Distinct macroscopic structures developed from solutions of chemical compounds and periodic proteins

Kiyotaka Shiba^{1,2+}, Takako Honma², Tamiko Minamisawa¹, Keiichi Nishiguchi² & Tetsuo Noda^{2,3}

¹Department of Protein Engineering and ²Department of Cell Biology, Japanese Foundation for Cancer Research, Toshima, Japan, and ³Department of Molecular Genetics, University of Tohoku School of Medicine, Aoba, Japan

By controlling the growth of inorganic crystals, macro-biomolecules, including proteins, play pivotal roles in modulating biomineralization. Natural proteins that promote biomineralization are often composed of simple repeats of peptide sequences; however, the relationship between these repetitive structures and their functions remains largely unknown. Here we show that an artificial protein containing a repeated peptide sequence allows NaCl, KCl, CuSO₄ and sucrose to form a variety of macroscopic structures, as represented by their dendritic configurations. Mutational analyses revealed that the physicochemical characteristics of the protein, not the peptide sequence *per se*, were responsible for formation of the dendritic structures. This suggests that proteins that modulate crystal growth may have evolved as repeat-containing forms at a relatively high rate. These observations could serve as the basis for developing new genetic programming systems for creation of artificial proteins able to modulate crystal growth from inorganic compounds, and may thus provide a new tool for nano-biotechnology.

EMBO reports 4, 148–153 (2003)

doi:10.1038/sj.embor.embor737

INTRODUCTION

The translation of concepts dealing with biomineralization into strategies for a 'bottom-up' approach to the synthesis of materials has recently come into the spotlight as one of the most important areas of research in nano-biotechnology. In biologically controlled mineralization, organic compounds—proteins, lipids, proteoglycans and polysaccharides—are known to control the structure, composition, shape and organization of inorganic crystals (Mann, 2001). Among these, proteins hold a pivotal position because a variety of methods for altering and engineering their structure and function have been established. Particularly advantageous is engineering at

the level of the gene, as a preparation with precise and reproducible molecular specifications is possible.

Recent studies on biomineralization have revealed the structures of natural proteins believed to be involved in modulating crystal growth. These proteins have been crudely categorized as 'framework proteins' that delineate space for crystal growth and 'hydrophilic proteins' that anchor the framework proteins and present an active nucleating surface (Mann, 2001). Among the former are MSI60 and MSI31 found in the nacreous layer of the pearl oyster *Pinctada fucata* (Sudo *et al.*, 1997), mucoperlin from the mollusc *Pinna nobilis* (Marin *et al.*, 2000) and the frustulins and HEP200 from the diatom *Cylindrotheca fusiformis* (Kröger *et al.*, 1996, 1997). Framework proteins such as these share a hydrophobic nature and often contain stretches of simple amino acids (Ala and Gly in MSI60 and MSI3) or repeats of a peptide sequence (13 repeats of 31 amino acids in mucoperlin). Hydrophilic proteins such as nacrein from *Pinctada fucata* (Miyamoto *et al.*, 1996) and nephropontin from humans (Shiraga *et al.*, 1992) often contain a large number of acidic residues (Glu and Asp) that can form electrostatic bonds with metal ions and are thus believed to play a role in controlling inorganic nucleation. Occasionally, these acidic residues appear as repeating units, as exemplified by the Gly-X-Asn (where X is Asp or Asn) repeat in nacrein (Miyamoto *et al.*, 1996). However, the details of the molecular mechanisms by which these acidic proteins govern crystal growth remain largely unknown.

We have been focusing on the periodicity perceivable in the structures of existing natural proteins. Their repetitive structures are compatible with the hypothesis that genes arose in the form of repeats of short DNA sequences (Ohno, 1984). The repetitive structure of proteins could therefore be a relic of their natal mode. However, some patterns of repetition are apparently related to the function of the proteins. For example, antifreeze proteins from insects are composed of tandem repeats of peptide sequences that fold into a regular β -helix structure within which Thr-X-Thr motifs are arranged at regular intervals. The spacing between the -OH groups of these Thr residues is a near-perfect match with the oxygen atoms in ice lattices and, based on this molecular complementarity, it is believed that the protein binds to the surface of ice crystals to inhibit their growth (Graether *et al.*, 2000; Liou *et al.*, 2000). Similar

¹Department of Protein Engineering and ²Department of Cell Biology, Cancer Institute, Japanese Foundation for Cancer Research, Toshima, Tokyo 170-8455, Japan

³Department of Molecular Genetics, University of Tohoku School of Medicine, Aoba, Sendai 980-8575, Japan

*Corresponding author. Tel: +81 3 3918 0111; Fax: +81 3 5394 3903; E-mail: kshiba@jfc.or.jp

Received 22 July 2002; accepted 13 December 2002

Published 24 January 2003

complementarity-based control of growth has been proposed to underpin the molecular mechanisms involved in the regulation of biomineralization by macro-biomolecules containing periodic structures (Addadi *et al.*, 2001).

We recently established a novel system in which artificial proteins containing periodic structures are created by polymerization of a microgene (Shiba *et al.*, 1997, 2002). In the course of characterizing these proteins, we found that one, protein 288, formed a macroscopic dendritic structure in conjunction with NaCl. Here we show that several macroscopic patterns can be formed when this protein is combined with other inorganic compounds.

RESULTS AND DISCUSSION

Properties of a repetitive artificial protein (protein 288)

In an earlier study we created the gene for protein 288 by polymerizing a 54-base-pair (bp) microgene (MG-14) in a head-to-tail manner using the microgene polymerization reaction (MPR) method (Shiba *et al.*, 2002). MG-14 was designed so that the peptide translated from the first coding frame of the sense strand was

similar to the type-1 copper-binding site of plastocyanin; peptides from the other five frames (two in the sense strand and three in the anti-sense strand) had no relationship to any existing protein (Shiba *et al.*, 2002). Because the MPR method randomly inserts or deletes nucleotides at end-joining junctions of microgenes, the reading frames of microgene polymers change randomly at junctions (Shiba *et al.*, 1997). The gene for protein 288 is composed of 6.4 repeats of the antisense strand of MG-14; a single nucleotide deletion at the junction between the fourth and fifth repeats changed its reading frame, producing a primary structure consisting of two blocks made up of two different repeated sequences (Fig. 1A). The calculated molecular mass of the protein is 15,162 Da and its pI 8.4. The sequence contains 7 acidic residues (Asp) and 12 basic residues (Arg and His) that are arranged at regular intervals, reflecting the repetitive structure of this protein (Fig. 1A). The charged residues, as well as many polar residues, suppress the net hydrophobicity of the protein, although it is noteworthy that a sequence of non-polar residues (Ile-Pro-Ala-Pro-Trp) is repeated in the first block of the protein and may form a hydrophobic patch, making the molecule block-copolymer-like in nature. A standard BLAST search (Altschul *et al.*, 1990) confirmed that the 116-amino-acid sequence of protein 288 does not show any significant similarity to any protein in the database (data not shown). A secondary-structure calculation program (Geourjon & Deleage, 1994) predicted that the protein is composed mainly of an extended coil with six β -turns at Pro-Tyr and Pro-Gly positions.

Protein 288 was purified from recombinant bacterial cells using a standard protocol in the presence of denaturant. After removal of the denaturant by dialysis, the protein turned out to be insoluble in most physiological buffers; only under acidic conditions—for example, in Tris-acetate (pH 4.0) or 10% formic acid—was the protein soluble, and the solubilized protein formed a film-like structure when dried on a solid support at 25 °C (Fig. 1B). The infrared spectrum of the film (cast on KRS-5 crystal from a 10% formic acid solution) had a broad absorbance peak centred at 1,644 cm^{-1} (Fig. 1C), which is in the amide band I region and is indicative of varying patterns of hydrogen bonding (Surewicz & Mantsch, 1988).

Dendrite development from protein 288 and NaCl

During the course of the experiments on film formation, we observed that macroscopic dendritic structures developed from a protein solution containing 100 mM NaCl and 50 mM Tris-acetate (pH 4.0) (Fig. 2). In the absence of protein, only the faceted, blocky shapes of NaCl crystals were formed (Fig. 2E), but in the presence of 20–180 μM protein 288, dendritic structures in the sub-millimetre range grew on the glass support (Fig. 2A–C). These structures were not formed from NaCl solutions containing BSA, lysozyme or any other artificial protein tested (see Fig. 2D and Supplementary Information). Observations from high-magnification microscopy revealed that a thin layer with an uneven texture was spread over the bottom of the glass support on which the dendritic structures developed (Fig. 3A). Using a confocal laser-scanning microscope we further showed that the film-like layer had a rough surface, and that the dendritic structures arose from that layer. It is proposed that the dendritic architecture is composed of NaCl crystal; we do not yet know whether the structure also contains protein 288.

Fig. 1 Properties of protein 288. (A) Primary structure. The amino-terminal MRGSHHHHHH peptide tag derived from the expression vector pKS583 (Shiba *et al.*, 2002) is omitted. White letters in grey boxes indicate charged residues. (B) The protein formed a film-like structure after drying from a solution of 10% formic acid. In this picture, the film was swollen and shrivelled by adding a small amount of Coomassie Brilliant Blue solution. (C) The infrared spectrum of the film showed a broad peak centred at 1,643 cm^{-1} , which is in the amide band I region.

Fig. 2 | Development of dendritic structures. The structures were developed from 50 μ l of solution containing 100 mM NaCl, 50 mM Tris-acetate (pH 4.0) and the following concentrations of protein: 180 μ M protein 288 (A), 45 μ M protein 288 (B), 20 μ M protein 288 (C) no protein (E) or 40 μ M BSA (D). The micrographs were obtained using a stereomicroscope (Wild M10; Leica).

It is intriguing that slight changes in the experimental conditions resulted in the formation of qualitatively different dendritic structures, several examples of which are shown in Fig. 4 (and see Supplementary Information). Crystals having a diffusion-limited aggregation (DLA)-like structure (Fig. 4A), which is often observed in both living and non-living non-equilibrium systems (Ball, 1999), predominated when high concentrations of protein 288 were dried slowly. Structures having a chiral morphology (Fig. 4B, C)—that is, those composed of twisted branches that could be a manifestation of a microscopic chirality of one or more factors present in the system (Ben-Jacob & Levine,

Fig. 3 | Characterization of dendritic structures. (A) High-magnification micrograph showing a bough region. The picture was taken using a microscope equipped with a 20 $\times$ objective lens (Axioplan 2; Zeiss). In the picture, the protein film formed on a plastic plate that had been scratched with a needle (arrows). (B) Reconstituted 3D image of a dendritic crystal on the film. Fifty reflected images, each 0.5 μ m wide, were collected using a confocal laser scanning microscope (beam wavelength 488 nm; LSM 510; Zeiss), and a 3D image was constructed from the data (using Topography software; Zeiss).

2001) — predominated when succinate-NaOH (pH 4.0) buffer was used. The structures that developed on (3-aminopropyl)triethoxysilane-coated glass showed regular branching at 60° angles (Fig. 4D) that was reminiscent of hexagonal snow crystals. The fact that distinct patterns were produced by subtle changes in the experimental conditions is characteristic of non-equilibrium growth processes and is exemplified by the array of branched patterns produced from a simple Hele–Shaw cell apparatus (Ball, 1999; Ben-Jacob *et al.*, 2000). The formation of these dendritic structures is therefore likely to be the result of complex interactions between protein 288, NaCl and the supporting material.

Mutational analyses of protein 288

To understand the mechanism of formation of the ramified structures requires knowledge of how the specific sequence of the protein affects the structure. We therefore carried out a series of similar experiments using variants of protein 288 which were

created by introducing random mutations into the coding frame of the protein with an error-prone polymerase chain reaction (PCR) method (Bartel & Szostak, 1993). Four variants were then selected for analysis: protein 288-001 (Fig. 5A) has 11 amino-acid changes and one termination codon that truncates the carboxy terminus of the protein by seven residues. Protein 288-003 (Fig. 5B) has 11 amino-acid changes, including three positively charged residues (Arg and Lys); it lacks three of the six predicted β -turns in protein 288 (at Pro-Tyr and Pro-Gly positions) and its hydrophilicity is increased from that of protein 288 by the gain of several charged and polar residues. Protein 288-011 (Fig. 5C) has four amino-acid alterations and one additional frame-shift that results in the last repeat being encoded by the third frame. Its net charge is shifted into the acidic range by the loss of two Arg residues and its hydrophobicity is slightly increased compared with that of protein 288. Protein 288-009 (Fig. 5D) has two additional frame-shifts, causing the carboxy-terminal half of the protein to be composed of repeats in the third frame, which is not used in protein 288; the net charge is unchanged from that of protein 288, but it is the most hydrophobic of these proteins, including protein 288. Multiple-sequence alignment showed that the sequence of protein 288-003 is the closest to the wild-type sequence, followed, in order of decreasing similarity, by

proteins 288-001, 288-011 and 288-009 (see Supplementary Information). Growth of dendritic structures using these variants was examined using solutions containing each protein at 100 μ M, 100 mM NaCl and 50 mM Tris-acetate (pH 4.0).

As shown by the micrographs in Fig. 5, development of dendritic structures is dramatically impaired for proteins 288-001 and 288-003, indicating that certain physicochemical qualities are required for pattern formation. Interestingly, although the sequences of proteins 288-011 and 288-009 are markedly changed from that of protein 288 due to alteration of their reading frames, the ability to form dendritic structures was retained. This indicates that the specific sequence of protein 288 *per se* is not required for the development of the dendritic structures; rather it is its physicochemical characteristics—for example, the repetition of structures, its hydrophobicity and block-copolymer-like nature—that are important.

Fig. 4 | Various dendritic structures formed from a solution of protein 288 and NaCl. (A) A diffusion-limited aggregation (DLA)-like structure formed from a solution containing 500 μ M protein 288, 100 mM NaCl and 50 mM Tris-acetate (pH 4.0) on a non-treated plastic microplate (Iwaki Glass). The sample was stained with thioflavine T (final concentration in the initial solution was 40 μ M), and the 485 nm emission excited by 435 nm illumination was recorded (Wild M10; Leica). (B) A chiral structure was formed from a solution of 125 μ M protein 288, 100 mM NaCl, and 50 mM succinate-NaOH (pH 3.0) on a plastic microplate. (C) A brush-like structure was formed from a solution containing 500 μ M protein 288, 100 mM NaCl and 50 mM succinate-NaOH (pH 3.0) on a plastic microplate. (D) A sextile branch was formed from 50 μ l of solution containing 10 μ M protein 288, 100 mM NaCl, 1 mM EDTA and 50 mM Tris-acetate (pH 4.0) on a (3-aminopropyl)triethoxysilane-coated cover glass (Matsunami).

Fig. 5 | Dendritic structures formed with derivatives of protein 288. (A) Protein 288-001. (B) Protein 288-003. (C) Protein 288-011. (D) Protein 288-009. The primary structures of the proteins (without the amino-terminal peptide tag derived from the pKS589 expression vector) are shown on the left. Residues that are the same as in the parent protein are shown as grey lower case; mutated residues are shown as black upper case; charged residues are shown as white letters in black boxes.

Fig. 6 | Microscopic structures formed with protein 288 and various salts. All structures were formed by placing 50 μ l of solution onto a siliconized coverslip (HR3-215; Hampton Research). (A) and (B) Stereomicroscopic images of crystals formed from 100 mM KCl and 50 mM Tris-acetate (pH 4.0), without (A) or with (B) 36 μ M protein 288. (C) A stereomicrograph of a bulky sucrose crystal formed in the absence of protein 288. (D) Micrograph obtained using cross-polarized light showing crystals formed from 100 mM sucrose, 50 mM Tris-acetate (pH 4.0) and 36 μ M protein 288. (E, F) Images obtained using a stereomicroscope (E) and a monocular microscope equipped for cross-polarized illumination (F) showing a deposit formed from 100 mM CuSO_4 , 50 mM Tris-acetate (pH 4.0) and 18 μ M protein 288. (G, H) Images obtained using a stereomicroscope (G) and a monocular microscope (H) showing crystals formed from 100 mM NaCl, 50 mM trimethylamine-formate (pH 4.0) and 9 μ M protein 288.

Microscopic structures developed with other compounds

We next examined the formation of macroscopic structures when protein 288 is combined with other compounds. We first used 100 mM KCl (lattice constant: 6.29 Å); like NaCl (lattice constant: 5.65 Å), its crystal class is cubic, and as with NaCl (Fig. 4C), the combination of protein 288 and KCl resulted in the formation of branched structures dominated by orthogonal and 60° angles (Fig. 6B), indicating that the protein interacts with a

specific plane of the crystals. In the absence of the protein, the solution formed bulky KCl crystals (Fig. 6A). A distinct dendritic structure was also formed when the protein was combined with sucrose (Fig. 6D). This pattern was similar to previously observed dendritic crystals of sucrose (Powers, 1969) and was not formed in the absence of the protein (Fig. 6D). When 100 mM copper sulphate was included in the protein solution, only irregular deposit patterns were observed (Fig. 6E). Observations with polarized light, however, revealed that periodic structures, which are another type of macroscopic structure, emerged from this solution (Fig. 6E). Finally, we tested the combination of NaCl and protein 288 in 50 mM trimethylamine-formate (pH 4.0) and observed the formation of composite crystals comprised of bulky criss-crosses and thin needles (Fig. 6G and H) that were probably formed from NaCl and sodium formate, respectively. Thus, protein 288 has the potential to develop a wide variety of qualitatively different macroscopic structures through its interaction with different inorganic compounds.

DISCUSSION

An artificially created repeat-containing protein, protein 288, that shares no sequence similarity with any known naturally occurring protein enables minerals to form a wide range of dendritic structures. Mutational analyses showed that this ability does not result from the peptide sequence *per se*, but from the physicochemical characteristics of the protein. These observations suggest that proteins with the ability to modulate crystal growth could have evolved as forms containing repeats at a relatively high rate. In addition, they could serve as the basis for the development of new experimental systems for creating proteins that modulate the mineralization of inorganic compounds.

METHODS

Purification of protein 288. Protein 288 was expressed from plasmid pYT288 in *Escherichia coli* and purified as described previously (Shiba *et al.*, 2002). The purified protein was dialysed against 50 mM Tris-acetate (pH 4.0) containing 1 mM EDTA and 100 mM NaCl.

Growth of dendritic structures. The standard conditions for formation of macroscopic structures with protein 288 were as follows: a 50 μ l sample of a solution of protein 288 in 50 mM Tris-acetate (pH 4.0) containing 100 mM NaCl was put into one well of a 96-well microplate, after which the solution was left to evaporate at 25 °C. The macroscopic structures grew rapidly as the solution dried up (see Supplementary Information).

Infrared measurement. The IR spectrum was obtained using a Shimadzu FTIR-8200A spectrophotometer. Sample film was prepared by casting 50 μ l of protein 288 solution (5 mg ml⁻¹) in 10% formic acid on a KRS-5 (thallium bromo-iodide) crystal plate (Shimadzu).

Construction of derivatives of protein 288. Derivatives of protein 288 were obtained by mutation of pYT288 DNA using an error-prone PCR method (Bartel & Szostak, 1993). DNAs encoding the derivatives of protein 288 were then cloned into the pKS589 expression vector, which was constructed by cloning a 1.7-kb *SalI*–*HindIII* fragment of pKS583 (Shiba *et al.*, 2002) into the same sites in pQE-9 (Qiagen).

Supplementary information is available at *EMBO reports* Online (<http://www.emboreports.org>).

ACKNOWLEDGEMENTS

We thank K. Matsuda, R. Kobayashi and N. Shigesada for discussions. This work was supported in part by a HFSP Research grant.

REFERENCES

- Addadi, L., Weiner, S. & Geva, M. (2001) On how proteins interact with crystals and their effect on crystal formation. *Z. Kardiol.*, **90** (suppl. 3), 92–98.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Ball, P. (1999) *The self-made tapestry: pattern formation in nature*. Oxford University Press, New York.
- Bartel, D.P. & Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411–1418.
- Ben-Jacob, E. & Levine, H. (2001) The artistry of nature. *Nature*, **409**, 985–986.
- Ben-Jacob, E., Cohen, I. & Levine, H. (2000) Cooperative self-organization of microorganisms. *Adv. Phys.*, **49**, 395–554.
- Geourjon, C. & Deleage, G. (1994) SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.*, **7**, 157–164.
- Graether, S.P. *et al.* (2000) β -Helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature*, **406**, 325–328.
- Kröger, N., Bergsdorf, C. & Sumper, M. (1996) Frustulins: domain conservation in a protein family associated with diatom cell walls. *Eur. J. Biochem.*, **239**, 259–264.
- Kröger, N., Lehmann, G., Rachel, R. & Sumper, M. (1997) Characterization of a 200-kDa diatom protein that is specifically associated with a silica-based substructure of the cell wall. *Eur. J. Biochem.*, **250**, 99–105.
- Liou, Y.C., Tocilj, A., Davies, P.L. & Jia, Z. (2000) Mimicry of ice structure by surface hydroxyls and water of a β -helix antifreeze protein. *Nature*, **406**, 322–324.
- Mann, S. (2001) *Biomineralization—principles and concepts in bioinorganic materials chemistry*. Oxford University Press, Oxford.
- Marin, F. *et al.* (2000) Mucins and molluscan calcification. Molecular characterization of mucoperlin, a novel mucin-like protein from the nacreous shell layer of the fan mussel *Pinna nobilis* (Bivalvia Pteriomorpha). *J. Biol. Chem.*, **275**, 20667–20675.
- Miyamoto, H. *et al.* (1996) A carbonic anhydrase from the nacreous layer in oyster pearls. *Proc. Natl Acad. Sci. USA*, **93**, 9657–9660.
- Ohno, S. (1984) Repeats of base oligomers as the primordial coding sequences of the primeval earth and their vestiges in modern genes. *J. Mol. Evol.*, **20**, 313–321.
- Powers, H.E.C. (1969) Sucrose crystals: inclusions and structure. *Sugar Tech. Rev.*, **1**, 85–190.
- Shiba, K., Takahashi, T. & Noda, T. (1997) Creation of libraries with long open reading frames by polymerization of a micro-gene. *Proc. Natl Acad. Sci. USA*, **94**, 3805–3810.
- Shiba, K., Takahashi, Y. & Noda, T. (2002) On the role of periodism in the origin of proteins. *J. Mol. Biol.*, **320**, 833–840.
- Shiraga, H. *et al.* (1992) Inhibition of calcium oxalate crystal growth *in vitro* by uropontin: another member of the aspartic acid-rich protein superfamily. *Proc. Natl Acad. Sci. USA*, **89**, 426–430.
- Sudo, S. *et al.* (1997) Structures of mollusc shell framework proteins. *Nature*, **387**, 563–564.
- Surewicz, W.K. & Mantsch, H.H. (1988) New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim. Biophys. Acta*, **952**, 115–130.