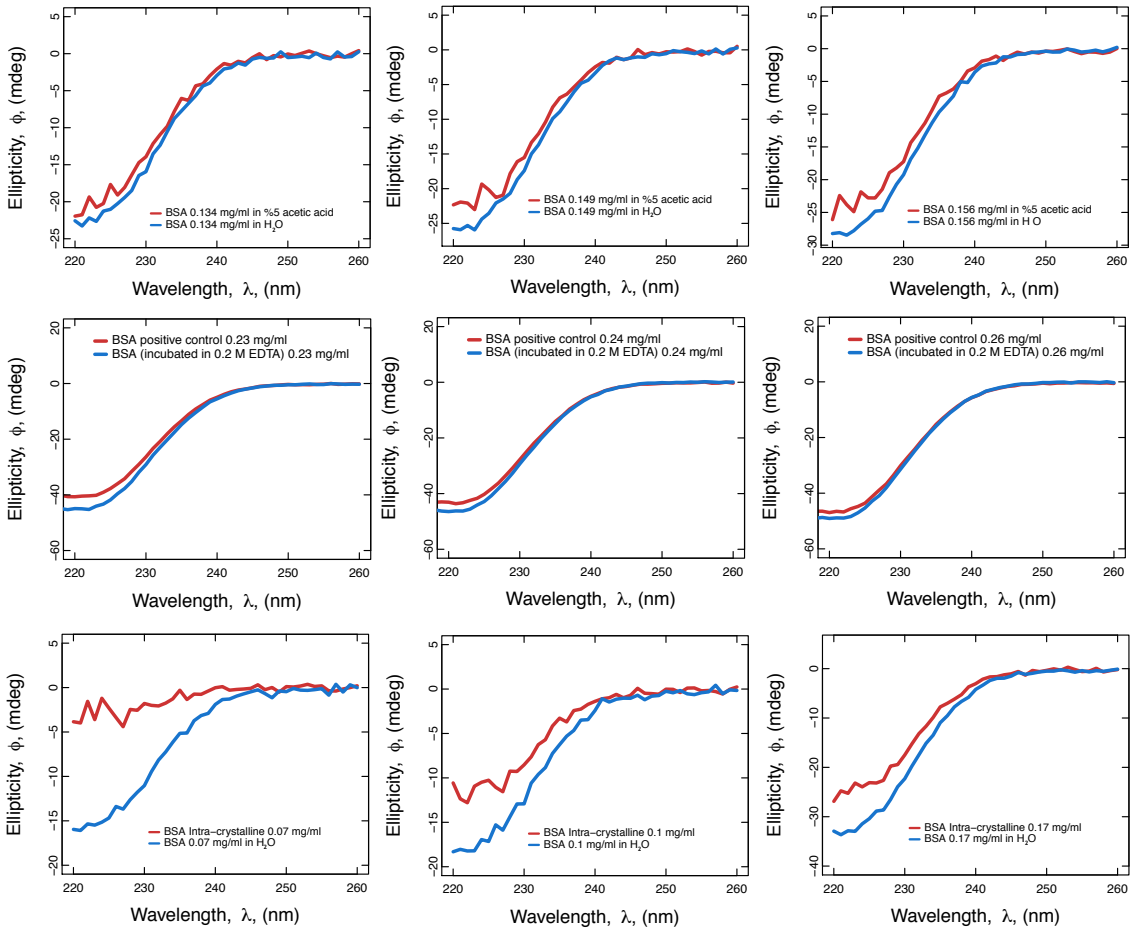


Figure 4.9: The CD spectra of the effect of the reconstitution buffers on protein structure. Top row shows the three replicates of BSA incubated in 5% (v/v) acetic acid solution and BSA in H_2O as a positive control. The spectra were collected from 220 to 260 nm as acetic acid spectrum below 220 nm causes significant amount of noise. The middle row shows the three replicates of BSA incubated in 0.2 M EDTA solution and BSA in H_2O as a positive control. EDTA was chosen as the reconstitution buffer because although the acetic acid reconstitution was shown to affect protein structure minimally, intra-crystalline reconstitution experiments using acetic acid were found to be inconsistent. The bottom row shows the CD spectra of the three replicates of BSA incorporated $CaCO_3$ reconstituted in 5% (v/v) acetic acid. Although the samples are replicates, concentrations show considerable differences which makes acetic acid reconstitution inconsistent.

tertiary structure was not affected by overnight EDTA incubation and it could be used (at least for BSA) to assess the effect of incorporation without an interference from the reconstitution step.

After analyzing the effect of EDTA reconstitution on BSA, the diphtheria antibody was tested as the overarching goal of the present work, that is, to stabilize the vaccine immunogens in $CaCO_3$ at high temperatures without refrigeration. The positive control diphtheria antibody was prepared in H_2O as recommended in the diphtheria antibody manual provided by the WHO/NIBSC who kindly provided the antibody used in this work. After the 0.2 M EDTA incubation overnight, the buffer was changed to H_2O and the CD and intrinsic tryptophan fluorescence measurements were done as explained above. Figure 4.11 shows the CD and intrinsic tryptophan fluorescence spectra. Although more noisy than BSA samples, the CD spectra suggest that there is no significant change in the secondary structure upon incubation of the diphtheria antibody in EDTA. The CAPITO analysis in 4.11 shows that the positive control and EDTA incubated diphtheria antibody lie within the upper part of the native state. As mentioned above, the extent to which the function of the diphtheria antibody was affected by EDTA incubation could be tested further (particularly for the intra-crystalline protein) as a spectroscopic analysis is not sufficient alone. In addition, intrinsic tryptophan spectra show differences between positive control and EDTA incubated diphtheria antibody. The effect of any structural deviation (compared to the positive control) on the function of the diphtheria antibody could be further tested *in vitro* (personal communication with P. Stickings from the NIBSC).

4.3.4 Structural Analysis of $CaCO_3$ Incorporated Intra-crystalline BSA

After testing the effect of EDTA reconstitution on the protein structure, the intra-crystalline incorporation experiments were conducted for BSA, diphtheria antibody and lysozyme. Because it was shown that BSA is not affected structurally upon overnight EDTA incubation, a change in the protein structure could be attributed to the incorporation process after reconstitution in EDTA. Figure 4.12 shows the CD and intrinsic tryptophan fluorescence spectra of the intra-crystalline BSA samples. The consistency of the replicates in terms of protein concentration shows that EDTA is more consistent as a reconstitution buffer compared to the 5% (v/v) acetic acid. In addition the intra-crystalline concentration is much higher in case of EDTA reconstitution as is given in Table 4.1.

The CD spectra of the intra-crystalline BSA suggest that, although not significant, there is some degree of change in the secondary structure. Compared to the EDTA incubated samples (see 4.10), intra-crystalline BSA samples show divergence from the positive control. The CAPITO analysis of the CD data suggest that the BSA intra-crystalline samples were not af-

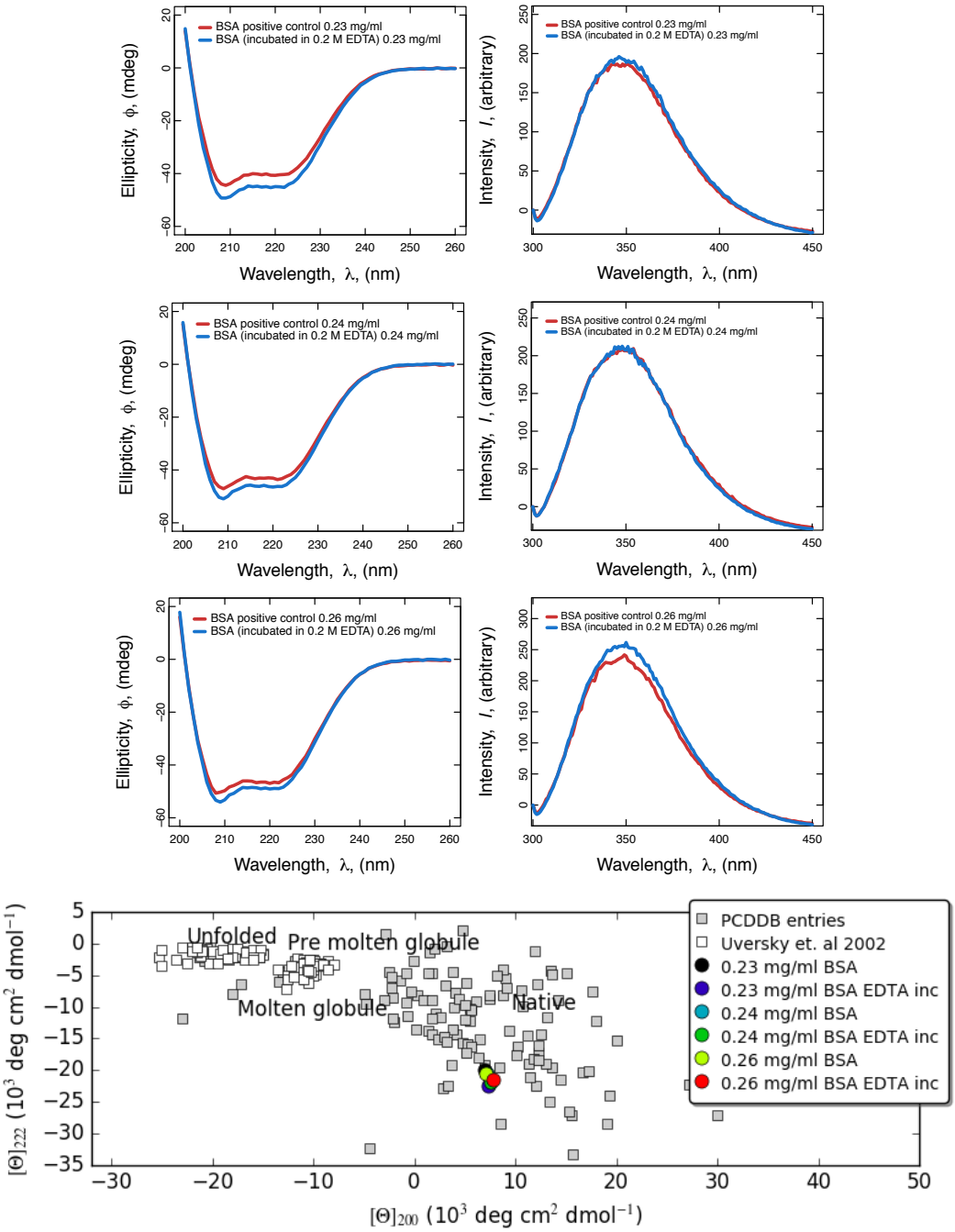


Figure 4.10: The CD and intrinsic tryptophan fluorescence spectra of the positive control and 0.2 M EDTA incubated BSA samples (three replicates are shown). The same samples used for both CD and intrinsic tryptophan fluorescence measurements and presented next to each other for comparison. The incubation was done in 0.2 M EDTA overnight. A buffer exchange step was conducted as EDTA interferes with the CD signal in the UV range. At the bottom, the CAPITO CD data analysis is shown as the ellipticity values of 200 nm versus 220 nm (see the text for details).

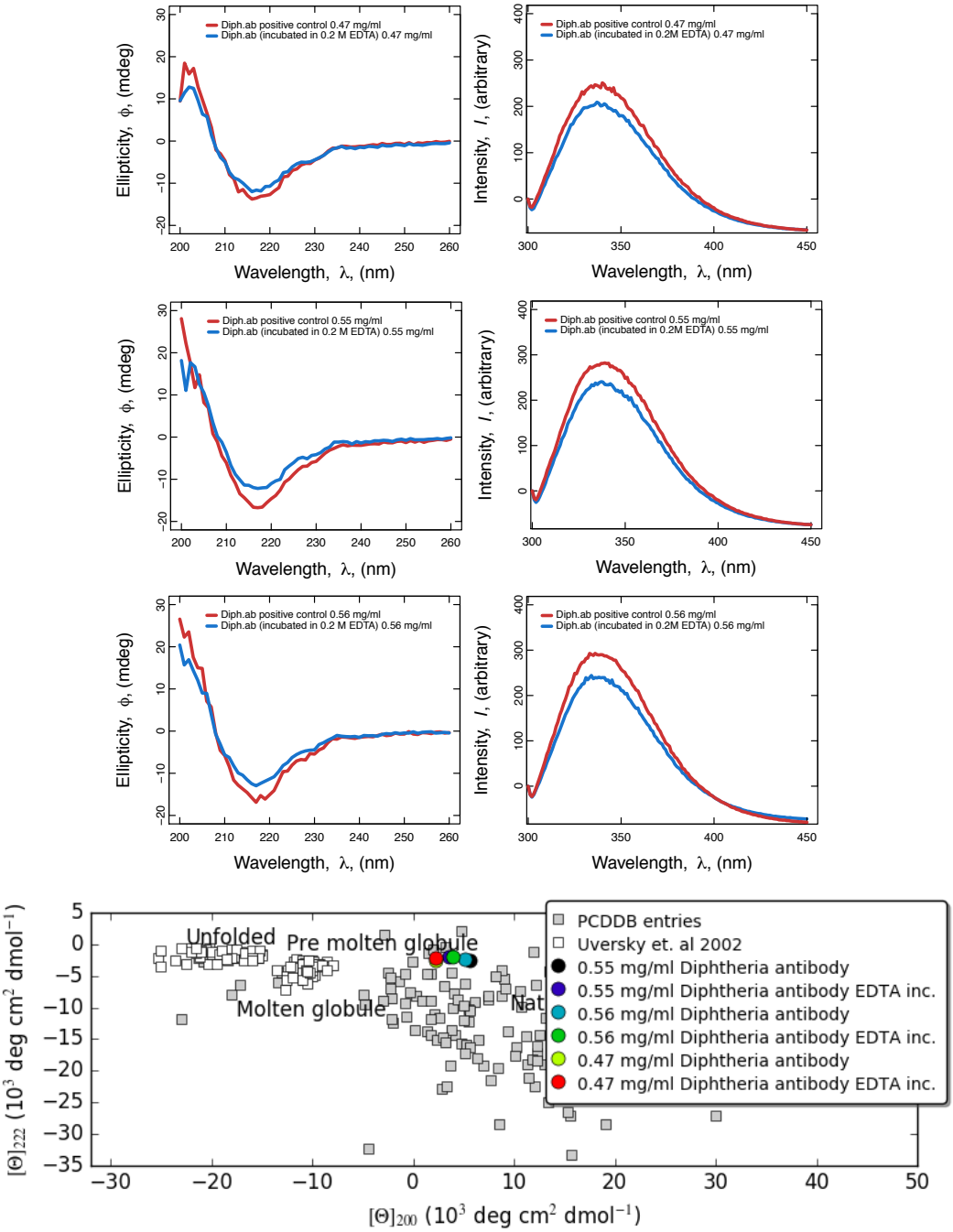


Figure 4.11: The CD and intrinsic tryptophan fluorescence spectra of the positive control and 0.2 M EDTA incubated diphtheria antibody samples (three replicates are shown). The same samples used for both CD and intrinsic tryptophan fluorescence measurements and presented next to each other for comparison. The incubation was done in 0.2 M EDTA overnight. A buffer exchange step was conducted as EDTA interferes with the CD signal in the UV range. At the bottom, the CAPITO CD data analysis is shown as the ellipticity values of 200 nm versus 220 nm (see the text for details).

Protein	Added (in 10 mL)	Recons. buffer	Extracted (0.5 mL)
BSA	0.5 mg	EDTA (0.2 M)	0.42 ± 0.04 mg/mL (n=3)
BSA	0.5 mg	Acetic acid (5%)	0.077 ± 0.02 mg/mL (n=3)

Table 4.1: The amount of BSA upon reconstitution either in acetic acid (5%) or in EDTA (0.2 M). The incorporation was done in 10 mL of crystallization solution by adding 1 mL of 0.5 mg/mL protein and the reconstituted protein was concentrated to 0.5 mL.

fected considerably yet all intra-crystalline samples (dark blue, green and red dots in CAPITO plot) are slightly located above the positive controls. (Note that in the incubation tests all samples were much closer to each other.) However, the intrinsic tryptophan fluorescence spectra show significant change in intensity of the intra-crystalline samples compared to the positive control. This suggest that the tertiary structure was changed likely because of the incorporation process.

4.3.5 Incorporation of the Diphtheria Antibody and Lysozyme into $CaCO_3$ as Intra-crystalline Proteins

After using BSA as a model intra-crystalline protein, two other proteins were tested: lysozyme and diphtheria. The rationale for using lysozyme is the fact that it is found in the eggshell as was shown above (Section 4.3.1). This suggests that lysozyme could be incorporated in $CaCO_3$. The diphtheria antibody was used to test the hypothesis of the present work, that is, using $CaCO_3$ incorporation for long term vaccine preservation. The incorporated amount of the proteins varies substantially (Table 4.2). In the incorporation experiments, 0.5 mg/mL protein was added into a 10 mL crystallization solution and before reconstitution, the crystals were treated in 5% NaClO solution for 3 h to remove the proteins and any other organic contaminants from the surface of $CaCO_3$. This cleaning step also ensures that only the intra-crystalline proteins are present in the final protein solution. The concentration of the incorporated proteins with standard deviations and incorporation efficiencies are presented in Table 4.2. The BSA is efficiently incorporated into $CaCO_3$ as 84% of the protein added into the solution could be extracted back from the crystal lattice. In the case of the diphtheria antibody and lysozyme, the efficiencies clearly indicates that these proteins are not incorporated and the concentrations of the intra-crystalline proteins are negligible.

4.3.6 Co-incorporation of Lysozyme Using BSA as a Chaperone Protein

As explained above, lysozyme is found in the eggshell but not incorporated *in vitro* (as presented in Table 4.2). This suggests that lysozyme in the eggshell might be incorporated during

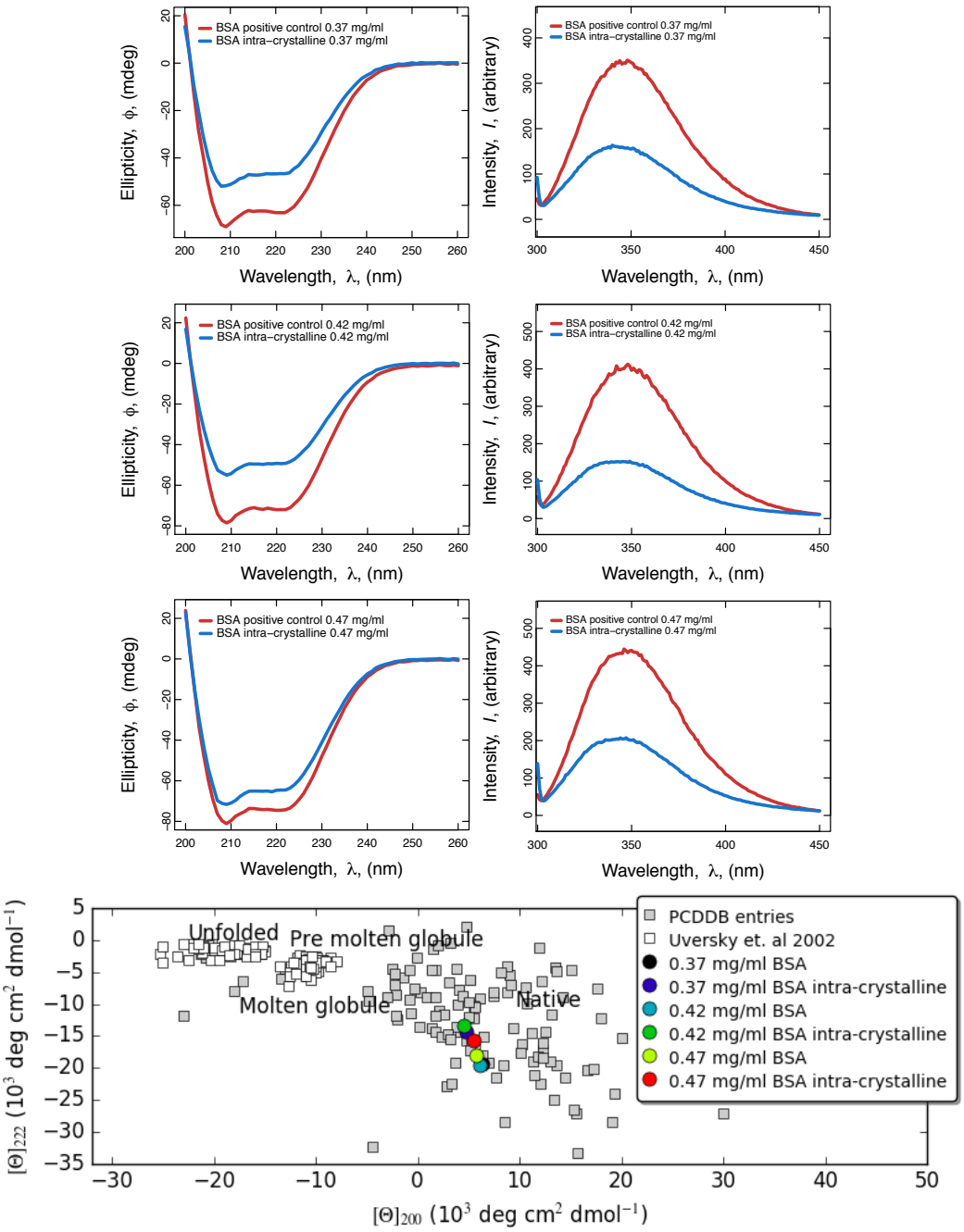


Figure 4.12: The CD and intrinsic tryptophan fluorescence spectra of the positive control and intra-crystalline BSA samples (three replicates are shown). The same samples used for both the CD and intrinsic tryptophan fluorescence measurements and presented next to each other for comparison. Before reconstitution, 3 h of 5% NaClO treatment was done to remove the proteins on $CaCO_3$ surfaces and to ensure that only the intra-crystalline proteins were tested. A buffer exchange step was conducted as EDTA interferes with the CD signal in the UV range. At the bottom, the CAPITO CD data analysis is shown as the ellipticity values of 200 nm versus 220 nm (see the text for details).

Protein	Added (in 10 mL)	Extracted (0.5 mL)	Efficiency (%)
BSA	0.5 mg	0.42 ± 0.04 mg/mL (n=3)	84
Lysozyme	0.5 mg	0.002 ± 0.001 mg/mL (n=3)	0.54
Diphtheria	0.5 mg	0.03 ± 0.0002 mg/mL (n=3)	0.6

Table 4.2: The amount of protein added in 10 mL of crystallization solution was 0.5 mg/mL for all three model proteins. After overnight reconstitution in 0.2M EDTA, buffers were changed to H_2O and protein solutions were concentrated to 0.5 mL using ultrafiltration centrifugal filters.

Protein	Added (in 10 mL)	Extracted (0.5 mL)	Efficiency (%)
BSA	0.25 mg/mL	0.1 mg/mL (n=1)	40
Lysozyme	0.25 mg/mL	0.004 mg/mL (n=1)	1.6

Table 4.3: The amount of the extracted proteins from co-incorporation of BSA and lysozyme. BSA and lysozyme were mixed and added together in a 10 mL crystallization solution. After crystallization and reconstitution, intra-crystalline solution was loaded into CM Sepharose ion exchange column to separate BSA from lysozyme. Each protein fraction after separation was then concentrated to 0.5 mL. One of the four replicates was tested.

incorporation of other proteins (i.e. incorporated along with OC-17 if it acts a chaperone protein). To test the hypothesis of 'co-incorporation', BSA and lysozyme were added together in a crystallization solution. After crystallization and reconstitution, the intra-crystalline solution loaded into an ion exchange column to separate the BSA and lysozyme (see Section 4.2.2.5 for details). The amount of the extracted proteins from the co-incorporation of BSA and lysozyme is given in Table 4.3.

The lysozyme concentration measured to be 0.004 mg/mL which suggests that BSA has no effect on lysozyme incorporation. In addition, the amount of the incorporated BSA was observed to decrease significantly (from 0.42 mg/mL when it is incorporated alone to 0.1 mg/mL in the presence of lysozyme). This result suggests that any protein that is incorporated in $CaCO_3$ with high efficiency does not necessarily act as a chaperone protein. OC-17 needs to be tested for its ability to act as a chaperone. This could be tested using lysozyme because as explained in the previous chapter (see Chapter 3, Section 3.3.1), OC-17 and lysozyme are eluted together during the eggshell protein extract purification. The lysozyme-OC-17 solution could be used directly in an incorporation experiment for the analysis of lysozyme incorporation in the presence of OC-17. (Although the instability problem of the lysozyme-OC-17 solution should be considered, see Chapter 3 Section 3.3.1)

4.4 Discussion

In order to test the hypothesis that protein incorporation in $CaCO_3$ might be used for vaccine stabilization at high temperatures, the stability of the proteins in $CaCO_3$ lattice after high temperature storage must be shown. The preliminary analysis of OC-17 extracted from the heat treated eggshells suggests that the incorporation in eggshell crystals (at least in the case of OC-17) could preserve the protein structure. However, this might be unique for the eggshell protein incorporation *in vivo* and OC-17 might not be stable if incorporated in $CaCO_3$ *in vitro*. The stability of OC-17 in synthetic $CaCO_3$ needs to be tested. Because OC-17 was only partially purified, three commercial proteins (BSA, lysozyme and diphtheria antibody) were used to test the incorporation and protein structure upon reconstitution of the proteins from the host $CaCO_3$.

The reconstitution buffer used to release the intra-crystalline proteins should have minimal or ideally no effect on the protein structure. EDTA was selected as the buffer because the acetic acid reconstitution was found to be problematic for two reasons. First, the replicates dissolved using 5% (v/v) acetic acid were found to have significant differences in their concentrations which makes this reconstitution method inconsistent. The EDTA reconstitution was found to be consistent in terms of the replicate concentrations. Second, although the incorporation experiments were done under the same conditions, when EDTA was used for reconstitution instead of acetic acid, the intra-crystalline protein concentration was found to be considerably higher (0.07 ± 0.02 mg/mL versus 0.42 ± 0.04 mg/mL). The reason for the loss of proteins during acetic acid reconstitution might be acid-trigged precipitation of the proteins. Similarly, a virus incorporation study has reported the effect of acetic acid dissolution [194]. Although the virus particles have been detected in $CaCO_3$, after dissolution using 0.5 M acetic acid, the number of virus particles decreased drastically such that the particles could not be detected even in gel electrophoresis or in TEM analysis [194]. Although less likely, bubble formation during acetic acid dissolution (because of CO_2 release upon acid-$CaCO_3$ reaction) could also trigger protein precipitation which may explain the low amount of intra-crystalline proteins. On the other hand, EDTA has the advantage of dissolving $CaCO_3$ at pH 7.4 which is an ideal option (particularly for vaccine reconstitution) and is safe for human use at the concentration used in the present work [241, 242].

Among the model proteins tested, BSA was the only protein incorporated in $CaCO_3$ with high efficiency (84%). The lysozyme and diphtheria antibody were incorporated with less than 1% efficiency. The mechanism of incorporation of a given protein or any other molecule is likely to be governed by specific chemical interactions between the target molecule and $CaCO_3$ [39, 44, 91, 224]. The Incorporation mechanism of BSA might be similar to that of ovalbumin which is found in the eggshell [109] although its incorporation mechanism is unknown. Note

that BSA and ovalbumin are closely related proteins, the former is the albumin of the bovine serum and the latter is the albumin of the egg white. It is known that the incorporation of positively charged proteins (i.e. proteins with isoelectric points higher than the reaction pH) is non-trivial based on the fact that the biological intra-crystalline proteins (e.g. nacre proteins rich in aspartic and glutamic acid) are highly negatively charged [40, 243–245]. With respect to the negatively charged proteins, ovalbumin and BSA are expected to be incorporated but the diphtheria antibody is also negatively charged at the reaction conditions (e.g. pH of crystallization solution) and yet is not incorporated. Similarly, lysozyme is not incorporated but in this case the reason might be the electrostatic repulsion between the lysozyme and Ca^{2+} ions as reported in the literature [244, 246, 247]. It should be noted that lysozyme is an eggshell intra-crystalline protein as was also shown in the present work (see Section 4.3.1) and also in the literature [232]. This confirms that lysozyme could be incorporated in $CaCO_3$. Lysozyme loading into porous $CaCO_3$ has been reported [191, 244] but in these cases lysozyme is not an intra-crystalline component. The mechanism of lysozyme incorporation in the eggshell could happen from the effect of 'chaperone' roles of some of the other intra-crystalline proteins.

Apart from the incorporation efficiency of proteins in $CaCO_3$, there are two crucial issues that need to be addressed. The first one is the structure of the intra-crystalline proteins upon reconstitution. The second one is the effect of high temperature on the intra-crystalline proteins. Because the EDTA reconstitution was shown to have a minimal effect on the secondary and tertiary structures, analysis of the incorporated BSA samples using CD and intrinsic tryptophan fluorescence shows that the incorporation process does affect the protein structure (at least in the case of BSA). For this reason, heat treatment of BSA incorporated $CaCO_3$ samples was not conducted further. The stability of the intra-crystalline proteins needs to demonstrated before any high temperature tests on the incorporated proteins.

The stability and incorporation ability of the intra-crystalline proteins seem to depend on how the crystallization reaction is conducted. For instance, *in vivo* crystallization (i.e. eggshell formation) includes copious amount of different proteins (see [108]) some of which are inherently unable to be incorporated, such as lysozyme. Yet during the formation of the eggshell, many proteins including lysozyme are incorporated even with high efficiency (see Section 3.3.1 for lysozyme). This suggests that the eggshell formation involves proteins that act as chaperones to assist the incorporation of other protein during the shell formation. A natural chaperone protein might be the OC-17. Although the function of OC-17 is likely to transform amorphous $CaCO_3$ to calcite [33, 143], it could assist other proteins to be incorporated. This possibility is not unreasonable because *in vitro* dye incorporation in $CaCO_3$ has been shown using amino acids [248]. The possibility of OC-17 to act as a chaperone could be tested using the lysozyme and OC-17 mixture in the 'co-incorporation' experiments.

The hypothesis that the protein incorporation in $CaCO_3$ could be used for vaccine storage without refrigeration is based on the fact that in nature some of the intra-crystalline proteins seem to be preserved for eons in the avian eggshell crystals [21],[22],[23],[24]. For instance, the ostrich eggshell intra-crystalline protein SCAs have been studied to understand their survival mechanism for millions of years bound in the ostrich eggshell [19]. The simulations of the intra-crystalline proteins (see [33, 143] for OC-17 and [19] for SCAs) need crystal structures of these proteins which have been obtained using X-ray crystallography of the eggshell extracted proteins. This suggests that these proteins may not significantly denature during the incorporation in the eggshell or dissolution from the eggshell crystals. In addition, in the case of the fossil intra-crystalline samples, the eggshells have been excavated from the archeological sites have been exposed to temperature fluctuation during thousands of years and yet still the intra-crystalline proteins resist degradation [19] (also see [26, 30]). The preservation of intra-crystalline proteins for thousand of years in nature might not be unique to the avian eggshells and it might be possible to mimic this phenomenon using OC-17 or other C-type lectin-like eggshell proteins. In order to mimic this phenomenon, the confirmation of the function of OC-17 experimentally may be useful. A viable option to test the proposed function of the OC-17 for its interaction with $CaCO_3$ (i.e. function proposed through MD simulations [33]) is to use highly sensitive biophysical methods, such as FRET, through which OC-17-$CaCO_3$ interaction might be studied (see Chapter 5, Section 5.2.1 for details).

Determining the cDNA sequence [149] of OC-17 enabled cloning of this protein in *E. coli* and similarly the cDNAs of SCAs need to be determined to express these proteins in expression vectors. The SCAs might be more efficient compared to OC-17 in terms of the binding energy in $CaCO_3$ and possibly they may act as chaperone proteins. In addition, a genetic fusion of OC-17 to a target protein, such as GFP, might be expressed and the ability of OC-17 in incorporating the GFP could be detected. The fluorescence activity of GFP upon reconstitution would indicate an unperturbed protein 3D structure.

Alternatively, it might be feasible to apply the vaccine preservation method proposed here to produce thermostable vaccines *in vivo*. Because it was shown that OC-17 seems to be structurally intact upon heat treatment in the eggshell, it could be possible to genetically fuse a vaccine protein with a purification tag upstream or downstream to the cDNA of OC-17 in a chicken embryo) using the CRISPR-Cas genome editing tool to produce an intra-crystalline vaccines directly synthesized in the eggshell. This might allow incorporation of vaccine proteins in natural eggshells which could then be reconstituted and purified on-site from the eggshell powder.

Chapter 5

Conclusions & Future Work

5.1 Conclusions

The thermostability of vaccines at high temperatures (>45 °C) outside the cold-chain is still a major problem in global health. Novel methods to address this problem are needed because of their potential impact on human health. Although intra-crystalline protein incorporation has been known for decades, biomimetic protein incorporation *in vitro* is still a nascent field of research. In this work, eggshell samples from different species were analyzed in terms of their mechanical and chemical properties. Because the major high-concentration proteins are intra-crystalline ones in the eggshell, the effect of these proteins on eggshells could be inferred through sensitive mechanical and chemical tests. It is concluded that nanoindentation and IR mapping reveals differences between eggshells from different species. From 14 species studied here, 3 species, kookaburra, tinamou and rhea could offer currently unknown yet potentially high-affinity intra-crystalline proteins based on the mechanical and chemical properties of eggshells of these species.

It was shown that OC-17 could be purified by separating lysozyme from eggshell extract using Ivy protein. Although OC-17 purification has been reported in the literature, it was concluded in this work that purification methods described for OC-17 are neither reliable nor efficient. In addition, the methods reported in the literature could potentially denature OC-17 which would hamper subsequent experiments on OC-17 in particular if the protein is to be used to probe its function or to be used as a chaperone protein which requires unfolded protein. The first recombinant expression of OC-17 (i.e first intra-crystalline eggshell protein cloning) was conducted in the present work which opens up a number of possibilities. Although the expression needs to be improved further, cloning of OC-17 allows to synthesize the protein without the need of purification from eggshell which was concluded to be non-trivial. Cloning also allows the introduction of point mutations which could be key to probe the function of

OC-17 using biophysical techniques, such as FRET-FLIM.

A key issue that needs to be addressed about intra-crystalline proteins is the 3D structure of these proteins upon dissolution of their inorganic host. The 3D structures of model proteins were analyzed both for the effect of reconstitution conditions (i.e the effect dissolution buffer) and for the incorporation process itself. It can be concluded that *in vitro* protein incorporation is not always feasible for a given protein as was the case for diphtheria antibody and lysozyme. It is worth noting that *in vivo* lysozyme incorporation does occur in the eggshell. This suggests that there is much to unveil about *in vivo* protein incorporation to further improve *in vitro* incorporation of a target protein. In addition, even if protein incorporation occurs, the 3D structure of the incorporated protein could change during the process. A change in the tertiary structure was observed for BSA. On the contrary, an intra-crystalline protein, OC-17, is thought to preserve its conformation during incorporation. This is based on the fact that OC-17 X-ray 3D structure has been determined using eggshell-extracted protein. Similarly, fossil eggshell proteins have been analyzed to understand their survival mechanism in $CaCO_3$ lattice. Temperature fluctuations had undoubtedly happened many times in millions of years until the fossil eggshell samples are excavated. However, it could be concluded that these changes in temperatures do not entirely degrade proteins as they can still be identified in the fossil record. Fossil eggshell intra-crystalline proteins suggest that it could be possible to engineer this natural phenomenon to incorporate a target protein, such as a vaccine antigen, *in vitro* to keep it stable without refrigeration even in the hot climates.

It is also concluded that moving towards applications of intra-crystalline protein incorporation could be done in two ways. First, because it is not always possible to incorporate a target protein in $CaCO_3$, a chaperone protein might be needed. Eggshell contains a myriad of proteins some of which are unable to be incorporated themselves, such as lysozyme. This suggests that some proteins might also act as chaperones, that is, assisting other proteins to be incorporated in growing $CaCO_3$ lattice. A candidate for this role could be OC-17. The ability of OC-17 to act as a chaperone needs to be tested experimentally. Second, harnessing *in vivo* incorporation could be a more efficient way compared to *in vitro* incorporation. More specifically, instead of conducting protein incorporation *in vitro* using both the purified target protein and a chaperone protein, it might be possible to produce a target protein in the eggshell *in situ*. This might allow obtaining a target protein from the eggshell directly as a recombinant protein synthesized *in vivo*. This process would need genome editing to insert the target protein gene as well as a purification tag to the relevant locus (i.e down-stream to OC-17 with a self-cleavable peptide between OC-17 and the target protein to separate them after translation). Reconstitution and isolation of the target protein could be optimized for on-site applications. As the preliminary results of this work suggest, eggshell proteins could

be inherently stable at high temperatures inside $CaCO_3$ lattice.

In conclusion, both incorporation and reconstitution of a target protein could be a novel strategy to keep vaccines stable at high temperatures until they are needed. This biologically inspired vaccine stabilization method could therefore address one of the problems in global health. Production of heat stable vaccines would be beneficial in countries where the lack of infrastructure hampers vaccination efforts. Delivering vaccines to people who need them most would help save many more lives and decrease the burden of diseases through achieving a better vaccination coverage.

5.2 Future Work

5.2.1 Probing the Interaction of OC-17 with $CaCO_3$ via FRET-FLIM TCSPC

FRET (also known as Förster resonance energy transfer by the name Theodor Förster who originally described it) is a process in which energy transfer occurs from one fluorophore to the other under specific conditions [233, 234, 249]. In order for the phenomenon to occur, one fluorophore should be in the excited state and second fluorophore should be in close proximity with in a range of 2-10 nm [233]. During FRET a non-radiative transfer of energy should occur meaning that no light is given off [234]. FRET is used to analyze the interaction of two samples, for instance two proteins, if one of the proteins has a fluorophore acting as donor and the second has another fluorophore that could act as an acceptor [234]. The basic principle of FRET is that if the first fluorophore is excited, the other could absorb the energy and then emit at its emission wavelength. In this way, an emission from second fluorophore is observed although it is not excited by an energy source, i.e other than the non-radiative energy from the first fluorophore [234]. This means an emission could only occur if the samples interact with each other so that they could come close enough for the energy transfer to occur. A requirement for FRET analysis is that the emission spectra of one fluorophore should overlap substantially with the spectra of the second fluorophore [234]. The efficiency of FRET depends on the distance of the fluorophores and it reduces with a relation to the sixth power of the distance [233]. In addition, dipoles of fluorophores should be parallel to each for the FRET to occur so if they are perpendicular no FRET is observed [233].. A schematic explanation of FRET rules are depicted in Fig 5.1 (Ref. [234]).

There are five methods to harness FRET, namely, acceptor bleaching, sensitized emission, spectral imaging, anisotropy and fluorescence lifetime imaging (FLIM). Among these, the most sensitive method is FLIM [234]. Furthermore, it is possible to measure the percentage

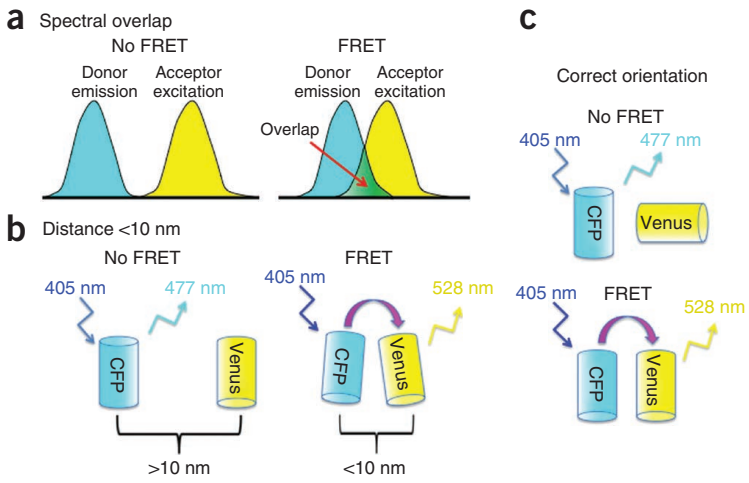


Figure 5.1: Schematic representation of conditions that allow FRET. a. Donor and acceptor spectra should overlap for efficient FRET, without an overlap no FRET is observed. b. The distance between donor and acceptor should not exceed 10 nm, otherwise only the emission of donor is observed. c. Dipoles of the donor and acceptor should be parallel to each other for FRET to occur. Perpendicular arrangement gives rise to donor emission alone [234].

of interacting and non-interacting donor population using FLIM, a measurement that is not possible with the other four [234].

The method to measure the fluorescence lifetime is the time-correlated single photon counting (TCSPC) [249]. TCSPC uses a pulsed laser at high frequency to excite the donor fluorophore [249]. A detector measures fluorescence decay time after pulse excites a photon to be emitted from the donor [249]. A decay curve of lifetime (in nanoseconds range) versus photon count is plotted as seen in Fig 5.2(a). A fitting procedure is done to calculate lifetime after correction for the instrument response function [249]. If coupled with confocal laser scanning microscope, it is possible to image the lifetime for each pixel location as is seen in Fig 5.2(b).

A decrease in the lifetime of the donor occurs in the presence of an acceptor as it quenches the fluorescence. A shorter lifetime is observed for the donor when FRET occurs compared to donor emission alone without an acceptor. FLIM-FRET could be used to analyze the interaction of OC-17 with growing amorphous calcium carbonate clusters. The rationale for proposing this type of experiment is based on the fact that quenching should also occur if the donor interacts with the molecule. More specifically, if OC-17 interacts with amorphous calcium carbonate, then during fluorescence emission, labelled OC-17 might show a shorter lifetime compared to its own lifetime alone, i.e. in the absence of amorphous phase.

In order to probe the function of OC-17 further, it could be worth testing two-label FRET-

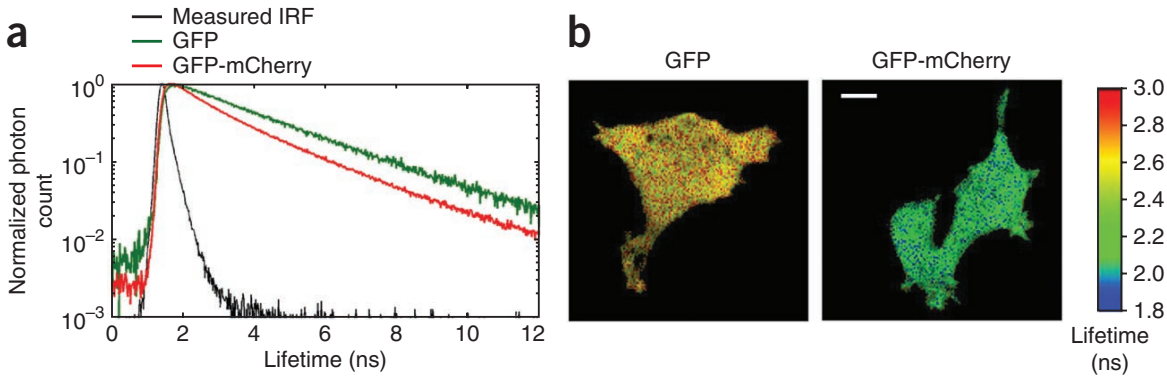


Figure 5.2: (a) The fluorescence lifetime measurement as a function of total photon count. GFP alone has a longer lifetime compared to the presence of an acceptor, GFP-mCherry. IRF denotes instrument response function which is used for correcting lifetime measurements. (b) The FLIM images show spatial distribution of lifetime value in cells, left, expressing GFP alone and right, FRET positive GFP-mCherry probe. Note that the color scale shows lifetime where it is longer for yellow-red compared to green. Not to be confused with green fluorescence of protein [234].

FLIM with not only labelling OC-17 but also growing $CaCO_3$ clusters. This could be done using a fluorescent dye that is known to label amorphous $CaCO_3$. Calcein is known to label amorphous $CaCO_3$ [250] and the dye requires calcium ions for its fluorescent activity. Because $CaCO_3$ growth is done with mixing Ca^{2+} and $CO_3{}^{2-}$ as it occurs in nature unlike in $(NH_4)_2CO_3$ diffusion method, the crystallization solution containing calcein could be used with labelled OC-17. The label of OC-17 can be chosen to be the donor and calcein to be the acceptor. A possible combination is C3 oxacyanine or Cy2. It should be noted that C3 oxacyanine is not commercially available as an NHS-ester and C3 oxacyanine was found to not label model proteins. Cy2 NHS-ester can be used and if there is an interaction between $CaCO_3$ and OC-17, then excitation of Cy2 labelled OC-17 would cause FRET. This experimental design could be an alternative to one-label FRET.

5.2.2 Genetic Fusion of OC-17 to Fluorescent Proteins or Vaccine Immunogens

Incorporation of proteins with using OC-17 as a chaperone could be a straightforward way to test incorporation ability of OC-17. This strategy, though, might be improved further by genetically fusing the target protein to OC-17. The recombinant expression of the target protein and OC-17 together can decrease the number of steps for testing. In other words, rather than

expressing and purifying OC-17 and conducting co-incorporation crystallization afterwards, it may be more desirable to express and purify target protein OC-17 fusion in one step. This method might also increase the incorporation efficiency because of the presence OC-17 of within the target protein.

Cloning and expression of OC-17 in *E. coli* was shown to be feasible. This means that the plasmid construct for OC-17 synthesis can be used to introduce a target protein gene to be fused with OC-17. Optimized conditions of OC-17 expression could allow the expression of target-OC-17 fusion under the similar growth conditions without further optimization. Cloning a fusion is thought to be straightforward because the insertion of the target gene could be conducted in one step in a PCR using Gibson assembly.

As mentioned in Chapter 4 (see the Discussion) it might be possible to test both the incorporation ability and protein structure/activity using a GFP-OC-17 fusion. If expression of this fusion is achieved, then it could allow analysis of GFP-OC-17 fusion for GPF fluorescent activity before and after incorporation into $CaCO_3$. The effect of reconstitution needs to be analyzed to ensure that EDTA does not cause loss of function to GFP activity. If the activity of GFP is retained after incorporation and reconstitution, the effect of heat-treatment can also be detected in a straightforward manner. GFP fused OC-17 could be the first step to move towards more complex vaccine immunogen fusions to OC-17.

5.2.3 Accelerated Protein Degradation Tests and Measurement of the Activity of Diphtheria Antibody

Intra-crystalline vaccine preservation for real-world applications needs to provide a heat stability at 45 °C (or above) for a 6-month period. If incorporation of the diphtheria antibody is achieved using OC-17 without compromising secondary and tertiary structures, then accelerated degradation studies could be done to assess antibody activity. Accelerated degradation tests allow vaccines to be tested for long term high temperature storage. It is possible to conduct diphtheria activity tests using commercially available kits yet the heat stability is crucial for the hypothesis presented in the present work so the accelerated degradation tests need to be conducted. It should be added that, the experiments proposed in this section (i.e accelerated degradation and *in vitro* antibody binding affinity tests) could be conducted at the WHO vaccine pre-qualification laboratory NIBSC if future work allows incorporation of diphtheria antibody with a high enough efficiency to be feasible for real world applications.

References

[1] N. Angier. *The Canon: The Beautiful Basics of Science*. Scribe Publications, 2008.

[2] Amanda Kay McVety. *The Rinderpest Campaigns: A Virus, Its Vaccines, and Global Development in the Twentieth Century*. Global and International History. Cambridge University Press, 2018.

[3] D.R. Hopkins. *The Greatest Killer: Smallpox in History*. University of Chicago Press, 2002.

[4] Thucydides. *History of the Peloponnesian War. Books I and II / Thucydides : translated by C.F. Smith.* Loeb classical library ; 108. Harvard University Press, Cambridge, MA, 2003.

[5] Woodward J. Timonius E. An account, or history, of the procuring the small pox by incision, or inoculation; as it has for some time been practised at constantinople. *Philosophical Transactions Of The Royal Society*, 29(339):72–82, 1714.

[6] Isobel Grundy. *Lady Mary Wortley Montagu / Isobel Grundy.* Oxford University Press, Oxford, 2001.

[7] J. Rhodes. *The End of Plagues: The Global Battle Against Infectious Disease*. MacSci. St. Martin's Press, 2013.

[8] Kai Kupferschmidt. Life-saving diphtheria drug is running out. *Science*, 355(6321):118–119, JAN 13 2017.

[9] M. Wadman. *The Vaccine Race: Science, Politics, and the Human Costs of Defeating Disease*. Penguin Publishing Group, 2017.

[10] D Baxby. Edward Jenner's inquiry; a bicentenary analysis. *Vaccine*, 17(4):301–307, JAN 28 1999.

[11] Edward Jenner. *An inquiry into the causes and effects of the variolae vacciae*. Sampson Low, London, 1798.

[12] N. Bacaer. *Daniel Bernoulli, d'Alembert and the inoculation of smallpox (1760). In: A Short History of Mathematical Population Dynamics*. Springer, London, 2011.

[13] T. Körner. *The Pleasures of Counting*. Cambridge University Press, 1996.

[14] WHO. Temperature sensitivity of vaccines. *Immunization, Vaccines and Biologicals. Geneva:World Health Organization/IVB/06.10*, 2006.

[15] John Lloyd and James Cheyne. The origins of the vaccine cold chain and a glimpse of the future. *Vaccine*, 35(17):2115–2120, APR 19 2017.

[16] K. Bartlett. *The Health of Nations: The Campaign to End Polio and Eradicate Epidemic Diseases*. Oneworld Publications, 2018.

[17] World Bank WHO, UNICEF. State of the world's vaccines and immunization,. *WHO Library Cataloguing-in-Publication Data, Geneva, 3rd ed.*, 2009.

[18] Ozan S. Kumru, Sangeeta B. Joshi, Dawn E. Smith, C. Russell Middaugh, Ted Prusik, and David B. Volkin. Vaccine instability in the cold chain: Mechanisms, analysis and formulation strategies. *Biologicals*, 42(5):237–259, SEP 2014.

[19] Beatrice Demarchi, Shaun Hall, Teresa Roncal-Herrero, Colin L. Freeman, Jos Woolley, Molly K. Crisp, Julie Wilson, Anna Fotakis, Roman Fischer, Benedikt M. Kessler, Rosa Rakownikow Jersie-Christensens, Jesper V. Olsen, James Haile, Jessica Thomas, Curtis W. Marean, John Parkington, Samantha Presslee, Julia Lee-Thorp, Peter Ditchfield, Jacqueline F. Hamilton, Martyn W. Ward, Chunting Michelle Wang, Marvin D. Shaw, Terry Harrison, Manuel Dominguez-Rodrigo, Ross D. E. MacPheel, Amandus Kwekason, Michaels Ecker, Liora Kolska Horwitz, Michael Chazan, Roland Kroger, Jane Thomas-Oates, John H. Harding, Enrico Cappellini, Kirsty Penkman, and Matthew J. Collins. Protein sequences bound to mineral surfaces persist into deep time. *Elife*, 5, SEP 27 2016.

[20] AS Brooks, PE Hare, JE Kokis, GH Miller, RD Ernst, and F Wendorf. Dating pleistocene archaeological sites by protein diagenesis in ostrich eggshell. *Science*, 248(4951):60–64, APR 6 1990.

[21] S Weiner, HA Lowenstam, and l Hood. Characterization of 80 million Year Old Mollusk Shell Proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 73(8):2541–2545, 1976.

[22] M Salamon, N Tuross, B Arensburg, and S Weiner. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13783–13788, SEP 27 2005.

[23] John R. M. Stewart, Richard B. Allen, Andrew K. G. Jones, Kirsty E. H. Penkman, and Matthew J. Collins. ZooMS: making eggshell visible in the archaeological record. *Journal of Archaeological Science*, 40(4):1797–1804, APR 2013.

[24] Charlotte L. Oskam, James Haile, Emma Mclay, Paul Rigby, Morten E. Allentoft, Maia E. Olsen, Camilla Bengtsson, Gifford H. Miller, Jean-Luc Schwenninger, Chris Jacomb, Richard Walter, Alexander Baynes, Joe Dortch, Michael Parker-Pearson, M. Thomas P. Gilbert, Richard N. Holdaway, Eske Willerslev, and Michael Bunce. Fossil avian eggshell preserves ancient DNA. *Proceedings Of The Royal Society Series B-Biological Sciences*, 277(1690):1991–2000, JUL 7 2010.

[25] PH Abelson. Amino acids in fossils. *Science*, 119(3096):576, 1954.

[26] GA Sykes, MJ Collins, and DI Walton. The significance of a geochemically isolated intracrystalline organic fraction within biominerals. *Organic Geochemistry*, 23(11-12):1059–1065, NOV-DEC 1995.

[27] Mary Higby Schweitzer, Elena R. Schroeter, and Michael B. Goshe. Protein Molecular Data from Ancient (> 1 million years old) Fossil Material: Pitfalls, Possibilities and Grand Challenges. *Analytical Chemistry*, 86(14):6731–6740, JUL 15 2014.

[28] GH Miller, PB Beaumont, HJ Deacon, AS Brooks, PE Hare, and AJT Jull. Earliest modern humans in southern Africa dated by isoleucine epimerization in ostrich eggshell. *Quaternary Science Reviews*, 18(13):1537–1548, 1999.

[29] Molly Crisp, Beatrice Demarchi, Matthew Collins, Michael Morgan-Williams, Emily Pilgrim, and Kirsty Penkman. Isolation of the intra-crystalline proteins and kinetic studies in Struthio camelus (ostrich) eggshell for amino acid geochronology. *Quaternary Geochronology*, 16(SI):110–128, APR 2013.

[30] DS Kaufman and GH Miller. Overview of Amino-Acid Geochronology. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology*, 102(2):199–204, JUN 1992.

[31] JP Reyes-Grajeda, A Moreno, and A Romero. Crystal structure of ovocleidin-17, a major protein of the calcified Gallus gallus eggshell - Implications in the calcite mineral growth pattern. *Journal of Biological Chemistry*, 279(39):40876–40881, SEP 24 2004.

[32] Rayana R. Ruiz-Arellano, Francisco J. Medrano, Abel Moreno, and Antonio Romero. Structure of struthiocalcin-1, an intramineral protein from Struthio camelus eggshell, in two crystal forms. *Acta Crystallographica Section D-Structural Biology*, 71(4):809–818, APR 2015.

[33] Colin L. Freeman, John H. Harding, David Quigley, and P. Mark Rodger. Structural Control of Crystal Nuclei by an Eggshell Protein. *Angewandte Chemie-International Edition*, 49(30):5135–5137, 2010.

[34] GE Fantner, E Oroudjev, G Schitter, LS Golde, P Thurner, MM Finch, P Turner, T Gutsmann, DE Morse, H Hansma, and PK Hansma. Sacrificial bonds and hidden length: Unraveling molecular mesostructures in tough materials. *Biophysical Journal*, 90(4):1411–1418, FEB 2006.

[35] J Gosline, M Lillie, E Carrington, P Guerette, C Ortlepp, and K Savage. Elastic proteins: biological roles and mechanical properties. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, 357(1418):121–132, FEB 28 2002.

[36] K Autumn, YA Liang, ST Hsieh, W Zesch, WP Chan, TW Kenny, R Fearing, and RJ Full. Adhesive force of a single gecko foot-hair. *Nature*, 405(6787):681–685, JUN 8 2000.

[37] H.A. Lowenstam and S. Weiner. *On Biomineralization*. Oxford University Press, 1989.

[38] A Berman, L Addadi, A Kvick, L Leiserowitz, M Nelson, and S Weiner. Intercalation of Sea-Urchin Proteins in Calcite - Study of a Crystalline Composite-Material. *Science*, 250(4981):664–667, NOV 2 1990.

[39] L Addadi and S Weiner. Interactions Between Acidic Proteins And Crystals - Stereochemical Requirements In Biomineralization. *Proceedings of the National Academy of Sciences of the United States of America*, 82(12):4110–4114, 1985.

[40] Fiona C. Meldrum and Helmut Coelfen. Controlling Mineral Morphologies and Structures in Biological and Synthetic Systems. *Chemical Reviews*, 108(11):4332–4432, NOV 2008.

[41] M Panheleux, M Bain, MS Fernandez, I Morales, J Gautron, JL Arias, SE Solomon, M Hincke, and Y Nys. Organic matrix composition and ultrastructure of eggshell: a comparative study. *British Poultry Science*, 40(2):240–252, MAY 1999.

[42] Louise N. Johnson. The structure and function of lysozyme. *Science Progress*, 54(215):367–385, 1966.

[43] Y Nys, MT Hincke, JL Arias, JM Garcia-Ruiz, and SE Solomon. Avian eggshell mineralization. *Poultry and Avian Biology Reviews*, 10(3):143–166, 1999.

[44] L Addadi, J Moradian, E Shay, NG Maroudas, and S Weiner. A Chemical Model For The Cooperation Of Sulfates And Carboxylates In Calcite Crystal Nucleation Relevance To Biomineralization. *Proceedings of the National Academy of Sciences of the United States of America*, 84(9):2732–2736, MAY 1987.

[45] Liliana Marin-Garcia, Bernardo A. Frontana-Uribe, Juan Pablo Reyes-Grajeda, Vivian Stojanoff, Hugo Javier Serrano-Posada, and Abel Moreno. Chemical recognition of carbonate anions by proteins involved in biomineralization processes and their influence on calcite crystal growth. *Crystal Growth and Design*, 8(4):1340–1345, APR 2008.

[46] K Mann and F Siedler. Ostrich (Struthio camelus) eggshell matrix contains two different C-type lectin-like proteins. Isolation, amino acid sequence, and posttranslational modifications. *Biochimica Et Biophysica Acta-Proteins And Proteomics*, 1696(1):41–50, JAN 14 2004.

[47] K Mann and F Siedler. Amino acid sequences and phosphorylation sites of emu and rhea eggshell C-type lectin-like proteins. *Comparative Biochemistry and Physiology B-Biochemistry and Molecular Biology*, 143(2):160–170, FEB 2006.

[48] RZ Wang, Z Suo, AG Evans, N Yao, and IA Aksay. Deformation mechanisms in nacre. *Journal of Materials Research*, 16(9):2485–2493, SEP 2001.

[49] AG Evans, Z Suo, RZ Wang, IA Aksay, MY He, and JW Hutchinson. Model for the robust mechanical behavior of nacre. *Journal of Materials Research*, 16(9):2475–2484, SEP 2001.

[50] CM Zaremba, AM Belcher, M Fritz, YL Li, S Mann, PK Hansma, DE Morse, JS Speck, and GD Stucky. Critical transitions in the biofabrication of abalone shells and flat pearls. *Chemistry of Materials*, 8(3):679–690, MAR 1996.

[51] Nils Kroeger. The Molecular Basis of Nacre Formation. *Science*, 325(5946):1351–1352, SEP 11 2009.

[52] Ulrike G. K. Wegst, Hao Bai, Eduardo Saiz, Antoni P. Tomsia, and Robert O. Ritchie. Bioinspired structural materials. *Nature Materials*, 14(1):23–36, JAN 2015.

[53] PL Davies and CL Hew. Biochemistry of Fish Antifreeze Proteins. *Faseb Journal*, 4(8):2460–2468, MAY 1990.

[54] Y Yeh and RE Feeney. Antifreeze proteins: Structures and mechanisms of function. *Chemical Reviews*, 96(2):601–617, MAR-APR 1996.

[55] K Drickamer. C-type lectin-like domains. *Current Opinion in Structural Biology*, 9(5):585–590, OCT 1999.

[56] AN Zelensky and JE Gready. The C-type lectin-like domain superfamily. *Febs Journal*, 272(24):6179–6217, DEC 2005.

[57] Rachel Gueta, Amir Natan, Lia Addadi, Steve Weiner, Keith Refson, and Leeor Kronik. Local atomic order and infrared spectra of biogenic calcite. *Angewandte Chemie-International Edition*, 46(1-2):291–294, 2007.

[58] A.L. Peck. *Generation of Animals: With an English Translation by A.L. Peck*. The Loeb classical library. Harvard University Press, 1953.

[59] W. Harvey and G. Whitteridge. *Disputations Touching the Generation of Animals*. Blackwell Scientific Publications, 1981.

[60] T. Birkhead. *The Most Perfect Thing: Inside (and Outside) a Bird's Egg*. Bloomsbury Publishing, 2016.

[61] Claire N. Spottiswoode. The most perfect thing, explained. *Science*, 356(6344):1234–1235, JUN 23 2017.

[62] M.E. Hauber, J. Bates, B. Becker, and J.S. Weinstein. *The Book of Eggs: A Lifesize Guide to the Eggs of Six Hundred of the World's Bird Species*. Book of Series. Ivy Press, 2014.

[63] David J. Varricchio, Jason R. Moore, Gregory M. Erickson, Mark A. Norell, Frankie D. Jackson, and John J. Borkowski. Avian Paternal Care Had Dinosaur Origin. *Science*, 322(5909):1826–1828, DEC 19 2008.

[64] Graeme D. Ruxton, Geoffrey F. Birchard, and D. Charles Deeming. Incubation time as an important influence on egg production and distribution into clutches for sauropod dinosaurs. *Paleobiology*, 40(3):323–330, SUM 2014.

[65] Mary Caswell Stoddard, Ee Hou Yong, Derya Akkaynak, Catherine Sheard, Joseph A. Tobias, and L. Mahadevan. Avian egg shape: Form, function, and evolution. *Science*, 356(6344):1249+, JUN 23 2017.

[66] R.W. Burley and D.V. Vadehra. *The avian egg: Chemistry and Biology*. Wiley-Interscience publications. Wiley, 1989.

[67] Dominic Vella, Amin Ajdari, Ashkan Vaziri, and Arezki Boudaoud. Indentation of Ellipsoidal and Cylindrical Elastic Shells. *Physical Review Letters*, 109(14), OCT 5 2012.

[68] Maxwell T. Hincke, Yves Nys, Joel Gautron, Karlheinz Mann, Alejandro B. Rodriguez-Navarro, and Marc D. McKee. The eggshell: structure, composition and mineralization. *Frontiers in Bioscience-Landmark*, 17:1266–1280, JAN 1 2012.

[69] LW Hillier, W Miller, E Birney, W Warren, RC Hardison, CP Ponting, P Bork, DW Burt, MAM Groenen, ME Delany, JB Dodgson, AT Chinwalla, PF Cliften, SW Clifton, KD Delehaunty, C Fronick, RS Fulton, TA Graves, C Kremitzki, D Layman, V Magrini, JD McPherson, TL Miner, P Minx, WE Nash, MN Nhan, JO Nelson, LG Oddy, CS Pohl, J Randall-Maher, SM Smith, JW Wallis, SP Yang, MN Romanov, CM Rondelli, B Paton, J Smith, D Morrice, L Daniels, HG Tempest, L Robertson, JS Masabanda, DK Griffin, A Vignal, V Fillon, L Jacobbson, S Kerje, L Andersson, RPM Crooijmans, J Aerts, JJ van der Poel, H Ellegren, RB Caldwell, SJ Hubbard, DV Grafham, AM Kierzek, SR McLaren, IM Overton, H Arakawa, KJ Beattie, Y Bezzubov, PE Boardman, JK Bonfield, MDR Croning, RM Davies, MD Francis, SJ Humphray, CE Scott, RG Taylor, C Tickle, WRA Brown, J Rogers, JM Buerstedde, SA Wilson, L Stubbs, I Ovcharenko, L Gordon, S Lucas, MM Miller, H Inoko, T Shiina, J Kaufman, J Salomonsen, K Skjoedt, GKS Wong, J Wang, B Liu, J Wang, J Yu, HM Yang, M Nefedov, M Koriabine, PJ deJong, L Goodstadt, C Webber, NJ Dickens, I Letunic, M Suyama, D Torrents, C von Mering, EM Zdobnov, K Makova, A Nekrutenko, L Elnitski, P Eswara, DC King, S Yang, S Tyekucheva, A Radakrishnan, RS Harris, F Chiaromonte, J Taylor, JB He, M Rijnkels, S Griffiths-Jones, A Ureta-Vidal, MM Hoffman, J Severin, SMJ Searle, AS Law, D Speed, D Waddington, Z Cheng, E Tuzun, E Eichler, ZR Bao, P Flicek, DD Shteynberg, MR Brent, JM Bye, EJ Huckle, S Chatterji, C Dewey, L Pachter, A Kouranov, Z Mourelatos, AG Hatzigeorgiou, AH Paterson, R Ivarie, M Brandstrom, E Axelsson, N Backstrom, S Berlin, MT Webster, O Pourquie, A Reymond, C Ucla, SE Antonarakis, MY Long, JJ Emerson, E Betran, I Dupanloup, H Kaessmann, AS Hinrichs, G Bejerano, TS Furey, RA Harte,

B Raney, A Siepel, WJ Kent, D Haussler, E Eyras, R Castelo, JF Abril, S Castellano, F Camara, G Parra, R Guigo, G Bourque, G Tesler, PA Pevzner, A Smit, LA Fulton, ER Mardis, and RK Wilson. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018):695–716, DEC 9 2004.

[70] Wesley C. Warren, David F. Clayton, Hans Ellegren, Arthur P. Arnold, LaDeana W. Hillier, Axel Kuenstner, Steve Searle, Simon White, Albert J. Vilella, Susan Fairley, Andreas Heger, Lesheng Kong, Chris P. Ponting, Erich D. Jarvis, Claudio V. Mello, Pat Minx, Peter Lovell, Tarciso A. F. Velho, Margaret Ferris, Christopher N. Balakrishnan, Saurabh Sinha, Charles Blatti, Sarah E. London, Yun Li, Ya-Chi Lin, Julia George, Jonathan Sweedler, Bruce Southey, Preethi Gunaratne, Michael Watson, Kiwoong Nam, Niclas Backstroem, Linnea Smeds, Benoit Nabholz, Yuichiro Itoh, Osceola Whitney, Andreas R. Pfenning, Jason Howard, Martin Voelker, Bejamin M. Skinner, Darren K. Griffin, Liang Ye, William M. McLaren, Paul Flicek, Victor Quesada, Gloria Velasco, Carlos Lopez-Otin, Xose S. Puente, Tsviya Olender, Doron Lancet, Arian F. A. Smit, Robert Hubley, Miriam K. Konkel, Jerilyn A. Walker, Mark A. Batzer, Wanjun Gu, David D. Pollock, Lin Chen, Ze Cheng, Evan E. Eichler, Jessica Stapley, Jon Slate, Robert Ekblom, Tim Birkhead, Terry Burke, David Burt, Constance Scharff, Iris Adam, Hugues Richard, Marc Sultan, Alexey Soldatov, Hans Lehrach, Scott V. Edwards, Shiaw-Pyng Yang, XiaoChing Li, Tina Graves, Lucinda Fulton, Joanne Nelson, Asif Chinwalla, Shunfeng Hou, Elaine R. Mardis, and Richard K. Wilson. The genome of a songbird. *Nature*, 464(7289):757–762, APR 1 2010.

[71] Y Nys, J Gautron, JM Garcia-Ruiz, and MT Hincke. Avian eggshell mineralization: biochemical and functional characterization of matrix proteins. *Comptes Rendus Palevol*, 3(6-7):549–562, OCT 2004.

[72] Matej Balaz. Eggshell membrane biomaterial as a platform for applications in materials science. *Acta Biomaterialia*, 10(9, SI):3827–3843, SEP 2014.

[73] Y Nys, J Zawadzki, J Gautron, and AD Mills. Whitening Of Brown-Shelled Eggs - Mineral-Composition Of Uterine Fluid And Rate Of Protoporphyrin Deposition. *Poultry Science*, 70(5):1236–1245, MAY 1991.

[74] JL Arias, MS Fernandez, JE Dennis, and AI Caplan. Collagens Of The Chicken Eggshell Membranes. *Connective Tissue Research*, 26(1-2):37–45, 1991.

[75] Kritsda Kaweewong, Wunwiboon Garnjanagoonchorn, Wannee Jirapakkul, and Sittiruk

Roytrakul. Solubilization and Identification of Hen Eggshell Membrane Proteins During Different Times of Chicken Embryo Development Using the Proteomic Approach. *Protein Journal*, 32(4):297–308, APR 2013.

[76] DJ Fink, AI Caplan, and AH Heuer. Eggshell Mineralization - A Case-Study Of A Bioprocessing Strategy. *Mrs Bulletin*, 17(10):27–31, OCT 1992.

[77] Patrycja Nowak-Sliwinska, Tatiana Segura, and M. Luisa Iruela-Arispe. The chicken chorioallantoic membrane model in biology, medicine and bioengineering. *Angiogenesis*, 17(4):779–804, OCT 2014.

[78] Bernd Bayerlein, Paul Zaslansky, Yannicke Dauphin, Alexander Rack, Peter Fratzl, and Igor Zlotnikov. Self-similar mesostructure evolution of the growing mollusc shell reminiscent of thermodynamically driven grain growth. *Nature Materials*, 13(12):1102–1107, DEC 2014.

[79] J Gautron, MT Hincke, and Y Nys. Precursor matrix proteins in the uterine fluid change with stages of eggshell formation in hens. *Connective Tissue Research*, 36(3):195–210, 1997.

[80] J Gautron, MT Hincke, K Mann, M Panheleux, M Bain, MD McKee, SE Solomon, and Y Nys. Ovocalyxin-32, a novel chicken eggshell matrix protein - Isolation, amino acid sequencing, cloning, and immunocytochemical localization. *Journal of Biological Chemistry*, 276(42):39243–39252, OCT 19 2001.

[81] JL Arias, DJ Fink, SQ Xiao, AH Heuer, and AI Caplan. Biomineralization And Eggshells - Cell-Mediated Acellular Compartments Of Mineralized Extracellular-Matrix. In *International Review Of Cytology - A Survey Of Cell Biology*, volume 145 of *International Review Of Cytology-A Survey Of Cell Biology*, pages 217–250. Academic Press Inc, 1993.

[82] Megan Rose-Martel, Jingwen Du, and Maxwell T. Hincke. Proteomic analysis provides new insight into the chicken eggshell cuticle. *Journal Of Proteomics*, 75(9):2697–2706, MAY 17 2012.

[83] L. North, D. Labonte, M. L. Oyen, M. P. Coleman, H. B. Caliskan, and R. E. Johnston. Interrelated chemical-microstructural-nanomechanical variations in the structural units of the cuttlebone of Sepia officinalis. *APL Materials*, 5(11), NOV 2017.

[84] S. Nedomova, L. Severa, and J. Buchar. Influence of hen egg shape on eggshell compressive strength. *International Agrophysics*, 23(3):249–256, 2009.

[85] M. Weller, T. Overton, F. Armstrong, and J. Rourke. *Inorganic Chemistry*. Oxford University Press, 2018.

[86] GM Bond, VD Scott, and RG Board. Correlation of Mechanical Properties of Avian Eggshells with Hatching Strategies. *Journal of Zoology*, 209(2):225–237, JUN 1986.

[87] S. James Reynolds, Graham R. Martin, and Phillip Cassey. Is sexual selection blurring the functional significance of eggshell coloration hypotheses? *Animal Behaviour*, 78(1):209–215, JUL 2009.

[88] A. Lazarus, H. C. B. Florijn, and P. M. Reis. Geometry-Induced Rigidity in Nonspherical Pressurized Elastic Shells. *Physical Review Letters*, 109(14), OCT 5 2012.

[89] Alice Nasto, Amin Ajdari, Arnaud Lazarus, Ashkan Vaziri, and Pedro M. Reis. Localization of deformation in thin shells under indentation. *Soft Matter*, 9(29):6796–6803, 2013.

[90] Kristin M. Poduska, Lior Regev, Elisabetta Boaretto, Lia Addadi, Steve Weiner, Leeor Kronik, and Stefano Curtarolo. Decoupling Local Disorder and Optical Effects in Infrared Spectra: Differentiating Between Calcites with Different Origins. *Advanced Materials*, 23(4):550+, JAN 25 2011.

[91] Nico A. J. M. Sommerdijk and Gijsbertus de With. Biomimetic CaCO3 Mineralization using Designer Molecules and Interfaces. *Chemical Reviews*, 108(11):4499–4550, NOV 2008.

[92] J. J. Kruzic, D. K. Kim, K. J. Koester, and R. O. Ritchie. Indentation techniques for evaluating the fracture toughness of biomaterials and hard tissues. *Journal of the Mechanical Behavior of Biomedical Materials*, 2(4):384–395, AUG 2009.

[93] B de Ketelaere, T Govaerts, P Coucke, E Dewil, J Visscher, E Decuypere, and J De Baerdemaeker. Measuring the eggshell strength of 6 different genetic strains of laying hens: techniques and comparisons. *British Poultry Science*, 43(2):238–244, MAY 2002.

[94] P Coucke, E Dewil, E Decuypere, and J De Baerdemaeker. Measuring the mechanical stiffness of an eggshell using resonant frequency analysis. *British Poultry Science*, 40(2):227–232, MAY 1999.

[95] H Lin, K Mertens, B Kemps, T Govaerts, B De Ketelaere, J De Baerdemaeker, E Decuypere, and J Buyse. New approach of testing the effect of heat stress on eggshell

quality: mechanical and material properties of eggshell and membrane. *British Poultry Science*, 45(4):476–482, AUG 2004.

[96] A Rodriguez-Navarro, O Kalin, Y Nys, and JM Garcia-Ruiz. Influence of the microstructure on the shell strength of eggs laid by hens of different ages. *British Poultry Science*, 43(3):395–403, JUL 2002.

[97] AMH Ahmed, AB Rodriguez-Navarro, ML Vidal, J Gautron, JM Garcia-Ruiz, and Y Nys. Changes in eggshell mechanical properties, crystallographic texture and in matrix proteins induced by moult in hens. *British Poultry Science*, 46(3):268–279, JUN 2005.

[98] Libor Severa, Jiri Nemecek, Sarka Nedomova, and Jaroslav Buchar. Determination of micromechanical properties of a hen's eggshell by means of nanoindentation. *Journal of Food Engineering*, 101(2):146–151, NOV 2010.

[99] Dimitra Athanasiadou, Wenge Jiang, Dina Goldbaum, Aroba Saleem, Kaustuv Basu, Michael S. Pacella, Corinna F. Boehm, Richard R. Chromik, Maxwell T. Hincke, Alejandro B. Rodriguez-Navarro, Hojatollah Vali, Stephan E. Wolf, Jeffrey J. Gray, Khanh Huy Bui, and Marc D. McKee. Nanostructure, osteopontin, and mechanical properties of calcitic avian eggshell. *Science Advances*, 4(3), MAR 2018.

[100] Stephan E. Wolf, Corinna F. Boehm, Joe Harris, Benedikt Demmert, Dorrit E. Jacob, Mihail Mondeshki, Encarnacion Ruiz-Agudo, and Carlos Rodriguez-Navarro. Nonclassical crystallization in vivo et in vitro (I): Process-structure-property relationships of nanogranular biominerals. *Journal of Structural Biology*, 196(2, SI):244–259, NOV 2016. 13th International Biomineralization Symposium, Granada, Spain, Sep 16-19, 2015.

[101] Carlos Rodriguez-Navarro, Encarnacion Ruiz-Agudo, Joe Harris, and Stephan E. Wolf. Nonclassical crystallization in vivo et in vitro (II): Nanogranular features in biomimetic minerals disclose a general colloid-mediated crystal growth mechanism. *Journal of Structural Biology*, 196(2, SI):260–287, NOV 2016. 13th International Biomineralization Symposium, Granada, Spain, Sep 16-19, 2015.

[102] M Pines, V Knopov, and A Bar. Involvement of osteopontin in egg shell formation in the laying chicken. *Matrix Biology*, 14(9):765–771, DEC 1995.

[103] JE Dennis, SQ Xiao, M Agarwal, DJ Fink, AH Heuer, and AI Caplan. Microstructure of matrix and mineral components of eggshells from white leghorn chickens (Gallus gallus). *Journal of Morphology*, 228(3):287–306, JUN 1996.

[104] DT Denhardt and XJ Guo. Osteopontin - a Protein with Diverse Functions. *Faseb Journal*, 7(15):1475–1482, DEC 1993.

[105] N Hansen. Hall-Petch relation and boundary strengthening. *Scripta Materialia*, 51(8):801–806, OCT 2004.

[106] Karlheinz Mann, Boris Macek, and Jesper V. Olsen. Proteomic analysis of the acid-soluble organic matrix of the chicken calcified eggshell layer. *Proteomics*, 6(13):3801–3810, JUL 2006.

[107] Karlheinz Mann and Matthias Mann. The proteome of the calcified layer organic matrix of turkey (Meleagris gallopavo) eggshell. *Proteome Science*, 11, AUG 27 2013.

[108] Congjiao Sun, Guiyun Xu, and Ning Yang. Differential label-free quantitative proteomic analysis of avian eggshell matrix and uterine fluid proteins associated with eggshell mechanical property. *Proteomics*, 13(23-24):3523–3536, DEC 2013.

[109] MT Hincke. Ovalbumin is a component of the chicken eggshell matrix. *Connective Tissue Research*, 31(3):227–233, 1995.

[110] Y. C. Chien, M. T. Hincke, H. Vali, and M. D. McKee. Ultrastructural matrix-mineral relationships in avian eggshell, and effects of osteopontin on calcite growth in vitro. *Journal of Structural Biology*, 163(1):84–99, JUL 2008.

[111] MT Hincke, CPW Tsang, M Courtney, V Hill, and R Narbaitz. Purification and Immunochemistry of a Soluble Matrix Protein of the Chicken Eggshell (Ovocleidin-17). *Calcified Tissue International*, 56(6):578–583, JUN 1995.

[112] Joel Gautron, Emi Murayama, Alain Vignal, Mireille Morisson, Marc D. McKee, Sophie Rehault, Valerie Labas, Maya Belghazi, Mary-Laure Vidal, Yves Nys, and Maxwell T. Hincke. Cloning of ovocalyxin-36, a novel chicken eggshell protein related to lipopolysaccharide-binding proteins, bactericidal permeability-increasing proteins, and plunc family proteins. *Journal of Biological Chemistry*, 282(8):5273–5286, FEB 23 2007.

[113] K Mann, MT Hincke, and Y Nys. Isolation of ovocleidin-116 from chicken eggshells, correction of its amino acid sequence and identification of disulfide bonds and glycosylated Asn. *Matrix Biology*, 21(5):383–387, AUG 2002.

[114] MT Hincke, J Gautron, CPW Tsang, MD McKee, and Y Nys. Molecular cloning and ultrastructural localization of the core protein of an eggshell matrix proteoglycan, ovocleidin-116. *Journal of Biological Chemistry*, 274(46):32915–32923, NOV 12 1999.

[115] Karlheinz Mann, Jesper V. Olsen, Boris Macek, Florian Gnad, and Matthias Mann. Phosphoproteins of the chicken eggshell calcified layer. *Proteomics*, 7(1):106–115, JAN 2007.

[116] Claire Bardet, Sidney Delgado, and Jean-Yves Sire. MEPE evolution in mammals reveals regions and residues of prime functional importance. *Cellular and Molecular Life Sciences*, 67(2):305–320, JAN 2010.

[117] I. C. Dunn, N. T. Joseph, M. Bain, A. Edmond, P. W. Wilson, P. Milona, Y. Nys, J. Gautron, M. Schmutz, R. Preisinger, and D. Waddington. Polymorphisms in eggshell organic matrix genes are associated with eggshell quality measurements in pedigree Rhode Island Red hens. *Animal Genetics*, 40(1):110–114, FEB 2009.

[118] MT Hincke, J Gautron, K Mann, M Panheleux, MA McKee, M Bain, SE Solomon, and Y Nys. Purification of ovocalyxin-32, a novel chicken eggshell matrix protein. *Connective Tissue Research*, 44(1):16–19, 2003. 7th International Conference on the Chemistry and Biology of Mineralized Tissues, Ponte Vedra, FL, Nov 04-09, 2001.

[119] Cristianne M. M. Cordeiro, Hamed Esmaili, George Ansah, and Maxwell T. Hincke. Ovocalyxin-36 Is a Pattern Recognition Protein in Chicken Eggshell Membranes. *Plos One*, 8(12), DEC 31 2013.

[120] Maxwell T. Hincke, Yung-Ching Chien, Louis C. Gerstenfeld, and Marc D. McKee. Colloidal-gold immunocytochemical localization of osteopontin in avian eggshell gland and eggshell. *Journal of Histochemistry and Cytochemistry*, 56(5):467–476, MAY 2008.

[121] M Pines, V Knopov, and A Bar. Involvement of osteopontin in egg shell formation in the laying chicken. *Matrix Biology*, 14(9):765–771, DEC 1995.

[122] H Lis and N Sharon. Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chemical Reviews*, 98(2):637–674, MAR-APR 1998.

[123] E Garcia Hernandez, RA Zubillaga, A Rojo Dominguez, A Rodriguez Romero, and A Hernandez Arana. New insights into the molecular basis of lectin-carbohydrate interactions: A calorimetric and structural study of the association of hevein to oligomers

of N-acetylglucosamine. *Proteins-Structure Function And Genetics*, 29(4):467–477, DEC 1997.

[124] WI Weis, ME Taylor, and K Drickamer. The C-type lectin superfamily in the immune system. *Immunological Reviews*, 163:19–34, JUN 1998.

[125] BG Fry and W Wuster. Assembling an arsenal: Origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Molecular Biology And Evolution*, 21(5):870–883, MAY 2004.

[126] W Gronwald, MC Loewen, B Lix, AJ Daugulis, FD Sonnichsen, PL Davies, and BD Sykes. The solution structure of type II antifreeze protein reveals a new member of the lectin family. *Biochemistry*, 37(14):4712–4721, APR 7 1998.

[127] KV Ewart, ZJ Li, DSC Yang, GL Fletcher, and CL Hew. The ice-binding site of Atlantic herring antifreeze protein corresponds to the carbohydrate-binding site of C-type lectins. *Biochemistry*, 37(12):4080–4085, MAR 24 1998.

[128] L Patard, V Stoven, B Gharib, F Bontems, JY Lallemand, and M DeReggi. What function for human lithostathine? Structural investigations by three-dimensional structure modeling and high-resolution NMR spectroscopy. *Protein Engineering*, 9(11):949–957, NOV 1996.

[129] D Bimmler, R Graf, GA Scheele, and TW Frick. Pancreatic stone protein (lithostathine), a physiologically relevant pancreatic calcium carbonate crystal inhibitor? *Journal of Biological Chemistry*, 272(5):3073–3082, JAN 31 1997.

[130] JA Bertrand, D Pignol, JP Bernard, JM Verdier, JC Dagorn, and JC FontecillaCamps. Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation. *Embo Journal*, 15(11):2678–2684, JUN 3 1996.

[131] M De Reggi, B Gharib, L Patard, and V Stoven. Lithostathine, the presumed pancreatic stone inhibitor, does not interact specifically with calcium carbonate crystals. *Journal of Biological Chemistry*, 273(9):4967–4971, FEB 27 1998.

[132] G Krampitz and G Graser. Molecular Mechanisms Of Biomineralization In The Formation Of Calcified Shells. *Angewandte Chemie-International Edition*, 27(9):1145–1156, SEP 1988.

[133] K Mann and F Siedler. The amino acid sequence of ovocleidin 17, a major protein of the avian eggshell calcified layer. *Biochemistry And Molecular Biology International*, 47(6):997–1007, JUN 1999.

[134] R Lakshminarayanan, JS Joseph, RM Kini, and S Valiyaveettil. Structure-function relationship of avian eggshell matrix proteins: A comparative study of two major eggshell matrix proteins, ansocalcin and OC-17. *Biomacromolecules*, 6(2):741–751, MAR-APR 2005.

[135] Colin L. Freeman, John H. Harding, David Quigley, and P. Mark Rodger. Simulations of Ovocleidin-17 Binding to Calcite Surfaces and Its Implications for Eggshell Formation. *Journal Of Physical Chemistry C*, 115(16):8175–8183, APR 28 2011.

[136] K Mann. Isolation of a glycosylated form of the chicken eggshell protein ovocleidin and determination of the glycosylation site. Alternative glycosylation/phosphorylation at an N-glycosylation sequon. *Febs Letters*, 463(1-2):12–14, DEC 10 1999.

[137] M.E. Taylor and K. Drickamer. *Introduction to Glycobiology*. Oxford University Press, 2006.

[138] M Karplus and JA McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, SEP 2002.

[139] JA Mccammon, BR Gelin, and M Karplus. Dynamics Of Folded Proteins. *Nature*, 267(5612):585–590, 1977.

[140] A Laio and M Parrinello. Escaping free-energy minima. *Proceedings of The National Academy Of Sciences Of The United States Of America*, 99(20):12562–12566, OCT 1 2002.

[141] L Addadi, S Raz, and S Weiner. Taking advantage of disorder: Amorphous calcium carbonate and its roles in biomineralization. *Advanced Materials*, 15(12):959–970, JUN 17 2003.

[142] S Weiner, I Sagi, and L Addadi. Choosing the crystallization path less traveled. *Science*, 309(5737):1027–1028, AUG 12 2005.

[143] Colin L. Freeman, John H. Harding, David Quigley, and P. Mark Rodger. How does an amorphous surface influence molecular binding? - ovocleidin-17 and amorphous calcium carbonate. *Physical Chemistry Chemical Physics*, 17(26):17494–17500, 2015.

[144] Paolo Raiteri and Julian D. Gale. Water Is the Key to Nonclassical Nucleation of Amorphous Calcium Carbonate. *Journal of the American Chemical Society*, 132(49):17623–17634, DEC 15 2010.

[145] Colin L. Freeman, John H. Harding, David Quigley, and P. Mark Rodger. Protein binding on stepped calcite surfaces: simulations of ovocleidin-17 on calcite {31.16} and {31.8}. *Physical Chemistry Chemical Physics*, 14(20):7287–7295, 2012.

[146] J. Keeler and P. Wothers. *Why Chemical Reactions Happen*. OUP Oxford, 2003.

[147] Laurie B. Gower. Biomimetic Model Systems for Investigating the Amorphous Precursor Pathway and Its Role in Biomineralization. *Chemical Reviews*, 108(11):4551–4627, NOV 2008.

[148] Gordon D. Brown, Janet A. Willment, and Lauren Whitehead. C-type lectins in immunity and homeostasis. *Nature Reviews Immunology*, 18(6):374–389, JUN 2018.

[149] Quan Zhang, Long Liu, Feng Zhu, ZhongHua Ning, Maxwell T. Hincke, Ning Yang, and ZhuoCheng Hou. Integrating De Novo Transcriptome Assembly and Cloning to Obtain Chicken Ovocleidin-17 Full-Length cDNA. *Plos One*, 9(3), MAR 27 2014.

[150] Adam C. English, Stephen Richards, Yi Han, Min Wang, Vanesa Vee, Jiaxin Qu, Xiang Qin, Donna M. Muzny, Jeffrey G. Reid, Kim C. Worley, and Richard A. Gibbs. Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *Plos One*, 7(11), NOV 21 2012.

[151] B. Lewin. *Genes 8*. Pearson Prentice Hall, 2004.

[152] A. Hernandez-Hernandez, M. L. Vidal, J. Gomez-Morales, A. B. Rodriguez-Navarro, V. Labas, J. Gautron, Y. Nys, and J. M. Garcia Ruiz. Influence of eggshell matrix proteins on the precipitation of calcium carbonate (CaCO3). *Journal Of Crystal Growth*, 310(7-9, SI):1754–1759, APR 2008. 15th International Conference on Crystal Growth, Salt Lake City, UT, AUG 12-17, 2007.

[153] Wonhee Lee, Warren Fon, Blake W. Axelrod, and Michael L. Roukes. High-sensitivity microfluidic calorimeters for biological and chemical applications. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36):15225–15230, SEP 8 2009.

[154] Dipika M. Matthias, Joanie Robertson, Michelle M. Garrison, Sophie Newland, and Carib Nelson. Freezing temperatures in the vaccine cold chain: A systematic literature review. *Vaccine*, 25(20):3980–3986, MAY 16 2007.

[155] John Lloyd, Patrick Lydon, Ramzi Ouhichi, and Michel Zaffran. Reducing the loss of vaccines from accidental freezing in the cold chain: The experience of continuous temperature monitoring in Tunisia. *Vaccine*, 33(7):902–907, Feb 11 2015.

[156] Louise C. Ivers. Eliminating Cholera Transmission in Haiti. *New England Journal of Medicine*, 376(2):101–103, JAN 12 2017.

[157] Pascal Leuenberger, Stefan Ganscha, Abdullah Kahraman, Valentina Cappelletti, Paul J. Boersema, Christian von Mering, Manfred Claassen, and Paola Picotti. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*, 355(6327):812+, FEB 24 2017.

[158] MY Sherman and AL Goldberg. Cellular defenses against unfolded proteins: A cell biologist thinks about neurodegenerative diseases. *Neuron*, 29(1):15–32, JAN 2001.

[159] LJ Rothschild and RL Mancinelli. Life in extreme environments. *Nature*, 409(6823):1092–1101, FEB 22 2001.

[160] KO Stetter. Extremophiles and their adaptation to hot environments. *Febs Letters*, 452(1-2):22–25, JUN 4 1999.

[161] TD Brock and H Freeze. Thermus Aquaticus Gen N And Sp N A Nonsporulating Extreme Thermophile. *Journal Of Bacteriology*, 98(1):289–&, 1969.

[162] Bernadette Modell and Matthew Darlison. Global epidemiology of haemoglobin disorders and derived service indicators. *Bulletin of the World Health Organization*, 86(6):480–487, Jun 2008.

[163] CB Anfinsen. Principles that Govern Folding of Protein Chains. *Science*, 181(4096):223–230, 1973.

[164] D.T. Haynie. *Biological Thermodynamics*. Cambridge University Press, 2008.

[165] Dexiang Chen and Debra Kristensen. Opportunities and challenges of developing thermostable vaccines. *Expert Review of Vaccines*, 8(5):547–557, MAY 2009.

[166] LaToya Jones Braun, Anil Tyagi, Shalimar Perkins, John Carpenter, David Sylvester, Mark Guy, Debra Kristensen, and Dexiang Chen. Development of a freeze-stable formulation for vaccines containing aluminum salt adjuvants. *Vaccine*, 27(1):72–79, JAN 1 2009.

[167] Simona Zipursky, Liliane Boualam, Dah Ould Cheikh, Jacqueline Fournier-Caruana, Djabar Hamid, Mathias Janssen, Umit Kartoglu, Genevieve Waeterloos, and Olivier Ronveaux. Assessing the potency of oral polio vaccine kept outside of the cold chain during a national immunization campaign in Chad. *Vaccine*, 29(34):5652–5656, AUG 5 2011.

[168] Ariane Halm, Idrissa Yalcouye, Mady Kamissoko, Tenemakan Keita, Ndoutabe Modjirom, Simona Zipursky, Umit Kartoglu, and Olivier Ronveaux. Using oral polio vaccine beyond the cold chain: A feasibility study conducted during the national immunization campaign in Mali. *Vaccine*, 28(19):3467–3472, APR 26 2010.

[169] Lara J. Wolfson, Francois Gasse, Shook-Pui Lee-Martin, Patrick Lydon, Ahmed Magan, Abdelmajid Tibouti, Benjamin Johns, Raymond Hutubessya, Peter Salama, and Jean-Marie Okwo-Bele. Estimating the costs of achieving the WHO-UNICEF Global Immunization Vision and Strategy, 2006-2015. *Bulletin of the World Health Organization*, 86(1):27–39, JAN 2008.

[170] PATH Vaccine and Pharmaceutical Technologies Group. Summary of stability data for licensed vaccines. *Reported by Working in Tandem, Ltd. for the PATH*, 2012.

[171] SN Ho, HD Hunt, RM Horton, JK Pullen, and LR Pease. Site-directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction. *Gene*, 77(1):51–59, APR 15 1989.

[172] BI Dahiyat and SL Mayo. De novo protein design: Fully automated sequence selection. *Science*, 278(5335):82–87, OCT 3 1997.

[173] JC Lee and SN Timasheff. The Stabilization of Proteins by Sucrose. *Journal of Biological Chemistry*, 256(14):7193–7201, 1981.

[174] N. Grasmeijer, M. Stankovic, H. de Waard, H. W. Frijlink, and W. L. J. Hinrichs. Unraveling protein stabilization mechanisms: Vitrification and water replacement in a glass transition temperature controlled system. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 1834(4):763–769, APR 2013.

[175] Maarten A. Mensink, Henderik W. Frijlink, Kees van der Voort Maarschalk, and Wouter L. J. Hinrichs. How sugars protect proteins in the solid state and during drying (review): Mechanisms of stabilization in relation to stress conditions. *European Journal of Pharmaceutics and Biopharmaceutics*, 114:288–295, MAY 2017.

[176] SD Allison, B Chang, TW Randolph, and JF Carpenter. Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Archives of Biochemistry and Biophysics*, 365(2):289–298, MAY 15 1999.

[177] J.P. Hansen and I.R. McDonald. *Theory of Simple Liquids*. Elsevier Science, 2006.

[178] S. Mahajan, K.H.J. Buschow, R. Cahn, M.C. Flemings, B. Ilschner, E.J. Kramer, and P. Veyssiere. *Encyclopedia of Materials: Science and Technology*. Elsevier Science, 2001.

[179] Alexandra Simperler, Andreas Kornherr, Reenu Chopra, P. Arnaud Bonnet, William Jones, W. D. Samuel Motherwell, and Gerhard Zifferer. Glass transition temperature of glucose, sucrose, and trehalose: An experimental and in silico study. *Journal of Physical Chemistry B*, 110(39):19678–19684, OCT 5 2006.

[180] Robert Alcock, Matthew G. Cottingham, Christine S. Rollier, Julie Furze, Samodh D. De Costa, Marian Hanlon, Alexandra J. Spencer, Jared D. Honeycutt, David H. Wyllie, Sarah C. Gilbert, Migena Bregu, and Adrian V. S. Hill. Long-Term Thermostabilization of Live Poxviral and Adenoviral Vaccine Vectors at Supraphysiological Temperatures in Carbohydrate Glass. *Science Translational Medicine*, 2(19), FEB 17 2010.

[181] Minoru Sakurai, Takao Furuki, Ken-ichi Akao, Daisuke Tanaka, Yuichi Nakahara, Takahiro Kikawada, Masahiko Watanabe, and Takashi Okuda. Vitrification is essential for anhydrobiosis in an African chironornid, Polypedilum vanderplanki. *Proceedings of the National Academy of Sciences of the United States of America*, 105(13):5093–5098, APR 1 2008.

[182] John H. Crowe. Trehalose as a Chemical Chaperone: Fact and Fantasy. In *Molecular Aspects of the Stress Response: Chaperones, Membranes and Networks*, volume 594 of *Advances in Experimental Medicine and Biology*, pages 143–158. 2007.

[183] J.-P. Amorij, J. Meulenaar, W. L. J. Hinrichs, T. Stegmann, A. Huckriede, F. Coenen, and H. W. Frijlink. Rational design of an influenza subunit vaccine powder with sugar glass technology: Preventing conformational changes of haemagglutinin during freezing and freeze-drying. *Vaccine*, 25(35):6447–6457, AUG 29 2007.

[184] Thomas S. Walter, Jingshan Ren, Tobias J. Tuthill, David J. Rowlands, David I. Stuart, and Elizabeth E. Fry. A plate-based high-throughput assay for virus stability and vaccine formulation. *Journal of Virological Methods*, 185(1):166–170, OCT 2012.

[185] Maria Pelliccia, Patrizia Andreozzi, Jayson Paulose, Marco D'Alicarnasso, Valeria Cagno, Manuela Donalisio, Andrea Civra, Rebecca M. Broeckel, Nicole Haese, Paulo Jacob Silva, Randy P. Carney, Varpu Marjomaki, Daniel N. Streblow, David Lembo, Francesco Stellacci, Vincenzo Vitelli, and Silke Krol. Additives for vaccine storage to improve thermal stability of adenoviruses from hours to months. *Nature Communications*, 7, NOV 30 2016.

[186] Retraction for zhang et al., stabilization of vaccines and antibiotics in silk and eliminating the cold chain. *Proceedings of the National Academy of Sciences*, 113(26):E3810–E3810, 2016.

[187] Guangchuan Wang, Xiaofeng Li, Lijuan Mo, Zhiyong Song, Wei Chen, Yongqiang Deng, Hui Zhao, Ede Qin, Chengfeng Qin, and Ruikang Tang. Eggshell-inspired biomineralization generates vaccines that do not require refrigeration. *Angewandte Chemie*, 124(42):10728–10731.

[188] Caiyu Peng, Qinghe Zhao, and Changyou Gao. Sustained delivery of doxorubicin by porous CaCO3 and chitosan/alginate multilayers-coated CaCO3 microparticles. *Colloids and Surfaces A-Physicochemical and Engineering Aspects*, 353(2-3):132–139, JAN 15 2010.

[189] A Lucas-Girot, MC Verdier, O Tribut, JC Sangleboeuf, H Allain, and H Oudadesse. Gentamicin-loaded calcium carbonate materials: Comparison of two drug-loading modes. *Journal of Biomedical Materials Research Part B-Applied Biomaterials*, 73B(1):164–170, APR 2005.

[190] Marie-Luce De Temmerman, Jo Demeester, Filip De Vos, and Stefaan C. De Smedt. Encapsulation Performance of Layer-by-Layer Microcapsules for Proteins. *Biomacromolecules*, 12(4):1283–1289, APR 2011.

[191] Pengzhong Shi, Shan Luo, Brigitte Voit, Dietmar Appelhans, and Xingjie Zan. A facile and efficient strategy to encapsulate the model basic protein lysozyme into porous CaCO3. *Journal of Materials Chemistry B*, 6(25):4205–4215, JUL 7 2018.

[192] AI Petrov, DV Volodkin, and GB Sukhorukov. Protein-calcium carbonate coprecipitation: A tool for protein encapsulation. *Biotechnology Progress*, 21(3):918–925, MAY-JUN 2005.

[193] DV Volodkin, NI Larionova, and GB Sukhorukov. Protein encapsulation via porous CaCO3 microparticles templating. *Biomacromolecules*, 5(5):1962–1972, SEP-OCT 2004.

[194] Marieh B. Al-Handawi, Patrick Commins, Sourabh Shukla, Pascal Didier, Masahiko Tanaka, Gijo Raj, Frank A. Veliz, Renu Pasricha, Nicole F. Steinmetz, and Pance Naumov. Encapsulation of Plant Viral Particles in Calcite Crystals. *Advanced Biosystems*, 2(5), MAY 2018.

[195] M.F. Ashby. *Materials Selection in Mechanical Design*. Elsevier Science, 2016.

[196] A.C. Fischer-Cripps. *Nanoindentation*. Mechanical Engineering Series. Springer New York, 2013.

[197] M.F. Ashby and D.R.H. Jones. *Engineering Materials 1: An Introduction to Properties, Applications and Design*. Number v. 1 in Materials science and engineering. Elsevier Science, 2012.

[198] M.L. Oyen. *Handbook of Nanoindentation: With Biological Applications*. Pan Stanford, 2010.

[199] K.L. Johnson. *Contact Mechanics*. Cambridge University Press, 9th edition, 2003.

[200] David Labonte, Anne-Kristin Lenz, and Michelle L. Oyen. On the relationship between indentation hardness and modulus, and the damage resistance of biological materials. *Acta Biomaterialia*, 57:373–383, Jul 15 2017.

[201] M. L. Oyen. Nanoindentation of Biological and Biomimetic Materials. *Experimental Techniques*, 37(1):73–87, JAN-FEB 2013.

[202] WC Oliver and GM Pharr. An Improved Technique for Determining Hardness and Elastic-Modulus Using Load and Displacement Sensing Indentation Experiments. *Journal of Materials Research*, 7(6):1564–1583, JUN 1992.

[203] Sneddon, Ian N. The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *International Journal of Engineering Science*, 3(1):45–57, 1965.

[204] J Thurn and RF Cook. Simplified area function for sharp indenter tips in depth-sensing indentation. *Journal of Materials Research*, 17(5):1143–1146, May 2002.

[205] M. Hesse, H. Meier, and B. Zeeh. *Spectroscopic Methods in Organic Chemistry, 2nd Edition 2007*. Foundations series. Thieme.

[206] D.H. Williams and I. Fleming. *Spectroscopic Methods in Organic Chemistry, 6th Edition*. Mc Graw-HIll, 2008.

[207] J.M. Hollas. *Modern Spectroscopy*. Wiley, 2004.

[208] Matthew J. Baker, Julio Trevisan, Paul Bassan, Rohit Bhargava, Holly J. Butler, Konrad M. Dorling, Peter R. Fielden, Simon W. Fogarty, Nigel J. Fullwood, Kelly A. Heys, Caryn Hughes, Peter Lasch, Pierre L. Martin-Hirsch, Blessing Obinaju, Ganesh D. Sockalingum, Josep Sule-Suso, Rebecca J. Strong, Michael J. Walsh, Bayden R. Wood,

Peter Gardner, and Francis L. Martin. Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols*, 9(8):1771–1791, Aug 2014.

[209] Lior Regev, Kristin M. Poduska, Lia Addadi, Steve Weiner, and Elisabetta Boaretto. Distinguishing between calcites formed by different mechanisms using infrared spectrometry: archaeological applications. *Journal Of Archaeological Science*, 37(12):3022–3029, DEC 2010.

[210] Yael Politi, Yael Levi-Kalisman, Sefi Raz, Fred Wilt, Lia Addadi, Steve Weiner, and Irit Sagi. Structural characterization of the transient amorphous calcium carbonate precursor phase in sea urchin embryos. *Advanced Functional Materials*, 16(10):1289–1298, JUL 4 2006.

[211] J Aizenberg, G Lambert, L Addadi, and S Weiner. Stabilization of amorphous calcium carbonate by specialized macromolecules in biological and synthetic precipitates. *Advanced Materials*, 8(3):222–&, Mar 1996.

[212] L Brecevic and AE Nielsen. Solubility of amorphous calcium-carbonate. *Journal of Crystal Growth*, 98(3):504–510, Nov 1989.

[213] E Beniash, J Aizenberg, L Addadi, and S Weiner. Amorphous calcium carbonate transforms into calcite during sea urchin larval spicule growth. *Proceedings of the Royal Society B-Biological Sciences*, 264(1380):461–465, Mar 22 1997.

[214] GF Xu, N Yao, IA Aksay, and JT Groves. Biomimetic synthesis of macroscopic-scale calcium carbonate thin films. Evidence for a multistep assembly process. *Journal of the American Chemical Society*, 120(46):11977–11985, NOV 25 1998.

[215] RW Zimmerman. The Effect of Microcracks on the Elastic-moduli of Brittle Materials. *Journal of Materials Science Letters*, 4(12):1457–1460, 1985.

[216] B. Bhushan. *Nanotribology and Nanomechanics: An Introduction*. Springer Berlin Heidelberg, 2008.

[217] DS Harding, WC Oliver, and GM Pharr. Cracking During Nanoindentation and Its Use in the Measurement of Fracture Toughness. In Baker, SP and Ross, CA and Townsend, PH and Volkert, CA and Borgesen, P, editor, *Thin Films: Stresses and Mechanical Properties V*, volume 356 of *Materials Research Society Symposium Proceedings*, pages 663–668. Mat Res Soc, 1995.

[218] George Z. Voyiadjis and Rick Peters. Size effects in nanoindentation: an experimental and analytical study. *Acta Mechanica*, 211(1-2):131–153, APR 2010.

[219] ME Broz, RF Cook, and DL Whitney. Microhardness, toughness, and modulus of Mohs scale minerals. *American Mineralogist*, 91(1):135–142, JAN 2006.

[220] Miki E. Kunitake, Lauren M. Mangano, John M. Peloquin, Shefford P. Baker, and Lara A. Estroff. Evaluation of strengthening mechanisms in calcite single crystals from mollusk shells. *Acta biomaterialia*, 9(2):5353–5359, Feb 2013.

[221] Branislav Igic, Kim Braganza, Margaret M. Hyland, Heather Silyn-Roberts, Phillip Cassey, Tomas Grim, Jarkko Rutila, Csaba Moskat, and Mark E. Hauber. Alternative mechanisms of increased eggshell hardness of avian brood parasites relative to host species. *Journal of the Royal Society Interface*, 8(64):1654–1664, NOV 7 2011.

[222] L. Christidis, W. Boles, and CSIRO Publishing. *Systematics and Taxonomy of Australian Birds*. CSIRO Pub., 2008.

[223] Shirly Borukhin, Leonid Bloch, Tzvia Radlauer, Adrian H. Hill, Andrew N. Fitch, and Boaz Pokroy. Screening the Incorporation of Amino Acids into an Inorganic Crystalline Host: the Case of Calcite. *Advanced Functional Materials*, 22(20):4216–4224, OCT 23 2012.

[224] Yi-Yeoun Kim, Joseph D. Carloni, Beatrice Demarchi, David Sparks, David G. Reid, Miki E. Kunitake, Chiu C. Tang, Melinda J. Duer, Colin L. Freeman, Boaz Pokroy, Kirsty Penkman, John H. Harding, Lara A. Estroff, Shefford P. Baker, and Fiona C. Meldrum. Tuning hardness in calcite by incorporation of amino acids. *Nature Materials*, 15(8):903+, AUG 2016.

[225] FH Arnold. Design by directed evolution. *Accounts of Chemical Research*, 31(3):125–131, MAR 1998.

[226] Philip A. Romero and Frances H. Arnold. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology*, 10(12):866–876, DEC 2009.

[227] A.J. Schwartz, M. Kumar, B.L. Adams, and D.P. Field. *Electron Backscatter Diffraction in Materials Science*. Springer US, 2013.

[228] V Monchois, C Abergel, J Sturgis, S Jeudy, and JM Claverie. Escherichia coli ykfE ORFan gene encodes a potent inhibitor of C-type lysozyme. *Journal of Biological Chemistry*, 276(21):18437–18441, MAY 25 2001.

[229] Chantal Abergel, Vincent Monchois, Deborah Byrne, Sabine Chenivesse, Frederique Lembo, Jean-Claude Lazzaroni, and Jean-Michel Claverie. Structure and evolution of the Ivy protein family, unexpected lysozyme inhibitors in Gram-negative bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15):6394–6399, APR 10 2007.

[230] JP Reyes-Grajeda, D Jauregui-Zuniga, A Rodriguez-Romero, A Hernandez-Santoyo, VM Bolanos-Garcia, and A Moreno. Crystallization and preliminary X-ray analysis of ovocleidin-17 a major protein of the Gallus gallus eggshell calcified layer. *Protein and Peptide Letters*, 9(3):253–257, JUN 2002.

[231] ASB Edge. Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochemical Journal*, 376(2):339–350, DEC 1 2003.

[232] MT Hincke, J Gautron, M Panheleux, J Garcia-Ruiz, MD McKee, and Y Nys. Identification and localization of lysozyme as a component of eggshell membranes and eggshell matrix. *Matrix Biology*, 19(5):443–453, SEP 2000.

[233] Yuansheng Sun, Richard N. Day, and Ammasi Periasamy. Investigating protein-protein interactions in living cells using fluorescence lifetime imaging microscopy. *Nature Protocols*, 6(9):1324–1340, SEP 2011.

[234] Joshua A. Broussard, Benjamin Rappaz, Donna J. Webb, and Claire M. Brown. Fluorescence resonance energy transfer microscopy as demonstrated by measuring the activation of the serine/threonine kinase Akt. *Nature Protocols*, 8(2):265–281, FEB 2013.

[235] SM Kelly, TJ Jess, and NC Price. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 1751(2):119–139, AUG 10 2005.

[236] Christoph Wiedemann, Peter Bellstedt, and Matthias Goerlach. CAPITO-a web server-based analysis and plotting tool for circular dichroism data. *Bioinformatics*, 29(14):1750–1757, JUL 15 2013.

[237] J.R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer US, 2007.

[238] VN Uversky. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, 11(4):739–756, APR 2002.

[239] Shun ichi Kidokoro and Shigeyoshi Nakamura. Chapter fifteen - iatc, dsc, and ppc analysis of reversible and multistate structural transition of cytochrome c. In Andrew L.

Feig, editor, *Calorimetry*, volume 567 of *Methods in Enzymology*, pages 391 – 412. Academic Press, 2016.

[240] Vladimir N. Uversky and Oleg B. Ptitsyn. Further evidence on the equilibrium "pre-molten globule state": Four-state guanidinium chloride-induced unfolding of carbonic anhydrase b at low temperature. *Journal of Molecular Biology*, 255(1):215 – 228, 1996.

[241] The World Health Organization. Edetic Acid (EDTA) in Drinking-Water. *In 'Guidelines for drinking-water quality: Health criteria and other supporting information' 2nd ed.*, Addendum to Vol. 2, 2003.

[242] Yamarik TA. Lanigan RS. Final report on the safety assessment of EDTA, Calcium Disodium EDTA, diammonium EDTA, Dipotassium EDTA, Disodium EDTA, TEA-EDTA, Tetrasodium EDTA, Tripotassium EDTA, Trisodium EDTA, HEDTA, and Trisodium HEDTA. *International Journal of Toxicology*, 21(5, 2):95–142, OCT 2002.

[243] L Addadi and S Weiner. Control and design principles in biological mineralization. *Angewandte Chemie-International Edition*, 31(2):153–169, FEB 1992.

[244] Leila N. Hassani, Francois Hindre, Thomas Beuvier, Brice Calvignac, Nolwenn Lautram, Alain Gibaud, and Frank Boury. Lysozyme encapsulation into nanostructured CaCO3 microparticles using a supercritical CO2 process and comparison with the normal route. *Journal of Materials Chemistry B*, 1(32):4011–4019, 2013.

[245] Bram Cantaert, Yi-Yeoun Kim, Henning Ludwig, Fabio Nudelman, Nico A. J. M. Sommerdijk, and Fiona C. Meldrum. Think Positive: Phase Separation Enables a Positively Charged Additive to Induce Dramatic Changes in Calcium Carbonate Morphology. *Advanced Functional Materials*, 22(5):907–915, MAR 7 2012.

[246] Kang Zhao, Meng Wang, Xiaoqiang Wang, Congmeng Wu, Hai Xu, and Jian R. Lu. Crystal Growth of Calcite Mediated by Ovalbumin and Lysozyme: Atomic Force Microscopy Study. *Crystal Growth & Design*, 13(4):1583–1589, APR 2013.

[247] Izabela Polowczyk, Anna Bastrzyk, and Marta Fiedot. Protein-Mediated Precipitation of Calcium Carbonate. *Materials*, 9(11), NOV 2016.

[248] Bartosz Marzec, David C. Green, Mark A. Holden, Alexander S. Cote, Johannes Ihli, Saba Khalid, Alexander Kulak, Daniel Walker, Chiu Tang, Dorothy M. Duffy, Yi-Yeoun Kim, and Fiona C. Meldrum. Amino Acid Assisted Incorporation of Dye Molecules within Calcite Crystals. *Angewandte Chemie-International Edition*, 57(28):8623–8628, JUL 9 2018.

[249] W Becker, A Bergmann, MA Hink, K Konig, K Benndorf, and C Biskup. Fluorescence lifetime imaging by time-correlated single-photon counting. *Microscopy Research and Technique*, 63(1):58–66, JAN 1 2004.

[250] Netta Vidavsky, Sefi Addadi, Julia Mahamid, Eyal Shimoni, David Ben-Ezra, Muki Shpigel, Steve Weiner, and Lia Addadi. Initial stages of calcium uptake and mineral deposition in sea urchin embryos. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1):39–44, JAN 7 2014.

Appendix A

Supplementary Data for Cloning Experiments

Ivy gene sequence (474-nucleotide) is given below.

ATGGGCAGGATAAGCTCGGGAGGAATGATGTTTAAGGCAATAA
CGACAGTCGCCGCTCTGGTCATCGCCACCAGTGCAATGGCGC
AGGATGATTTAACCATTAGCAGCCTTGCAAAGGGCGAAACCA
CCAAAGCTGCATTTAATCAGATGGTACAAGGGCATAAGCTGCC
TGCCTGGGTGATGAAAGGCGGTACTTATACTCCCGCACAAACC
GTAACGTTGGGAGATGAGACGTATCAGGTGATGAGCGCGTGCA
AACCGCATGACTGTGGCTCGCAACGTATCGCTGTGATGTGGTCC
GAGAAATCTAATCAGATGACGGGGCTGTTCTCGACTATTGATGA
GAAAACGTCGCAAGAGAAACTCACCTGGCTGAATGTGAACGAT
GCGCTTTCGATTGATGGTAAAACGGTGTTGTTCGCGGCGTTGAC
CGGCAGCCTGGAAAACCATCCGGATGGCTTTAATTTTAAATAA

Plasmid sequencing result for the ivy region is given below. Ivy gene is shown in bold and 6XHis-tag is shown in italic.

ANNNNNNNGAGCGGNNANATTTCNCACAGAATTCATTAAAGAGG
AGAAATTAACT

ATGGGCAGGATAAGCTCGGGAGGAATGATGTTTAAGGCAATAA
CGACAGTCGCCGCTCTGGTCATCGCCACCAGTGCAATGGCGCA
GGATGATTTAACCATTAGCAGCCTTGCAAAGGGCGAAACCACCA
AAGCTGCATTTAATCAGATGGTACAAGGGCATAAGCTGCCTGCC

TGGGTGATGAAAGGCGGTACTTATACTCCCGCACAAACCGTAAC
GTTGGGAGATGAGACGTATCAGGTGATGAGCGCGTGCAAACCGC
ATGACTGTGGCTCGCAACGTATCGCTGTGATGTGGTCCGAGAAAT
CTAATCAGATGACGGGGCTGTTCTCGACTATTGATGAGAAAACGT
CGCAAGAGAAACTCACCTGGCTGAATGTGAACGATGCGCTTTCGA
TTGATGGTAAAACGGTGTTGTTCGCGGCGTTGACCGGCAGCCTGG
AAAACCATCCGGATGGCTTTAATTTTAAA

AGAGGATCG*CATCACCATCACCATCAC*
GGATCCGCATGCGAGCTCGGTACCCCGGGTCGACCTGCAGCCAAGCTT
AATTAGCTGAGCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCAGA
ACTCCATCTGGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTT
ATTGGTGAGAATCCAAGCTAGCTTGGCGAGATTTTCAGGAGCTAAGGA
AGCTAAAATGGAGAAAAAAATCACTGGATATACCACCGTTGATATATCC
CAATGGCATCGTAAAGAACATTTTGAGGCATTTCAGTCAGTTGCTCAAT
GTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTTTTTAAAGAC
CGTAAAGAAAAATAAGCACAAGTTTTATCCGGCCTTTATTCACATTCTT
GCCCGCCTGATGAATGCTCATCCGGAATTTCGTATGGCAATGAAAGACG
GTGAGCTGGTGATATGGGGATAGTGTTCACCCTTGTTACACCGTTTTCCA
TGAGCAAACTGAAACGTTTTCATCGCTCTGGAGTGAATACCACGACGA
TTTCCGGCAGTTTCTACACATATATTCGCAAGAAGGGGGCNTGTTACGG
TGAAAACCTGGCCTATTTCCCTAAAGGGTTTATTGAAAANTNGNTTTTT
CGTNTCACCCAATCCCTGGGNTGAATTTTACCNAGTTTTGGATTTAAAC
GGGGGCCAATNNGGGACAACTTNNTTCGCCCCCGGTTTTCCCNNGGGG
GCAAAATTTNAACGCCAGGGGCANNAAGGGCCA

OC17 gene used in this study is given below. Note that some amino acids were optimized for E.Coli expression.

ATCCGGATGGTTGTGGTCCTGGTTGGGTTCCGACACCTGGTGGTTGTCTGG
GTTTTTTTAGCCGTGAACTGAGCTGGTCACGTGCAGAAAGCTTTTGTCGTC
GTTGGGGTCCGGGTAGCCATCTGGCAGCAGTTCGTAGCGCAGCCGAACTG
CGTCTGCTGGCAGAACTGCTGAATGCAAGCCGTGGTGGTGATGGTAGCGG
TGAAGGTGCAGATGGTCGTGTTTGGATTGGTCTGCATCGTCCGGCAGGTAG
CCGTAGCTGGCGTTGGAGTGATGGCACCGCACCGCGTTTTGCAAGCTGGCA
TCGTACCGCAAAAGCACGTCGTGGTGGTCGTTGTGCAGCACTGCGTGATGA

AGAAGCATTTACCAGCTGGGCAGCACGTCCGTGTACCGAACGTAATGCATTT
GTTTGTAAAGCAGCAGCC

Plasmid sequencing result for the OC17 region is given below. OC17 gene is shown in bold and 6XHis-tag is shown in italic.

NNNNNNNNNNNNNTNNNTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGA
TATACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCG
CTGCCCAGCCGGCGATGGCCAGCGCATGGTCACATCCGCAGTTTGAAAAAG
GTGGTGGTAGCGGTGGTGGTTCAAGTGCTTGGAGCCACCCTCAGTTCGAAA
AAAGCGCA

GATCCGGATGGTTGTGGTCCTGGTTGGGTTCCGACACCTGGTGGTTGT
CTGGGTTTTTTTAGCCGTGAACTGAGCTGGTCACGTGCAGAAAGCTTT
TGTCGTCGTTGGGGTCCGGGTAGCCATCTGGCAGCAGTTCGTAGCGCA
GCCGAACTGCGTCTGCTGGCAGAACTGCTGAATGCAAGCCGTGGTGGT
GATGGTAGCGGTGAAGGTGCAGATGGTCGTGTTTGGATTGGTCTGCAT
CGTCCGGCAGGTAGCCGTAGCTGGCGTTGGAGTGATGGCACCGCACCG
CGTTTTGCAAGCTGGCATCGTACCGCAAAAGCACGTCGTGGTGGTCGT
TGTGCAGCACTGCGTGATGAAGAAGCATTTACCAGCTGGGCAGCACGT
CCGTGTACCGAACGTAATGCATTTGTTTGTAAAGCAGCAGCC*CATCACCA*
CCATCACCAT

TAATGACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCA
TAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG
GAACTATATCCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCG
CGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCT
AGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCCCGTTCGCCGGGCT
TTCCCCGTCAAGCTCTAAATCGGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCT
TTACGGNCCCTCCGACNCCAAAAAACTTGAATTAGGGGAAGGGTTC

Appendix B

Amino Acid Abbreviations

Name of amino acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V