Glycosylation: A “Last Word” in the Protein-Mediated Biomineralization Process

John Spencer Evans

Laboratory for Chemical Physics, Department of Craniofacial Medicine and Skeletal Biology, New York University College of Dentistry, 345 E. 24th Street, New York, NY 10010, USA; jse1@nyu.edu; Tel.: +1-347-753-1955

Received: 12 August 2020; Accepted: 12 September 2020; Published: 16 September 2020

Abstract: Post-translational modifications are one way that biomineral-associated cells control the function and fate of proteins. Of the ten different types of post-translational modifications, one of the most interesting and complex is glycosylation, or the covalent attachment of carbohydrates to amino acid sidechains Asn, Ser, and Thr of proteins. In this review the author surveys some of the known biomineral-associated glycoproteins and summarizes recent in vitro recombinant protein experiments which test the impact of glycosylation on biomineralization protein functions, such as nucleation, crystal growth, and matrix assembly. These in vitro studies show that glycosylation does not alter the inherent function of the polypeptide chain; rather, it either accentuates or attenuates functionality. In essence, glycosylation gives the cell the “last word” as to what degree a biomineralization protein will participate in the biomineralization process.

Keywords: biomineralization; glycosylation; proteins

1. Introduction

Over the last forty years there has been a concerted effort to understand how organisms craft biomineralized skeletal structures for survival [1–3]. This effort has focused along two lines. First, how do mineral crystals or amorphous minerals form under biological conditions? Recent evidence points to a mineral precursor nucleation process involving nanoparticle synthesis followed by particle assembly into larger mineral mesoscale structures [4–6]. Second, what agents are biosynthetically created by these same organisms to manage the mineral formation process? With regard to the latter, it has been well documented that the genomes of biomineralizing organisms code for families of proteins that are mineral-specific and unique with regard to primary sequence construction and structure [7–10]. The appearance of these proteins in the extracellular matrix during mineral formation is a clear attempt by cells to regulate the nucleation and assembly stages that lead to the final mineral product of the skeletal elements that are necessary for organism survival. Thus, to understand how biominerals form into larger, useful structures, we must understand the role or function that these proteins play in nucleation and particle assembly.

In the majority of eukaryotic organisms, the overall complexity of the biomineral proteomes is augmented by a process known as post-translational modification [11–13]. In essence, once a nascent protein polypeptide chain is produced on the ribosomal complexes, in some cases the cells express enzymes that perform further covalent modifications of certain amino acid sidechains on the protein, thereby altering the functionality of these sidechains. These covalent modifications occur in compartments that are separate from the cell cytoplasm (e.g., Golgi apparatus, rough endoplasmic reticulum (RER), intracellular vesicles) [11–13]. A summary of common post-translational modifications (Table 1) [12] indicates that certain amino acid sidechains are targeted by cells for covalent modification. These modifications are performed by intracellular enzymes and in the majority of cases the finalized
protein product is transported to its intended destination using vesicles [11–13]. In essence, genomic information is translated into proteomic information, and in some instances, this proteomic information can further be modified and thus diversified in terms of structure and ultimately function [11–13].

Table 1. Common post-translational modifications of proteins.

| Post-Translational Modification | Type of Modification | Site(s) of Modification |
|---------------------------------|----------------------|-------------------------|
| Disulfide bond formation        | Thiol oxidation to form -S-S- bond | Cys                     |
| Hydroxylation                   | Addition of -OH group | Pro, Glu                 |
| Ubiquitination                  | Addition of ubiquitin protein(s) | Lysine                  |
| Lipidation                      | Esterification to lipid group | Cys, Lys, N-terminal Gly |
| SUMOylation                     | Small Ubiquitin-like Modifier protein | Lys                     |
| Acetylation                     | Addition of acetyl group | N-terminus, Lys, Ser    |
| Methylation                     | Addition of methyl group | Lys, Arg, termini       |
| Phosphorylation                 | Addition of phosphate group | Ser, Thr, Tyr           |
| Glycosylation                   | Addition of carbohydrate group(s) | Asn, Thr, Ser           |
| Nitration                       | Addition of nitrogen | Tyr                     |
| Acylation                       | Addition of acyl chain | Cys, Gly, Ser, Thr, Lys |

Perhaps the most complex post-translational modification process is glycosylation, or the addition of one or more carbohydrate monomers (known as monosaccharides) to specific amino acid sidechains on a protein, thus converting the polypeptide into a glycoprotein [11–16]. There are three classifications of glycoproteins, depending on which amino acids serve as attachment points for carbohydrates [11–16]: (1) O-linked, where the oligosaccharide attachment occurs on Ser and/or Thr residues and is performed in the Golgi apparatus; (2) N-linked, where the oligosaccharide attachment occurs on Asn and is performed within the endoplasmic reticulum (ER); (3) hybrid, in which a glycoprotein has O-linked (Ser, Thr) glycans and N-linked (Asn) glycans.

Several features contribute to the overall complexity of glycosylation [11–16]: (a) The number of carbohydrate groups added to a single amino acid sidechain site can vary; (b) the number and type of amino acid sites for attachment on a given protein can vary; (c) at a given attachment point on a protein, the carbohydrate groups can be constructed as linear or branching chains; (d) the hydroxyl-rich carbohydrate groups themselves can be modified by the addition of chemical groups, such as carboxylate, sulfate, N-acetyl amino, and hydroxyl. Thus, unlike other post-translational modifications, glycosylation represents a unique opportunity for the cell to combine two very different macromolecular building blocks (amino acids, carbohydrates) into one macromolecule, which in turn may have a significant impact on the function and distribution of this protein class within a biomineralizing system.

For the purposes of this review, the focus will be on glycosylation and the impact that this post-translational modification has on known biomineralization processes. The review will begin by identifying notable well-studied biomineralization glycoproteins [17–25] and briefly touch upon their roles in their respective mineralization processes. Then, a discussion of recent in vitro studies [26–30] will follow, which investigated the effects of glycosylation on biomineralization protein mineralization functions (e.g., nucleation, crystal growth, particle assembly and protein–protein interactions). Finally, suggestions as to the direction of future studies of glycosylated biomineralization proteins will be offered.

2. Glycoproteins in Biomineralization: An Overview

Table 2 provides a summary of specific mineral matrix proteins that have been identified as glycoproteins and report the complete amino acid sequence [17–25]. Admittedly, this table is sparse,
and at the time of this writing very few biomineral-associated glycoproteins have complete protein sequence data or oligosaccharide composition/sequence data available. Note that some studies have identified glycoproteins in the extracellular matrices of different organisms [31–35], but to date these proteins have not been sequenced nor rigorously characterized. The majority of the identified biomineralization glycoproteins are found in association with calcium-based biominerals [17–25,31–35]; however, it should be acknowledged that glycoproteins may eventually be identified in other non-calcium based biominerals, such as magnetite (Fe$_3$O$_4$) [36] or silicates (SiO$_4$) [37].

Table 2. Biomineral-associated glycoproteins.

| Protein | Organism       | Tissue  | Associated Mineral Phase         |
|---------|----------------|---------|----------------------------------|
| SpSM30 A-F | S. purpuratus (sea urchin) | Embryonic spicule | Magnesium Calcite (CaMgCO$_3$) |
| AP24    | H. rufescens (abalone) | Shell nacre | Aragonite (CaCO$_3$)             |
| Enamelin | Vertebrates     | Tooth enamel | Hydroxyapatite (CaPO$_4$)        |
| SIBLING Family | Vertebrates | Bone, tooth dentine | Hydroxyapatite (CaPO$_4$)        |
| EDIL3   | Avian           | Eggshell | Calcite (CaCO$_3$)               |
| MFGE8   | Avian           | Eggshell | Calcite (CaCO$_3$)               |
| Proteoglycans | Vertebrates | Bone, tooth dentine | Hydroxyapatite (CaPO$_4$)        |

To provide some examples of the roles that glycoproteins play in biomineralization, we will briefly describe the proteins in Table 1. Note that in only a few examples the oligosaccharide chain attachment and composition are known at this time [26,27].

2.1. Enamelin

This is a glycoprotein found in the tooth enamel of vertebrates [17,18]. This protein plays a role in hydroxyapatite (HAP) formation from the amorphous precursor, amorphous calcium phosphate (ACP) [17]. This protein has two groups of oligosaccharides chains consisting of fucose, galactose, mannose, N-acetylglucosamine, and N-acetylneuraminic acid [18]. Enamelin combines with another enamel matrix protein, amelogenin, to form a protein–protein complex that stabilizes ACP and modulates HAP crystal growth during tooth formation [18].

2.2. EDIL3, MFGE8

These two proteins are found in avian eggshells [19] and, although the polypeptide chains have been sequenced, they have not been fully characterized with regard to their oligosaccharide content or sequence, it is known that they bind to amorphous calcium carbonate (ACC)—containing matrix vesicles—and guide these vesicles to the mineralization front, where calcite crystals form from the ACC particles [19].

2.3. Proteoglycans

These are a family of complex macromolecules that are composed of glycosaminoglycan (GAG) chains covalently attached to a core protein through a tetrasaccharide linker [20,21]. Proteoglycans act as polysaccharides rather than proteins as 95% of their weight is composed of glycosaminoglycans. The glycosaminoglycan chains consist of alternating hexosamine and hexuronic acid or galactose units [20,21]. There are also glycopeptide linkage regions that connect the polysaccharide chains to the core proteins that contain N- and/or O-linked oligosaccharides. Although found in the extracellular matrix of many tissues, PGs comprise a significant portion of bone and tooth dentine HAP-containing extracellular matrices and are believed to be involved in ion and water sequestration in these matrices [20,21].
2.4. SIBLING Family

There is a family of proteins found in bone and tooth dentine that are known as the SIBLING proteins (small integrin binding ligand N-glycosylated) [22,23]. These proteins, all have Arg, Gly, Asp RGD-cell binding domains, are anionic, and all are glycosylated [22,23]. At present, there is scant information regarding the N-linked oligosaccharide chain composition, sequence, or attachment location. SIBLINGS are found in multiple HAP-containing tissues in addition to bone and dentine and are multifunctional: cell signaling, hydroxyapatite binding, and mineral formation. The SIBLING proteins are osteopontin (bone sialoprotein 1), dentin matrix protein 1 (DMP1), bone sialoprotein (BSP2), matrix extracellular phosphoglycoprotein (MEPE) and the products of the dspp gene, dentin sialoprotein (DSP) and dentin phosphoprotein (DPP) [22,23].

2.5. SpSM30A-F

In the developing sea urchin Strongylcentrotus purpuratus embryo, the first skeletal element that emerges is the spicule [9], a stirrup-like structure that initially forms from ACC and transforms into mesocrystal calcite [27,29,30]. The matrix of the spicule is formed via many spicule matrix proteins (denoted as SpSM) [9] of which a subset of six isoforms, known as SpSM30A-F, are known to be glycosylated [24,27]. These proteins are known to stabilize ACC, inhabit the intracrystalline regions of mesocrystal calcite [4,5] and most likely contribute to the fracture resistance of the spicule itself [1,27]. The SpSM30 proteins are known to interact with the major spicule matrix protein, SpSM50, and these interactions are important for the assembly of the spicule matrix [27,29,30].

2.6. AP24

In the formation of the aragonitic nacre layer in the shells of mollusks, there exists families of proteins that inhabit the interior regions of aragonite crystals [6–8] and are termed intracrystalline proteins [25]. These proteins modify the material properties of the aragonite crystal and convey fracture resistance and ductility to these crystals, thus strengthening the shell itself [1,6,25,26]. In the Pacific red abalone, Haliotis rufescens, a family of intracrystalline proteins (the AP series) have been identified [25,26], with one member of this family, AP24, identified as a glycoprotein [25,26]. Subsequent studies confirmed that AP24 acts as a blocker of calcite formation, which then allows the metastable aragonite to form in the presence of extracellular Mg(II) [25,26].

3. The Impact of Glycosylation on Protein Function

Does the attachment of oligosaccharides affect the molecular behavior of a polypeptide chain? To answer this question, one could envision a comparative study wherein the function of an unglycosylated variant of a given protein is contrasted against that of a glycosylated variant, with each possessing the identical primary sequence. Here, the only variable would be the presence (or absence) of oligosaccharide chains.

Recently, this type of study was executed on two proteins, AP24 (aragonite nacre layer, Pacific red abalone H. rufescens [25] and SpSM30B/C (calcitic spicule matrix, S. purpuratus, purple sea urchin) [24]. Both proteins have been the subject of in vitro glycosylation studies in insect cells, where it was discovered that AP24 and SpSM30B/C belong to the hybrid classification—i.e., they consist of N- and O-linked linear and branching oligosaccharide chains [26,27]. Interestingly, the glycosylated variants of AP24 and SpSM30B/C both contain anionic monosialylated, bisialylated and monosulfated, bisulfated monosaccharides [26,27]. Given that both proteins inhabit a Ca(II)-rich environment in vivo, the anionic monosaccharides could serve as putative sites for Ca(II)-protein or mineral–protein interactions. To a certain extent, both proteins are similar in function: they are involved in the formation of the organic matrix, forming hydrogel particles that assemble mineral nanoparticles [26,27]. In addition, both protein hydrogels become occluded within calcium carbonates and modify the material and surface properties of the minerals they inhabit [26,27].
In the following section, we review these studies and their comparative use of two recombinant variants: (1) a non-glycosylated variant expressed in E. coli bacteria, and (2) a glycosylated variant expressed in baculovirus-infected sf9 insect cells [26,27]. By using these two variations within parallel mineralization and biophysical studies, it was possible to measure the contributions of oligosaccharide chains to the function of each protein [26,27].

3.1. The Nacre Glycoprotein AP24

In this in vitro study the protein was expressed in bacterial and insect cells as a single polypeptide. In Sf9 cells, the recombinant form of AP24 (denoted as rAP24G) is expressed with variations in glycosylation that create microheterogeneity in protein molecular masses [26]. The overall molecular mass of the oligosaccharide component was found to range from 650 Da to 6.5 kDa. It was observed that both rAP24G and the non-glycosylated variant (denoted as rAP24NG) aggregate to form protein hydrogels, with rAP24NG exhibiting a higher aggregation propensity compared to rAP24G [26]. With regard to functionality, both rAP24G and rAP24NG exhibit similar behavior within in vitro calcium carbonate mineralization assays and Ca(II) potentiometric titrations that measure prenucleation cluster appearance and ACC formation/transformation [26]. An interesting difference was noted in these studies: rAP24G modifies crystal growth directions and is a stronger nucleation inhibitor, whereas rAP24NG exhibits higher mineral phase stabilization and nanoparticle containment [26]. Hence, oligosaccharides may modulate certain functions of the nacre glycoprotein AP24 but have little effect on other intrinsic functionalities.

3.2. The Spicule Matrix Glycoprotein SpSM30B/C

Similarly, the spicule matrix protein SpSM30B/C is expressed in insect and bacterial cells as a single polypeptide. The recombinant glycosylated form (rSpSM30B/C-G) also contains variations in glycosylation that create microheterogeneity in rSpSM30B/C molecular masses [27]. The overall molecular mass of the oligosaccharide component was found to range from 1.2 kDa to 3.6 kDa to 7.5 kDa. In terms of aggregation propensities and hydrogel formation, the bacteria expressed non-glycosylated variant (rSpSM30B/C-NG) has a lower aggregation propensity compared to the glycosylated rSpSM30B/C-G variant. Both variants promote faceted growth and create surface texturing of calcite crystals in vitro, with rSpSM30B/C-G promoting these effect with higher intensity (Figure 1) [27].

![Figure 1](image_url). SEM images of in vitro calcium carbonate mineralization assay samples, following the protocol described in [27]. (A) Negative control, no protein added; (B) + rSpSM30B/C, non-glycosylated, 1.5 µM; (C) + rSpSM30B/C-G, glycosylated, 1.5 µM. Note faceted nanotexturing produced by both proteins, which is more pronounced in the presence of the glycosylated variant in (C). White arrow in (C) denotes protein hydrogel deposit that forms within the mineralization assay. Scalebars = 2 µm.

3.3. How Does Glycosylation Impact Function?

From these two studies we note a trend where glycosylation does not change the intrinsic function of the polypeptide chain; rather, the attachment of anionic oligosaccharide moieties either (1) attenuates specific functions or has no effect (AP24) or (2) accentuates protein functionality [SpSM30B/C].
Other studies with multiple glycoproteins will hopefully confirm this trend or provide evidence of other effects that oligosaccharides impose upon polypeptides.

4. The Impact of Glycosylation on Protein–Protein Interaction (Matrix Formation)

In addition to modulating the mineral formation process, a key role of biomineralization proteins is the assembly and organization of multiple proteins to form an organic matrix within which the nucleation and crystal growth processes take place [1,2,7]. We pose the question: how does glycosylation affect protein–protein interactions that dominate the matrix formation process? To address this question, investigations were conducted on molluscan (AP7, AP24, H. rufescens) [6,25,28] and sea urchin (SpSM50, SpSM30B/C, S. purpuratus) [6,24,29,30] recombinant two-protein systems. In both organisms it is known that each pair of proteins co-exist in vivo within the extracellular matrix [6,24].

4.1. AP7–AP24 Complex

It is known that AP7 forms a complex with AP24 in the nacre layer [25]. Using sensitive quartz crystal microbalance with dissipation (QCM-D) measurements, this complex formation was confirmed and it was found that both the glycosylated and non-glycosylated variants of recombinant AP24 bound to recombinant AP7 but with different quantities and binding kinetics (Figure 2). Interestingly, non-glycosylated recombinant AP24 underwent a conformational change when binding to AP7, but the glycosylated variant did not [28]. Moreover, the binding of AP7 with non-glycosylated and glycosylated variants of AP24 was found to be Ca(II)-dependent and -independent, respectively (Figure 2) [28]. Thus, AP7 and AP24 protein complexes form as a direct result of polypeptide–polypeptide chain recognition and not polypeptide–oligosaccharide recognition. However, the presence of anionic oligosaccharides on AP24 appears to modulate the intensity of AP7–AP24 protein–protein interactions and potentially stabilizes the AP24 conformation upon binding to AP7. As shown in Figure 3, both proteins have numerous surface-accessible regions or domains where interactions might take place.

![Figure 2](image-url)

**Figure 2.** QCM-D experiments of immobilized rAP7 (500 nM in 10 mM HEPES pH 8.0) exposed to either 500 nM rAP24G (solid lines) or 500 nM rAP24NG (dashed lines) in 10 mM HEPES, pH 8.0. Plots show the third harmonic frequency (F3, blue) and dissipation (D3, red) observed under each scenario. For more information on the QCM-D method and experimental protocol, please refer to [28]. rAP7 is adsorbed onto the poly-L-Lys coated QCM-D chip, then unbound rAP7 is washed off and then rAP24G (glycosylated) or rAP24NG (non-glycosylated) variants are introduced into the flowcell. Deflection in frequency and dissipation result from rAP24 protein adsorbing onto the immobilized rAP7 layer on the chip, with amplitudes of the deflection proportional to the amount of protein bound. The time-dependent introduction of proteins is noted on the plots by arrows. These experiments were repeated and found to be reproducible.
With SpSM50 in large abundance, there is the possibility that other SpSM proteins interact with recombinant forms of SpSM50 and glycosylated and non-glycosylated variants of SpSM30B/C. The glycosylation requirement clearly indicates that the SpSM50 polypeptide sequence recognizes and binds to the glycan moieties on the surface of SpSM30B/C. As shown in Figure 4, the SpSM50 sequence contains a conserved C-type lectin domain, which is known to bind the AP7–AP24 study described above, these interactions were found to be Ca(II)-independent for both variants [29,30]. The results were quite dramatic: the formation of a SpSM50–SpSM30B/C complex requires glycosylation and, in contrast to SpSM50 protein, in ribbon representation. The protocol for structure prediction is provided in [27]. AP24 is represented without glycan groups. Note that each protein has surface-accessible domains or regions which could serve as sites for protein–protein interaction.

4.2. SpSM50–SpSM30B/C Complex

SpSM50 is the major matrix protein of the sea urchin spicules in *S. purpuratus* embryos with other SpSM proteins, such as the six SpSM30A-F isoforms comprising smaller amounts in the matrix [9,23]. With SpSM50 in large abundance, there is the possibility that other SpSM proteins interact with SpSM50 to form the matrix and control mineralization. This was tested in a recent in vitro study, where recombinant forms of SpSM50 and glycosylated and non-glycosylated variants of SpSM30B/C were investigated for their ability to form protein–protein complexes [29,30]. The results were quite dramatic: the formation of a SpSM50–SpSM30B/C complex requires glycosylation and, in contrast to the AP7–AP24 study described above, these interactions were found to be Ca(II)-independent for both variants [29,30]. The glycosylation requirement clearly indicates that the SpSM50 polypeptide sequence recognizes and binds to the glycan moieties on the surface of SpSM30B/C. As shown in Figure 4, the SpSM50 sequence contains a conserved C-type lectin domain, which is known to bind carbohydrates, and presumably it is this domain that would interact with the glycan groups of SpSM30B/C [9,23,29,30].

![Figure 3.](image_url) **Figure 3.** INTFOLD-predicted three-dimensional structures of *H. rufescens* AP7 and AP24 proteins, in ribbon representation. The protocol for structure prediction is provided in [27]. AP24 is represented without glycan groups. Note that each protein has surface-accessible domains or regions which could serve as sites for protein–protein interaction.

![Figure 4.](image_url) **Figure 4.** INTFOLD-predicted three-dimensional structures of *S. purpuratus* SpSM50 protein, in ribbon representation. The protocol for structure prediction is provided in references 27 and 29. Note that the SpSM50 protein possesses a surface-accessible C-type lectin carbohydrate binding domain, which presumably acts as a site for interaction with SpSM30B/C glycan groups.

5. Summary and Future Directions

From the foregoing, we can observe that glycosylation provides an additional degree of control over extracellular protein function by either accentuating or attenuating the intrinsic functionality of the polypeptide sequence. In a sense, the cell can have the “last word” as to the degree of participation...
within the biomineralization process. In some cases (e.g., AP24), the oligosaccharides stabilize the conformation of the glycoprotein, which is a known trait of N-linked oligosaccharides [12–16]. The author proposes that glycosylation can serve several purposes vis a vis the biomineralization process: (1) “tweak” or “tune” protein mineralization function to suit the situation or need; (2) act as a site for molecular recognition and binding with other matrix proteins; (3) conformationally stabilize a protein, thereby enhancing functionality; (4) create additional anionic sites for ionic (e.g., Ca(II)), mineral, or water interactions; (5) invoke cell activation or deactivation via binding to outer membrane receptor proteins. Clearly, there may be other benefits that arise from glycosylation, and thus this process represents a powerful method that cells can exploit to create skeletal elements under ambient or extreme conditions [1,2].

The author believes that the biomineralization field is still in its infancy with regard to understanding the role that glycoproteins and their associated oligosaccharides play in the skeletal formation process. To make progress in this area, the author proposes several key issues that need to be addressed, which are elaborated upon in Sections 5.1–5.4, below.

5.1. A More Aggressive Approach to Glycoprotein Isolation and Identification

Simply put, the genomics of biomineralization have advanced quite rapidly [8,9], but the proteomics and the identification of post-translational modifications to these proteins lag to a certain extent, especially when compared to the advances in glyobiology within the fields of immunology and other medical branches [13,14]. It is the author’s opinion that this is not due to limitations in methodologies, technology, or skill; rather, it is due to at least two factors: (1) the unwillingness of laboratories to pursue these intensive and costly projects and (2) insufficient grant funding to permit these projects to move forward. It is hoped that this situation will change for the better over time.

5.2. Improvements in Glycoprotein Purification and Structure Determination

For a variety of reasons, glycoproteins can be difficult to purify to homogeneity for structural determinations [13–16,38–40]. Further, oligosaccharide chains and the protein region(s) to which they are attached are typically conformationally labile, making structural determination by X-ray crystallography or NMR to be highly problematic, which, in turn, makes it nearly impossible to establish protein structure–function relationships [14–16]. Currently, molecular modeling (i.e., energy minimization, molecular dynamics) is the only route to obtain structural protein–oligosaccharide information, albeit in qualitative form [41]. Thus, there is a need for new methodologies to obtain glycoproteins in a highly purified form and to decipher the three-dimensional structure of the oligosaccharide-polypeptide chain complex.

5.3. Improvements in Glycoprotein Localization

One can identify the location of proteins in situ within the extracellular matrix using monoclonal or polyclonal antibody recognition of protein epitopes [39]. In the case of glycoproteins, this becomes a more complicated issue, since the antibodies raised to epitopes on glycoproteins might be specific to only the polypeptide chain, to certain oligosaccharide chains, or to both [39]. Given that some glycoproteins often exhibit variations in glycosylation [26,27], the in situ identification of glycoproteins using antibodies may not be so straightforward. In such cases, it may be prudent to synthesize select protein sequence regions and/or glycan chains and use these for antibody generation. Further, improved methods of in situ glycoprotein detection will propel the field further and allow interpretations of biomimeral formation in the presence of matrix-specific glycoproteins.

5.4. Improvements in Understanding the Role of Variations in Glycosylation

The fact that in some in vitro systems there is variation in the degree of oligosaccharide chain completion and site attachment [26,27] creates a diverse pool of glycoproteins coded for by a single gene. Is this a flaw of the cellular system, or is this deliberate and with a purpose? If deliberate, how
does the cell exploit this for control over biomineralization? This remains unknown and should be more thoroughly explored to determine its importance.

**Funding:** This review was funded by the Life Sciences Division, U.S. Army Research Office, under contract W911NF-16-1-0262.

**Acknowledgments:** This report represents Contribution Number 99 from the Laboratory of Chemical Physics, New York University.

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**References**

1. Studart, A.R. Towards high-performance bioinspired composites. *Adv. Mater.* 2012, 24, 5024–5044. [CrossRef] [PubMed]

2. Lowenstam, H.A.; Weiner, S. *On Biomineralization*; Oxford University Press: New York, NY, USA, 1989; pp. 1–134, ISBN 0-19-504977-2.

3. Mann, S. *Biomineralization: Principles and Concepts in Bioinorganic Materials Chemistry*; Oxford University Press: New York, NY, USA, 2001; pp. 6–9, 24–108.

4. Wallace, A.F.; Hedges, L.O.; Fernandez-Martinez, A.; Raiteri, P.; Gale, J.D.; Waychunas, G.A.; Whitelam, S.; Banfield, J.F.; De Yoreo, J.J. Microscopic evidence for liquid-liquid separation in supersaturated calcium carbonate solutions. *Science* 2013, 341, 885–889. [CrossRef]

5. De Yoreo, J.J.; Gilbert, P.U.P.A.; Sommerdijk, N.A.J.M.; Penn, R.L.; Whitelam, S.; Joester, D.; Zhang, H.; Rimer, J.D.; Navrotsky, A.; Banfield, J.F.; et al. Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* 2015, 349, 498–508. [CrossRef] [PubMed]

6. Perovic, I.; Chang, E.P.; Verch, A.; Rao, A.; Colfen, H.; Kroeger, R.; Evans, J.S. An oligomeric C-RING nacre protein influences pre-nucleation events and organizes mineral nanoparticles. *Biochemistry* 2014, 53, 7259–7268. [CrossRef]

7. Liu, X.; Li, J.; Xiang, L.; Sun, J.; Zheng, G.; Zhang, G.; Wang, H.; Xie, L.; Zhang, R. The role of matrix proteins in the control of nacreous layer deposition during pearl formation. *Proc. R. Soc. B* 2012, 279, 1000–1007. [CrossRef] [PubMed]

8. Zhang, G.; Fang, X.; Guo, X.; Li, L.; Luo, R.; Xu, F.; Yang, P.; Zhang, L.; Wang, X.; Qi, H.; et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012, 490, 49–54. [CrossRef] [PubMed]

9. Sea Urchin Genome Sequencing Consortium; Sodergren, E.; Weinstock, G.M.; Davidson, E.H.; Cameron, R.A.; Gibbs, R.A.; Angerer, R.C.; Arnone, M.I.; Burgess, D.; et al. The genome of the sea urchin Strongylocentrotus purpuratus. *Science* 2006, 314, 941–952. [CrossRef]

10. Tirichine, L.; Rastogi, A.; Bowler, C. Recent progress in diatom genomics and epigenomics. *Curr. Op. Plant Biol.* 2017, 36, 46–55. [CrossRef]

11. Uversky, V.N. Post-Translational Modification. In *Brenner’s Encyclopedia of Genetics*, 2nd ed.; Stanley, M., Kelly, H., Eds.; Academic Press: Cambridge, MA, USA, 2013; pp. 425–430.

12. Seo, J.; Lee, K.J. Post-translational modifications and their biological functions: Proteomic analysis and systematic approaches. *J. Biochem. Mol. Biol.* 2004, 37, 35–44. [CrossRef]

13. Wang, Y.C.; Peterson, S.E.; Loring, J.F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* 2014, 24, 143–160. [CrossRef]

14. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* 2019, 15, 346–366. [CrossRef] [PubMed]

15. Krasnova, L.; Wong, C.H. Understanding the chemistry and biology of glycosylation with glyccan synthesis. *Ann. Rev. Biochem.* 2016, 85, 599–630. [CrossRef] [PubMed]

16. Dwek, R.A. Chapter 7: Glycobiology, a quantum leap in carbohydrate chemistry. *Found. Mod. Biochem.* 1996, 2, 153–202.

17. Iijima, M.; Fan, D.; Bromley, K.; Sun, Z.; Moradian-Oldak, J. Tooth enamel proteins enamelin and amelogenin cooperate to regulate the growth morphology of octacalcium phosphate crystals. *Cryst. Growth Des.* 2011, 19, 4815–4822. [CrossRef] [PubMed]
18. Yamakoshi, Y. Carbohydrate moieties of porcine 32 kDa enamelin. *Calcif. Tissue Int.* 1996, 56, 323–330. [CrossRef]
19. Stapane, L.; Roy, N.; Hincke, M.T.; Gautron, J. The glycoproteins EDIL3 and MFGE8 regulate vesicle-mediated eggshell calcification in a new model for avian biomineralization. *J. Biol. Chem.* 2019, 294, 14256–14545. [CrossRef]
20. Couchman, J.R.; Pataki, C.A. An introduction to proteoglycans and their localization. *J. Histochem. Cytochem.* 2012, 60, 885–897. [CrossRef]
21. Ozawa, H.; Hoshi, K.; Amizuka, N. Current concepts in bone biomineralization. *J. Oral Biosci.* 2008, 50, 1–14. [CrossRef]
22. Staines, K.A.; MacRae, V.E.; Farquharson, C. The importance of the SIBLING family of proteins on skeletal mineralization and bone remodeling. *J. Endocrinol.* 2012, 214, 241–255. [CrossRef]
23. Bellahcene, A.; Castronovo, V.; Ogbureke, K.U.E.; Fisher, L.W.; Fedarko, N.S. Small integrin-binding ligand N-linked glycoproteins (SIBLINGs): Multifunctional proteins in cancer. *Natl. Rev. Cancer* 2008, 8, 212–226. [CrossRef]
24. Killian, C.; Croker, L.; Wilt, F.H. SpSM30 gene family expression in embryonic and adult biomineralized tissues of the sea urchin, *strongylocentrotus purpuratus*. *Gene Expr. Patterns* 2010, 10, 135–139. [CrossRef] [PubMed]
25. Michenfelder, M.; Fu, G.; Lawrence, C.; Weaver, J.C.; Wustman, B.A.; Taranto, L.; Evans, J.S.; Morse, D.E. Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains. *Biopolymers* 2003, 70, 522–533. [CrossRef] [PubMed]
26. Chang, E.P.; Perovic, I.; Rao, A.; Colfen, H.; Evans, J.S. Insect cell glycosylation and its impact on the functionality of a recombinant intracrystalline nacre protein, AP24. *Biochemistry* 2016, 55, 1024–1035. [CrossRef] [PubMed]
27. Jain, G.; Pendola, M.; Rao, A.; Colfen, H.; Evans, J.S. A model sea urchin spicule matrix protein self-associates to form mineral-modifying protein hydrogels. *Biochemistry* 2016, 55, 4410–4421. [CrossRef]
28. Juan-Comas, J.; Jung, Y.S.; Johnson, S.; Evans, J.S. A complicated relationship: Glycosylation, Ca(II), and primary sequence affect the interactions and kinetics between two model mollusk shell intracrystalline nacre proteins. *Biochemistry* 2020, 59, 346–350. [CrossRef]
29. Jain, G.; Pendola, M.; Koutsoumpeli, E.; Johnson, S.; Evans, J.S. Glycosylation fosters interactions between model sea urchin spicule matrix proteins. Implications for embryonic spiculogenesis and biomineralization. *Biochemistry* 2018, 57, 3032–3035. [CrossRef]
30. Pendola, M.; Jain, G.; Huang, Y.-C.; Gebauer, D.; Evans, J.S. Secrets of the sea urchin spicule revealed: Protein cooperativity is responsible for ACC transformation, intracrystalline incorporation, and guided mineral particle assembly in biocomposite material formation. *ACS Omega* 2018, 3, 11823–11830. [CrossRef]
31. Le Pabic, C.; Marie, A.; Marie, B.; Percot, A.; Bonnaud-Ponticelli, L.; Lopez, P.J.; Luquet, G. First proteomic analysis of the dorsal and ventral parts of the *Sepia officinalis* cuttlebone. *J. Proteom.* 2017, 150, 63–73. [CrossRef]
32. Albeck, S.; Weiner, S.; Addadi, L. Polysaccharides of intracrystalline proteins modulate calcite growth in vitro. *Chem. A Eur. J.* 1996, 2, 278–284. [CrossRef]
33. Arivalagan, J.; Yarra, T.; Marie, B.; Sleight, V.A.; Duvernois-Berthet, E.; Clark, M.S.; Marie, A.; Berland, S. Insights from the shell proteome: Biomineralization to adaptation. *Mol. Biol. Evol.* 2017, 34, 66–77. [CrossRef]
34. Tweedie, E.P.; Coblenz, F.E.; Shafer, T.H. Purification of a soluble glycoprotein from the uncalfed ecysidal cuticle of the blue crab *Callinectes sapidus* and its possible role in initial mineralization. *J. Expt. Zool.* 2004, 297, 2589–2598. [CrossRef] [PubMed]
35. Kamiya, H.; Jimbo, M.; Yako, H.; Muramoto, K. Participation of the C-type hemolymph lectin in mineralization of the acorn barnacle *Megabalanus rosa*. *Mar. Biol.* 2002, 140, 1235–1240.
36. Kroger, N.; Bergsdorf, C.; Sumper, M. A new calcium binding glycoprotein family constitutes a major diatom cell wall component. *EMBO J.* 1994, 13, 4676–4683. [CrossRef] [PubMed]
37. Nemoto, M.; Ren, D.; Herrera, S.; Pan, S.; Tamura, T.; Inagaki, K.; Kisailus, D. Integrative transcriptomic and proteomic analyses of a molecular mechanism of radular teeth biomineralization in *Cryptochiton Stelleri*. *Sci. Rep.* 2019, 9. [CrossRef]
38. Gerard, C. Purification of glycoproteins. *Methods Enzymol.* 1990, 182, 529–539.
39. Dang, H.V.; Chan, Y.P.; Park, Y.J.; Snijder, J.; Da Silva, S.C.; Vu, B.; Yan, L.; Feng, Y.R.; Rockx, B.; Geisbert, T.W.; et al. An antibody against the F glycoprotein inhibits Nipah and Hendra virus infections. *Nat. Struct. Mol. Biol.* 2019, 26, 980–987. [CrossRef]

40. Malaker, S.A.; Pedram, K.; Ferracane, M.J.; Bensing, B.A.; Krishnan, V.; Pett, C.; Yu, J.; Woods, E.C.; Kramer, J.R.; Westerlind, U.; et al. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. USA* 2019, 116, 7278–7287. [CrossRef]

41. Mallajosyula, S.S.; Jo, S.; Im, W.; MacKerell, A.D., Jr. Molecular dynamics simulations of glycoproteins using CHARMM. *Methods Mol. Biol.* 2015, 1273, 407–429.

© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).