



HAL
open science

Mollusc shellomes: past, present and future.

Frédéric Marin

► **To cite this version:**

Frédéric Marin. Mollusc shellomes: past, present and future.. *Journal of Structural Biology*, 2020, 212 (1), pp.107583. 10.1016/j.jsb.2020.107583 . hal-02923864v1

HAL Id: hal-02923864

<https://hal.science/hal-02923864v1>

Submitted on 6 Jan 2021 (v1), last revised 15 Nov 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mollusc shellomes: past, present and future

by Frédéric Marin¹

¹ UMR CNRS 6282 Biogéosciences - Université de Bourgogne - Franche-Comté, 6
Boulevard Gabriel - 21000 DIJON - France

email: frederic.marin@u-bourgogne.fr

Abstract

In molluscs, the shell fabrication requires a large array of secreted macromolecules including proteins and polysaccharides. Some of them are occluded in the shell during mineralization process and constitute the shell repertoire. The protein moieties, also called shell proteomes or, more simply, 'shellomes', are nowadays analyzed via high-throughput approaches. Applied on about thirty genera, these latter have evidenced the huge diversity of shellomes from model to model. They also pinpoint the recurrent presence of functional domains of diverse natures. Shell proteins are not only involved in guiding the mineral deposition, but also in enzymatic and immunity-related functions, in signaling or in coping with many extracellular molecules such as saccharides. Many shell proteins exhibit low complexity domains, the function of which remains unclear. Shellomes appear as self-organizing systems that must be approached from the point of view of complex systems biology: at supramolecular level, they generate emergent properties, *i.e.*, microstructures that cannot be simply explained by the sum of their parts. We develop a conceptual scheme that reconciles the plasticity of the shellome, its evolvability and the constrained frame of microstructures. Other perspectives arising from the study of shellomes are discussed as well.

Keywords: mollusc, shell, biomineral, matrix, shellomics, emergent property

I. Introduction: shellomes

Biom mineralization refers to the dynamic process of formation of mineralized hard parts by living systems. It concerns several phyla across the tree of life, from bacteria and archaea to eucaryotic organisms, including protists, chlorophyll plants, algae, fungi and metazoans (Knoll, 2003). Among these latter, representatives of the phylum Mollusca produce a large array of mineralized structures, such as the radula in chitons, gizzard plates, equilibration organs (statoliths, statoconia), love dart in some pulmonate snails, calcified eggs capsules (all listed in Lowenstam and Weiner, 1989). They comprise also natural concretion or pearls (Vasiliu, 2015) and even amorphous granules used for detoxification (Simkiss, 1977). However, the most known and main type of biom mineralized object synthesized by molluscs is the shell.

The shell is typically an organo-mineral composite, where the mineral phase, calcium carbonate, represents the dominant fraction, and the organic part, a minor one called the shell matrix, yielding 1% or less of the shell weight. It is secreted by the mantle epithelium during calcification and sandwiched in the mineral phase. This matrix - classically retrieved after dissolution of the mineral phase by acid or by a calcium chelator, such as EDTA - has been the focus of a huge number of biochemical characterizations (Krampitz et al., 1976; Weiner et al., 1983). They indicate that the shell matrix is composed predominantly of proteins and saccharides, among which chitin. Small peptides, pigments, metabolites and lipids constitute the other components (Marin et al., 2012). Of all this set of organics, proteins only have been the subject of deepened analyses: they are the focus of the present paper.

In a dozen years, high-throughput approaches, in particular "shellomics", *i.e.* proteomics applied to shell proteins, a terminology of which we are the instigators (Marie et al., 2009), have radically modified our knowledge and our perception of the

shell matrix, giving access to the complete protein shell repertoire, the "shellome" (Marin et al., 2012). Nowadays, the shell repertoire is often perceived as the "molecular toolbox" for constructing a shell. However, we feel that this expression is misleading, because the synthesis of such a structure requires more than the extracellular components occluded in the shell: it also calls for a battery of nuclear, cytoplasmic, membrane-bound and extracellular components incorporated or not into this calcified structure, all these components being encoded by genes that form a gene regulatory network (GRN). The shellome is only a part of this molecular machinery, the terminal part of this network, the tip of the iceberg, so to speak (Marin et al., 2016) but, somehow, this is the most accessible part and the least elusive.

The aim of this paper is to summarize some recent findings on shellomes based solely on high-throughput techniques and to revisit how this repertoire could function. The second aim is to emphasize future research lines that stem from this new knowledge acquired by shellomics. This article picks up where our previous review article (Marin et al., 2016) left off, by focusing exclusively on molluscs and expanding the knowledge that we have acquired since then on this phylum. The main focus are the shell proteins retrieved by dissolution of the shell. We assume that this repertoire comprises key-ingredients that regulate mineral deposition. We also assume that the shell synthesis - from a physiological viewpoint - is predominantly an epithelial cell-driven process (Simkiss and Wilbur, 1989): in this mainstream view, the calcifying extracellular matrix is secreted by mantle epithelium cells via a classical vesicular pathway (exocytosis), self-assembles extracellularly and interacts with the ionic precursors, prenucleation clusters or nanometric amorphous granules that crystallize, get organized into mesocrystals, which are themselves packed in well-defined microstructures. However, we are fully aware that alternative views exist that should be

seriously considered: hemocytes - free circulating cells involved in defense mechanisms, tissue repair and apoptosis - are also involved in the process of shell formation. This hypothesis, published 16 years ago (Mount et al., 2004), is periodically revived and recent findings give consistency to the idea that hemocytes, beyond playing solely a role in shell repair, are also part of the cellular machinery that builds a shell "in steady state". In addition, hemocytes may contribute to deliver, in a coordinated manner, together with mantle epithelial cells, the matrix components, as some recent papers have shown (Li et al., 2016; Song et al., 2019). We do not exclude neither the possibility of a key-role played by exosomes, which may discharge intracellular components (cytoplasmic/nuclear) in the extrapallial space for helping to mineralize the shell (Zhang et al., 2012). These cellular processes should be reexamined, their contribution quantified and integrated in a general shell calcification model that does not exist yet. Whatever the cellular mechanism and the respective contributions of the mantle cells and the hemocytes, this does not modify the central tenets of the present paper.

II. Shellomes and their functional domains

II.1. High-throughput approaches

Until 2008, most of the approaches employed for obtaining the primary structure of shell proteins in molluscs were classical biochemistry or molecular biology. In very few cases, proteins were purified and fully sequenced but most of the time, they were digested or not, partly sequenced and oligonucleotide probes were developed for fishing the transcript, allowing obtaining the corresponding protein sequence. In any case, these reductionist approaches identified proteins "one-per-one", resulting in a limited number of fully-sequenced proteins - less than 50 - in a dozen years (1996-2008), in a disparate set of mollusc species (Marin et al., 2008). It is clear that these approaches favored only

the major proteins of the shell matrix mixture: framework proteins, potential nucleators or CaCO₃-interacting proteins. They completely ignored minor or ultraminor proteins (many of them not always visible on a electrophoresis gel), like enzymes, signaling molecules or immunity-related proteins that may be also key-players in biomineralization. In summary, the "one-per-one" approach did not have any chance to encapsulate the big picture of the functioning of shell repertoires.

Correlative to the decrease of sequencing costs, the increasing use of high-throughput techniques - namely transcriptomics and proteomics - has brought about a drastic change in the shell matrix protein landscape. A such, the first paper published by Jackson and coworkers (Jackson et al., 2006), based on transcriptomics of the ass's-ear abalone, can be considered as milestone work that have opened perspectives. High-throughput approaches have thus identified a wealth of new proteins and subsequently revealed novel functions, not previously envisaged in biomineralization. Above all, shellomics has pinpointed the cross-talk established between the calcifying mantle cells, the shell matrix and the mineralizing front. In short, the use of high-throughput approaches has emphasized the urgent need to revisit molecular cell physiology in biomineralization, in addition to orienting the research field towards complex system biology and emergent properties.

While the use of shellomics marks a real improvement of our knowledge on molluscan shell repertoires, this approach can be flawed by two series of factors: first, the intrinsic properties of the matrix proteins, second, analytical bias. Among the first, the complexity of the mixture may render the digestion less efficient; in an earlier paper, we have shown that treating the matrix in separate fractions ("matrix decomplexification") leads to higher number of hits (Immel et al., 2016). Extensive cross-linking of the shell matrix components is a problem as well as post-translational

modifications. The abundance of long stretches of low complexity domains, not cleaved by the trypsin digestion, does not help either. This may lead to the under-representation of some long domains / full proteins in the proteomic results. Analytical bias include the cleaning procedure, the digestion and the analysis *per se*. About the first one, the manner shells are cleaned impacts directly the number of identified proteins. The example of *Lottia gigantea* speaks for itself: when thoroughly cleaned, the shell was shown to contain few tens of major proteins (Marie et al., 2013); when cleaned gently, more than 300 proteins were obtained (Mann et al., 2012; Mann and Edsinger, 2014). Where is the limit, then? What should be considered as contaminants? The need to use different digestion approaches to improve qualitatively the results is another point to underline, as elegantly demonstrated with the shell matrix of the green ormer (Bédouet et al., 2012). At last, most proteomic investigations performed so far are qualitative and put all identified proteins on the same level (presence / absence), whatever their abundance in the mixture. Truly quantitative proteomics - because of its higher cost - has been tried only in few cases (Mann and Edsinger, 2014). It brings however the most reliable picture of shell protein repertoires, given the possibility that one protein may induce different effects, depending on its concentration in the matrix mixture and on its state, linked to the insoluble phase or in solution. The classical example are polyanionic proteins, which may serve as nucleating agent and promote crystal growth when attached on an insoluble substrate, but act as crystal growth inhibitors when present in solution at high concentration. We believe that threshold effects (concentration) and states (insoluble vs soluble) are key-regulators that fine-tune the system.

II.2. So many data, so little coverage

Table I, based on the compilation of 76 papers published between 2006 and 2020 (not cited here), summarizes the different biological molluscan models that have been investigated for their shellome covered by high-throughput techniques, *i.e.*, genomics, transcriptomics or proteomics. In addition, some of the data are still unpublished. The data represent however very different states of knowledge: in seven cases (nautilus, cuttlefish, landsnail, zebra mussel, etc...), proteomics alone was employed on shell extracts, without the support of transcriptomics/genomics. Peptide sequences only were obtained, giving a very partial coverage of the shellomes. In six cases, proteomics was applied but based on an existing and publicly available transcriptome. In a couple of additional cases, such as for the mytilid *Pteria penguin*, one or more transcriptomes were generated from which putative shell protein sequences were deduced via *in silico* investigation, but without accompanying shellomics. In most of the studies however, an integrated approach combined one or more transcriptomes and one or more shellomes, acquired in parallel from the same set of specimens. At last, three models, the limpet *Lottia*, the Akoya pearl oyster *Pinctada fucata* and the edible Pacific cupped oyster *Crassostrea gigas* (renamed *Magallana gigas* since 2017) benefit from a full coverage, comprising a genome, one or more transcriptomes and one or more shellomes.

Bivalves represent the 'best-covered' mollusc class, with 21 genera and 30 species. However, considering the size of this clade (about 12,000 living species) and its huge diversity, this represents very little: on the 46 superfamilies commonly accepted by taxonomists, the shellomes of few representatives of only ten of them have been studied and whole sections have not been explored yet. For reasons easy to understand, the models of economic interest, including the pearl oyster and its different geographical species (Japanese, Australian and Polynesian) and the edible species - the mussel and the Pacific cupped oyster - take the lion's share. They belong to the pteriomorphian

subclass, and represent nacro-prismatic and foliated microstructures. The small subclass of paleoheterodont bivalves, comprising mostly freshwater mussels, is relatively well represented by 5 nacre-forming unionoid genera (*Unio*, *Cristaria*, *Hyriopsis*, *Elliptio*, *Villosa*). One has to note that the research on these species is also driven by economic interest, the genera *Cristaria* and *Hyriopsis* being exploited in China for their ability to make pearls. Heterodont bivalves - the most diversified today's clade - are modestly represented by six genera.

Gastropods, the biggest molluscan class (from 80,000 to more than 120,000 living species), are covered only by 11 genera, comprising 15 species. The limpet genus *Lottia* and the nacreous abalone *Haliotis* concentrate half of the studies on gastropod shellomes. The limpet, which genome was the first to be sequenced, is a representative of the most basal clade, patellogastropod order, while the nacreous abalone occupies a basal position within vetigastropods. On the other spectrum of gastropod phylogeny, one finds Heterobranchs representatives, including land (*Cepaea*, *Helix*) and freshwater (*Lymnaea*) snails and, marine representatives: the sea hare (*Aplysia*) with an internal shell and planctonic gastropods, pteropods. The huge clade Caenogastropoda (60% of the gastropod diversity) is only represented by two genera (*Pomacea*, *Babylonia*). Obviously, the picture is far from complete and whole gastropod orders are absent.

Cephalopods are represented by four genera, *i.e.*, four species including the nacreous-shelled nautilus and three coleoids: the cuttlefish (*Sepia*) the ram's horn squid (*Spirula*), both having an internal shell and the paper nautilus (*Argonauta*, data not yet published). In this last case, the shell is a nature oddity and an apomorphy of argonautids, since it is not homologous to true shell. It is indeed an eggcase secreted by specialized arms of females only.

So far, no shell repertoires of scaphopods (tusk shells), of monoplacophorans and of polyplacophorans (chiton) have been published. Shell-less molluscs (Solenogastrea, Caudofoveata), that secrete spines or sclerites, are not the subjects of any study.

In this brief overview, most of the studies have a clear scope to gain knowledge on the shellome in order to understand the process of shell formation *per se* or for evolutionary purposes. Additional studies use this repertoire as molecular markers for studying the impact of environmental changes - in particular OA (Ocean Acidification) - on shell calcification. It is puzzling to observe that shellomes can change rather drastically when environmental parameters are modified (Timmins-Schiffman et al., 2014; Wei et al., 2015), which suggests, among other findings, that they have a great plasticity. Interestingly, most of the repertoires are obtained from adult specimens. However, few studies focus on larvae at different developmental stages. Zhao et al. (2018) have recently shown that the shell larval secretory repertoire of the edible oyster *C. gigas* is very different from that of adult stages. This finding is rather opposite to that observed in the freshwater gastropod *Lymnaea stagnalis* by Herlitze et al. (2018), who, by using ISH techniques, have shown that larvae express most of the transcripts that are later expressed in adult mantle tissues. Clearly, the study of the different developmental transitions (trochophore --> veliger --> juvenile) in term of shellome expression is a key-issue for the future.

II.3. Functional domains for fabricating a shell

One of the first thing that molecular biologists do when confronted to whole shell repertoires consists in associating protein sequences to molecular functions. This is classically done by blasting sequences against large datasets. In the case of shell protein, this task is complicated due to absence of homology with known proteins. Secondly,

many shell proteins exhibit a modular organization of their primary structure, suggesting their ability to perform very different functions. Thus, instead of trying to classify proteins according to their function - an impossible task - it is more relevant to establish a classification of functional domains. A functional domain is a subunit of a full-length protein sequence and corresponds to a conserved module that exists independently from the rest of the sequence. A domain has usually a well-defined tertiary structure (which itself is constrained by the succession of secondary structures along the sequence), required for its functionality. In some cases, there is a complete superimposition between domains and proteins when these latter are made of a single domain and thus, supposed to perform a single function: for example perlucin, a C-type lectin domain-containing protein of the nacre abalone (Mann et al., 2000).

Figure 1 is an attempt to summarize some functional domains commonly encountered in mollusc shellomes. The functional domains can be divided in two broad categories: those clearly identified by using Blastp or CD-search (Conserved Domains) at NCBI, owing to their sequence similarity with known functional domains; those not identified as such or which do not fall in classical domain categories. This second category comprises all low complexity domains (LCDs) or repetitive low complexity domains (RLCDs). Figure 1 groups domains according to their molecular properties, signatures or functions and links them to biological / cellular processes. One given domain can belong to different categories (it can be an enzyme and a sugar-interacting molecule, for example) and be involved in different biological processes: signaling and framework structuring, or framework structuring and protection *sensu lato*. clearly, many domains have overlapping functions.

II.4. Domains of identified functions

These domains cover a large set of functions, briefly exposed hereunder. It is however important to notice that, while being easily targeted from sequence analysis, very few of them have given rise to *in vitro* functional assays. To our knowledge, only carbonic anhydrase, tyrosinase and chitin-binding activities were measured, either from recombinant proteins, or from bulk or semi-purified shell extracts.

Among the most prominent domains, those that are typical of extracellular matrices (ECMs) or ECM-binding molecules. They include non-exhaustively collagen-like, von Willebrand type A, thrombospondin-like, decorin, fibronectin-like, laminin, filament-like, SPARC (secreted protein, acidic, rich in Cys), EGF-like, IGF-BP, Sushi-like (CCP modules), ependymin-related, mucin-like, zona pellucida. Von Willebrand type A domains are commonly found in many, if not all, matrices and can bind to many protein ligands, including collagens. Most of these actors play a role in structuring the 3D framework, by assembling into sheets, fibers, or gels. They are usually accompanied by ECM-binding domain-containing proteins, such as integrins. Some ECM members (EGF-like, IGF-BP) or ECM-binding members have a signaling function.

The very heterogeneous group of enzymatic domains comprises carbonic anhydrase (CA), tyrosinase, peroxidase, three enzymes known for a long time in their involvement in shell mineralization (Timmermans, 1969), but also "less evident" players: cyclophilin, arginine kinase, glutamine amino transferase, laccase, diverse proteases and a large set of enzymes interacting with the sugar moieties. CA domains are involved in the conversion of carbon dioxide into bicarbonate and these domains exhibit a large plasticity since they can be associated to different domains, in general of low complexity types (Le Roy et al., 2014). Tyrosinase domains (catechol oxidase) are involved in cross-linking ('sclerotization'), in shell pigmentation (by catalyzing the synthesis of melanins) and in defense mechanisms (encapsulation of parasites).

Peroxidases may be involved in periostracum formation by cross-linking fibrous proteins to form insoluble protease-resistant polymers (Herlitze et al., 2018). Laccase domains have a similar function. Cyclophilins (also defined as peptidyl-prolyl cis-trans isomerase) may act as extracellular chaperones, by allowing correct folding of other matrix proteins. The set of protease domains (SCP, cathepsin, metalloproteinase) may be involved in matrix degradation, remodeling and maturation. The respective roles of arginine kinase and glutamine amino transferase domains are less clear in the shell formation context.

The third category comprises domains that interact with the saccharidic moieties of the matrix. Generally, the shell saccharides are neglected by most of the shellome studies but everyone agrees to say that they are important in shell formation, starting with chitin. Although the quantity of chitin in shell needs to be reevaluated down in some models, as a recent paper has shown (Agbaje et al., 2018), this polymer of N-acetylglucosamine plays anyway a structural role and is supposed to anchor many matrix proteins. Shellomics has identified a bunch of enzymes involved in chitin formation (chitin synthase) and remodeling and degradation (chitinase, chitin deacetylase, chitotriosidase, chitobiase). Other enzymatic or non-enzymatic chitin-binding domains have been also detected, in particular peritrophin A. Lectins - sugar-binding proteins - have been found as well, like ficolin or perlucin. Many shell-associated lectins are calcium-dependent (C-type lectins). More than having a structural role as framework constituents, many of them may exert a function in immunity and protection, and potentially in signaling. Finally, polysaccharides other than chitin are also presumably 'detected' by shellomics, via their interacting protein partners, like sialic acid-binding protein or diverse types of glycosyltransferases.

Domains involved in divalent cation-binding are heterogeneous. They comprise calcium-binding and iron/copper-binding domains. Calcium-binding domains are of three types, EF-hand, Asp/Glu-rich and ependymin-related, corresponding to very different biological functions. EF-hands are high affinity-low capacity calcium-binding domains characterized by a 3D structure that traps one calcium cation. It is unlikely that such domains, observed for example in calmodulin and calreticulin are involved in providing the cationic precursors of calcium carbonate. They would rather be involved in signaling, or in interacting with other matrix members. Asp/Glu-rich, on the contrary, are low affinity-high capacity calcium-binding proteins and they are probably involved directly in calcium carbonate mineralization (see below). Ependymin domains are very often detected in shellomes but, besides binding calcium ions, their biological function in biomineralization remains elusive. Iron/copper-binding ions domains comprise that of ferritin, of transferrin (both Fe-binding) and of hephaestin (Fe/Cu-binding). Although frequently detected in mollusc shellomes (Oudot et al., 2020), their exact function is not clarified yet.

Domains related to immunity and protective mechanisms comprise a set of protease inhibitors, such as Kazal-1/2, Kunitz, serpin (serine protease inhibitor), TIMP (tissue inhibitors of metalloproteinase), VIT, CD109. These domains were not suspected before the use of shellomics. However, they are found in every shellome without exception. They are supposed to belong to matrix protecting system that prevents its extracellular degradation. Other members of the immunity-related system are MG2 (α -2 macroglobulin), immunoglobulin-like domains, or lipocalin-like, which are very abundant in some models (Arivalagan et al., 2016). Diverse lectin domains may also be part of this protective system.

Most of the shell repertoires studied so far comprise also a number of

cytoplasmic and nuclear proteins, suspected *a priori* to be cellular contaminants. However, their persistence in skeletal tissues that have been thoroughly bleached with sodium hypochlorite before extraction leads us to reconsider this point of view and estimate that they may be, after all, part of the shellome. They include cytoskeletal proteins and their associated partners as well as nuclear proteins. The cytoskeletal proteins are actin and tubulin, and their binding partners, like myosin (binding to actin via its head domain) or elongation factors while the nuclear ones are histones or histone-related. Actin has been shown to be associated to ECM in mineralizing / demineralizing vesicles in cartilage (Holliday et al., 2020), while, in molluscs, Weiss and coworkers demonstrated that there is a molecular link between the cytoskeleton and the calcifying extracellular matrix, via a chitin synthase that exhibits a myosin head domain (Weiss et al., 2006). The detection of histones, *i.e.*, very basic proteins (lysine/arginine-rich), in matrices associated to calcium carbonate biominerals is frequent. It is generally believed that they may function as antimicrobial agents, for example in eggshell (Réhault-Godbert et al., 2011), but also in molluscs, with molluskin, an antimicrobial peptide derived from histone H2A (Sathyan et al., 2012). Histones / histone-like domains would then be considered as a part of the protective system, together with the proteins that exhibit immunity-related domains.

II.5. The puzzling case of LCDs / RLCDs: a new paradigm in biology?

Beside the functional domains briefly exposed above, the big deal that makes shell protein repertoire so peculiar are the LCDs/RLCDs, in other words, the Low Complexity Domains, or Repetitive Low Complexity Domains, also called "compositionally biased regions". These domains are characterized by the predominance of one or few amino acid residues. They can constitute a major part of all

domains identified in a shellome, not only from a qualitative viewpoint but also quantitatively. Many major shell proteins exhibit indeed one or more low complexity domain. Although there is no superimposition between the two concepts, many of the LCDs/RLCDs are intrinsically disordered regions (IDRs), *i.e.*, regions that do not exhibit any specific secondary or tertiary structures. A recent estimate across UniProt database indicates that about one third of compositionally biased regions in proteins have both 'substantial intrinsic disorder and structure' (Harrison, 2018).

The most known of these LCDs/RLCDs are aspartic acid-rich ones. Identified in the seventies (Weiner & Hood, 1975), confirmation of their existence via full sequence acquisition came much later, with MSP-1 and aspein (Sarashina and Endo, 1998; Tsukamoto et al., 2004). Nowadays, shellomics has shown that many shell proteins possess long sequences enriched in aspartic acid residues. These domains have a strong affinity for calcium ion, bound with moderate affinity. They also have strong affinity for calcium carbonate crystal surfaces. It is still believed that they may act as nucleators and that they can inhibit crystal growth, when in solution. It is also very likely that they play additional functions by creating, when concentrated, a polyanionic microenvironment favorable to concentrate calcium ion. Serine-rich or, alternatively threonine-rich domains are also well-known LCDs. Because they exhibit multiple phosphorylation sites, they can exert similar effects as the Asp-rich domains.

Other well-known LCDs/RLCDs are the hydrophobic ones. They are usually enriched in glycine or in alanine, more rarely in valine, leucine or isoleucine. Similarly to Asp-rich members, proteins enriched in hydrophobic domains were known for decades in the insoluble mollusc shell matrices, since they were collectively defined as "silk fibroin-like" proteins. In old models, they were believed to act as structural framework, "molds" in which crystals could growth. This "static" view was challenged by the

dynamic concept that hydrophobic proteins are synthesized as a gel in which crystals grow until reaching confluence. Then, between crystals, this gel solidifies, cross-links and gets insoluble. We cannot exclude the possibility that hydrophobic domains, during mineral formation, contribute to expel water molecules from the system and promote crystal formation or render anhydrous amorphous minerals, thus stabilizing them.

Other LCDs/RLCDs have been discovered: basic ones (lysine- or arginine-rich). Identified by a "one-per-one approach" about fifteen years ago (Yano et al., 2006; Zhang et al., 2006), shematrin and KRMP domains were confirmed in many shellomics studies. Their 'polycationic' charge properties make them putative candidates for interacting with Asp-rich domains, but also with bicarbonate and carbonate ions. If partly cleaved, basic fragments may as well exhibit bactericidal properties. The other types of LCDs/RLCDs comprise methionine-, cysteine-, asparagine-, glutamine- and proline-rich domains. The functional significance of these domains is unclear. In non-biomineralizing models, many domains of these types are supposed to play protein-protein interaction function and signaling. For example, Q-rich domains (glutamine) are known for activating transcription factors. However, the functions of all these LCDs/RLCDs are very elusive: their ubiquity in all shell repertoires and their diversity are puzzling. To give an example, we recently obtained the transcriptome of the mantle cells supposedly involved in the secretion of the prismatic calcitic layer of the fan mussel *Pinna nobilis* (Marin, Jackson, unpublished data). Among the hundreds of transcripts identified, a large part of them encode proteins with LCDs/RLCDs that are suspected to be proteins of the calcitic prism shellome. The question arises as to whether all these proteins are necessary for the formation of mineralized structures which appear relatively simple. If so, then, what are their respective functions?

New concepts are emerging that try to answer this question. If validated by bench experiments, they would constitute a major change of paradigm in protein biochemistry that considers proteins as the functional units of the cell. In a provocative paper published one year ago (Pancsa et al., 2019), it was suggested that the well-established paradigm that associates one protein to one function (and one structural conformation) may simply not apply to proteins that exhibit intrinsically-disordered regions (IDRs). Using different molecular systems as examples - among which biomineralization - these authors suggest that IDR-containing proteins, taken separately, do not have a function by themselves, but that novel functions emerge when these proteins 'form dynamic and non-stoichiometric supramolecular assemblies'. In short, by acting collectively, the IDR-containing proteins would form peculiar microenvironment, such as a liquid-liquid phase separation, gels, (something approaching the PILP process (Gower and Odom, 2000)). These supramolecular labile constructs would generate mineralized structures that 'cannot be simply described by, or predicted from the properties of the isolated single proteins'. In short, Pancsa and coworkers plead for considering biomineralizations as emergent systems and for investigating them at supramolecular level, instead of trying to assign a property of the whole mineralized structure to properties of the proteins that constitute it. We completely share this point of view (Marin et al., 2008, p 263) and feel that this concept is remarkably illustrated by the example of mollusc shell microstructures.

III. The conundrum of shell microstructures morphogenesis: from continuity of the shellomes to discontinuity of microstructures.

III.1. Shell microstructures: diversity and constraints

Indeed, all mollusc shells consist of the superimposition of few mineralized layers - usually from 2 to 4 - each of them being characterized by a specific ordering of the crystallites that constitute this layer. All crystallite arrangement types are grouped under a generic terminology, "shell microstructures". Six broad types of microstructures were defined for all shell-bearing molluscs, based on their overall morphology: laminar, prismatic, crossed, homogeneous, helical and spherulitic (Carter and Clark, 1985). The four first subdivide in many subtypes: for example, the "laminar" type groups all microstructures with flat units oriented parallel or nearly parallel to the general depositional surface (Carter and Clark, 1985); it comprises not only the well-known "nacreous" microstructure, but also the semi-nacreous, the lamello-fibrillar, the crossed-bladed, foliated and semi-foliated ones. The "prismatic" type describes 'elongated crystalline objects that are rectilinear or curved and which opposite long sides are parallel' (Carter and Clark, 1985). It also gathers very different sub-types such as the simple, the fibrous, the spherulitic and the composite prismatic ones, all of them being subdivided again; the "crossed" type reunifies microstructures with at least two oblique directions of their elongate structural units relative to the depositional surface. It encompasses the classical crossed-lamellar, the complex crossed-lamellar, the dissected crossed-prismatic plus other sub-types. The "homogeneous" type characterizes microstructures with no apparent organization of their crystallites, for which can be distinguished homogeneous *sensu stricto* and granular sub-types depending on the grain size. Finally, the two last types are rare: helical structures are restricted to pteropods, a class of minute planctonic gastropods, while spherulitic are observed in shell repair processes of many molluscs.

What makes the world of shell microstructures fascinating is that these microstructures always combine with each other in superimposed layers. While it is not

clear why molluscs proceed in such a way to construct their shell, it is generally believed that making layers of different mechanical properties has a lot of evolutionary advantages, in particular avoiding crack propagation. This innovation, almost as old as the shell itself - somewhere in the Lower Cambrian (Kouchinsky, 2000) - is remarkably illustrated by bivalves, the emblematic class for studying shell microstructures and their combinations (Taylor et al., 1969; 1973; Carter, 1990). Bivalves exhibit a certain diversity of combinations of microstructures, 47, according to Uozumi and Suzuki (1981). In the Venerid family alone, 12 combinations of 3-4 layers of 5 different microstructures were distinguished (Shimamoto, 1986). Some combinations are extremely frequent: nacre (internal layer) is most of the time associated to prisms; foliated layers are often associated to prisms too, crossed-lamellar to complex crossed-lamellar or to homogeneous/granular. Some other combinations are extremely rare or simply do not exist, like nacre and crossed-lamellar.

To some extent, microstructures and their combinations carry valuable taxonomic information, utilized by palaeontologists in parallel to other morphological characters. There is no question about the fact that shell microstructures are constrained: one given clade exhibits always the same combination of microstructures and this combination is rather stable at genus or family levels, over several million years. The pattern of shell microstructures, when superimposed to a phylogenetic tree (Taylor et al., 1973, Carter, 1990), marks clear separations between major clades at the ordinal, superfamily or family levels. For example, nuculoids, the most "primitive" bivalve order, *i.e.* the sister-group of all other bivalve orders - are always nacro-prismatic with (mostly) lenticular nacre. Veneroids - usually considered as 'modern' bivalves because many today's clades have accomplished their radiation in the Mesozoic - only comprise specimens with 'crossed' or 'homogeneous' types, but never

nacro-prismatic, contrary to unionoid bivalves that are invariably nacro-prismatic but entirely aragonitic. Foliated calcitic microstructures are restricted to all pectinoids (scallops) and some ostreoids.

However, the signal given by microstructures and their combinations is not phylogenetic. It is a complex signal, blurred by what appears to be multiple convergent evolutions. Attempts were done to sketch phylogenetic trees of bivalve microstructures (Taylor et al., 1973; Uozumi and Suzuki, 1981), but without consensus. In the nineties, crystallographic properties (diffraction patterns visualized by pole figures) were associated to shell microstructures in a "crystallography-based phylogeny" (Chateigner et al., 2000). This produced interesting results, but in a limited number of specimens. This attempt was not extended further to a larger number of representatives of all mollusc classes.

Above all, some bivalve clades, supposedly monophyletic, exhibit heterogeneity of their microstructural patterns: the pteriomorphid subclass comprises orders/superfamilies that are dominantly nacro-prismatic or foliated, together with arcoids that exhibit exclusively crossed-lamellar and complex crossed-lamellar microstructures. This case illustrates the absence of simple correspondence between clades of high taxonomic rank (order, superorder, subclass) and microstructural patterns: two clades, phylogenetically distant from each other, may have very similar microstructures. Reversely, closely related clades may have completely dissimilar microstructures: within the Ostreida order, pinnids are invariably nacro-prismatic and ostreids, foliated. In this context, what do the skeletal repertoires teach us about a potential link between repertoires and microstructures?

III.2. Linking shell matrices and microstructures: the impossible quest

How epithelial mantle cells control microstructures morphogenesis is a long standing question that necessarily involves shell protein repertoires. Cause-and-effect relationships between shell matrices and microstructures are not new but were raised logically from the beginning of modern biochemical analyzes, when it became obvious that the matrix was the "sculptor of microstructures" so to speak. This question was tackled first by obtaining amino acid compositions of bulk matrices from separated layers: a profusion of articles mentioned the difference in amino acid compositions of dissociated shell layers of the same species (Grégoire, 1972). Because they are easy to dissociate, nacreous and calcitic prismatic layers were often the focus of such analyzes, more than any other microstructures. It was notably found that calcitic prism matrices were enriched in acidic amino acids (Asp and Glu) in comparison to nacre ones (Hare, 1963). At that time, it was difficult to assess whether these differences were due to the microstructures (prisms vs. nacre) or to the mineralogy (calcite vs. aragonite). Later on, with the development of fractionation techniques, the comparisons between microstructures were extended to chromatography (Weiner, 1983; Samata, 1990) and electrophoresis (Marin et al., 1994) and both sets of techniques evidenced biochemical differences. At last, serological comparisons made from polyclonal antibodies elicited against soluble matrices of dissociated layers gave puzzling results, when these antibodies were cross-tested (Marin et al., 1994) or tested against several matrices extracted from specimens of well-defined microstructures (Marin et al., 1999, 2011). Each time, patterns that owe nothing to chance were obtained in relation to microstructures but these patterns could not be explained easily, mainly because of some technical limitations inherent in the use of polyclonal antibodies.

With the publication of an increasing number of mollusc shell protein sequences during the 2000s owing to the "one per one" molecular biology approach, it became

apparent that some proteins were microstructure specific. In the inventory of shell proteins published twelve years ago (Marin et al., 2008), many of them were supposed to be associated to a single microstructure. However, complete demonstrations of belonging to one or other of the microstructural types have only been carried out in few cases: mucoperlin is nacre-specific (Marin et al., 2000); prismalin-14, aspein, shematrins and KRMPs (Suzuki et al., 2004; Tsukamoto et al., 2004; Yano et al., 2006; Zhang et al., 2006) seem prism-specific; other proteins, like nacrein, are found in the two layers. In 2008, the graph plotting the shell proteins in a diagram where the x axis is the isoelectric points and Y axis, the molecular weights, evidenced a partition of proteins according to microstructures and/or mineralogy (Marin et al., 2008). Shell proteins associated to calcitic prisms were either more acidic or more basic than that associated to nacre (aragonite). However, the reason of this partition was unclear.

Finally, shellomics was applied on separated layers, in particular on macro-prismatic shells. On the Polynesian pearl oyster, it was observed that prism and nacre shellomes were constituted of very different repertoires (45 prisms specific proteins vs 30 nacre specific, overlap of 3 proteins between the two layers, Marie et al., 2012). On the Japanese pearl oyster, a relatively similar result was found with however a higher overlap between the two layers (17 proteins on a total of 72, Liu et al., 2015). On the same species, an in-depth shellomics generated 366 proteins, with 127 nacre-specific, 132 prisms-specific, and 107 overlapping proteins (Du et al., 2017). On the mussel genus *Mytilus*, two studies on nacre and prism layers also showed that the two repertoires possess layer-specific protein markers and shared proteins (Gao et al., 2015; Liao et al., 2015). At last, on the abalone, the most comprehensive analysis identified 297 nacre proteins, 350 prisms ones (total: 448 with 199 overlaps; Mann et al. (2018). Reduced to major proteins, these numbers fall to 51 for nacre, 43 for prisms and 17 shared.

Interestingly, many of these latter were detected with notably different abundances in the two layers. All the analyses performed so far converge to the idea that prism and nacre have relatively different shellomes with, however, some shared protein tools. It is then logical and tempting to assert that what makes the microstructural differences is due to the layer-specific proteins that may work in synergy to produce the microstructure. It is to note that such analyses were performed solely on few nacro-prismatic shells. We do not have any idea if similar conclusions can be drawn from other combinations of microstructures such as composite prismatic and crossed-lamellar, or crossed-lamellar and homogeneous.

III.3. Shell microstructure: explained solely by physico-chemical / thermodynamic principles?

Taking the opposite view of what has been exposed above, a current of thought tends to deny - or at least attenuate - the role of the shell matrix in generating the complex shapes of microstructures. On the contrary, it gives the full focus to physico-chemical laws and concepts. This current of thought - largely taken up by biomimetics or bioinspired chemistry - is rooted in the pioneering work of D'Arcy Thompson, "On Growth and Form" (1917, revised ed. 1942). This scottish naturalist, arguably considered as the founding father of biomathematics, was himself influenced by precursor chemical studies of Rainey (1857) and P. Harting (1872) who showed that complex mineralized shapes could be produced abiotically, from relatively simple colloid media such as gelatin or albumin containing calcium chloride, in which sodium carbonate was added. In his experimental design, Harting described calcospherites that, at later stages, formed pavements of polygones, rather similar in their morphology, as

the ones found in the simple prism type shell microstructure, when observed in section orthogonal to the prism elongation axis.

These ideas were also explored by different authors (Grigor'ev; 1965; Ubukata, 1994; Checa et al., 2006). Competition for space between neighboring crystals has often been inferred as a powerful mechanism either for generating elongated prism-like crystals that look like microstructures observed in some bivalve shells (Ubukata, 1994) or for explaining why nacre tablets orient progressively their b-axis parallel to the direction of propagation of the lamella (Checa et al., 2006). In the case of prisms, relatively simple descriptors of the initial state, like the shape of the surface on which crystals grow (flat or uneven), the density of nuclei on this surface or the kinetics of crystal growth can modulate the shape of the produced crystals. From a completely different perspective, generating shell microstructures can be modeled on computers: cellular automata (Wolfram, 2002), based on iterative rules, can also produce 'pattern'. Used for mimicking the color patterns observed on conid gastropod shells (Meinhard, 2009), algorithms can be developed to generate 3D microstructures as well.

In the last few years, a series of remarkable papers has rekindled the approaches from materials physics, by combining analytical techniques (EBSD, synchrotron-based microtomography, XRD) and mathematical modeling tools commonly employed for describing man-made materials, which are based on thermodynamics, geometric and kinetics considerations (Bayerlein et al., 2014; Zöllner et al., 2017; Reich et al., 2019; Schoeppler et al., 2019). For example, the growth of the calcitic prisms of the fan mussel *Pinna nobilis* has been described according to the theories of normal grain growth. It is striking to observe that the phenomenon of 'prism coarsening' - as prisms grow inwards from the periostracal layer and perpendicular to it - is well mimicked by mathematical models. The coarsening (corresponding to a reduction of the grain boundary area, *i.e.*, to

a decrease of the free energy of the system) results from the reduction of the interface area between the prisms and the periprismatic organic sheath. The coarsening is expressed by the progressive shrinkage and disappearance of some prisms counterbalanced by the growth in diameter of some others. Questions that arise are: can these concepts that accurately describe the growth of 'simple' microstructures apply to other more complex microstructures, such as crossed-lamellar? What is the role of the organic matrix then? How to reconcile the model applied to the prisms of *Pinna nobilis*, briefly described above, and the fact that the mantle tissues of this model organism expresses hundreds of different transcripts (Marin, Jackson, unpublished data) encoding secreted proteins, most of which exhibiting low complexity domains? Clearly, a gap must be filled between physics and biology for apprehending the genesis of shell microstructures. In the next paragraph, we tentatively try to fill this gap and unify the two approaches.

III.4. Alternative view: microstructures, emergent properties and attractors

We propose an alternative view of what could be the shell matrices, as shown by Figure 2. This representation postulates that microstructures are emergent properties, *i.e.*, properties that cannot be solely explained by the characteristics of their separated constituents, in particular shell matrix proteins taken one after the other. It considers that proteins of the shellomes are elements of a regulatory network and that they work in a cooperative way to produce a given microstructure. Our scheme is intuitive and hypothetical. It tries to reconcile some of our experimental observations deduced from shellomics, namely that two protein repertoires associated with similar microstructure (nacre for example) can be rather different (in spite of overlaps) from the point of view of their protein sequences or of the presence of specific functional domains. This scheme

also takes into account the reverse possibility, *i.e.* that rather similar protein repertoires may lead to very different microstructures. The model that we propose is a construction of the mind, a concept for the moment purely theoretical, but which has the advantage of being visual - easy to catch - and that can be tested, when a large number of secretory shell repertoires will be published and 'normalized'. We insist on the fact that normalization will be a prerequisite step in order to bring shellomes to the same level of knowledge, so that they can be compared all together (this is absolutely not the case, today). Normalizing all shellomes will take into account major, minor and ultraminor proteins in the matrices. This prerequisite will drastically limit bias. One technical improvement will be to employ quantitative proteomics - rarely applied today - to have an accurate idea of the amount of each protein and functional domain in the shellome. As underlined in § II.1, we believe that the regulation of mineral deposition is not only a question of presence/absence of a given protein or domain in the shellome but deals also with its abundance in the reactive mixture.

A first step will consist of integrating (*i.e.* coding) all the biochemical characteristics of each shellome in the form of a 'character matrix' (array), similarly to what is done in multivariate analyses. These characteristics may have the same weight; alternatively, some may be pondered, with a higher weight, such as the percentage of matrix (per gram of shell powder), the ratio between soluble and insoluble fraction, the associated mineralogy (calcite *vs.* aragonite). Secondly, each protein of a given shellome will be described by its relative abundance in the mixture and by its primary structure: this latter information can be condensed to the succession of functional domains, from N- to C-termini. To this end, the conventional pfam code (with five-digit numbers) for domain recognition can be used (PF00264 corresponds to tyrosinase domain, PF00194, to CA domains, etc...). Additional sequence descriptors should be designed accurately for

LCDs/RLCDs., something left to do. Complementary information, like the classical descriptive parameters of each protein (mass, isoelectric point) can be added too. This encoding process will be performed for all proteins of a shell layer proteome. One would ultimately end up with a planar representation (expressing the highest variance), referred to as "*the general microstructural field*", in which each dot, integrating all the data of the character matrix, would represent a unique secretory repertoire of one shell layer of a given mollusc species. In this plan, the dots close to each other correspond to very similar matrices, the distant dots, to very different matrices. Dots can be grouped in clouds or areas, according to their belonging to a given microstructure. These areas, or "fields" have any shape, delimited according to the greater or lesser dispersion of the dots they contain. Each field corresponds to a given microstructure: in part A of Figure 2, one sees the "mother-of-pearl field", the "prism field", the "foliated field", the "crossed lamellar field". For the simplicity of the graph, only these four microstructures were represented.

Part B of Figure 2 shows the same figure but with a topography, *i.e.*, in three dimensions: each microstructural field then behaves like a funnel, the basal outlet of which is plumb with the barycenter of the cloud of dots. Each funnel is separated from its neighbor by a crest. This representation expresses the fact that no matter where one is in a given microstructural field, one always ends up with the formation of a single microstructure. Thus, considering the "nacre" field, if a bead is rolled from point "a" or from point "b", which are diametrically opposite (which means that they correspond to extremely different matrices), the ball will follow the slope of the funnel and end up in exactly the same place. This reflects the fact that, within one given microstructural field (the nacreous one for example), regardless of the starting point (*i.e.*, the secretory repertoire) and the path taken (*i.e.*, the biochemical reactions between the partners of

the repertoire), the end result is always the same, the formation of a well-defined microstructure. In the example, points "a" and "c" are relatively close, reflecting partial similarities of the shellomes. However, two beads, rolled from these two points, will end up in different funnels, *i.e.*, different microstructures. Such a vision conciles two antagonist concepts: first, the large plasticity of the shellome; there are multiple starting points and biochemical trajectories for reaching the same end result; second (which stems from the first), the shellome plasticity is "constrained" by attractors. In other words, in this model, the microstructures are governed by 'attractors' in the mathematical sense of the term. These attractors, for the moment undefined, can be - among other 'objects' - thermodynamic and physical constraints, such as crystal growth kinetics. While it is easy to go from the nacre field to the prism one (a short shift in the graph, corresponding to a relatively slight change in the shellome composition), there are also insurmountable barriers, and one cannot go from the nacre field to the crossed-lamellar one. Such a representation also explains that from a quasi-continuous set of points that fill the microstructural field, one ends up with discontinuous, discrete microstructures, without intermediaries. As shown in the insert C of figure 2, the general frame exposed here can be refined with the existence of microstructural fields comprising not one but two or three outlets (funnels), which reflect the existence of close microstructural subtypes. This may be the case for granular and homogeneous microstructures on the one hand, or for crossed-lamellar and complex crossed-lamellar ones, on the other hand. These microstructures are twinned and it is likely that they belong to the same microstructural field, in term of shellomes resemblance.

IV. Conclusion: other territories to explore, discover or revisit by studying shellomes

IV.1. Shellome macroevolution

As molluscs started to mineralize their shell around the Proterozoic/Cambrian transition ("the Cambrian explosion", or P/C transition), more than 540 years ago, it is logical to assume that shellomes emerged synchronically. How molecular functions were recruited for shell calcification, how they evolved across the Phanerozoic are questions that are central in biomineralization studies (Marin et al., 2014). Owing to shellomics, major advances have been made in a decade. We just begin to understand some aspects of shell repertoires macroevolution, as witnessed by the publication of a dozen fundamental articles in the last few years (Kocot et al., 2016; McDougall and Degnan, 2018; Mann et al., 2018; Song et al., 2019). In particular, shellomics has allowed comparing and visualizing repertoires between them (via Venn representations or via Circoletto diagrams) and sketching evolutionary scenarios over the Phanerozoic times. These fascinating aspects require in themselves a full review and an informed discussion, far beyond the scope of this paper. We simply sum up here few important factual clues that cover two aspects of shellome evolution: the 'oldity' of functional domains, the molecular mechanisms by which shell proteins evolve.

About the first aspect, a nuanced answer is provided that embraces two phases of recruitment: on the one hand, some functional domains may have been recruited early for shell fabrication and may be part of a 'core' toolkit, essential for biomineralization (Arivalagan et al., 2017): functional domains such as CA, tyrosinase, peroxidase, chitin-binding, Von Willebrand Factor A, C-type lectin, are obviously ancient. Presumably in existence in the Proterozoic and shared by all bilaterians (Aguilera et al., 2017), they would have been coopted the first time for mineralization around the P/C transition. An early recruitment of functional domains does not mean that they were equally spread in all lineages and 'left as is'. These early-recruited domains have been submitted to a very

complex evolutionary history, marked by losses, secondary recruitments, independent expansions in some lineages and multiple convergent evolutions. The best illustrations of these complex secondary evolutions are provided by carbonic anhydrase (Le Roy et al., 2014) and tyrosinase (Aguilera et al., 2014). However, an early recruitment of functions may not represent the dominant evolutionary event about mollusc shellomes.

Indeed, on the other hand, one of the most unexpected outcomes of shellome comparisons suggests that numerous functional domains are of 'recent' origin, *i.e.*, are lineage-specific, which means that a part of the shellome is constructed from rapidly evolving genes (Jackson et al., 2006; Kocot et al., 2016). This has been shown via a phylostratigraphic approach (Aguilera et al., 2017) but also via the compared analysis of orthologue genes (such as lustrin) within a genus (Jackson et al., 2017). It seems that numerous - if not all - LCD/RLCD-containing proteins, evolve rapidly, since they do not possess homologues in neighboring lineages. It is fascinating to observe that this rapidly evolving part may quantitatively be dominant in the shellome of several studied models, when compared to conserved, well-identified, domains. In other words, this means that the construction of microstructures that are stable and invariant over geological times is largely based on 'molecular tools' that are themselves not 'conserved', a very counterintuitive concept. This is an interesting conclusion that, once again, pleads in favor of the plasticity of the matrix and of its ability to regulate the formation of microstructures in a constant manner over geological times, in spite of the evolvability of the shellome. In brief, the evolutionary view of a rapidly evolving shellome may also reflect the presence of attractors that constrain, force or pull the system towards the synthesis of well-defined microstructures, as sketched by the funnels of Figure 2.

The second aspect, the molecular mechanism of shell protein evolution, is also being progressively deciphered. Suspected for more than a decade for shell matrix

proteins (Marin et al., 2008), exon-shuffling - which results in swapping functional modules - is a powerful mechanism for inventing novel genes. Other mechanisms are also involved, such as gene duplications, domain recruitment, replication slippage (Kocot et al., 2016). Alternative splicing, followed by independent evolution of the variants, is also a way to increase the plasticity of the shellome (Herlitzke et al., 2018). Other putative molecular evolutionary mechanisms include also massive losses (which might explain why some shellomes, like the ones of the abalone *Haliotis*, are so different from that of the patellogastropod *Lottia*), and, at the margin, horizontal gene transfer (from symbiont/bacteria). A lot more work needs to be done before we can quantify precisely which of these mechanisms predominate and tentatively reconstitute, family per family the evolutionary trajectory of key functional domains in mollusc shell biomineralization.

IV.2. Shell remodeling, shellome maturation

Many consider shells as dead tissues that only grow by addition of new mineralized layers (incremental growth) until the death of the animal they contain: obviously, because shells are acellular tissues, they do not have the capability to constantly remodel, like bone tissues. But shells have however a certain plasticity and capacity to remodel. Firstly, in anaerobic conditions, molluscs can slightly re-dissolve a very thin inner part of their shell to reabsorb calcium ions, when needed. Secondly, this plasticity is particularly expressed when shells break. When a non-lethal break occurs in a shell, the animal implements a shell repair strategy of emergency: synthesis of a periostracal-like layer to fill the hole and maintain the closure and integrity of the interface between the mantle tissue and the shell itself, rapid mineralization - usually different from that of normal microstructures, such as spherulites - to consolidate the repair zone (Fleury et

al., 2008). From a 'shellomics viewpoint', shell repair - studied in two cases, the Japanese pearl oyster (Chen et al., 2019) and the edible mussel (Hüning et al., 2016) - is marked in the first case by the secretion of a prism-like repertoire and in the second one, by the expression of a number of transcripts encoding different tyrosinases, CA, peroxidase and above all, chitin metabolizing enzymes. In both cases, it appears that mantle cells (assisted by hemocytes) are able to adapt their response to unusual situations and modulate the biosynthesis of key-skeletal proteins.

Besides the plasticity of the shell confronted to a repair situation, one may infer that the matrix modifies, while the animal lives, which is another form of plasticity. The matrix is not simply occluded and stays inert in the mineral phase but may evolve as well. The slight dissolution mentioned above (in anaerobic phases) releases a little bit of matrix, some components of which may act as cell signaling molecules. However, the main event that may occur is the maturation of the shellome. We believe that some concepts, borrowed from vertebrate extracellular matrices, may apply to mollusc shells as well, and more generally, to a large variety of metazoan calcified structures. More than twenty years ago was invented the concept of 'matrikin' (Maquart et al., 1999), soon after renamed 'matricryptin' (Davis et al., 2000). Matrikins are peptides that result from the directed and controlled enzymatic proteolysis of some macromolecules of the extracellular matrix. They are subsequently released in the medium where they play regulatory functions on neighboring cells (paracrine effects). Collagen matrikins, generated by the action of MMPs (matrix metalloproteinases) have been shown to mediate cell proliferation, migration or apoptosis (Kisling et al., 2019). Many mollusc shellomes contain peptidases that may play the required proteolytic function for generating matrikins that would subsequently activate mantle cells. However, to our

knowledge, nobody has ever tried to investigate the potential existence of matrikins in shellomes, but this may be tested via shellomics.

IV.3. Palaeoshellomics: ancient shell proteins and their diagenesis

Shell diagenesis and fossilization is the last point that we briefly discuss here. As shell proteins are occluded in the mineral phase, they have a rather good potential of preservation over time and this property has been evidenced through numerous studies in the last fifty years (Hare and Abelson, 1980; Demarchi et al., 2016). However, until recently, due to limited knowledge on shellome at primary structure level, the search for ancient proteins was very often a dead-end, in the sense that limited information was extracted from fossil shells proteins. With the growing body of knowledge on shellomes, the research field of shell protein diagenesis can be entirely revisited via a 'palaeoshellomics approach' and questions that were kept open until now can be answered, either by reinvestigating fossil and subfossil materials or by performing laboratory diagenesis experiments. One can wonder for example until which geological period fossil proteins that carry exploitable sequence information can be retrieved. Pliocene? Miocene? Cretaceous? Is it possible, by using fossil specimens of the same lineage (at genus or family levels) as the extant one used as control, to calibrate protein evolution rates? With the knowledge of complete shellomes, it becomes also possible to track the diagenetic behavior of each shell protein, taken individually, by artificial diagenesis experiments. Our first pilot experiment on nacre (Parker et al., 2015) proved the validity of the concept: we observed for example that some nacre proteins were persistent (nacrein, Pif, shematin-like-2) after ten days of heating at 100°C while some other proteins disappeared, suggesting a differential degradation pattern of shell proteins. Even the most stable proteins were submitted to diagenetic effects, since they

were identified by a decreasing number of peptides as the diagenetic experience progressed, suggesting partial degradation along the sequence. An analysis of well-dated subfossil clam (*Tridacna*) shells showed that this probably occurs indeed (Marin et al., 2018). Clearly, this type of approach has a great potential and should be extended to several types of shells. At last, shellomics builds bridges with humanities, especially archeology and opens new vistas in this field: recent palaeoshellomics analyses of minute manufactured pearls found in neolithic and mesolithic sites (danish, german and romanian) established the biological origin - freshwater mussels - of these objects, and consequently showed that there were no needs of long distance trading of shell raw materials across Europe, but rather the exploiting of local natural resources (Sakalauskaite et al., 2019). This example illustrates perfectly that hidden information can be extracted and exploited, even from the tiniest archeological remains.

To conclude, shellomics on mollusc shells exhibits a large potential in many research fields: material physics, environment sciences, evolutionary and complex systems biology, palaeontology and sedimentology, taphonomy, archaeology. Shellomes are complex secretory products that are far from having revealed all their secrets.

Acknowledgments

This work is financially supported by ANR MOBi (ANR-18-CE02-0014, *Interface Organique/Inorganique dans les Biominéraux*, J. Perrin coordinator), by INTERRVIE funds and by recurrent budget from UMR CNRS6282 Biogeosciences. The author thanks Grégoire Pierre, graphic designer (<http://www.gregoirepierre.com>) for shaping Figure

2. He also thanks Prof. Peter Westbroek (Leiden University), who, long time ago, inspired some ideas developed in this paper.

References

- Agbaje, O.B.A., Ben Shir, I., Zax, D.B., Schmidt, A., Jacob, D.E., 2018. Biomacromolecules within shells: is chitin abundant? *Acta Biomater.* 80, 176-187.
- Aguilera, F., McDougall, C., Degnan, B.M., 2014. Evolution of the tyrosinase gene family in bivalve molluscs: independent expansion of the mantle gene repertoire. *Acta Biomaterialia* 10, 3855-3865.
- Aguilera, F., McDougall, C., Degnan, B.M., 2017. Co-option and *de novo* gene evolution underlie molluscan shell diversity. *Mol. Biol. Evol.* 34 (4), 779-792.
- Arivalagan, J., Marie, B., Sleight, V.A., Clark, M.S., Berland, S., Marie, A., 2016. Shell matrix proteins of the clam, *Mya truncata*: roles beyond shell formation through proteomic study. *Mar. Genomics* 27, 69-74.
- Arivalagan, J., Yarra, T., Marie, B., Sleight, V.A., Duvernois-Berthet, E., Clark, M.S., Marie, A., Berland, S., 2017. Insights from the shell proteome: biomineralization to adaptation. *Mol. Biol. Evol.* 34 (1), 66-77.
- Bayerlein, B., Zaslansky, P., Dauphin, Y., Rack, A., Fratzl, P., Zlotnikov, I., 2014. Self-similar mesostructure evolution of the growing mollusc shell reminiscent of thermodynamically driven grain growth. *Nature Mater.* 13, 1102-1107.
- Bédouet, L., Marie, A., Berland, S., Marie, B., Auzoux-Bordenave, S., Marin, F., and Milet, C., 2012. Proteomic strategy for identifying mollusc shell proteins using mild chemical degradation and trypsin digestion of insoluble organic shell matrix: a pilot study on *Haliotis tuberculata*. *Mar. Biotechnol.* 14(4), 446-458.
- Carter, J.G., 1990. *Skeletal Biomineralization: Patterns, Processes, and Evolutionary Trends*. Van Nostrand Reinhold, New York.
- Carter, J.G., and Clark, G.R.II, 1985. Classification and phylogenetic significance of mollusk shell microstructures, in Broadhead, T.W. (Ed.), *Mollusk, Note for a Short Course*. Studies in Geology 13, Dpt. Geological Sciences, University of Tennessee Press, Tennessee, pp. 50-71.
- Chateigner, D., Hedegaard, C., Wenk, H.R., 2000. Mollusc shell microstructures and crystallographic textures. *J. Struct. Geol.* 22 (11-12), 1723-1735.
- Checa, A.G., Okamoto, T., Ramirez, J., 2006. Organization pattern of nacre in Pteriidae (Bivalvia: Mollusca) explained by crystal competition. *Proc. R. Soc. B* 273, 1329-1337.

- Chen, Y., Liu, C., Li, S.G., Liu, Z.W., Xie, L.P., Zhang, R.Q., 2019. Repaired shells of the pearl oyster largely recapitulate normal prismatic layer growth: a proteomics study of shell matrix proteins. *ACS Biomat. Sci. Engin.* 5 (2), 519-529.
- Davis, G.E., Bayless, K.J., Davis, M.J., Meininger, G.A., 2000. Regulation of tissue injury response by the exposure of matricryptic sites within extracellular matrix molecules. *Am. J. Pathol.* 156 (5), 1489-1498.
- Demarchi, B., Hall, S., Roncal-Herrero, T., Freeman, C.L., Woolley, J., Crisp, M.K., Wilson, J., Fotakis, A., Fischer, R., et al., 2016. Protein sequences bound to mineral surfaces persist into deep time. *Elife* 5, e17092.
- Du, X.D., Fan, G.Y., Jiao, Y., Zhang, H., Guo, X.M., Huang, R.L., Zheng, Z., Bian, C., Deng, Y.W., Wang, Q.H., et al., 2017. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. *Gigasciences* 6, 8.
- Fleury, C., Marin, F., Marie, B., Luquet, G., Thomas, J., Josse, C., Serpentini, A., Lebel, J. M., 2008. Shell repair process in the green ormer *Haliotis tuberculata*: a histological and microstructural study. *Tissue Cell* 40, 207-218.
- Gao, P., Liao, Z., Wang, X.X., Bao, L.F., Fan, M.H., Li, X.M., Wu, C.W., Xia, S.W., 2015. Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. *Plos One* 10 (7), e0133913.
- Gower, L.B., Odom, D.J., 2000. Deposition of calcium carbonate films by a polymer-induced liquid-precursor (PILP) process. *J. Crystal Growth* 210 (4), 719-734.
- Grégoire, C., 1972. Structure of the molluscan shell, in Florkin, M., Scheer B.T. (Eds.), *Chemical Zoology, Vol. VII: Mollusca*, Academic Press, New York, pp. 45-102.
- Grigor'ev, D.P., 1965. *Ontogeny of Minerals*. Israel Program for Scientific Translation, Jerusalem.
- Hare, P.E., 1963. Amino acids in the proteins from aragonite and calcite in the shells of *Mytilus californianus*. *Science* 139, 216-217.
- Hare, P.E., Hoering, T.C., King, K., 1980. *Biogeochemistry of Amino Acids*. Wiley, New York.
- Harrison, P.M., 2018. Compositionally biased dark matter in the protein universe. *Proteomics* 18 (21-22), 1800069.
- Harting, P., 1872. *Recherche de morphologie synthétique sur la production artificielle de quelques formations calcaires organiques*, van der Post, Amsterdam.
- Herlitze, I., Marie, B., Marin, F., Jackson, D.J., 2018. Molecular modularity and asymmetry

- of the molluscan mantle revealed by a gene expression atlas. *Gigascience* 7, 6.
- Holliday, L.S., De Faria, L.P., Rody, W.J., 2020. Actin and actin-associated proteins in extracellular vesicles shed by osteoclasts. *Int. J. Mol. Sci.* 21 (1), 158.
- Hüning, A.K., Lange, S.M., Ramesh, K., Jacob, D.E., Jackson, D.J., Panknin, U., Gutowska, M.A., Philipp, E.E.R., Rosenstiel, P., Lucassen, M., Melzner, F., 2016. A shell regeneration assay to identify biomineralization candidate genes in mytilid mussels. *Mar. Genomics* 27, 57-67.
- Immel, F., Broussard, C., Catherinet, B., Plasseraud, L., Alcaraz, G., Bundeleva, I., Marin, F., 2016. The shell of the invasive bivalve species *Dreissena polymorpha*: biochemical, elemental and textural investigations. *Plos One* 11 (5), e0154264.
- Jackson, D.J., McDougall, C., Green, K., Simpson, F., Wörheide, G., Degnan, B.M., 2006. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol.* 4, 40-49.
- Jackson, D.J., Reim, L., Randow, C., Cerveau, N., Degnan, B.M., Fleck, C., 2017. Variation in orthologous shell-forming proteins contribute to molluscan shell diversity. *Mol. Biol. Evol.* 34 (11), 2959-2969.
- Kisling, A., Lust, R.M., Katwa, L.C., 2019. What is the role of peptide fragments of collagen I and IV in health and disease? *Life Sci.* 228, 30-34.
- Knoll, A.H., 2003. Biomineralization and evolutionary history. Rev, in: Dove, P.M., De Yoreo, J.J., Weiner, S (Eds.), *Biomineralization. Rev. Mineral. Geochem* 54, Mineral. Soc. Amer., Geochem. Soc, pp. 329-356.
- Kocot, K.M., Aguilera, F., McDougall, C., Jackson, D.J., Degnan, B.M., 2016. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. *Front. Zool.* 13, 23.
- Kouchinsky, A., 2000. Shell microstructures in early Cambrian molluscs. *Acta Palaeontol. Pol.* 45, 119–150.
- Krampitz, G., Engels, J., Cazaux, C., 1976. Biochemical studies on water-soluble proteins and related components of gastropod shells, in Watabe, N., Wilbur, K.M., (Eds.), *The Mechanisms of Mineralization in the Invertebrates and Plants*, University of South Carolina Press, Columbia, South Carolina, pp.155-173.
- Le Roy, N., Jackson, D.J., Marie, B., Ramos-Silva, P., and Marin, F., 2014. The evolution of metazoan alpha-carbonic anhydrases and their roles in CaCO₃ biomineralization. *Front. Zool.* 11, 75.

- Li, S.G., Liu, Y.J., Liu, C., Huang, J.L., Zheng, G.L., Xie, L.P., Zhang, R.Q., 2016. Hemocytes participate in calcium carbonate crystal formation, transportation and shell regeneration in the pearl oyster *Pinctada fucata*. *Fish Shellfish. Immunol.* 51, 263-270.
- Liao, Z., Bao, L.F., Fan, M.H., Gao, P., Wang, X.X., Qin, C.L., Li, X.M., 2015. In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *J. Proteomics* 122, 26-40.
- Liu, C., Li, S.G., Kong, J.J., Liu, Y.J., Wang, T.P., Xie, L.P., Zhang, R.Q., 2015. In-depth proteomic analysis of the shell matrix proteins of *Pinctada fucata*. *Sci. Rep.* 5, 17269.
- Lowenstam, H.A., Weiner, S., 1989. *On Biomineralization*. Oxford University Press, New York.
- Mc Dougall, C., Degnan, B.M., 2018. The evolution of mollusc shells. *Dev. Biol.* 7, e313.
- Mann, K., Weiss, I.M., Andre, S., Gabius, H.J., Fritz, M., 2000. The amino-acid sequence of the abalone (*Haliotis laevis*) nacre protein perlucin - Detection of a functional C-type lectin domain with galactose/mannose specificity. *Eur. J. Biochem.* 267 (16), 5257-5264.
- Mann, K., Edsinger-Gonzales, E., Mann, M., 2012. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proteom. Sci.*, 10: 28.
- Mann, K., Edsinger-Gonzales, E., 2014. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantification and phosphoproteome analysis. *Proteome Sci.* 12, 28.
- Mann, K., Cerveau, N., Gummich, M., Fritz, M., Mann, M., Jackson, D.J., 2018. In-depth proteomic analyses of *Haliotis laevis* (greenlip abalone) nacre and prismatic organic shell matrix. *Proteome Sci.* 16, 11.
- Maquart, F.X., Siméon, A., Pasco, S., Monboisse, J.C., 1999. Regulation of cell activity by the extracellular matrix: the concept of matrikines. *J. Soc. Biol.* 193 (4-5), 423-428.
- Marie, B., Le Roy, N., Marie, A., Dubost, L., Milet, C., Bédouet, L., Becchi, M., Zanella-Cléon, I., Jackson, D., Degnan, B., Luquet, G., Marin, F., 2009. Nacre evolution: a proteomic approach, in: Kisalius, D., Estroff, L., Gupta, H.S., Landis, W.J., Zavattieri, P.D., (Eds.), *Structure-Property Relationships in Biomineralized and Biomimetic Composites*, *Mater. Res. Soc. Symp. Proc.* 1187, 3-8.

- Marie, B., Joubert, C., Tayalé, A., Zanella-Cléon, I., Belliard, C., Piquemal, D., Cochenne-Laureau, N., Marin, F., Gueguen, Y., Montagnani, C. 2012. Different secretory repertoires control the biomineralization processes of prisms and nacre deposition of the pearl oyster shell. *Proc. Natl. Acad. Sci. USA* 109 (51), 20986-20991.
- Marie, B., Jackson, D.J., Ramos-Silva, P., Zanella-Cléon, I., Guichard, N., Marin, F., 2013. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS J.* 280 (1), 214-232.
- Marin, F., Muyzer, G., Dauphin, Y., 1994. Electrophoretic and immunological characterization of shell soluble matrices from two living periomorphid bivalves, *Pinna nobilis* L. et *Pinctada margaritifera* (L.). *C. R. Acad. Sci. Paris* 318, Sér. II, 1653-1659.
- Marin, F., Gillibert, M., Wesbroek, P., Muyzer, G., Dauphin, Y., 1999. Evolution: disjunct degeneration of immunological determinants. *Geol. Mijnbouw* 78 (2), 135-139.
- Marin, F., Corstjens, P., De Gaulejac, B., De Vrind-De Jong, E., Westbroek, P., 2000. Mucins and molluscan calcification: molecular characterization of mucoperlin, a novel acidic mucin-like protein of the nacreous shell-layer of the fan mussel *Pinna nobilis* (Bivalvia, Pteriomorphia). *J. Biol. Chem.* 275 (27), 20667-20675.
- Marin, F., Luquet, G., Marie, B., Medakovic, D., 2008. Molluscan shell proteins: primary structure, origin and evolution. *Curr. Top. Dev. Biol.* 80, 209-276.
- Marin, F., Narayanappa, P., Motreuil, S. 2011. Acidic shell proteins of the Mediterranean fan mussel *Pinna nobilis*. in: Müller, W.E.G. (Ed.), *Molecular Biomineralization: Aquatic Organisms Forming Extraordinary Materials*. *Prog. Mol. Subcell. Biol.* 52, 353-395.
- Marin, F., Le Roy, N., Marie, B., 2012. The formation and mineralization of mollusk shell. *Front. Biosci. (Schol. Ed.)* 4, 1099-1125.
- Marin, F., Le Roy, N., Marie, B., Ramos-Silva, P., Bundeleva, I., Guichard, N., Immel, F., 2014. Metazoan calcium carbonate biomineralizations: macroevolutionary trends - challenges for the coming decade. *Bull. Soc. Géol. France* 185 (4), 217-232.
- Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., and Medakovic, D., 2016. Organic matrices in metazoan calcium carbonate skeletons: composition, functions, evolution. *J. Struct. Biol.* 196, 98-106.
- Marin, F., Chmiel, A., Takeuchi, T., Bundeleva, I., Durllet, C., Samankassou, E., Medakovic, D., 2018. Skeletal Organic Matrices in Molluscs: Origin, Evolution, Diagenesis, in:

- Endo, K., Kogure, T., Nagasawa, H. (Eds.), *Biom mineralization - From Molecular and Nano-Structural Analyses to Environmental Science*, Springer, Singapore, pp. 325-332.
- Meinhardt, H., 2009. *The Algorithmic Beauty of Sea Shells*. Springer, Berlin, 3rd Edition.
- Mount, A.S., Wheeler, A.P., Par adkar, R.P., Snider, D., 2004. Hemocyte-mediated shell mineralization in the eastern oyster. *Science* 304 (5668), 297-300.
- Oudot, M., Neige, P., Shir Ben I., Schmidt, A., Strugnell, J.M., Plasseraud, L., Broussard, C., Hoffmann, R., Lukeneder, A., Marin, F., 2020. The shell matrix and microstructure of the Ram's Horn squid: molecular and structural characterization. *J. Struct. Biol.* in press.
- Panca s, R., Schad, E., Tantos, A., Tompa, P., 2019. Emergent functions of proteins in non-stoichiometric supramolecular assemblies. *Biochim. Biophys. Acta - Prot. Proteom.* 1867 (10), 970-979.
- Parker, A., Immel, F., Guichard, N., Broussard, C. and Marin, F., 2015. Thermal stability of nacre proteins of the Polynesian pearl oyster: a proteomic study. *Key Eng. Mater.* 672, 222-233.
- Rainey, G., 1857. On the elementary formation of the skeletons of animals, and other hard structures formed in connection with living tissues. *Brit. For. Med. Ch. Rev.* XX, 451-476 (published separately with additions, 8vo, London, 1858).
- Ré hault-Godbert, S., Hervé-Grépinet, V., Gautron, J., Cabau, C., Nys, Y., Hincke, M., 2011. Molecules involved in chemical defence of the chicken egg, in: Nys, Y., Bain, M., van Immerseel, F. (Eds.), *Improving the Safety and Quality of Eggs and Egg Products - Egg Chemistry, Production and Consumption*. Woodhead Publishing Series in Food Science, Technology and Nutrition, pp. 183-208.
- Reich, E., Schoeppler, V., Lemanis, R., Lakin, E., Zolotoyabko, E., Zöllner, D., Zlotnikov, I., 2019. Morphological and textural evolution of the prismatic ultrastructure in mollusc shells/ a comparative study of Pinnidae species. *Acta Biomater.* 85, 272-281.
- Sakalauskaite, J., Andersen, S., Biagi, P., Borrello, M.A., Cocquerez, T., Colonese, A.C., Dal Bello, F., Girod, A., Heumüller, M., Koon, H., Mandili, G., Medana, C., Penkman, K.E.H., Plasseraud, L., Schlichtherle, H., Taylor, S., Tokarski, C., Thomas, J., Wilson, J., Marin, F., Demarchi, B., 2019. 'Palaeoshellomics' reveals the use of freshwater mother-of-pearl in prehistory. *eLife* 8, e45644.

- Samata, T., 1990. Ca-binding glycoproteins in molluscan shells with different types of ultrastructure. *Veliger* 33, 190–201.
- Sarashina, I., Endo, K., 1998. Primary structure of a soluble matrix protein of scallop shell: Implications for calcium carbonate biomineralization. *Am. Mineral.* 83, 1510–1515.
- Sathyan, N., Philip, R., Chaithanya, E.R., Anil Kumar, P.R., 2012. Identification and molecular characterization of molluskin, a histone-H2A-derived antimicrobial peptide from molluscs. *ISRN Mol. Biol* 2012, 219656.
- Schoeppler, V., Lemanis, R., Reich, E., Pusztai, T., Granazy, L., Zlotnikov, I., 2019. Crystal growth kinetics as an architectural constraint on the evolution of molluscan shell. *PNAS* 116 (41), 20388-20397.
- Shimamoto, M., 1986. Shell microstructure of the Veneridae (Bivalvia) and its phylogenetic implications. *Sci. Rep., 2nd ser. Tohoku Univ.* 56, 1-39.
- Simkiss, K., 1977. Biomineralization and detoxification. *Calcif. Tiss. Res.* 24, 199-200.
- Simkiss, K., Wilbur, K.M., 1989. *Biomineralization, Cell biology and Mineral Deposition.* Academic Press, New York.
- Song, X.R., Liu, Z.Q., Wang, L.L., Song, L.S., 2019. Recent advances of shell matrix proteins and cellular orchestration in marine molluscan shell biomineralization. *Front. Mar. Sci.* 6, UNSP 41.
- Suzuki, M., Murayama, E., Inoue, H., Ozaki, N., Tohse, H., Kogure, T., Nagasawa, H., 2004. Characterization of Prismaticin-14, a novel matrix protein from the prismatic layer of the Japanese pearl oyster (*Pinctada fucata*). *Biochem. J.* 382, 205–213.
- Taylor, J.D., Kennedy, W.J., Hall, A., 1969. The shell structure and mineralogy of the Bivalvia. Introduction. *Nuculacea-Trigonacea. Bull. Br. Mus. (Nat. Hist.) Zool. Lond.* 3 (Suppl.), 1-125.
- Taylor, J.D., Kennedy, W.J., Hall, A. (1973). The shell structure and mineralogy of the Bivalvia. II. *Lucinacea-Clavagellacea. Conclusions. Bull. Br. Mus. (Nat. Hist.) Zool. Lond.* 22, 253–294.
- Thompson, D.A.W, 1942. *On Growth and Form.* University Press, Cambridge (new edition).
- Timmermans, L.P.M., 1969. Studies on shell formation in molluscs. *Neth. J. Zool.* 19, 413-523.

- Timmins-Schiffman, E., Coffey, W.D., Hua, W., Nunn, B.L., Dickinson, G.H., Roberts, S.B., 2014. Shotgun proteomics reveals physiological response to ocean acidification in *Crassostrea gigas*. *BMC Genomics* 15, 951.
- Tsukamoto, D., Sarashina, I., Endo, K., 2004. Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochem. Biophys. Res. Commun.* 320, 1175–1180.
- Ubukata, T., 1994. Architectural constraints on the morphogenesis of prismatic structure in Bivalvia. *Palaeontology* 37, 241–261.
- Uozumi, S., Suzuki, S., 1981. The evolution of shell structure in the Bivalvia, in: Habe, T., Omori, M. (Eds.), *Study of Molluscan Paleobiology*, Prof. M. Omori Memorial Volume, pp. 63-77.
- Vasiliu, A., 2015. Natural pearls. *Key Eng. Mater.* 672, 80-102.
- Wei, L., Wang, Q., Ning, X.X., Mu, C.K., Wang, C.L., Cao, R.W., Wu, H.F., Cong, M., Li, F., Ji, C.L., Zhao, J.M., 2015. Combined metabolome and proteome analysis of the mantle tissue from Pacific oyster *Crassostrea gigas* exposed to elevated pCO₂. *Comp. Biochem. Physiol., D Genom. Proteom.* 13, 16-23.
- Weiner, S., 1983. Mollusk shell formation: Isolation of two organic matrix proteins associated with calcite deposition in the bivalve *Mytilus californianus*. *Biochemistry* 22, 4139–4145.
- Weiner, S., Hood, L., 1975. Soluble proteins of the organic matrix of mollusc shells: A potential template for shell formation. *Science* 190, 987–989.
- Weiner, S., Traub, W., Lowenstam, H.A., 1983. Organic matrix in calcified exoskeletons, in: Westbroek, P., De Jong, E.W. (Eds.), *Biom mineralization and Biological Metal Accumulation*, D. Reidel Publishing Company, Dordrecht, pp. 205-224.
- Weiss, I.M., Schonitzer, V., Eichner, N., Sumper, M., 2006. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. *FEBS Lett.* 580 (7), 1846-1852.
- Wolfram, S., 2002. *A New Kind of Science*. Wolfram Media.
- Yano, M., Nagai, K., Morimoto, K., Miyamoto, H., 2006. Shematrin: A family of glycine rich structural proteins in the shell of the pearl oyster. *Comp. Biochem. Physiol.* 144, 254–262.
- Zhang, C., Xie, L., Huang, J., Liu, X., Zhang, R., 2006. A novel matrix protein family participating in the prismatic layer formation of pearl oyster, *Pinctada fucata*.

Biochem. Biophys. Res. Commun. 344, 735–740.

Zhang, G.F., Fang, X.D., Guo, X.M., Li, L., Luo, R.B., Xu, F., Yang, P.C., Zhang, L.L., Wang, X.T., Qi, H.G. et al., 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490 (7418), 49-54.

Zhao, R., Takeuchi, T., Luo, Y.J., Ishikawa, A., Kobayashi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Iwanaga, S., Nagai, K., Satoh, N., Endo, K., 2018. Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Mol. Biol. Evol.* 35 (11), 2751-2761.

Zöllner, D., Reich, E., Zlotnikov, I., 2017. Morphogenesis of biomineralized calcitic prismatic tissue in Mollusca fully described by classical hierarchical grain boundary motion. *Cryst. Growth Des.* 17, 5023-5027.

Legends to figures

Figure 1. The most commonly found functional domains in mollusc shellomes. Domains are listed and grouped according to molecular functions or features when the function is unknown. This is the case of most LCDs/RLCDs. Some domains can belong to different categories: chitinase is an enzyme and a saccharide-interacting protein. The most external circles correspond to main cellular/physiological roles in biomineralization.

Figure 2. Hypothetical view on shell microstructures in molluscs. A. In a plan representing a hypothetical "microstructural field", each protein repertoire associated to a given shell layer of a given species is represented by a dot, and dots are grouped in microstructure areas. Dots close to each other reflect resemblances in the protein repertoires, while distant dots mean dissimilar repertoires. For clarity, the graph here comprises only four main microstructures, nacreous, prismatic, foliated and crossed-lamellar. B. Same representation in 3D. Each microstructural field is funnel-shaped, indicating that whatever the position of the dots within this field, the end-product is the microstructure. Beads put at points "a" and "b", although relatively distant, will roll down in the same funnel, while beads placed in "a" and "c", rather close, make their way in two adjacent funnels. This view reconciles the plasticity of the shellomes and the fact that the outcome of their "molecular activity" is constrained in giving a rather limited set of microstructure. C. The insert shows that each field does not necessarily contains only one funnel, but two or more. This example illustrates the case of closely-related microstructures, such as homogeneous and granular (or crossed-lamellar and complex crossed-lamellar) that can be grouped within one area. More explanations are given in § III.4.

Table 1. List of mollusc species for which high-throughput approaches (genomics, transcriptomics, proteomics) were applied in relation to shell formation. The list is sorted according to taxonomy. List of abbreviations: Monoplac. = Monoplacophora; Naut. = Nautiloidea; Col. = Coleoidea; Pteriom. = Pteriomorpha; Pal. = Palaeoheterodonta; Het. = Heterodonta; Patel. = Patellogastropoda; Vetig. = Vetigastropoda; Caen. = Caenogastropoda; Heter. = Heterobranchia. Some species, like the oyster *C. gigas* or the pearl oyster *P. fucata* are well-covered by a genome, different transcriptomes and shellomes, while other species are simply covered by a single shellome (*Helix aspersa*, *Sepia officinalis*).