BioMine-DB: A database for metazoan biomineralization proteins

Biomineralization is the process by which living organisms construct hard skeletons creating complex structures that range from specialized tissues such as bone or teeth to ecosystems such as coral reefs. Biominerals are composed of both inorganic minerals and proteins, which give them extra hardness and special attributes. Biomineralization proteins are also known to be associated with multiple bone disorders and are therefore of biomedical importance. Herein we describe BioMine-DB, a biomineralization centric protein database. Availability and implementation: BioMine-DB can be accessed at http://biomine.net, SQL dump, FASTA files and source code are available for download as well at https://github.com/bishoyh/biomineDB "

Journal: PeerJ
Title(s): BioMine-DB: A database for metazoan biomineralization proteins

Authors: Bishoy S Kamel, Christian Voolstra[2] and Mónica Medina[1]
Affiliations: [1]Department of Biology, Pennsylvania State University, University Park, PA, USA
[2]Red Sea Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Introduction

Biomineralization is a process in which minerals form inside or outside the cells of a variety of organisms (Lowenstam & Weiner 1989; Simkiss & Wilbur 1989). In animals, these minerals are primarily calcium carbonates and calcium phosphates (Knoll 2003). The majority of biominerals formed in bones, shells, skeletons and spicules are composed of mineral crystals, however all biominerals contain various amounts of other proteins that give these minerals extraordinary properties. The cell orchestrates the mineral formation process through the expression and translocation of proteins that nucleate the crystals either intracellularly or extracellularly. More importantly, the cell has to inhibit mineral formation and crystal growth in unwanted sites (Kawasaki et al. 2009; Marin et al. 1996). Both nucleation and inhibition can be achieved through multiple cellular mechanisms. For example, the cell will produce enzymes that modify proteins by breaking them into smaller peptides (Qin et al. 2004) thus changing their function. The cell is able to tightly regulate the biomineralization process by molecular modification (e.g. adding sugars or other moieties) and regulation of ion transport across membranes (Qin et al. 2004; Saavedra 1994; Sarashina et al. 2006). Such modifications to biomineralization-associated proteins will determine how they interact with other proteins, other cells, and with the biomineral in general. Biominerals are essential to the survival of a broad range of animal taxa because they deliver protection against predation, act as energy storage, provide support and unique optical properties (Addadi et al. 2006). In particular, biomineralization plays a pivotal role in multiple human diseases

and other pathological phenomena such as coronary artery calcification (Atlan et al. 1997; Collette et al. 2010; Fisher et al. 2001; Lopez et al. 1992; Rousselle & Heymann 2002; Salih et al. 1996; Wallin et al. 2001; Westbroek & Marin 1998; Yang et al. 2002). A growing interest in bio-inspired materials has generated a large body of work that uses proteins and other biological scaffolds for *in vitro* mineralization and synthetic materials (Chiu et al. 2012; Perry et al. 2009).

The process of biomineralization is ubiquitous throughout the animal tree. Such distribution has generated speculation about the origin of metazoan biomineralization and its evolutionary history. Biomineralization is a complex process that relies on multiple cellular pathways (Knoll 2003; Marin et al. 1996). Many of the studied biomineralization proteins are part of other important processes such as cell adhesion, extracellular matrix organization and immune functions (Bryden et al. 1999; Clendenon et al. 2009). This evidence favors the idea that biomineralization independently evolved in multiple phyla using pre-existing pathways in the eumetazoan ancestor. It can also be argued that biomineralization was present in the early eumetazoan ancestor yet various parts of the pathway were lost in several animal lineages. According to fossil evidence and when mapped onto a phylogeny, carbonate skeletons seem to have evolved at least 20 different times in metazoans (Knoll 2003). If biomineralization evolved multiple times, it is relevant to understand which components of the process exactly underwent innovations. Since biomineralization is an active process, it requires 1) targeted localization of calcium and carbonate, 2) an organic matrix as a template for the mineral nucleation, 3) growth, and 3) efficient inhibitors in order to stop undesired calcification or even formation of the mineral (Jackson et al. 2010; Marin et al. 1996). When all these different requirements are taken into account, it seems unlikely that such diverse biochemical processes involved in metazoan biomineralization evolved independently more than 20 times. A process such as transport is quite conserved across animal lineages and it shows a clear history of gene duplication events (Dean et al. 2001; Saier et al. 2009). Such complexity presents us with a conundrum. While the biomineralization process, with different minerals and methods of calcification and clear evolutionary novelties, is found across multiple animal phyla (Jackson et al. 2007a; Jackson et al. 2006; Jackson et al.

2009; Jackson et al. 2007b; Marin et al. 2000; Marin & Luquet 2004; Marin et al. 2008; Marin et al. 1996), it also remains that many parts of biomineralization pathways must be conserved.

As a first step in tackling such questions in the evolution of animal biomineralization, we have created a database to accumulate, annotate and curate biomineralization proteins and protein-coding sequences. The database aims to serve the community by bridging the gap between the few identified biomineralization proteins, and the unannotated plethora of Expressed Sequence Tags (ESTs), draft genome gene models and next-generation sequencing datasets. We employed various bioinformatics techniques using domain-based searches to collect and identify novel biomineralization proteins in metazoans. We hope that due to the increasing surge of sequence information along with broad phylogenetic representation in the public domain, a clearer picture of the evolutionary history of biomineralization proteins will emerge, rendering BioMine-DB as a dynamic platform to answer not only fundamental questions in animal evolution but also about the process of biomineralization in particular lineages.

Methods

Biomineralization proteins list

We carried out a wide primary literature and database survey in order to compile a list of proteins that are functionally implicated biomineralization in animals. Specifically, we included data from scleractinian corals, calcareous sponges, gastropod and bivalve molluscs, crustaceans, echinoderms, and vertebrates. Additional sequences were collected from the AMIGO Gene Ontology database (Carbon et al. 2009). The complete biomineralization gene list is accessible through the BioMine-DB web application (http://biomine.net/). These already annotated sequences were used as a seed to search for related biomineralization proteins in undocumented taxa or new sequence databases, and were stored in a pre-computed BLAST search database. After building the list of candidate proteins, each protein could be traced back to an original publication where it was described.

Pfam domain search and protein homolog identification

To further improve the search strategy, using the Pfam database, we scanned for conserved protein domains in the proteins we gathered from primary literature (Finn et al. 2008). The identified domains in the already known biomineralization proteins were scanned against 6-frame translations of ESTs and protein sequences from dbEST and nr databases (NCBI) of the taxa Cnidaria, Mollusca, Echinodermata and Vertebrata using the HMMER 3 package (Eddy 2008; Eddy 2011). The tool used for the translation was sixpack from the EMBOSS package (Rice et al. 2000).

BLAST Searches

In addition to the domain searches, we conducted BLASTp searches for all the proteins in the seed database against nr and against the 6-frame translations of the dbEST for the selected taxa. For filtering the results, we only considered hits that match e-value < 0.000001 and bitscore > 150 to be significant.

BioMine-DB construction

In order, to organize all the data in a searchable platform, we constructed a web application that enables us to search the results and to submit new sequences into the database. BioMine-DB is written in PHP and Perl, and relies on MySQL for relational information Source code for BioMine-DB is under GPL v3 at http://github.com/bishoyh/biomineDB. The MySQL database contains the results of all the HMMER results in addition to the BLAST results and FASTA files of all sequences can be downloaded from the website.

Results

After assembling the initial list of biomineralization-related proteins, we identified putative homologs of given candidate genes in calcifying lineages (molluscs, cnidarians, arthropods and echinoderms). A protein domain search was initiated on our candidate list of 472 proteins, based on Pfam models using HMMER 3 (Eddy 2008). In the Pfam search 198 domain families were found to be linked to biomineralization. The search

results were stored in a relational database linking the detected domains with the taxonomic information, in addition detected orthologs for every particular protein.

Below we describe two potential scenarios for the use of BioMine-DB by the scientific community.

Use case 1:

1) A user prepares a list of proteins from a newly sequenced organism.
2) The user submits the protein list to BioMine-DB through the web interface.
3) BioMine-DB generates potential matching biomineralization proteins in the submitted dataset, together with the publications in which these similar proteins have been described.

Use case 2:

1) A user is already working with a known biomineralization protein and is doing functional work, i.e. the user is doing whole mount in situ gene expression research in a given organism and finds it hard to explain the observed expression pattern. Thus, the user thinks there could be other proteins involved.
2) The user submits his protein to BioMine-DB.
3) The user gets back a list of potential biomineralization from

Discussion and Conclusion

By combining thorough literature scrutiny with similarity searches we were able to construct a large dataset of biomineralization-related proteins. BioMine-DB proved useful in annotating sequence data from non-model organisms involved in the particular process of biomineralization. The ability to always link back to the primary literature provides a unique opportunity to the investigator to directly examine the experimental evidence that deemed a particular protein as biomineralization-associated. We believe this should fast-forward research in non-model systems by knowledge transfer from model species in biomineralization research. By providing a BLAST interface and downloadable versions, we are certain that biomineralization researchers can benefit from BioMine-DB, as there are no other resources that catalog and curate biomineralization proteins.

References

Addadi L, Joester D, Nudelman F, and Weiner S. 2006. Mollusk Shell Formation: a Source of New Concepts for Understanding Biomineralization Processes. *Chemistry - A European Journal* 12:980 – 987

Atlan G, Balmain N, Berland S, Vidal B, and Lopez É. 1997. Reconstruction of human maxillary defects with nacre powder: histological evidence for bone regeneration. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie* 320:253-258.

Bryden AA, Freemont AJ, Clarke NW, and George NJ. 1999. Paradoxical expression of E-cadherin in prostatic bone metastases. *BJU Int* 84:1032-1034. bju378 [pii]

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, and Lewis S. 2009. AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288-289. btn615 [pii]
10.1093/bioinformatics/btn615

Chiu D, Zhou W, Kitayaporn S, Schwartz DT, Murali-Krishna K, Kavanagh TJ, and Baneyx F. 2012. Biomineralization and size control of stable calcium phosphate core-protein shell nanoparticles: potential for vaccine applications. *Bioconjug Chem* 23:610-617. 10.1021/bc200654v

Clendenon SG, Shah B, Miller CA, Schmeisser G, Walter A, Gattone VH, 2nd, Barald KF, Liu Q, and Marrs JA. 2009. Cadherin-11 controls otolith assembly: evidence for extracellular cadherin activity. *Dev Dyn* 238:1909-1922. 10.1002/dvdy.22015

Collette NM, Genetos DC, Murugesh D, Harland RM, and Loots GG. 2010. Genetic evidence that SOST inhibits WNT signaling in the limb. *Dev Biol* 342:169-179. S0012-1606(10)00187-9 [pii]
10.1016/j.ydbio.2010.03.021

Dean M, Hamon Y, and Chimini G. 2001. The human ATP-binding cassette (ABC) transporter superfamily. *J Lipid Res* 42:1007-1017.

Eddy SR. 2008. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput Biol* 4:e1000069. 10.1371/journal.pcbi.1000069

Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* 7:e1002195. 10.1371/journal.pcbi.1002195

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, and Bateman A. 2008. The Pfam protein families database. *Nucl Acids Res* 36:D281-288. 10.1093/nar/gkm960

Fisher LW, Torchia DA, Fohr B, Young MF, and Fedarko NS. 2001. Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin. *Biochem Biophys Res Commun* 280:460-465. 10.1006/bbrc.2000.4146
S0006-291X(00)94146-9 [pii]

Jackson DJ, Macis L, Reitner J, Degnan BM, and Worheide G. 2007a. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. *Science* 316:1893-1895. 1141560 [pii]
10.1126/science.1141560

Jackson DJ, McDougall C, Green K, Simpson F, Worheide G, and Degnan BM. 2006. A rapidly evolving secretome builds and patterns a sea shell. *BMC Biol* 4:40.

Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C, Piquemal D, and Degnan BM. 2009. Parallel evolution of nacre building gene sets in molluscs. *Mol Biol Evol*:msp278. 10.1093/molbev/msp278

Jackson DJ, Thiel V, and Worheide G. 2010. An evolutionary fast-track to biocalcification. *Geobiology*. GBI236 [pii]
10.1111/j.1472-4669.2010.00236.x

Jackson DJ, Worheide G, and Degnan BM. 2007b. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. *BMC Evol Biol* 7:160. 1471-2148-7-160 [pii]
10.1186/1471-2148-7-160

Kawasaki K, Buchanan AV, and Weiss KM. 2009. Biomineralization in humans: making the hard choices in life. *Annual review of genetics* 43:119-142.

Knoll AH. 2003. Biomineralization and Evolutionary History. *Reviews in Mineralogy and Geochemistry* 54:329-356. 10.2113/0540329

Lopez E, Vidal B, Berland S, Camprasse S, Camprasse G, and Silve C. 1992. Demonstration of the capacity of nacre to induce bone formation by human osteoblasts maintained in vitro. *Tissue and Cell* 24:667-679.

Lowenstam HA, and Weiner S. 1989. *On Biomineralization*. New York: Oxford University Press.

Marin F, Corstjens P, de Gaulejac B, de Vrind-De Jong E, and Westbroek P. 2000. Mucins and Molluscan Calcification. MOLECULAR CHARACTERIZATION OF MUCOPERLIN, A NOVEL MUCIN-LIKE PROTEIN FROM THE NACREOUS SHELL LAYER OF THE FAN MUSSEL PINNA NOBILIS (BIVALVIA, PTERIOMORPHIA). *J Biol Chem* 275:20667-20675. 10.1074/jbc.M003006200

Marin F, and Luquet G. 2004. Molluscan Shell Proteins. *Comptes Rendus Palevol* 3:469–492.

Marin F, Luquet G, Marie B, and Medakovic D. 2008. Molluscan shell proteins: primary structure, origin, and evolution. *Curr Top Dev Biol* 80:209-276. S0070-2153(07)80006-8 [pii]
10.1016/S0070-2153(07)80006-8

Marin F, Smith M, Isa Y, Muyzer G, and Westbroek P. 1996. Skeletal matrices, muci, and the origin of invertebrate calcification. *Proc Natl Acad Sci U S A* 93:1554-1559.

Perry CC, Patwardhan SV, and Deschaume O. 2009. From biominerals to biomaterials: the role of biomolecule-mineral interactions. *Biochem Soc Trans* 37:687-691. 10.1042/BST0370687

Qin C, Baba O, and Butler WT. 2004. POST-TRANSLATIONAL MODIFICATIONS OF SIBLING PROTEINS AND THEIR ROLES IN OSTEOGENESIS AND DENTINOGENESIS. *Critical Reviews in Oral Biology & Medicine* 15:126-136.

Rice P, Longden I, and Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends in genetics* 16:276-277.

Rousselle AV, and Heymann D. 2002. Osteoclastic acidification pathways during bone resorption. *Bone* 30:533-540. S8756328202006725 [pii]

Saavedra RA. 1994. The roles of autophosphorylation and phosphorylation in the life of osteopontin. *BioEssays* 16:913-918. 10.1002/bies.950161210

Saier MH, Jr., Yen MR, Noto K, Tamang DG, and Elkan C. 2009. The Transporter Classification Database: recent advances. *Nucleic Acids Res* 37:D274-278. 10.1093/nar/gkn862

Salih E, Ashkar S, Gerstenfeld LC, and Glimcher MJ. 1996. Protein kinases of cultured osteoblasts: selectivity for the extracellular matrix proteins of bone and their catalytic competence for osteopontin. *J Bone Miner Res* 11:1461-1473. 10.1002/jbmr.5650111013

Sarashina I, Yamaguchi H, Haga T, Iijima M, Chiba S, and Endo K. 2006. Molecular Evolution and Functionally Important Structures of Molluscan Dermatopontin: Implications for the Origins of Molluscan Shell Matrix Proteins. *Journal of Molecular Evolution* 62:307-318. 10.1007/s00239-005-0095-2

Simkiss K, and Wilbur KM. 1989. *Biomineralization. Cell Biology and Mineral Deposition.* San Diego, CA: Academic Press.

Wallin R, Wajih N, Greenwood GT, and Sane DC. 2001. Arterial calcification: a review of mechanisms, animal models, and the prospects for therapy. *Med Res Rev* 21:274-301.

Westbroek P, and Marin F. 1998. A marriage of bone and nacre. *Nature* 392:861-862. 10.1038/31798

Yang RS, Tang CH, Ling QD, Liu SH, and Fu WM. 2002. Regulation of fibronectin fibrillogenesis by protein kinases in cultured rat osteoblasts. *Mol Pharmacol* 61:1163-1173.