Korean J. Chem. Eng., 33(8), 2406-2410 (2016)
DOI: 10.1007/s11814-016-0091-9

pISSN: 0256-1115
eISSN: 1975-7220

Recombinant production and biochemical characterization of a hypothetical acidic shell matrix protein in *Escherichia coli* for the preparation of protein-based $CaCO_3$ biominerals

Chaeyeon Son*, Wooho Song*, Dong Soo Hwang, Yong-Ki Hong***, Jin Joo****, and Yoo Seong Choi*,†**

*Department of Chemical Engineering and Applied Chemistry, Chungnam National University, Daejeon 34134, Korea
**Division of Integrative Biosciences and Biotechnology, POSTECH, Pohang 37673, Korea
***Department of Biotechnology, Pukyong National University, Busan 48513, Korea
****School of Applied Chemical Engineering, Kyungpook National University, Daegu 41566, Korea
(*Received 6 January 2016 • accepted 31 March 2016*)

Abstract–The biological structure of shells consists of highly organized calcium carbonate ($CaCO_3$) crystals, which have remarkable mechanical and biological properties compared to the pure mineral form of $CaCO_3$. It has been known that the organization of these biominerals is controlled by a relatively tiny amount of organic components such as shell matrix proteins. Here, we successfully produced a recombinant hypothetical acidic shell matrix protein in *Escherichia coli*, although the protein is composed of highly repetitive and biased amino acid sequences. About 15 mg/L purified protein with greater than 95% purity was obtained in the 400 mL lab scale flask culture. The protein was able to efficiently form a complex with calcium ions, and spherulitic calcite crystals were synthesized in the presence of the recombinant protein. We expect that biomineralization using the recombinant protein could not only overcome the limited amount of protein available for biomineralization studies and biomineral preparation in practical aspects, but also provide opportunities to enable biomimetic synthesis of notable bio-composites based on organic-inorganic complexation.

Keywords: Biomineralization, Calcium Carbonate, Shell Matrix Protein, Glycine-rich Protein, Recombinant

INTRODUCTION

Biominerals based on organic-inorganic complexation have become attractive due to their exquisite nanostructure, and superior mechanical and biological properties [1-3]. Specifically, calcium carbonate ($CaCO_3$) biominerals in molluscan shells and eggshells have been relatively well studied, which are dominantly composed of calcite and/or aragonite complexed with organic components. Shell matrix proteins are mainly involved in the biomineralization process of $CaCO_3$ shell formation for biological functions, including structural support and protection. Less than 5% of shell matrix proteins control different aspects of the shell formation processes, such as synthesis of transient amorphous minerals, evolution to crystalline phases, choice of $CaCO_3$ polymorph and organization of crystallites in complex shell textures [4,5]. Bioinformatic and proteomic studies have identified primary structures and putative function of these proteins [6-8]. However, biomineralization studies using shell matrix proteins have been mainly conducted by only using high amount of extractable native proteins and characteristic peptides of some shell matrix proteins [4,9,10]. The limited amount of purified shell matrix proteins has hampered the *in vitro* investigation of functional roles of each matrix components and synergetic effect of the isolated matrix components. Moreover, it has been relatively difficult to obtain recombinant proteins compared to conventional globular proteins, due to their biased and repetitive amino acid composition.

Thus, it is important to obtain high amount of shell matrix proteins with high purity, which accelerates the preparation of biomimetic biominerals with highly attractive mechanical and biological properties in biomaterials and biomedical fields. Interestingly, they have few compositional or structural homologs, although common characteristics are found, such as acidic, anionic functional groups, repetitive sequences and intrinsically disordered protein (IDP) domains [11,12]. In addition, the functional contribution of natural extracts under denatured condition in crystal nucleation and growth inhibition suggests that exact native three-dimensional folding structure cannot be crucial for showing functional properties of the matrix proteins [9,13,14], although stereochemical aspects of side chain in protein structure as well as amino acid side chain chemistry itself play important roles [11,15]. Until now, the roles of shell matrix proteins are not currently fully understood. Biomineralization studies using mass-producible model proteins could provide insights into the understanding of protein-based biomineralization and also expand potential applications of these biocomposites.

In the present study, we found a novel hypothetical protein sequence (named GG1234, NCBI reference sequence, XP_001234449) from the server of the National Center for Biotechnology Information (NCBI). The hypothetical protein has a theoretical pI value of 3.58 (acidic) and is composed of amino acid composition enriched with acidic and IDP promoting sequences [16] of glycine (23.6%), aspartate (21.5%), serine (7.4%) and alanine (6.0%) (Fig. 1(a)). Al-

†To whom correspondence should be addressed.
E-mail: biochoi@cnu.ac.kr

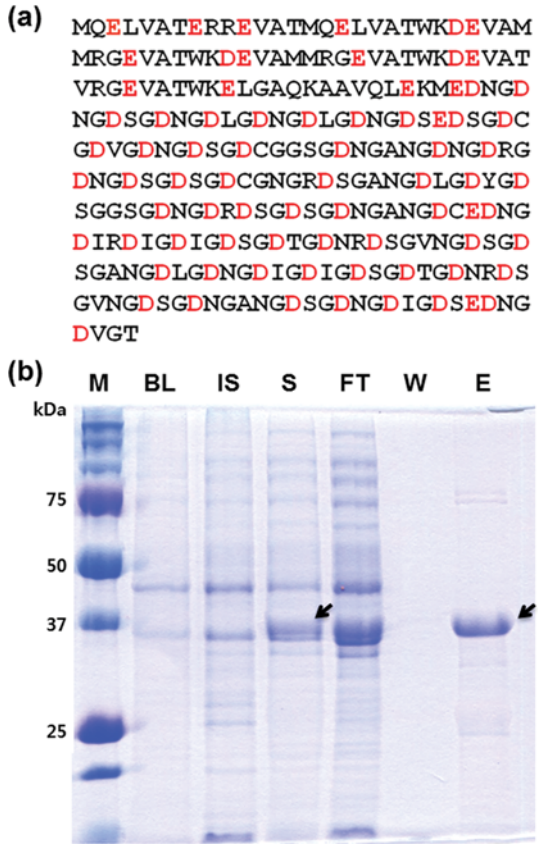

Fig. 1. (a) Schematic representation of the amino acid sequence of GG1234, where red letters indicate acidic amino acid residues. (b) Over-expression and purification of GG1234 in *E. coli*. Lanes: M, protein molecular weight marker; BL, *E. coli* whole cell lysate; IS, insoluble fraction of GG1234 expression in *E. coli*; S, soluble fraction of GG1234 expression in *E. coli*; FT, unbound flow-through in His_6-tag affinity chromatography; W, wash fraction; E, purified GG1234 protein from His_6-tag affinity chromatography.

though the sequence has been removed from the NCBI record after a recent update in the genome annotation process, the characteristics of primary structure similar to other shell matrix proteins suggest that this hypothetical protein could be a model protein for protein-based biomineralization studies, and could be used as a novel material for the fabrication of biominerals by the protein-mineral interaction. In this manner, we genetically redesigned GG1234 sequence to heterologously produce the protein in *Escherichia coli* (*E. coli*). It is considered that the recombinant production approach can overcome the limitation of obtaining enough amount of GG1234 protein. The features of GG1234 interactions with calcium ions (Ca^{2+}) were simply examined, and *in vitro* $CaCO_3$ mineralization in the presence of GG1234 was performed to determine the morphological effect on $CaCO_3$ crystals.

EXPERIMENTAL

1. Strains and Vector Construction

E. coli DH5α (Life Technologies, Carlsbad, CA, USA) cell line was used as a host for recombinant vector preparation, and *E. coli* BL21 (DE3) (Merck KGaA, Darmstadt, Germany) was used for the expression of GG1234. The *E. coli* cells were grown in Luria-Bertani (LB) medium with 50 μg ampicillin/mL (Sigma-Aldrich, St. Louis, MO, USA). The amino acid sequence of GG1234 (NCBI reference sequence XP_001234449) was obtained from an NCBI BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). GG1234 gene sequence was redesigned based on *E. coli* codon preference and to avoid gene sequence repeats, and then chemically synthesized (GenScript USA Inc., Piscataway, NJ, USA). Using *Nde*I and *Xho*I restriction digest, the sequence was introduced into the pET23b+ vector (Novagen, Darmstadt, Germany), which contains a strong T7 promoter and six-histidine (His_6) tag. Finally, the vector construct was confirmed by direct sequencing.

2. Over-expression and Purification of GG1234

The vector construct was introduced into *E. coli* BL21 (DE3) cells by electroporation. A single colony from a freshly streaked plate of these cells was grown in 400 mL of LB medium with 50 μg ampicillin/mL at 37 °C and 200 rpm until the OD_{600} was 0.8-1.0. Isopropyl-β-D-thiogalactopyranoside (IPTG; 0.02 mM final concentration; Sigma-Aldrich) was added to induce protein over-expression, and the cells were further incubated at 20 °C and 200 rpm for 20 h. Cells were harvested by centrifugation at 9,000 rpm for 10 min at 4 °C, and the pellets were stored at −80 °C. The expression level was analyzed by image analysis software (CLIQS; TotalLab Ltd., Newcastle upon Tyne, UK).

For Ni-NTA affinity purification, cells were resuspended in 40 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole; pH 8.0). The cells were lysed by using an ultrasonic cell disrupter (ULH-700S; Ulsso High-Tech Co., Cheongwon, Korea) at 30% power with a 3 sec on pulse and a 10 sec cooling period between each burst. The lysate was centrifuged at 9,000 rpm for 10 min at 4 °C. The soluble fraction was collected, and applied to Ni-nitrilotriacetic acid (Ni-NTA) resin (Qiagen, Germantown, MD, USA) for affinity purification. After the lysate-Ni-NTA mixture was incubated at 4 °C for 1 h, the mixture was loaded onto a column. The resin was washed three times with five column volumes of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole; pH 8.0). The GG1234 protein was eluted by using an elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM and imidazole; pH 8.0), and dialyzed in distilled water using a dialysis membrane with a 12-14 kDa molecular weight cut off (Spectrum Laboratories Inc., Los Angeles, CA, USA). Finally, the purified GG1234 protein was freeze-dried and stored at −80 °C.

3. Stains-all Assay and Turbidimetric Measurement of Ca^{2+}-induced GG1234 Agglomeration

The Stains-all staining was conducted based on a previously described method [17]. After SDS-PAGE, the gel was washed exhaustively in 25% isopropanol (Sigma-Aldrich) twice to remove SDS and then fixed overnight in 25% isopropanol. The gel was soaked in a fresh Stains-all solution (0.005% Stains-all (Sigma-Aldrich), 10% formamide (Sigma-Aldrich), 25% isopropanol and 15 mM Tris; pH 8.8) and stained for 24 h in the dark at room temperature with gentle agitation. After washing with distilled water several times, the gel was scanned using a gel scanner. Additionally, turbidimetry was performed based on a previously described

method [18] with the following modifications. Various concentrations of GG1234 solution were directly mixed with 20 mM $CaCl_2$ solution to analyze agglomeration at different Ca^{2+}-protein ratios. The samples were transferred to a flat-bottom 96-well plate (SPL Life Science, Pocheon, Korea), and the absorbance was measured at 600 nm with a microplate reader (SpectraMax 190; Molecular Device, Sunnyvale, CA, USA). This experiment was performed in triplicate.

4. *In vitro* $CaCO_3$ Crystallization

The effect of GG1234 on *in vitro* $CaCO_3$ crystallization was investigated by the rapid mixing of $CaCl_2$ and Na_2CO_3 solutions. Mineralization was conducted in a 24-well plate at 4 °C using glass cover slips. Various concentrations of GG1234 were prepared in 20 mM $CaCl_2$, and rapidly mixed with an equivalent volume of 20 mM Na_2CO_3 solution. After 20 h, the glass cover slips with precipitated $CaCO_3$ crystals were rinsed with distilled water and dried at room temperature. For morphological observations, the samples were gold-coated and imaged by using a scanning electron microscope (SEM) (SNE-4500M; SEC Co., Suwon, Korea). The X-ray diffraction (XRD) patterns of the obtained precipitates were analyzed and recorded by an X-ray diffractometer (D/MAX-2200; Rigaku Co., Tokyo, Japan) using Cu-Kα radiation (λ=1.5406 Å, 60 kV, 80 mA). X-ray photoelectron spectroscopy (XPS) was conducted with Thermo Scientific MultiLab 2000 (Thermo Fisher Scientific, Waltham, MA, USA). Transmission electron microscopy (TEM) and selected area electron diffraction (SAED) were performed on a TEM (Tecnai G2 F30; FEI, OR, USA).

RESULTS AND DISCUSSION

1. Over-expression and Purification of GG1234 Protein

The DNA sequence encoding GG1234 protein was optimized based on codon usage preference in *E. coli* and alteration of some major codons to other codons to avoid repetition, because GG1234 contains many repeat sequences (Fig. 1(a)). The sequence was introduced into the pET-23b vector for over-expression and easy purification of the protein due to the presence of a strong T7 promoter and C-terminal His_6-tag. The protein was optimally over-expressed at 20 °C with 0.02 mM IPTG induction for 20 h additional culture after induction, and produced as soluble form with about 30% of the total soluble protein in 400 mL of *E. coli* culture (Fig. 1(b)). And GG1234 was successfully purified by His_6-tag affinity chromatography under native condition. The final purity was greater than 95%, and the yield was approximately 15 mg/L for the 400 mL lab scale flask culture. Interestingly, the purified GG1234 (the theoretical molecular weight of 29 kDa) appeared with a size of approximately 37 kDa on SDS-PAGE, although the exact sequence of the protein was confirmed by gene sequencing and the protein was purified with the C-terminal purification tag. The unusually acidic properties (the calculated pI value of 3.58) of GG1234 might result in slower mobility than expected due to negative charge repulsion with SDS [19], and similar results were also observed for some acidic proteins such as ferredoxin and rPif97 [20,21].

2. Ca^{2+}-binding of GG1234

Because the interaction between Ca^{2+} and $CaCO_3$-biomineralization associated proteins is an important characteristic, the possible Ca^{2+}-binding ability of GG1234 was simply investigated by Stains-all staining. The cationic carbocyanine dye in Stains-all staining interacts with possible Ca^{2+}-binding anionic sites and produces a dye-protein complex with blue color, whereas most Ca^{2+}-unbound proteins stain red or pink [17]. The blue color development of GG1234 in Stains-all staining showed the high possibility of Ca^{2+}-binding, although bovine serum albumin (BSA), a control acidic protein (pI value of 4.7) [22] stained pink (Fig. 2(a)). In addition, we directly investigated the possible interaction of GG1234 with Ca^{2+} by measuring turbidity in the mixing of GG1234 and $CaCl_2$ solutions. Turbidity was increased with GG1234 concentration in the fixed concentration of $CaCl_2$ solution (Fig. 2(b)). GG1234 also formed Ca^{2+}-induced agglomeration similar to other calcification-related proteins such as osteonectin, phosphophoryn and rPif97 [18,20,23]. The results indicated that GG1234 could be involved in the morphology regulation of $CaCO_3$ crystal growth like other shell matrix proteins.

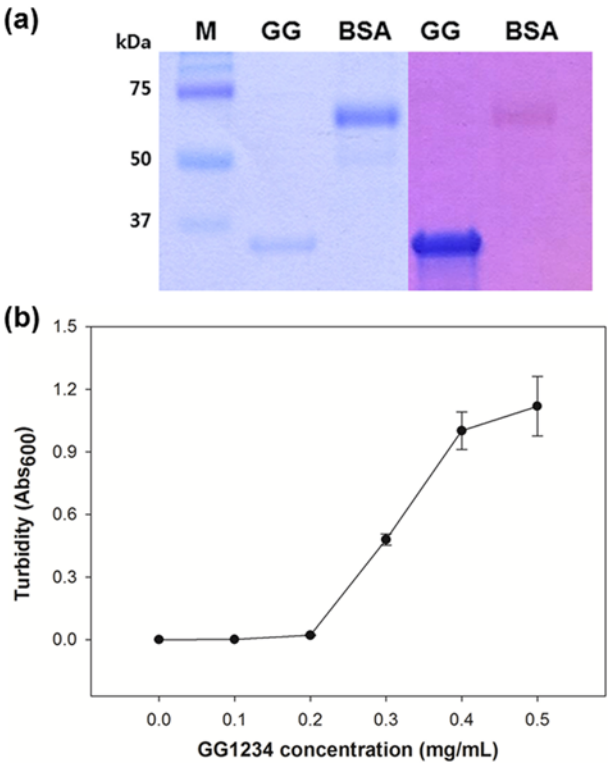

Fig. 2. (a) SDS-PAGE and Stains-all staining of GG1234. Lanes: M, protein molecular weight marker; GG, purified GG1234; BSA, bovine serum albumin as a negative control. (b) Ca^{2+}-induced agglomeration of GG1234. All experiments were performed in triplicate, and error bars represent standard deviations of the experimental errors in the turbidimetric measurement.

3. *In vitro* $CaCO_3$ Crystallization using GG1234

In vitro $CaCO_3$ crystallization using GG1234 was conducted by the rapid mixing of $CaCl_2$ and Na_2CO_3 solutions. $CaCO_3$ precipitates were collected at 20 h using various concentrations of GG1234 (0-1 mg/mL). From scanning electron microscope (SEM) and X-ray diffraction (XRD) analysis, characteristic calcite rhombohedra of approximately 30 μm were formed in the absence of GG1234 (Fig. 3(a)); however, the morphology was changed to sphere-like

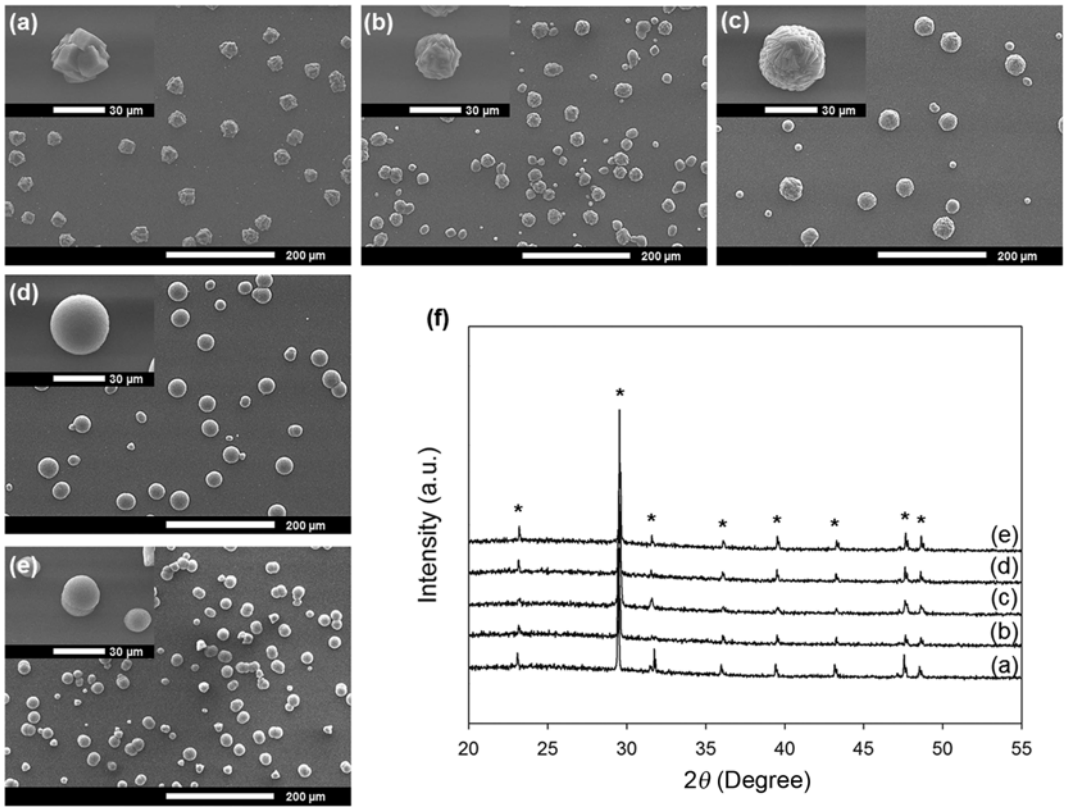

Fig. 3. Scanning electron microscope (SEM) images of $CaCO_3$ particles obtained by the rapid mixing of $CaCl_2$ and Na_2CO_3 solutions with no GG1234 (a), 50 μg/mL GG1234 (b), 100 μg/mL GG1234 (c), 150 μg/mL GG1234 (e) and 200 μg/mL GG1234, and (f) X-ray diffraction (XRD) spectra of the $CaCO_3$ particles in various concentrations of GG1234. Peaks indicated by 'asterisks' correspond to the calcite phase of the XRD data.

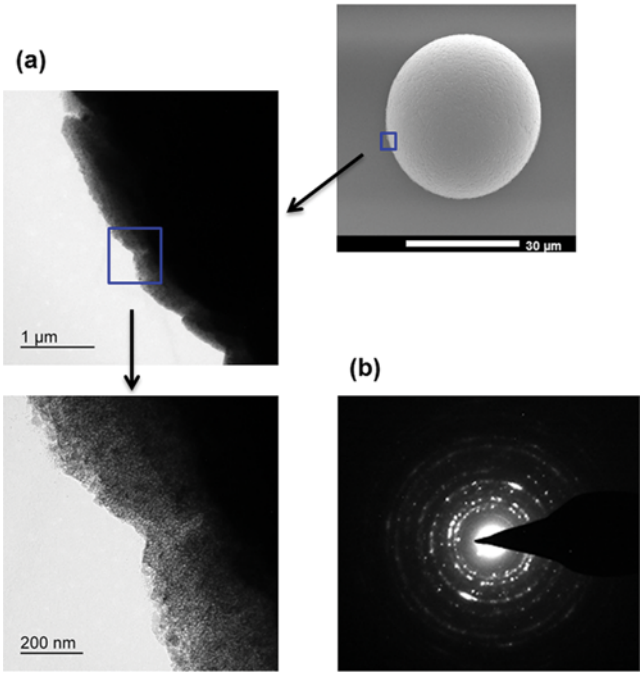

Fig. 4. Transmission electron microscopy (TEM) images (a) and the selected area electron diffraction pattern (c) of $CaCO_3$ particles obtained by the rapid mixing of $CaCl_2$ and Na_2CO_3 solutions with 150 μg/mL GG1234.

crystals (above 0.15 mg/mL GG1234) via partially rounded crystals with smooth edges by increasing the concentration of GG1234 (Fig. 3(b)-(e)). XRD analysis showed that all crystals were calcites, regardless of the morphologies (Fig. 3(f)). Transmission electron microscopy (TEM) revealed that the spherical particles were assembled through the aggregation of numerous primary nanoparticles (Fig. 4(a)), and the calcitic crystal was polycrystalline from the selected area electron diffraction pattern (Fig. 4(b)). However, these sphere-like calcite crystals were not observed in the *in vitro* $CaCO_3$ crystallization using various concentrations (0-1 mg/mL) of BSA (pI value of 4.7), L-aspartic acid and L-glutamic acid (data not shown). Although BSA, L-aspartic acid and L-glutamic acid as control organic components have similar functional groups with GG1234, calcitic rhombohedra were only formed due to relatively low morphological effect on $CaCO_3$ crystallization, which was broadly consistent with the previous observations [15,20]. The 8.3% of nitrogen atom components was detected on the surface of the particles crystallized by 0.15 mg/mL GG1234 in X-ray photoelectron spectroscopy (XPS), which revealed that GG1234 was absorbed on the $CaCO_3$ particles. Sphere-like or spherulitic calcite crystals were also found in some synthetic materials, corals and molluscan shells, and possible nucleation and growth mechanisms have been suggested [1,24-26]. Possible procedure of this crystallization could be suggested based on these mechanisms: Complexes of Ca^{2+} and anionic GG1234 are first formed by the interaction between

Ca^{2+} and negatively charged functional groups of GG1234. The addition of CO_3^{2-} forms stable calcite nanograins bound to GG1234 via amorphous $CaCO_3$ (ACC) by aggregation-based routes from the grains. Finally, they stack together to form sphere-like calcite crystals. Spherulitic crystals are found in many synthetic materials, corals and avian eggshells as well as diseases such Alzheimer's and kidney stones [27]. Similarly, it has been known that shell matrix proteins inhibit calcium carbonate crystallization in many cases, which leads to morphological changes of rhombohedral calcite crystals [28,29]. In this manner, the spherulitic calcite growth in the presence of GG1234 could be interesting, although the chemical nature of the protein and additional information needs to be further investigated.

CONCLUSION

We synthesized sphere-like calcite spherules using a recombinant hypothetical shell matrix protein, GG1234. Successful production of this recombinant protein with highly repetitive and biased amino acid sequences could not only address the need for large amount of protein for biomimetic preparation of biominerals, but also expand the synthesis of protein-based biominerals for biomedical applications. Recombinant proteins can be modified in a directed manner by introducing and/or changing amino acid sequences at the genetic level. Moreover, $CaCO_3$-based bio-composites may also be utilized as bio-seeds for hydroxyapatite deposition in biological systems [30]. It is expected that this approach could provide opportunities to fabricate notable biominerals in practical aspects.

ACKNOWLEDGEMENTS

This work is supported by the National Research Foundation of Korea Grant funded by the Ministry of Science, ICT and Future Planning, Korea (NRF-C1ABA001-2011-0029960 and NRF-2012R1A1A1003516; to Y.S.C.) and by 2015 research fund of Chungnam National University.

REFERENCES

1. N. K. Dhami, M. S. Reddy and A. Mukherjee, *Front Microbiol.*, **4**, 314 (2013).
2. B. Perito and G Mastromei, in *Molecular Biomineralization*, W. E. G. Muller Ed., Springer, New York (2011).
3. V. Achal, A. Mukherjee, D. Kumari and Q. Zhang, *Earth-Sci. Rev.*, **148**, 1 (2015).
4. F. Marin, G. Luquet, B. Marie and D. Medakovic, *Curr. Top. Dev. Biol.*, **80**, 209 (2008).
5. M. L. H. Rose and M. T. Hincke, *Cell. Mol. Life Sci.*, **66**, 2707 (2009).
6. H. Miyamoto, H. Endo, N. Hashimoto, K. Iimura, Y. Isowa, S. Kinoshita, T. Kotaki, T. Masaoka, T. Miki, S. Nakayama, C. Nogawa, A. Notazawa, F. Ohmori, I. Sarashina, M. Suzuki, R. Takagi, J. Takahashi, T. Takeuchi, N. Yokoo, N. Satoh, H. Toyohara, T. Miyashita, H. Wada, T. Samata, K. Endo, H. Nagasawa, S. Asakawa and S. Watabe, *Zool. Sci.*, **30**, 801 (2013).
7. K. Mann, B. Macek and J. V. Olsen, *Proteomics*, **6**, 3801 (2006).
8. J. S. Evans, *Bioinformatics*, **28**, 3182 (2012).
9. M. Suzuki, K. Saruwatari, T. Kogure, Y. Yamamoto, T. Nishimura, T. Kato and H. Nagasawa, *Science*, **325**, 1388 (2009).
10. A. Arakaki, K. Shimizu, M. Oda, T. Sakamoto, T. Nishimura and T. Kato, *Org. Biomol. Chem.*, **13**, 974 (2015).
11. A. Picker, M. Kellermeier, J. Seto, D. Gebauer and H. Colfen, *Z. Kristallogr.*, **227**, 744 (2012).
12. W. Magdalena, P. Dobryszycki and A. Ozyhar, in *Advanced topics in biomineralization*, J. Seto Ed., InTech (2012).
13. J. Sun and B. Bhushan, *RSC Adv.*, **2**, 7617 (2012).
14. L. Addadi, D. Joester, F. Nudelman and S. Weiner, *Chem. Eur. J.*, **12**, 980 (2006).
15. C. Briegel and J. Seto, in *Advanced topics in biomineralization*, J. Seto Ed., InTech (2012).
16. A. K. Dunker and Z. Obradovic, *Nat. Biotechnol.*, **19**, 805 (2001).
17. K. P. Campbell, D. H. MacLennan and A. O. Jorgensen, *J. Biol. Chem.*, **258**, 11267 (1983).
18. Y. Kuboki, H. Takita, T. Komori, M. Mizuno, E. Furu-uchi and K. Taniguchi, *Calcif. Tissue Int.*, **44**, 269 (1989).
19. A. Shirai, A. Matsuyama, Y. Yashiroda, A. Hashimoto, Y. Kawamura, R. Arai, Y. Komatsu, S. Horinouchi and M. Yoshida, *J. Biol. Chem.*, **283**, 10745 (2008).
20. S. Y. Bahn, B. H. Jo, B. H. Hwang, Y. S. Choi and H. J. Cha, *Cryst. Growth Des.*, **15**, 3666 (2015).
21. S. Lim, Y. S. Choi, D. G. Kang, Y. H. Song and H. J. Cha, *Biomaterials*, **31**, 3715 (2010).
22. S. Ge, K. Kojio, A. Takahara and T. Kajiyama, *J. Biomat. Sci. Polym. E.*, **9**, 131 (1998).
23. Y. Kuboki, R. Fujisawa, K. Aoyama and S. Sasaki, *J. Dent. Res.*, **58**, 1926 (1979).
24. C. Zhong and C. C. Chu, *Cryst. Growth Des.*, **10**, 5043 (2010).
25. W. Li and P. Y. Wu, *CrystEngComm*, **11**, 2466 (2009).
26. J. Harris, I. Mey, M. Hajir, M. Mondeshki and S. E. Wolf, *CrystEngComm*, **17**, 6831 (2015).
27. C. Zhong and C. C. Chu, *Cryst. Growth Des.*, **10**, 5043 (2010).
28. M. T. Hincke, Y. Nys, J. Gautron, K. Mann, A. B. Rodriguez-Navarro and M. D. McKee, *Front. Biosci.-Landmark*, **17**, 1266 (2012).
29. D. Ren, Q. Feng and X. Bourrat, *Micron*, **42**, 228 (2011).
30. X. Wang, H. C. Schroder and W. E. Muller, *Trends Biotechnol.*, **32**, 441 (2014).