A Synthesis Approach to Understanding Repeated Peptides Conserved in Mineralization Proteins

Kiyotaka Shiba*,†,‡ and Tamiko Minamisawa†

Department of Protein Engineering, Cancer Institute, Japanese Foundation for Cancer Research, and Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Koto-ku, Tokyo 135-8550, Japan,

Received June 11, 2007; Revised Manuscript Received July 13, 2007

We created artificial proteins that contained repeats of a short peptide motif, Asn-Gly-Asx. In nature this motif is repeated within shell proteins as an idiosyncratic domain, while in vitro it has been shown to suppress calcification. The motif was embedded within peptide sequences that did or did not have the ability to form secondary structures, which provided the motif with a variety of physicochemical properties. Although a short synthetic peptide containing the motif did not inhibit calcification in vitro, some of the artificial proteins carrying repeats of the motif did show robust suppression of calcification. Artificial proteins lacking the motif did not exhibit suppressive activity. Likewise, one construct containing multiple repeats of the motifs also did not exert an inhibitory effect on calcification. Apparently, carrying the Asn-Gly-Asx motif is not, by itself, sufficient for expression of its cryptic activity; instead, certain physicochemical properties of the polypeptides mediate its manifestation. We anticipate that syntheses using "motif programming", such as the one described here, will shed light on the origin of repetitive sequences as well as on the evolution of biomineralization proteins.

Introduction

Extant proteins contain a multitude of periodicities in both their primary and tertiary structures.[1,2] There are two possible scenarios that could explain the origin of the repetitiveness observed within the structures of proteins. One is that periodic structures have an advantage when exerting their function so that periodicity has been emphasized during the course of evolution. The second scenario is that periodic structure is merely a record of a protein's upbringing; i.e., repeating structures prevail because biomolecules tend to give birth to periodic structures,[3] irrespective of their contribution to function. In that context a repetitive structure comes first; function only comes later and does not necessarily make the most of the repetitive structure.

Proteins isolated from mineralized tissues, and thus somehow involved in biomineralization, are also repositories of repetitive structures (see Supporting Information for details). For example, nacrein from pearl oyster contains a domain comprised of repeats of a tripeptide sequence, while aspein, another protein from pearl oyster, contains repeats of a pentapeptide and its derivatives.[4] Similarly, SM50 from the sea urchin spicule contains mixtures of repeats of a hexapeptide and a heptapeptide,[5] while starmaker from the zebrafish otolith contains three tandem repeats of 13-amino-acid sequences in its N-terminal domain.[6] Most of the structures of mucoperlin from mussel shell and MAM22 from a magnetic bacterium are respectively filled with repeats of 31- and 34-amino-acid sequences.[7,8] Finally, MSP-1 from scallop shell and lustrin A from abalone nacre show convoluted repetitive structures.[9,10]

So how do these periodic structures contribute to the biomineralization process and are these repetitive structures are unique to biomineralization proteins? Although our knowledge of the molecular mechanism by which biomolecules promote mineralization of inorganic materials remains limited, it has been proposed that biomineralization proteins (i) provide nucleation sites (nucleators),[11] (ii) suppress crystal growth (suppressors),[12] and (iii) form frameworks for crystal growth (frameworker).[11,13,14] The repetitive structures of nucleators serve to align multiple nucleation sites. In that way, the protein may act as a template for epitaxial-like crystal growth. Weiner and Hood first proposed this protein template hypothesis in 1975, after they found that matrix proteins from shell contain blocks of Asp-Gly or Asp-Ser repeats and that the distance between the repeating Asp residues (∼7 Å when the peptide is fully extended) roughly matched the $Ca^{2+}-Ca^{2+}$ distances within the crystal lattices of aragonite and calcite.[11] Such molecular complementarity between protein and crystal surfaces also could explain the inhibitory action of suppressor proteins.[15] Several proteins have been proposed to bind to a specific crystal face within inorganic materials through structural complementarity, thereby modulating the appearance of crystals.[16] Thus the periodic structure of mineralization-related proteins is compatible with both the nucleator and the suppressor functions of these proteins. It also is noteworthy that sometimes the same protein can serve as either a nucleator or a suppressor, depending upon the assay conditions.[17−19]

Periodic structure is not an exclusive feature of biomineralization proteins. Others exhibiting periodicity include cytoskeletal proteins that often assemble into macroscopic supramolecular structures.[2] In this regard, the third role of mineralization proteins, the formation of frameworks for crystal growth, also may reflect their periodic structure. As mentioned earlier, however, we cannot rule out the possibility that the apparent periodicity observed in mineralization proteins is not at all related to the functions of these proteins. The repetitive structures could merely represent their evolutionary paths; i.e., long open reading frames frequently emerged from reiteration

* Author to whom correspondence should be addressed. E-mail: kshiba@jfcr.or.jp

† Japanese Foundation for Cancer Research.

‡ Japan Science and Technology Agency.

10.1021/bm700652b CCC: $37.00
Published on Web 08/01/2007

of short DNA sequences, and functions such as mineralization and self-assembly, among many others, are easily found within these long open reading frames.[20−22]

In the present work, we focused on the short sequence motif reiterated in the natural mineralization protein nacrein and synthesized artificial proteins assembled from the motif with the aim of determining how readily functional mineralization-related proteins can arise from repeats of this motif.

Experimental Procedures

Protein Construction and Purification. Construction, expression, and purification of artificial proteins essentially followed procedures that we described previously.[21,22] Details of the reaction conditions and vector construction are shown in the Supporting Information.

Circular Dichroism Measurements. Circular dichroism (CD) spectra were recorded with a JASCO J-820 spectrophotometer (Japan Spectroscopic, Tokyo) using a cell with a path length of 1 mm. Samples (5 μM) were prepared in water purified with a Milli-Q Biocel A10 System (Millipore, Billerica). The spectra from 190 to 280 nm were recorded at 25 °C.

In Vitro Monitoring of $CaCO_3$ Formation. We evaluated $CaCO_3$ formation based on pH changes using a method originally developed by Wheeler et al.[12] and modified by Weiss et al.[23] Because the assay temperature had a critical effect on the results, we used a temperature-controlled stirrer (Cool Stirrer SWC-900, Nissin, Tokyo) in which the reaction vessel was held within a temperature-controlled chamber, and all reaction reagents were equilibrated to the temperature used before starting the reaction. In addition, because buffer constituents interfered with the pH changes used to assess $CaCO_3$ formation, we only used proteins that could be dissolved in water, and those proteins were then extensively dialyzed before their use in the assays. The first step in the reaction was to incubate 3 mL of 20 mM $NaHCO_3$ in 3 mM TrisHCl (pH 8.7) in a 10 mL polystyrene bottle (Sansyo ϕ = 25 mm, Tokyo). One minute after pH monitoring was started, 3 mL of 20 mM $CaCl_2$ in 3 mM TrisHCl (pH 8.7), which was freshly prepared and pre-equilibrated at the indicated temperature, and 50 μL of the protein solution, which was freshly dialyzed against water, were added to the $NaHCO_3$ solution. The pH of the solution was then recorded every 1 s using a pH meter (model F22, Horiba Inc., Tokyo) with a Φ3 mm electrode (6069MP-10C).

Results

Experimental Design. Our focus in this study was the Asn-Gly-Asx (Asx = Asp or Asn) motif that has been identified in nacrein.[24] Nacrein was isolated as a major soluble organic matrix protein from the nacreous layer of pearls cultured in *Pinctada fucata*. Analysis of its cDNA revealed that the protein is composed of 447 amino acids and is a member of the carbonic anhydrase family of enzymes, which catalyze the interconversion of carbon dioxide and bicarbonate and are widely distributed among species. In particular, the enzymes expressed in organs engaged in calcification are thought to be central players during biomineralization.[24−27] Resolution of the three-dimensional structure of human carbonic anhydrase II, one of the closest homologues to nacrein, by X-ray crystallography revealed that the structure contains a twisted β-sheet composed of 10 β-strands (Supporting Information). Idiosyncratic to nacrein is the presence of a large inserted domain containing ~130 amino acids. From multiple alignments of related sequences, it can be discerned that the inserted domain is located somewhere between strands 7 and 8, starting in the vicinity of the active site and encircling to the backside of the active site (Supporting Information). It appears that this large inserted domain may flop over the active site or lie on the back of the enzyme so that it does not interfere with the enzymatic activity.

A unique feature of the inserted domain of nacrein form *P. fucata* is that it is composed of simple repeats of Asn-Gly-Asx (NGX) (Supporting Information), specifically 11 × NGN and 6 × NGD. Relative repeats, including 3 × NGY, 3 × NGE, and 1 × NGV also were detected. Interestingly, the numbers and sequences of the repeats in this nacrein isoform differ from those in related species. For example, a nacrein from the Australian pearl oyster *P. maxima* contains an inserted domain that is much longer (~245 amino acids) than the one from *P. fucata*.[28] The repeated sequences include 36 × NGN, 4 × NGD, 4 × NGY, and 4 t× NGG as well as repeats of a dipeptide (6 × GN) that is not found in *P. fucata*. Notably, Miyamoto et al. recently noticed that this repeated GN dipeptide dominates within the ~235-amino-acid inserted domain of the nacrein isolated from *Turbo marmoratus* (turban shell) (Supporting Information), which belongs to separate class of *Pinctada*. (The pearl oyster is a bivalve (Bivalvia), whereas the turban shell is a conch (Gastropoda).) They suggested that the different compositions of the inserted domains within the different nacreins are indicative of the evolvability of the inserted domain.[29]

Although the in vivo function of nacrein during calcification is not yet certain, indirect evidence suggests that the protein somehow modulates the process. For instance, Miyamoto et al., used *Escherichia coli* to express a recombinant protein containing the 17 repeats of the Asn-Gly-Asx motif from the inserted domain of the nacrein from *P. fucata* and showed that the purified protein suppressed calcification in vitro.[30] Often, however, isolated peptide motifs (synthesized peptides) fail to recapitulate the functionality expressed in the intact protein, which we found to be case with the Asn-Gly-Asx motif (see below). Conversely, peptide motifs often exert their cryptic functions only when they are displayed on proteins in a particular context.[31,32] With that in mind, we decided to create a set of artificial proteins containing repeats of the Asn-Gly-Asx motif with the aim of determining whether a functional protein, i.e., one that suppresses calcification, can be found among the created proteins and, if so, how readily they occur.

Creation of Artificial Proteins Comprised from the Asn-Gly-Asx Motif. We constructed a set of artificial proteins using our MolCraft system,[33] in which a custom-made microgene is rationally designed and then tandemly polymerized to form microgene repeats. Importantly, the conditions under which the microgene polymerization reaction (MPR; ref 21) is carried out allows changes in the reading frames to occur randomly at the junctions of the microgene units, so that the resultant microgene repeats encode combinatorial polymers of three peptides from three reading frames of the single microgene. By embedding different functional motifs (or sequences having certain physicochemical properties) within the different reading frames of the microgene, we are able to create a library of artificial proteins that contain various numbers of these motifs in various orders.[31,32]

We initially designed microgenes in which the first reading frames were filled with Asn-Gly-Asx repeats and then prepared libraries from the microgenes. However, we were never able to express any artificial proteins from these libraries. We tested a number of expression conditions using several *E. coli* strains and also tried an in vitro translation system, but none produced any detectable proteins (data not shown). Although we are still not certain why this protein failed to be expressed, we suggest that expression was somehow impaired by the skewed amino acid composition encoded by the microgene. Therefore, we next

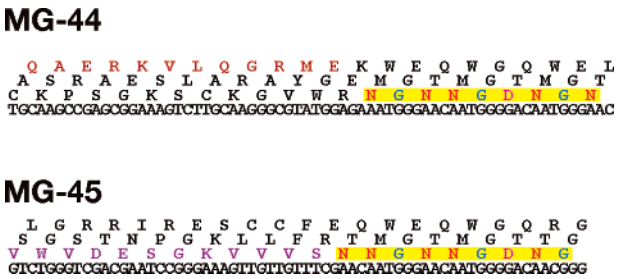

Figure 1. Designer microgenes used in this study. Peptides translated from three reading frames of a single microgene sequence are shown above the microgene sequence. The sequences highlighted by the yellow box were extracted from the inserted domain of *Pinctada* nacrein. The sequences shown in brown and purple are from existing proteins. Asparagine, asparate, and glycine in the motif are shown in red, magenta, and cyan, respectively.

Figure 2. Artificial proteins created by polymerization of designer microgenes using the MPR method. Clones #382 and #386 were derived from MG-44 polymers, whereas #458 and #448 were from MG-45. Color coding is the same as described in the legend to Figure 1 except all asparagine, asparate, and glycine residues are colored.

designed a second generation of microgenes by fusing together two smaller microgenes: one that we already knew would give a productive library and the other encoding three Asn-Gly-Asx repeats (Figure 1). The first 39 nucleotides of the 66-nucleotide microgene MG-44 encoded a peptide sequence related to the α-helical structure of an existing protein;[21] artificial proteins containing this sequence tend to be well-expressed and to have an α-helical structure.[22] The remaining 27 nucleotides encoded three Asn-Gly-Asx repeats. Similarly, the first 12 amino acids encoded by MG-45 corresponded to part of a 22-amino-acid sequence encoded by a microgene that we knew yielded artificial proteins having secondary structures.[22] From these two second-generation microgenes, we separately prepared artificial protein libraries and found that when expressed in *E. coli* most randomly selected clones produced large amounts of artificial proteins (Supporting Information). Thus, our inability to express the mineralization motif was remedied by appending extra sequences to the microgenes encoding the skewed amino acid compositions.

Characterization of the Artificial Proteins. We purified well-expressed artificial proteins and initially characterized their solubility under various conditions. Because the presence of

Table 1. Properties of Artificial Proteins

name	no. of residues	pI[a]	no. of motifs[b]	motifs[c] (%)	Asn/Asp (%)	Gly (%)	inhibition index[d]
#382	93	10.1	6	19	17	19	>9.5[e]
#386	239	9.7	13	29	13	18	2.2
#458	135	5.8	8	18	24	16	3.5
#448	92	9.3	0	0	3	13	2.2

[a] Calculated values.[45] [b] Numbers of "NGN" and "NGD." Overlaps were not allowed in calculation. [c] Numbers of motifs were divided by (numbers of resides/3). [d] Calculated by (time needed to drop pH to the halfway point with proteins)/(time needed without protein) in calcification experiments (Figure 4). [e] Calculated using the data of Figure S9B (Supporting Information).

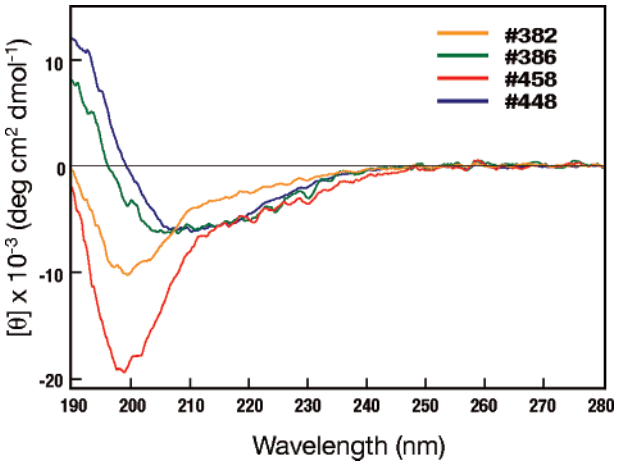

Figure 3. Far-UV CD spectra of artificial proteins. Protein samples (5 μM) were dissolved in water, and CD spectra were measured in a cell having a path length of 1 mm. For details see Experimental Procedures.

buffer proved unfavorable in the in vitro calcification assay (see below), we focused on four clones that were soluble in water (Figure 2). Clones #382 and #386 were derived from MG-44 and, respectively, contained 6 and 13 repeats of the triplet motif in the context of α-helix-preferring sequences. Clones #458 and #448 were derived from MG-45, one reading frame of which was related to part of a β-helical protein. Clone #458 contained 13 repeats of the triplet, whereas clone #448 contained no repeats of the triplet and was therefore used as a negative control. The properties of the amino acid compositions and pI's are summarized in Table 1.

Results of CD analyses are shown in Figure 3. The CD spectra for #382 and #458 indicated that these proteins do not have secondary structures under the conditions used. By contrast, the spectra of #386 and #448 suggested that both have some secondary structure, though the broad peaks were not indicative of any particular type. Given that #386 contains three copies of a peptide derived from a natural α-helical protein, one might expect these sequences to contribute to the formation of an α-helix. At the same time, #382 and #458 did not acquire the ability to form secondary structure, despite the fact that they, too, contained sequences related to natural proteins, while #448 had secondary structure though its sequence was not explicitly embedded with secondary structure. That these peptides exhibited both rationality and capriciousness in their structural programming is a hallmark of the combinatorial approach of MolCraft. Unstructured mineralization proteins reportedly can acquire secondary structure in the presence of Ca^{2+};[34] however, when we tested the effect of Ca^{2+} on the CD spectra of #382, we found that the peptide remained unstructured in the presence of Ca^{2+} (data not shown).

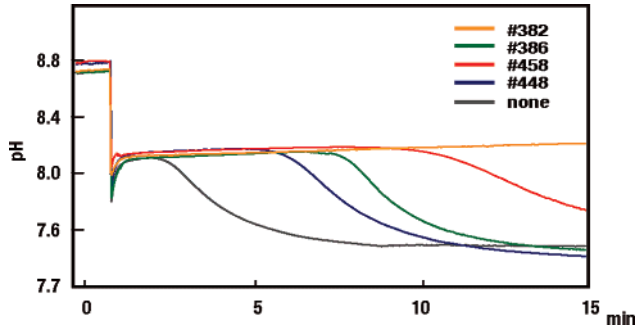

Figure 4. In vitro inhibition of $CaCO_3$ formation by artificial proteins was detected based on pH changes.

In Vitro Suppression of Calcification. We next assessed whether #382, #386, or #458 would suppress $CaCO_3$ formation in vitro by monitoring pH changes after mixing each of the artificial proteins with $NaHCO_3$ and $CaCl_2$. (The pH of the solution declines as $CaCO_3$ is formed.) This assay method was first developed by Wheele et al.[12] and has been widely used to monitor the suppressor and nucleator activities of biomineralization molecules.[8,23,30,35,36] In the present study we were careful to keep the reaction temperature constant (see Experimental Procedures) and to avoid carryover of buffer constituents from protein samples, which we found to have a significant effect on pH changes.

We first used the pH monitoring system to test the activity of a synthetic 9-mer peptide (NGN-NGD-NGN). Under the conditions used, the peptide caused no significant delay in the drop in pH, indicating that the isolated peptide motif was not sufficient to suppress $CaCO_3$ formation in this assay (Supporting Information). By contrast, two of the artificial proteins tested did impede the drop in pH (Figure 4). In particular #382 exerted a pronounced suppressor effect, completely inhibiting the formation of $CaCO_3$, as indicated by the absence of a drop in pH, even after incubation for 15 min at 25 °C. Clone #458 was the second most effective suppressor, whereas #386 and #448 had very little or no suppressive effect (scored by inhibitory index in Table 1).

Focusing on #382, we found that at the ambient temperature the suppressive effect of the protein was concentration-dependent; i.e., the time to the onset of the decline in pH lengthened and the rate of the subsequent decline slowed as the amount of #382 added to the assay was increased from 0 to 20 μg (Supporting Information). Moreover, the suppressive effects exerted by 20 μg of #382 were diminished as the assay temperature was increased from 25 to 40 °C (Supporting Information). Indeed, at 40 °C, both the time to onset and the rate of decline in pH were indistinguishable from those observed in a control experiment with no inhibitory protein (Supporting Information).

Among the three artificial proteins containing Asn-Gly-Asx repeats (#382, #386, and #458), #386 showed the least ability to suppress calcification. At the same time, CD analyses showed that both #386 and #448 had some secondary structure (Figure 3). This suggests the possibility that an unstructured state is the key feature that enables a peptide to act as a suppressor, not the presence of the motif. To test this possibility, we investigated the suppressive activities of two other unstructured and structured artificial proteins created from MG-45 and confirmed that these proteins did not suppress calcification (Supporting Information). In addition, we tested three unstructured artificial proteins created from microgenes other than MG-44 and MG-45 and found that they, too, failed to exhibit suppressor activity (Supporting Information). Apparently, an unstructured state is not sufficient to confer suppressor activity to a peptide; the presence of the Asn-Gly-Asx motif was required. That said, our data also showed that the presence of the Asn-Gly-Asx motif was not sufficient for suppression of $CaCO_3$ formation: Clone #386 contained the highest numbers of the motif per weight (Table 1) but showed little suppressive activity. Thus expression of the Asn-Gly-Asx motif's cryptic activity likely requires its display in the proper context of a polypeptide, which would be consistent with our earlier findings.[31,32]

At the moment, we do not know the physicochemical properties that underlie the emergence of the prominent inhibitory activity of the Asn-Gly-Asx motif. The calculated isoelectric points of #382 and #458 markedly differ, but both actively suppressed $CaCO_3$ formation (Table 1). That the polypeptide is in an unstructured state is one plausible condition, as the unstructured #382 and #458 showed much stronger suppressive activity than the structured #386. The fact that many natural biomineralization proteins adopt unstructured states is compatible with this idea, though it remains to be further tested.[37−39]

Discussion

Because they replicate in a semiconservative manner, unequal pairing followed by reiteration of sequence information is the inevitable nature of nucleic acids.[40] Consequently, the genomes of extant organisms are abundant with repetitive structures that may serve as a type of primordial soup from which new genes emerge. As sources of embryonic genes, repetitive sequences are advantageous in part because longer open reading frames can emerge at a higher frequency, as long as the repeating unit is devoid of a translation termination codon.[40] In addition, the proteins translated from repeated sequences would likely have a higher propensity to form secondary structures because of their periodicities.[22,40,41] The flowing generation of repetitive sequences never stops, and yet they never stay the same because most have no function and are therefore under no selection pressure. Once the products gain a particular set of biological functions, the genes are immobilized by the evolutionary pressure. If the structural solutions for a given function are scattered throughout sequence space, then the chances that repetitive sequences will gain that function would seem high.[42] At the same time, we would predict that the same biological function would be performed by various periodic structures having distinct repeating units. Antifreeze proteins are a good example of such convergent evolution; closely related species use distinct repeating units to realize periodic arrangement of threonine residues within these proteins.[15] Another intriguing example is involucrin, a protein expressed in keratinocytes. Involucrins from human and lemur have evolved their own appendix domains through different short repeating sequences that situate distinct sites within the protein.[43] Most likely, these domains evolved independently after separation of human and lemur, by selecting particular sequences as repeating units. The evolution of involucrin throws some light on the evolution of nacrein's idiosyncratic domain. Although closely related, *P. fucata* and *T. marmoratus* appear to have evolved their appendix domains separately by, respectively, reiterating the Asn-Gly-Asx and Gly-Asn unit. The resemblance of these repeating units may reflect the domain's function, i.e., modulation of calcification.

On the basis of top-down analyses of the structures of extant proteins, we can infer the evolutional paths of repeating

sequences. The synthesis approach, as adopted in this work, fully complements reductionist approaches by reconstituting an inferred story in a bottom-up manner.[44] Here, we started from a short peptide motif found as a reiterated form in biomineralization-related proteins and created artificial proteins containing multiple copies of the motif. A number of artificial proteins recapitulated the mineralization-related activity of the parental protein, though the isolated motif did not. From that observation, we conclude that modulation of calcification likely evolved from reiteration of the Asn-Gly-Asx motif with relative ease. Because codons for Asn-Gly-Asx should be found everywhere within genomes, genes related to calcification may readily emerge de novo.

This work and our earlier studies indicate that the emergence of function from motif repeats is dependent upon the physicochemical properties of the polypeptide in which they reside. Future elucidation of those physicochemical properties will deepen our knowledge of the mechanisms underlying biomineralization and the evolution of proteins.

Conclusions

We created artificial proteins that contained the motif repeating within shell proteins. In vitro calcification assays showed that a synthetic peptide containing the motif did not inhibit calcification, whereas some of the artificial proteins carrying repeats of the motif showed robust suppression of calcification. The data suggested that mineralization proteins likely evolved from reiteration of short sequences with relative ease.

Acknowledgment. The authors thank Dr. J. S. Evans for discussion and Ms. M. Sugiyama for her help with the matrix-assisted laser desorption ionization time-of-flight mass spectrometry analyses. This work is supported in part by a Human Frontier Science Program grant and a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan, to K.S.

Supporting Information Available. Experimental procedures, figures showing periodic structures of mineralization proteins, gel electrophoresis of artificial proteins, $CaCO_3$ formation assay, CD analyses, and Fourier transform infrared analyses. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Marcotte, E. M.; Pellegrini, M.; Yeates, T. O.; Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **1999**, *293*, 151−160.
(2) Andrade, M. A.; Perez-Iratxeta, C.; Ponting, C. P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **2001**, *134*, 117−131.
(3) Ohno, S. Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes. *J. Mol. Evol.* **1984**, *20*, 313−321.
(4) Tsukamoto, D.; Sarashina, I.; Endo, K. Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochem. Biophys. Res. Commun.* **2004**, *320*, 1175−1180.
(5) Katoh-Fukui, Y.; Noce, T.; Ueda, T.; Fujiwara, Y.; Hashimoto, N.; Higashinakagawa, T.; Killian, C. E.; Livingston, B. T.; Wilt, F. H.; Benson, S. C.; Sucov, H. M.; Davidson, E. H. The corrected structure of the SM50 spicule matrix protein of *Strongylocentrotus purpuratus*. *Dev. Biol.* **1991**, *145*, 201−202.
(6) Sollner, C.; Burghammer, M.; Busch-Nentwich, E.; Berger, J.; Schwarz, H.; Riekel, C.; Nicolson, T. Control of crystal size and lattice formation by starmaker in otolith biomineralization. *Science* **2003**, *302*, 282−286.
(7) Okuda, Y.; Denda, K.; Fukumori, Y. Cloning and sequencing of a gene encoding a new member of the tetratricopeptide protein family from magnetosomes of *Magnetospirillum magnetotacticum*. *Gene* **1996**, *171*, 99−102.
(8) Marin, F.; Corstjens, P.; de Gaulejac, B.; de Vrind-De Jong, E.; Westbroek, P. 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.* **2000**, *275*, 20667−20675.
(9) Sarashina, I.; Endo, K. The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. *Mar. Biotechnol.* **2001**, *3*, 362−369.
(10) Shen, X.; Belcher, A. M.; Hansma, P. K.; Stucky, G. D.; Morse, D. E. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *J. Biol. Chem.* **1997**, *272*, 32472−32481.
(11) Weiner, S.; Hood, L. Soluble protein of the organic matrix of mollusk shells: A potential template for shell formation. *Science* **1975**, *190*, 987−989.
(12) Wheeler, A. P.; George, J. W.; Evans, C. A. Control of calcium carbonate nucleation and crystal growth by soluble matrix of oyster shell. *Science* **1981**, *212*, 1397−1398.
(13) Berman, A.; Addadi, L.; Weiner, S. Interactions of sea-urchin skeleton macromolecules with growing calcite crystals−A study of intracrystalline proteins. *Nature* **1988**, *331*, 546−548.
(14) Du, C.; Falini, G.; Fermani, S.; Abbott, C.; Moradian-Oldak, J. Supramolecular assembly of amelogenin nanospheres into birefringent microribbons. *Science* **2005**, *307*, 1450−1454.
(15) Davies, P. L.; Baardsnes, J.; Kuiper, M. J.; Walker, V. K. Structure and function of antifreeze proteins. *Philos. Trans. R. Soc. London, Ser. B* **2002**, *357*, 927−935.
(16) DeOliveira, D. B.; Laursen, R. A. Control of calcite crystal morphology by a peptide designed to bind to a specific surface. *J. Am. Chem. Soc.* **1997**, *119*, 10627−10631.
(17) Romberg, R. W.; Werness, P. G.; Riggs, B. L.; Mann, K. G. Inhibition of hydroxyapatite crystal growth by bone-specific and other calcium-binding proteins. *Biochemistry* **1986**, *25*, 1176−1180.
(18) Linde, A.; Lussi, A.; Crenshaw, M. A. Mineral induction by immobilized polyanionic proteins. *Calcif. Tissue. Int.* **1989**, *44*, 286−295.
(19) Onuma, K. Effect of phosvitin on the nucleation and growth of calcium phosphates in physiological solutions. *J. Phys. Chem. B* **2005**, *109*, 8257−8262.
(20) Rey, M.; Ohno, S.; Pintor-Toro, J. A.; Llobell, A.; Benitez, T. Unexpected homology between inducible cell wall protein QID74 of filamentous fungi and BR3 salivary protein of the insect *Chironomus*. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6212−6216.
(21) Shiba, K.; Takahashi, T.; Noda, T. Creation of libraries with long open reading frames by polymerization of a microgene. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3805−3810.
(22) Shiba, K.; Takahashi, Y.; Noda, T. On the role of periodism in the origin of proteins. *J. Mol. Biol.* **2002**, *320*, 833−840.
(23) Weiss, I. M.; Kaufmann, S.; Mann, K.; Fritz, M. Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*. *Biochem. Biophys. Res. Commun.* **2000**, *267*, 17−21.
(24) Miyamoto, H.; Miyashita, T.; Okushima, M.; Nakano, S.; Morita, T.; Matsushiro, A. A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9657−9660.
(25) Gay, C. V.; Mueller, W. J. Cellular localization of carbonic anhydrase in avian tissues by labeled inhibitor autoradiography. *J. Histochem. Cytochem.* **1973**, *21*, 693−702.
(26) Kingsley, R.; Watabe, N. Role of carbonic anhydrase in calcification in the gorgonian *Leptogorgia virgulata* (Lamarck) (Coelenterata: Gorgonacea). *J. Exp. Zool.* **1987**, *241*, 171−180.
(27) Miyashita, T.; Takagi, R.; Miyamoto, H.; Matsushiro, A. Identical Carbonic anhydrase contributes to nacreous or prismatic layer formation in *Pinctada fucata* (Mollusca: Bivalvia). *Veliger* **2002**, *45*, 250−255.
(28) Kono, M.; Hayashi, N.; Samata, T. Molecular mechanism of the nacreous layer formation in *Pinctada maxima*. *Biochem. Biophys. Res. Commun.* **2000**, *269*, 213−218.
(29) Miyamoto, H.; Yano, M.; Miyashita, T. Similarities in the structure of nacrein, the shell-matrix protein, in a bivalve and a gastropod. *J. Mollus. Stud.* **2003**, *69*, 87−89.

(30) Miyamoto, H.; Miyoshi, F.; Kohno, J. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. *Zool. Sci.* **2005**, *22*, 311−315.
(31) Saito, H.; Honma, T.; Minamisawa, T.; Yamazaki, K.; Noda, T.; Yamori, T.; Shiba, K. Synthesis of functional proteins by mixing peptide motifs. *Chem. Biol.* **2004**, *11*, 765−773.
(32) Saito, H.; Minamisawa, T.; Shiba, K. Motif programming: A microgene-based method for creating synthetic proteins containing multiple functional motifs. *Nucleic Acids Res.* **2007**, *35*, e38.
(33) Shiba, K. MolCraft: A hierarchical approach to the synthesis of artificial proteins. *J. Mol. Catal. B: Enzym.* **2004**, *28*, 145−153.
(34) Hoang, Q. Q.; Sicheri, F.; Howard, A. J.; Yang, D. S. Bone recognition mechanism of porcine osteocalcin from crystal structure. *Nature* **2003**, *425*, 977−980.
(35) Zhang, Y.; Xie, L.; Meng, Q.; Jiang, T.; Pu, R.; Chen, L.; Zhang, R. A novel matrix protein participating in the nacre framework formation of pearl oyster, *Pinctada fucata*. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **2003**, *135*, 565−573.
(36) Marin, F.; Amons, R.; Guichard, N.; Stigter, M.; Hecker, A.; Luquet, G.; Layrolle, P.; Alcaraz, G.; Riondet, C.; Westbroek, P. Caspartin and calprismin, two proteins of the shell calcitic prisms of the mediterranean fan mussel *Pinna nobilis*. *J. Biol. Chem.* **2005**, *280*, 33895−33908.
(37) Michenfelder, M.; Fu, G.; Lawrence, C.; Weaver, J. C.; Wustman, B. A.; Taranto, L.; Evans, J. S.; Morse, D. E. Characterization of two molluscan crystal-modulating biomineralization proteins and identification of putative mineral binding domains. *Biopolymers* **2003**, *70*, 522−533.
(38) Wustman, B. A.; Weaver, J. C.; Morse, D. E.; Evans, J. S. Characterization of a Ca(II)-, mineral-interactive polyelectrolyte sequence from the adhesive elastomeric biomineralization protein lustrin A. *Langmuir* **2003**, *19*, 9373−9381.
(39) Kim, I. W.; Morse, D. E.; Evans, J. S. Molecular characterization of the 30-AA N-terminal mineral interaction domain of the biomineralization protein AP7. *Langmuir* **2004**, *20*, 11664−11673.
(40) Ohno, S. Evolution from primordial oligomeric repeats to modern coding sequences. *J. Mol. Evol.* **1987**, *25*, 325−329.
(41) Shiba, K.; Shirai, T.; Honma, T.; Noda, T. Translated products of tandem microgene repeats exhibit diverse properties also seen in natural proteins. *Protein Eng.* **2003**, *16*, 57−63.
(42) Shiba, K.; Honma, T.; Minamisawa, T.; Nishiguchi, K.; Noda, T. Distinct macroscopic structures developed from solutions of chemical compounds and periodic proteins. *EMBO Rep.* **2003**, *4*, 148−153.
(43) Tseng, H.; Green, H. Remodeling of the involucrin gene during primate evolution. *Cell* **1988**, *54*, 491−496.
(44) Benner, S. A. Act natural. *Nature* **2003**, *421*, 118.
(45) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005; pp 571−607.

BM700652B