# Primary structure of a soluble matrix protein of scallop shell: Implications for calcium carbonate biomineralization

## I. Sarashina and K. Endo*

Geological Institute, University of Tokyo, Tokyo 113-0033, Japan

## Abstract

Soluble proteins in the scallop (*Patinopecten yessoensis*) foliated calcite shell layer were characterized using biochemical and molecular biological techniques. SDS PAGE of these molecules revealed three major protein bands, 97 kD, 72 kD, and 49 kD in molecular weight, when stained with Coomassie Brilliant Blue. Periodic Acid Schiff staining and Stains-All staining indicated that these proteins are slightly glycosylated and may have cation-binding potential. N-terminal sequencing of the three proteins revealed that all three share the same amino acid sequence at least for the first 20 residues. A partial amino acid sequence of 436 amino acids of one of these proteins (MSP-1) was deduced by characterization of the complementary DNA encoding the protein. The deduced sequence is composed of a high proporton of Ser (31%), Gly (25%), and Asp (20%), typifying an acidic glycoprotein of mineralized tissues. The protein has a basic domain near the N-terminus and two highly conserved Asp-rich domains interspersed in three Ser and Gly-rich regions. In contrast with prevalent expectations, (Asp-Gly)n-, (Asp-Ser)n-, and (Asp-Gly-X-Gly-X-Gly)n-type sequence motifs do not exist in the Asp-rich domains, demanding revision of previous theories of protein-mineral interactions.

## Introduction

Minerals produced by organisms often have crystal shapes clearly different from those formed inorganically. Most such biominerals are a composite of inorganic crystals and organic molecules such as lipids, polysaccharides, and proteins, collectively known as the organic matrix. It is generally postulated that the elaborate fabrication of biominerals arises from specific molecular interactions at inorganic-organic interfaces (Mann et al. 1993), and that the organic matrix represents many of the important molecules involved in the interactions controlling crystal growth (e.g., Watabe and Wilbur 1960; Lowenstam 1981; Weiner 1984; Lowenstam and Weiner 1989).

Calcium carbonate is one of the most common biominerals, and its matrix molecules, especially of molluscan shells, have been studied to a considerable extent to unveil their roles in the mineralization processes. The matrix molecules have been classified conventionally into two types based on their solubility in aqueous solutions: the insoluble matrix is thought to be largely intercrystalline (Krampitz 1982) and provides a framework where mineralization is to occur, whereas the soluble matrix is known as intracrystalline or located on the intercrystalline matrix surfaces, but its functions are still poorly understood (Addadi and Weiner 1997).

Advocated functions of the mainly proteinaceous, soluble matrix of the molluscan shell in particular, include:

(1) induction of oriented nucleation (Weiner 1975; Weiner and Addadi 1991); (2) inhibition of crystal growth (Wheeler et al. 1981; Wheeler 1992); (3) control of aragonite-calcite polymorphism (Falini et al. 1996; Belcher et al. 1996); and (4) enhancement of mechanical properties of the crystals (Berman et al. 1988; Berman et al. 1990). Most of the evidence to support these hypotheses has been obtained through in vitro experiments. The weakness of these studies is that unpurified proteins or protein fractions of dubious homogeneity have been applied in the biochemical analyses and in the in vitro mineralization experiments, and that the stereochemical relationships between the organic and inorganic phases have been presumed without precise information of the fine structures of the proteins.

To understand the underlying mechanisms of the protein-mineral interactions, it seems essential first of all to know the primary structure of the proteins involved. However, only a limited number of amino acid sequences have been determined so far for the calcium carbonate matrix proteins. The available sequences comprise those from spicules of sea urchin emryo (Sucov et al. 1987; Katoh-Fukui et al. 1991; Katoh-Fukui et al. 1992; Benson and Wilt 1992), from pearl oyster shell layers (Miyamoto et al. 1996; Sudo et al. 1997), and from the nacre of a gastropod shell (Shen et al. 1997), in addition to partial sequences from brachiopod shells (Cusack et al. 1992), an oyster shell (Wheeler 1992), and gastrolith of a crayfish (Ishii et al. 1996; Ishii et al. 1998). Here we present a partial amino acid sequence of the molluscan shell pro-

---

* E-mail: endo@geol.s.u-tokyo.ac.jp

tein MSP-1, a major soluble matrix protein of the scallop foliated calcite shell layer, and discuss its bearing on the functions of matrix proteins in calcium carbonate biomineralization.

## MATERIALS AND METHODS

### Isolation of MSP-1

Specimens of the commercial scallop *Patinopecten yessoensis* were purchased locally. A single shell valve was thoroughly cleaned mechanically and incubated for 48 h at room temperature in a 10 vol% solution of sodium hypochlorite to destroy surface contaminants. After thorough washing with ultrapure water, the marginal portion of the shell, consisting only of the outer shell layer of foliated calcite, was crushed to fine fragments. The matrix proteins were extracted by dissolution of the shell flakes (100 g) in 3 liters of 0.5 $M$ EDTA (ethylenediaminetetraacetate), pH 8. The extraction was performed at 4 °C with continuous stirring for 72 h. The preparation was then filtered through cheesecloth to remove viscous insoluble materials and desalted by ultrafiltration using the Minitan tangential flow system (Millipore). In this procedure, the amount of EDTA was reduced to less than $10^{-6}$ mol, at which point the sample was concentrated to about 50 mL, then lyophilized.

### SDS PAGE analyses

The extracted macromolecules (1 mg per each well) were separated by SDS PAGE (Laemmli 1970), using slab gels (15 × 15 cm) of 2 mm thickness containing 12% polyacrylamide. After electrophoresis, gels were stained by Coomassie Brilliant Blue R or silver (Silver Stain kit; Bio-Rad) to visualize proteins, by Stains-All to visualize cation-binding proteins (Campbell et al. 1983), and by periodic acid and Schiff's reagent to visualize carbohydrates (Holden et al. 1971).

### N-terminal sequence determination

Following separation by SDS PAGE, the proteins were electroblotted onto polyvinylidene difluoride membrane (ProBlott; ABI) in Caps buffer (10 m$M$, pH 11) containing methanol (10 vol% solution), prior to staining with Coomassie Brilliant Blue R. N-terminal amino acid sequence analysis of the immobilized protein samples was by Edman degradation using an automated protein sequencer (Perkin-Elmer Applied Biosystems). Sequences were determined at least twice for each protein band reproduced by different SDS PAGE gels.

### RNA purification and RT-PCR

A nucleotide sequence in mRNA (messenger ribonucleic acids) codes for the terminal amino acid sequence found (as determined above) in *P. yessoensis* shell proteins. The nucleotide sequence is not uniquely determined by the amino acid sequence, as up to six three-base groups (codons) may code for the same amino acid. The actual sequence encoding the terminal amino acids is re-

quired to design unique primers to facilitate subsequent amplification of the DNA sequence encoding the entire protein.

We extracted the total RNA from the mantle tissue of a single specimen of *P. yessoensis,* using ISOGEN (Nippon Gene) and the single-step method for RNA isolation (Chomzynski 1993). The RNA (5 μg) was applied as a template for reverse transcription to prepare complementary DNA (cDNA) in a 20-μL reaction, primed with a "hybrid" primer, TCGAATTCGGATCC-GAGCTC(T)$_{17}$, using the SuperScript preamplification system (Life Technologies). The target cDNA sequence, encoding the N-terminal sequence of 20 amino acid residues determined for the purified MSP-1, was amplified by a method known as reverse transcription-polymerase chain reaction (RT-PCR). The sense primer and the antisense primer corresponded to the sequence encoding LDTDKD and NAAED (for one-letter abbrevation of amino acids, see the caption for Fig. 1), respectively, each being degenerate, containing oligonucleotides of all the possible sequences for each amino acid sequence. The reaction mixture (50 μL) contained 2 μL cDNA, 2 μ$M$ of each primer, 1 × Taq DNA polymerase buffer (Life Technologies), 3 m$M$ MgC$_{12}$, 100 μ$M$ dNTP, and 1 unit of Taq DNA polymerase (TOYOBO). A Cetus DNA Thermal Cycler (Perkin Elmer) was employed with an initial step of 94 °C for 3 min, then 30 cycles at 94 °C for 30 s, 52 °C for 30 s, 72 °C for 1 min, followed by a final extension step of 72 °C for 5 min. The resulting PCR products of 60 base pairs (bp) in length (corresponding to 20 amino acids), were sequenced directly by the chain termination method using the BigDye Terminator Cycle Sequencing Kit and an automated DNA sequencer (Perkin-Elmer Applied Biosystems).

### Amplification and sequencing of cDNA 3'-end

Three gene-specific sense primers, P1, P2, and P3 (Fig. 1), designed based on the sequence determined by the above method, and the antisense "adaptor" primer (PA), TCGAATTCGGATCCGAGCTC, were synthesized for the PCR amplification of the region between the point corresponding to the N-terminal end of the mature protein (MSP-1) and the 3'-end of the transcript (3' RACE: rapid amplification of cDNA ends protocol; Frohman 1990). The reactions using the primer pair of P1-PA were performed first, under conditions similar to those described above, except that the reactions were catalyzed by Ex Taq (TAKARA) in this case. A second and a third round of PCR reactions were performed with the P2-PA and P3-PA primer pairs, respectively, using the PCR products of the previous round of reactions as a template to verify specific amplification of the target cDNA fragment. PCR reactions with only one primer (P1, P2, P3, or PA) were included as negative controls. Unique PCR products were digested with the restriction enzyme BamHI and subcloned into the pUC 19 plasmid vector. The inserts encompassing the first 765 bp segment of the PCR products were sequenced for both strands using the M13 forward

```
      P1                    P2                         P3
  ┌CTTGATACTGATAAGGAT┐┌TTAGAATTTCATCTAGA┐T┌AGCTTACTAAATG┐AGCCGAAGAT   60
    L  D  T  D  K  D    L  E  F  H  L  D  S  L  L  N  A  A  E  D

  GGCGGTGGTGGCGATGCTGCTGGCGCCGAGGAGGCAGCACCAGCAGCAGATCTAAGCGGA  120
    G  G  G  G  D  A  A  G  A  E  E  A  A  P  A  A  D  L  S  G

  GGTAGCAAAGGAGGAAGCAAAGGAAGTAGCACAAGAAGTAGTAAGGGAGGTAGTAAGGGA  180
    G  S  K  G  G  S  K  G  S  S  T  R  S  S  K  G  G  S  K  G

  GGTAGTAAGGGAGGCAATGGTGGTGGAGATGCTGATGATTCAAGTAGCTCAAGCGGTTCT  240
    G  S  K  G  G  N  G  G  G  D  A  D  D  S  S  S  S  G  S

  GATTCGGGAAGTTCTGGAAGTGACGAAGAATCAGATGATTCCAGCTCTAGTTCCAGTTCT  300
    D  S  G  S  S  G  S  D  E  E  S  D  D  S  S  S  S  S  S

  GGTTCTGGTTCCGGCTCTGGCTCAGGTTCTGGCTCCGGCTCAAGCTCCAGCTCTGGCTCA  360
    G  S  G  S  G  S  G  S  G  S  G  S  G  S  S  S  S  G  S

  TCCAGCGATGGATCTGATGATGGTTCCGATGACGGGTCTGATTCAGGTGACGATGCTGAT  420
    S  S  D  G  S  D  D  G  S  D  D  G  S  D  S  G  D  D  A  D

  TCCGCTAATGCTGATGACCTTGATTCCAATGATTCCGATGATTCCGATAACTCCGGTTCC  480
    S  A  N  A  D  D  L  D  S  N  D  S  D  D  S  D  N  S  G  S

  AATGGCGAGTCTGACTCTGATAACTCCTCCTCCGACGATGGCGATGGTTCCGATTCCGGT  540
    N  G  E  S  D  S  D  N  S  S  S  S  D  D  G  D  G  S  D  S  G

  TCCGATTCTGGCAATGAGTAGTCCGATGACGCTTCTAATAATGATTCTGATGACTCC  600
    S  D  S  G  N  D  S  Q  S  D  D  A  S  N  N  D  S  D  D  S
                                                              P4
  GATGACTCGGATGATTCTTCTAATGATGTAAATGAATCAA┌TTCTGATGAATCTGGCC┐CT  660
    D  D  S  D  D  S  S  N  D  V  N  E  S  N  S  D  E  S  G  P

  GGAGGATATGGGAGGCAATGGACCAGCAGGCAATGGAGGCAAGGGACGCAAGGGAGGCAAT  720
    G  G  Y  G  G  N  G  P  A  G  N  G  G  K  G  R  K  G  G  N
                                        BamHI
  GGAGGAGGCAATGAAGGCAATGGAGGCAATGGAGGTGAT┌GGATCC┐AGTTCTAGTTCGAGC  780
    G  G  G  N  E  G  N  G  G  N  G  G  D  G  S  S  S  S  S

  TCTGGTTCCGGCTCTGGTTCTGGCTCCGGCTCTGGCTCAAGTTCTGGCTCAGGTTCTGGC  840
    S  G  S  G  S  G  S  G  S  G  S  G  S  G  S  G  S  G  S  G

  TCAGGTTCTGGGTCCGGCTCAAGCTCTAGCTCTAGCTCATCCGGCGATGGATCTGATGAT  900
    S  G  S  G  S  G  S  S  S  S  S  S  S  S  G  D  G  S  D  D

  GGTTCCGATGACGGGTCTGATGATGGTGACGATGGCGATAGTTCCGATGATGACCTT  960
    G  S  D  D  G  S  D  D  G  D  D  A  N  S  A  N  A  D  D  L

  GATTCCAATGATCCCGATGATTCCGATAACTCCGGTTCCAATGGCGAGTCTGACTCTGAT  1020
    D  S  N  D  P  D  D  S  D  N  S  G  S  N  G  E  S  D  S  D

  AACTCTTCCTCCGACGATGGCGATGGTTCCGATTCCGGTTCCGATTCCGGCGGAGATAGT  1080
    N  S  S  S  D  D  G  D  G  S  D  S  G  S  D  S  G  R  D  S

  CAGTCCGATGACGCTTCTAATAATGATTCTGATGACTCCGATGACTCTGATAATTCGTCT  1140
    Q  S  D  D  A  S  N  N  D  S  D  D  S  D  D  S  D  N  S  S

  ACTGGTACTGGCGAATCCGATTCCGATGAGTCTGGGCCTGGAGGATATGGAGGCAATGGA  1200
    T  G  T  G  E  S  D  S  D  E  S  G  P  G  G  Y  G  G  N  G

  CCAGCAGGCAATGGAGGTAAGGGACGCAAGGGAGGCAATGGAGGAGGCAATGGAGGCAAT  1260
    P  A  G  N  G  G  K  G  R  K  G  G  N  G  G  G  N  G  G  N

  GGCAATGGAGGCAATGGAGGTGATGGATCCAGTTCTAGTTCGAGCTCG  1308
    G  N  G  G  N  G  G  D  G  S  S  S  S  S  S
```

**FIGURE 1.** Partial cDNA sequence of MSP-1 and the deduced amino acid sequence. Numbers on the right indicate positions of the nucleotides in the MSP-1 cDNA sequence. The residues determined by Edman degradation are in bold. Sequences that correspond to the oligonucleotide primers (P1, P2, P3, and P4) are boxed. Cloning site is also boxed (BamHI). Potential N-glycosylation sites are underlined. One-letter abbreviation of amino acid residues (three-letter abbreviation in parenthesis): G = glycine (Gly); A = alanine (Ala); V = valine (Val); L = leucine (Leu); F = phenylalanine (Phe); P = proline (Pro); S = serine (Ser); T = threonine (Thr); N = asparagine (Asn); Q = glutamine (Gln); Y = tyrosine (Tyr); D = aspartate (Asp); E = glutamate (Glu); H = histidine (His); K = lysine (Lys); R = arginine (Arg).

**TABLE 1.** Amino acid compositions of MSP-1 (partial sequence), soluble matrix (bulk) of *Patinopecten yessoensis* (Kasai and Ohta 1981), and an hplc fraction of soluble matrix of the oyster *Crassostrea virginica* (Wheeler 1992)

| Amino acid | MSP-1 Scallop | Soluble fraction | |
|---|---|---|---|
| | | Scallop | Oyster |
| Asp | 19.5 | 25.2 | 31.5 |
| Thr | 0.9 | 1.6 | 0.5 |
| Ser | 31.0 | 22.4 | 28.3 |
| Glu | 3.0 | 6.0 | 4.2 |
| Pro | 1.4 | 2.4 | 0.9 |
| Gly | 25.0 | 26.3 | 30.0 |
| Ala | 4.6 | 6.5 | 0.9 |
| Cys | 0.0 | 1.4 | ND |
| Val | 0.2 | 1.1 | 0.2 |
| Met | 0.0 | 0.0 | ND |
| Ile | 0.0 | 0.7 | 0.1 |
| Leu | 1.8 | 1.6 | 0.1 |
| Tyr | 0.5 | 0.0 | 1.7 |
| Phe | 0.2 | 0.6 | 0.1 |
| Lys | 2.3 | 3.4 | 0.6 |
| Arg | 0.9 | 0.9 | 0.3 |
| His | 0.2 | 0.7 | ND |
| Trp | 0.0 | ND | ND |
| Asn | 8.0 | ND | ND |
| Gln | 0.5 | ND | ND |

*Note:* ND = no data.

and reverse primers. PCR products were also sequenced directly using an internal primer (P4: Fig. 1). Resulting sequence data were analyzed with the software GENE-TYX-MAC ver. 9 (Software Development). The DDBJ data bank (National Institute of Genetics, Mishima, Japan) was searched to find similar sequences.

## RESULTS

### Biochemical characterization of MSP-1

The extraction procedure yielded about 1 mg of soluble total organic molecules per 3 g of shell flakes. SDS PAGE of these molecules revealed three major protein bands of 97, 72, and 49 kDa in apparent molecular weight, as well as three minor bands of smaller sizes, when stained with Coomassie Brilliant Blue or silver. This pattern was replicated several times using different shell extract preparations. PAS and Stains-All stained each protein band red and blue, respectively, indicating that all these components are glycosylated and may have cation-binding potential. PAGE under non-denaturing conditions revealed a similar gel pattern as in SDS PAGE, confirming that these proteins are highly acidic.

N-terminal sequencing of the three major components revealed that all three share the same amino acid sequence at least for the first twenty residues (LDTDK DLEFH LDSLL NAAED, in one-letter abbrevation of amino acids). These components may have derived from the same transcript, but were subjected to different degrees of post-transcriptional processing or post-translational processing (such as glycosylation, phosphorylation, and C-terminal cleavage), or they may be encoded by different genes. At the moment, evidence is lacking to support any of these possibilities, and MSP-1 cannot be specified to a particular band in the SDS PAGE gel.

### Amino acid sequence of MSP-1

Figure 1 shows the nucleotide sequence for the part of the cDNA encoding the first 436 amino acid sequence of MSP-1. The deduced sequence contains a high proportion of serine, glycine, and aspartate residues (Table 1), in agreement with the bulk composition of the soluble matrix of the same species (Kasai and Ohta 1981), supporting the fact that MSP-1 represents the major component of the soluble matrix. The amino acid composition is also
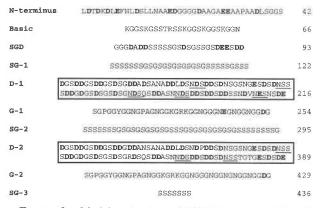
| | | |
|---|---|---|
| N-terminus | LDTDKDLEFHLDSLLNAAEDGGGGDAAGAEEAAPAADLSGGS | 42 |
| Basic | KGGSKGSSTRSSKGGSKGGSKGGN | 66 |
| SGD | GGGDADDSSSSSGSDSGSSGSDEESDD | 93 |
| SG-1 | SSSSSSSGSGSGSGSGSGSGSSSSSSGSSS | 122 |
| D-1 | DGSDDGSDDGSDSGDDADSANADDLDSNDSDDSDNSGSNGESDSDNSS<br>SDDGDGSDSGSDSGNDSQSDDASNNDSDDSDDSDDSSNDVNESNSDE | 216 |
| G-1 | SGPGGYGGNGPAGNGGKGRKGGNGGGNEGNGGNGGDG | 254 |
| SG-2 | SSSSSSSGSGSGSGSGSGSSSGSGSGSGSGSGSSSSSSSSG | 295 |
| D-2 | DGSDDGSDDGSDDGDDANSANADDLDSNDPDDSDNSGSNGESDSDNSS<br>SDDGDGSDSGSDSGRDSQSDDASNNDSDDSDDSDNSSTGTGESDSDE | 389 |
| G-2 | SGPGGYGGNGPAGNGGKGRKGGNGGGNGGNGNGGNGGDG | 429 |
| SG-3 | SSSSSSS | 436 |

FIGURE 2. Modular structure of MSP-1. The aspartate-rich domains are boxed. Acidic residues are in bold. Potential N-glycosylation sites are underlined. For one-letter abbreviation of amino acids, see caption of Figure 1.

| | | |
|---|---|---|
| D-1 | DGSDDGSDDGSDSGDDADSANADDLDSNDSDD | 154 |
| | ************ **** *********** ** | |
| D-2 | DGSDDGSDDGSDDGDDANSANADDLDSNDPDD | 327 |
| | | |
| D-1 | SDNSGSNGESDSDNSSSDDGDGSDSGSDSGND | 186 |
| | ****************************** * | |
| D-2 | SDNSGSNGESDSDNSSSDDGDGSDSGSDSGRD | 359 |
| | | |
| D-1 | SQSDDASNNDSDDSDDSDDSSNDVNESNSDE | 217 |
| | ***************** ** ** *** | |
| D-2 | SQSDDASNNDSDDSDDSDNSSTGTGESDSDE | 390 |

FIGURE 3. Sequence alignment of aspartate-rich domains (D-1 and D-2) of MSP-1. Identical amino acids are indicated by asterisks. Numbers on the right indicate positions of the amino acid residues in the MSP-1 sequence. Acidic residues are in bold. For one-letter abbreviation of amino acids, see caption of Figure 1.

comparable with that of a purified soluble fraction of oyster shell (Wheeler 1992; Table 1), and is in fact typical for a mollusc soluble matrix fraction (Lowenstam and Weiner 1989).

The predicted sequence revealed a modular structure with a basic domain near the N-terminus and two aspartate-rich domains ("D" domains) interspersed among three serine- and glycine-rich regions (Fig. 2), and eight possible N-glycosylation sites localized in the aspartate-rich domains (Figs. 1 and 2). Preliminary results indicate that the rest of the amplified cDNA coding region codes for repetition of similar sequence motifs.

The N-terminal domain of MSP-1 is 42 residues long, and an α-helix was predicted for this domain by the analyses of secondary structure using the method of Chou and Fasman (1978). The N-terminal domain is followed by a basic domain, containing five Lys-Gly-X-Y (X = Gly or Ser, Y = Ser or Asn) segments in tandem, intercalated by a Thr-Arg-Ser-Ser segment in the middle. Following the basic domain is a 27 residue serine and glycine-rich region ("SGD" domain), which is acidic due to the presence of seven aspartate and two glutamate residues.

The SG domains are solely composed of serine and glycine, dominated by (Ser-Gly)ₙ and (Ser)ₙ repeats. Among the sequences in the DDBJ data bank, this domain showed the highest sequence similarity with the "GS" domain of Lustrin A, a matrix protein of gastropod nacre (Shen et al. 1997), suggesting that they are evolutionarily and/or functionally related. The "GS" domain of Lustrin A is suggested to have an elastic property making the protein an extensor molecule (Shen et al. 1997). It seems possible that the SG domains of MSP-1 have a similar function. However, because the SG domains of MSP-1 lack the aromatic residues of tryptophan and phenylalanine, which are considered important in Lustrin A, their functions could well be different.

Each SG domain is followed by the D domain, the largest and seemingly most important of the revealed domains of MSP-1. The two D domains comprise 95 amino acids, containing 33–35 aspartate residues, and share a high degree of sequence similarity with each other (Fig. 3). This sequence conservation suggests its functional significance, which is most naturally thought to be that of calcium binding, with the calcium either in the solution or in the crystal, or both. A glycine-rich domain ("G" domain) follows the D domain. As the two G domains also exhibit a high degree of sequence similarity, they also may be important functionally.

Soluble proteins in foliated calcite of mollusc shells are phosphorylated (Borbas et al. 1991). Thus it is likely that MSP-1 is also phosphorylated because it has been extracted from the foliate calcite layer of *P. yessoensis*. The deduced partial sequence of MSP-1 contains a total of 137 serine, four threonine, and two tyrosine residues (Fig. 1), and any of these can be considered as a putative phosphorylation site.

## DISCUSSION

### Aspartate-rich domain and mineral-protein interactions

How do organisms produce delicately textured and yet rock-solid biominerals under physiological conditions? A key may be biopolymers, especially proteins, that interact with the ions in the media and the growing crystals. A considerable number of in vitro studies have demonstrated that such interactions between calcium carbonate and soluble matrix proteins are indeed possible.

Examples of such observations include inhibition of crystal growth by proteins (Wheeler et al. 1981), adsorption of proteins to specific crystal faces (Addadi and Weiner 1985; Berman et al. 1988; Berman et al. 1990; Albeck et al. 1993; Wierzbicki et al. 1994; Aizenberg et al. 1994; Aizenberg et al. 1995), and control of aragonite-calcite polymorphism by proteins (Belcher et al. 1996; Falini et al. 1996). But how do these proteins interact with the calcium carbonate to control crystal growth?

It is reasonable to assume that the negatively charged acidic amino acid residues and the positively charged basic residues in the matrix proteins interact with the calcium ions and carbonate ions, respectively, somehow pro-

viding specific templates for the nucleation of biominerals (Hare 1963), although no evidence has been reported for involvement of positively charged amino acids in these processes. In addition, the charged amino acids in these proteins can also interact with crystal planes and thus control growth.

Using the amino acid composition data on matrix proteins hydrolyzed partially or completely, Weiner and coworkers developed an hypothesis that predicts that the soluble matrix proteins adopt antiparallel β-sheet, containing (Asp-Y)$_n$ domains, where Y represents serine or glycine (Weiner 1975; Weiner 1983; Weiner and Traub 1984; Weiner and Addadi 1991). It was further hypothesized that the spacing (6.95 Å) between every second residue of the negatively charged aspartate is approximately similar to the distance (3.0–6.5 Å) between the calcium ions in the unit cells of aragonite and calcite, enabling the proteins to function as a template for mineralization (Weiner 1975).

This model was further extended to explain the crystallography of the three types of foliated calcite by a precise stereochemical fit between the amino acid residues of the matrix protein and the crystal lattices in the surface of a specific crystal face (Runnegar 1984). The spacing of calcium ions along the length of the laths is 19.3 Å in one type of foliated calcite, a distance that is matched by every sixth residue of a parallel β-sheet (19.44 Å). Taking account of the amino acid composition, the protein was predicted to have the repetitive sequence of (Asp-Gly-X-Gly-X-Gly)$_n$. The other two types of foliated calcite were specified by the repetitive sequence of (Asp-Y)$_n$ of either antiparallel or parallel β-sheet conformation (Runnegar 1984).

However, these hypotheses of the primary and secondary structures of the acidic matrix proteins have been based on indirect observations. The (Asp-Y)$_n$ domains are absent in a phosphoprotein of oyster shell (Wheeler 1992). The primary sequence of Nacrein, a major soluble protein of pearl nacre (Miyamoto et al. 1996), contains an acidic domain, but is not of the (Asp-Y)$_n$ type. Our results indicate that (Asp-Y)$_n$ type and (Asp-Gly-X-Gly-X-Gly)$_n$ type domains do not exist at least in the sequenced part of MSP-1. MSP-1 has aspartate-rich domains, which contain some regular motifs such as (Asp-Gly-Ser-Asp) and (Asp-Ser-Asp), but the overall arrangement of the acidic residues in the domains suggests that two-dimensional simple models may be insufficient to explain the interactive relationships at the protein-mineral interfaces.

## Implications for the polymorph formation

The calcium carbonate of mollusc shell occurs as calcite, aragonite, or vaterite (Wilbur 1960). Aragonite occurs in other animals as well as in plants, protists, and bacteria (Lowenstam and Weiner 1989). The problem of the formation of the much less stable aragonite instead of calcite ("the calcite-aragonite problem") is long standing and still unresolved (Wilbur 1960; Lowenstam and Wei-

ner 1989). Suggested explanations include: (1) presence of foreign ions, especially $Mg^{2+}$ (Simkiss et al 1982); (2) structure of organic matrix; (3) influence of carbonic anhydrase; and (4) temperature (Wilbur 1960).

Watabe and Wilbur (1960) reported that insoluble organic matrix extracted from mollusc shells can control calcite-aragonite polymorphism both in vitro and in vivo, but the evidence was not conclusive. Recent findings have demonstrated more definitively that soluble matrix proteins, rather than insoluble ones, can control the polymorphism in vitro (Belcher et al. 1996; Falini et al. 1996). The fact that soluble aragonitic proteins alone can switch the polymorph of growing crystals (Belcher et al. 1996), and that a carbonic anhydrase has been isolated as a major soluble matrix component of the aragonitic shell (Miyamoto et al. 1996), suggest that the carbonic anhydrase activity to produce local accumulation of carbonate ions at the site of crystal growth could be important in the formation of aragonite. The absence of a carbonic anhydrase domain in the major soluble proteins of the calcitic shell in this study tends to support this hypothesis.

## CONCLUSIONS

To understand the functions of matrix proteins, and biomineralization processes in general, it is necessary to determine primary, secondary, and tertiary structures; to localize the proteins precisely in the biominerals; and to perform refined in vitro experiments using pure proteins. The recombinant clones for the MSP-1 gene provide a basis for those experiments. Furthermore, to understand the functions of matrix proteins in vivo, it would be useful to produce "transgenic shellfish" with a certain matrix protein gene disrupted. The kind of information obtained in this study is considered as a prerequisite for such a venture.

## REFERENCES CITED

Addadi, L. and Weiner, S. (1985) Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. Proceedings of the National Academy of Science, 82, 4110–4114.
———(1997) A pavement of pearl. Nature, 389, 912–915.
Albeck, S., Aizenberg, J., Addadi, L., and Weiner, S. (1993) Interactions of various skeletal intracrystalline components with calcite crystals. Journal of American Chemical Society, 115, 11691–11697.
Aizenberg, J., Albeck, S., Weiner, S., and Addadi, L. (1994) Crystal-protein interactions studied by overgrowth of calcite on biogenic skeletal elements. Journal of Crystal Growth, 142, 156–164.
Aizenberg, J., Hanson, J., Ilan, M., Leiserowitz, L., Koetzle, T.F., Addadi, L., and Weiner, S. (1995) Morphogenesis of calcitic sponge spicules: a role for specialized proteins interacting with growing crystals. The FASEB Journal, 9, 262–268.
Belcher, A.M., Wu, X.H., Christensen, R.J., Hansma, P.K., Stucky, G.D.,

and Morse, D.E. (1996) Control of crystal phase switching and orientation by soluble mollusc-shell proteins. Nature, 381, 56–58.

Benson, S.C. and Wilt, F.H. (1992) Calcification of spicules in the sea urchin embryo. In E. Bonucci, Ed., Calcification in biological systems, p. 157–178. CRC Press, Cleveland, Ohio.

Berman, A., Addadi, L., and Weiner, S. (1988) Interactions of sea-urchin skeleton macromolecules with growing calcite crystals—a study of intracrystalline proteins. Nature, 331, 546–548.

Berman, A., Addadi, L., Kvick, A., Leiserowitz, L., Nelson, M., and Weiner, S. (1990) Intercalation of sea urchin proteins in calcite: study of a crystalline composite material. Science, 250, 664–667.

Borbas, J.E., Wheeler, A.P., and Sikes, C.S. (1991) Molluscan shell matrix phosphoproteins: Correlation of degree of phosphorylation to shell mineral microstructure and to in vitro regulation of mineralization. The Journal of Experimental Zoology, 258, 1–13.

Campbell, K.P., MacLennan, D.H., and Jorgensen, A.O. (1983) Staining of the $Ca^{2+}$-binding proteins, calsequestrin, calmodulin, troponin C, and S-100, with the cationic carbocyanine dye "Stains-all". The Journal of Biological Chemistry, 258, 11267–11273.

Chomzynski, P. (1993) A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. Biotechniques, 15, 532–537.

Chou, P.Y. and Fasman, G.D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Advances in Enzymology, 47, 45–148.

Cusack, M., Curry, G., Clegg, H., and Abbott, G. (1992) An intracrystalline chromoprotein from red brachiopod shells: implications for the process of biomineralization. Comparative Biochemistry and Physiology, 102B, 93–95.

Falini, G., Albeck, S., Weiner, S., and Addadi, L. (1996) Control of aragonite or calcite polymorphism by mollusk shell macromolecules. Science, 271, 67–69.

Frohman, M.A. (1990) RACE: Rapid amplification of cDNA ends. In M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White, Eds., PCR Protocols, p. 28–38. Academic Press, San Diego, California.

Hare, P.E. (1963) Amino acids in the proteins from aragonite and calcite in the shells of Mytilus californianus. Science, 139, 216–217.

Holden, K.G., Yim, N.C.F., Griggs, L.J., and Weissbach, J.A. (1971) Gel electrophoresis of mucous glycoproteins. I. Effect of gel porosity. Biochemistry, 10, 3105–3109.

Ishii, K., Yanagisawa, T., and Nagasawa, H. (1996) Characterization of a matrix protein in the gastrolith of the crayfish Procambarus clarkii. Bioscience Biotechnology Biochemistry, 60, 1479–1482.

Ishii, K., Tsutsui, N., Watanabe, T., Yanagisawa, T., and Nagasawa, H. (1998) Solubilization and Chemical characterization of an insoluble matrix protein in the gastroliths of a crayfish, Procambarus clarkii. Bioscience Biotechnology Biochemistry, 62, 291–296.

Kasai, H. and Ohta, N. (1981) Relationship between organic matrices and shell structures in recent bivalves. In T. Habe and M. Omori, Eds., Study of molluscan paleobiology, p. 101–106. Kokusai Insatsu, Tokyo (in Japanese).

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., and Davidson, E.H. (1991) The corrected structure of the SM50 spicule matrix protein of Strongylocentrotus purpuratus. Developmental Biology, 145, 201–222.

Katoh-Fukui, Y., Noce, T., Ueda, T., Fujiwara, Y., Hashimoto, N., Tanaka, S., and Higashinakagawa, T. (1992) Isolation and characterization of cDNA encoding a spicule matrix protein in Hemicentrotus pulcherrimus. International Journal of Developmental Biology, 36, 353–361.

Krampitz, G.P. (1982) Structure of the organic matrix in mollusc shells and avian eggshells. In G.H. Nancollas, Ed., Biological mineralization and demineralization, p 219–232. Springer-Verlag, Berlin, Heidelberg, New York.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227, 680–685.

Lowenstam, H.A. (1981) Minerals formed by organisms. Science, 211, 1126–1131.

Lowenstam, H.A. and Weiner, S. (1989) On Biomineralization. Oxford University Press, New York

Mann, S., Archibald, D.D., Didymus, J.M., Douglas, T., Heywood, B.R., Meldrum, F.C., and Reeves, N.J. (1993) Crystallization at inorganic-organic interfaces: biominerals and biomimetic synthesis. Science, 261, 1286–1292.

Miyamoto, H., Miyashita, T., Okushima, M., Nakano, S., Morita, T., and Matsushiro, A. (1996) A carbonic anhydrase from the nacreous layer in oyster pearls. Proceedings of the National Academy of Science, 93, 9657–9660.

Runnegar, B. (1984) Crystallography of the foliated calcite shell layers of bivalve molluscs. Alcheringa, 8, 273–290.

Shen, X., Belcher, A.M., Hansma, P.K., Stucky, G.D., and Morse, D.E. (1997) Molecular cloning and characterization of Lustrin A, a matrix protein from shell and pearl nacre of Haliotis rufescens. The Journal of Biological Chemistry, 272, 32472–32481.

Simkiss, K., Althoff, J., Anderson, C., Bennema, P., Boyde, A., Crenshaw, M.A., Fleisch, H., Hohling, H.-J., Howell, D.S., Kitano, Y., Nancollas, G.H., Newesely, H., Nielsen, A.E., Quint, P., and Termine, J.D. (1982) Mechanisms of mineralization (normal) group report. In G.H. Nancollas, Ed., Biological mineralization and demineralization, p. 351–363. Springer-Verlag, Berlin.

Sucov, H.M., Benson, S., Robinson, J.J., Britten, R.J., Wilt, F., and Davidson, E.H. (1987) A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II Structure of the gene and derived sequence of the protein. Developmental Biology, 120, 507–519.

Sudo, S., Fujikawa, T., Nagakura, T., Ohkubo, T., Sakaguchi, K., Tanaka, M., Nakashima, K., and Takahashi, T. (1997) Structures of mollusc shell framework proteins. Nature, 387, 563–564.

Watabe, N. and Wilbur, K.M. (1960) Influence of the organic matrix on crystal type in molluscs. Nature, 188, 334.

Weiner, S. (1975) Soluble protein of the organic matrix of mollusk shells: a potential template for shell formation. Science, 190, 987–989.

———(1983) Mollusk shell formation: Isolation of two organic matrix proteins associated with calcite deposition in the bivalve Mytilus californianus. Biochemistry, 22, 4139–4145

———(1984) Organization of organic matrix components in mineralized tissues. American Zoologist, 24, 945–951.

Weiner, S. and Addadi, L. (1991) Acidic macromolecules of mineralized tissues: the controllers of crystal formation. Trends in Biochemical Science, 16, 252–256.

Weiner, S. and Traub, W. (1984) Macromolecules in mollusc shells and their functions in biomineralization. Philosophical Transactions of Royal Society of London, B304, 425–434.

Wheeler, A.P. (1992) Phosphoproteins of oyster (Crassostrea virginica) shell organic matrix. In S. Suga and N. Watabe, Eds., Hard tissue mineralization and demineralization, p. 171–187. Springer-Verlag, Tokyo.

Wheeler, A.P., George, J.W., and Evans, C.A. (1981) Control of calcium carbonate nucleation and crystal growth by soluble matrix of oyster shell. Science, 212, 1397–1398.

Wierzbicki, A., Sikes, C.S., Madura, J.D., and Drake, B. (1994) Atomic force microscopy and molecular modeling of protein and peptide binding to calcite. Calcified Tissue International, 54, 133–141.

Wilbur, K.M. (1960) Shell structure and mineralization in molluscs. In R.F. Sognnaes, Ed., Calcification in biological systems, p. 15–40. American Association for the advancement of Science, Washington, D.C.