Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Nucleant-mediated protein crystallization with the application of microporous synthetic zeolites

Michihiro Sugahara, Yukuhiko Asada, Yuko Morikawa, Yuichi Kageyama and Naoki Kunishima*

RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo-cho, Sayo-gun, Hyogo 679-5148, Japan

Correspondence e-mail: kunisima@spring8.or.jp

Received 15 November 2007
Accepted 11 April 2008

PDB References: PH1033 (condition I), 1wmm; PH1033 (condition II), 2zbn; PH1033 (condition II from MS 5A), 2hd9; TTHB142, 2dpn.

Protein crystallization is still a major bottleneck in structural biology. As the current methodology of protein crystallization is a type of screening, it is usually difficult to crystallize important target proteins. It was thought that hetero-epitaxic growth from the surface of a mineral crystal acting as a nucleant would be an effective enhancer of protein crystallization. However, in spite of almost two decades of effort, a generally applicable hetero-epitaxic nucleant for protein crystallization has yet to be found. Here we introduce the first candidate for a universal hetero-epitaxic nucleant, microporous zeolite: a synthetic aluminosilicate crystalline polymer with regular micropores. It promotes a form-selective crystal nucleation of proteins and acts as a crystallization catalyst. The most successful zeolite nucleant was molecular sieve type 5A with a pore size of 5 Å and with bound Ca^{2+} ions. The zeolite-mediated crystallization improved the crystal quality in five out of six proteins tested. It provided new crystal forms with better resolution in two cases, larger crystals in one case, and zeolite-dependent crystal formations in two cases. The hetero-epitaxic growth of the zeolite-mediated crystals was confirmed by a crystal-packing analysis which revealed a layer-like structure in the crystal lattice.

1. Introduction

X-ray crystallography is one of the most powerful techniques for the three-dimensional (3D) structural determination of proteins. One of the major bottlenecks in this process is the production of high-quality crystals. The most widely used method for improving crystal quality is the large-scale screening of crystallization conditions. In order to facilitate the process of crystal screening, a number of automated crystallization facilities have been developed by structural genomics initiatives (Stevens, 2000; Sulzenbacher *et al.*, 2002; Sugahara & Miyano, 2002; Watanabe *et al.*, 2002; Hui & Edwards, 2003; Adachi *et al.*, 2004; Shah *et al.*, 2005; Miyatake *et al.*, 2005; Hiraki *et al.*, 2006). In spite of these efforts, the crystallization of proteins with poor crystallizability still remains a major problem. Thus, in order to overcome this difficulty, a new effective methodology for improving protein crystallizability is needed. The poor crystallizability of difficult target proteins may be due to the proteins' low stability and/or flexible molecular surface properties, thereby providing crystals with poor diffraction quality in many cases. Alternatively, even perfectly stable and 'firm' protein cannot be crystallized if it does not have a molecular surface for favourable crystal packing. In this report, the term 'poor crystallizability' is associated with the latter statement, where extrinsic control of

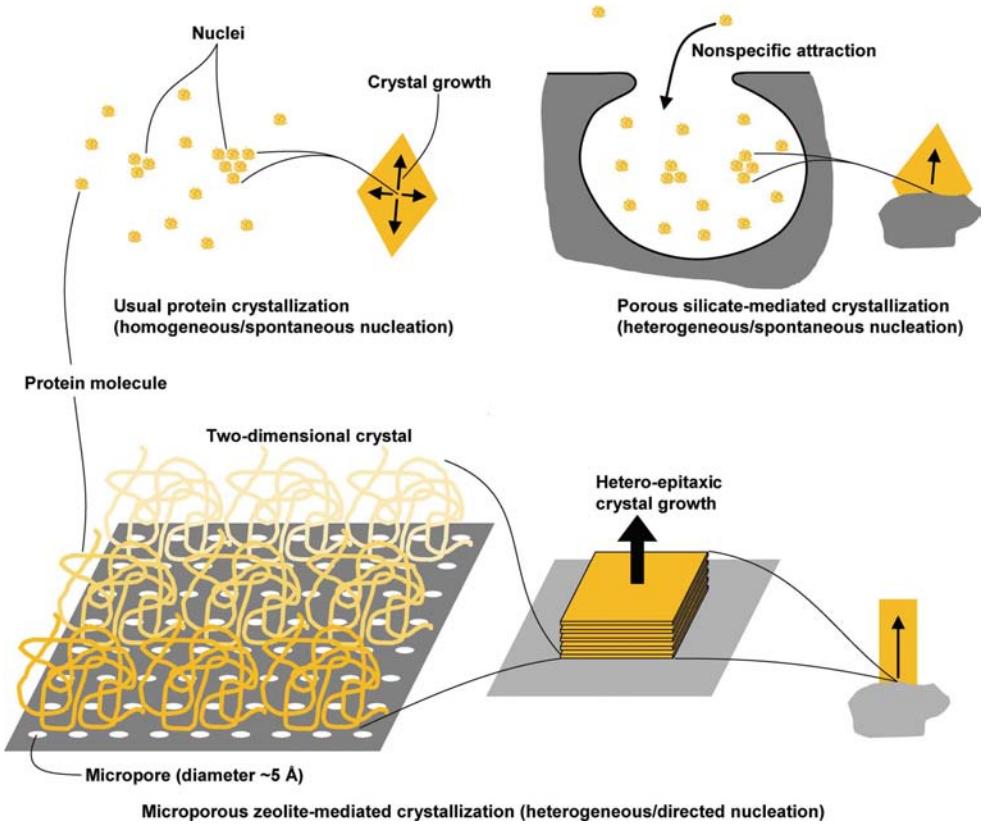

Figure 1
Schematic diagram of hetero-epitaxic nucleant for protein crystallization.

certain aspects of crystallization can enhance crystal nucleation.

Protein crystallization is initiated by nucleation whereby a number of protein molecules associate to make a small crystal nucleus. Crystal nucleation may be categorized as homogeneous/heterogeneous and spontaneous/directed. In homogeneous nucleation, crystal nuclei are spontaneously formed from the supersaturated protein aqueous solution, indicating spontaneous nucleation. In contrast, in heterogeneous nucleation, a different type of solid substance referred to as the 'nucleant' promotes nucleation, and this is thought to improve the crystallizability of proteins and to minimize the protein quantity required for crystallization (McPherson & Shlichta, 1988; Chayen *et al.*, 2001, 2006; Chayen & Saridakis, 2001; Edwards *et al.*, 1994; Hemming *et al.*, 1995; Ray & Bracker, 1986; Leung *et al.*, 1989; Punzi *et al.*, 1991; Wood *et al.*, 1992; Falini *et al.*, 2002). Whether nucleation is spontaneous or directed depends upon the type of nucleant used and how it is used. The most useful heterogeneous nucleants currently available are porous silicon (Chayen *et al.*, 2001) and mesoporous bioactive gel-glass (Chayen *et al.*, 2006), which generally promote the nucleation of protein crystals. The porous silicates attract protein molecules in their pores whose dimensions are larger than those of protein molecules, thereby encouraging proteins to form crystalline nuclei (Chayen *et al.*, 2006). Thus, the porous silicates facilitate heterogeneous and spontaneous nucleation which is not enough to innovate the screening of protein crystallization because it only elevates the local concentration of proteins. The hetero-epitaxic growth of protein crystals (McPherson & Shlichta, 1988; Edwards *et al.*, 1994; Hemming *et al.*, 1995) produces 3D crystals of a target protein from a stacking of two-dimensional (2D) crystals that are formed by specific and regular interactions with a crystal surface of another substance, indicating heterogeneous and directed nucleation (Fig. 1).

The first suggestion that mineral crystals could be used as epitaxic nucleants in protein crystallization was made by McPherson & Shlichta in 1988 (McPherson & Shlichta, 1988). Although 50 different mineral substances were examined, they were not found to be generally applicable in protein crystallization. Another approach to hetero-epitaxic growth in protein crystallization was attempted by Edwards *et al.* using a ligand array on the lipid layers (Edwards *et al.*, 1994; Hemming *et al.*, 1995). The layers of biotinated lipids specifically interact with tetrameric streptavidin molecules to form 2D crystals and subsequently promote hetero-epitaxic growth to form 3D crystals of streptavidin. Unfortunately, the applicability of this lipid seeding method is limited due to the difficulty in the preparation of lipid-linked ligands. Here, we introduce a simple and quick approach to protein crystallization using microporous synthetic zeolite molecular sieves (MS) as a hetero-epitaxic nucleant. The MS is a cage-structured crystalline oxide material where a covalent tetrahedral framework of silicon and oxygen atoms makes a seamless polymer with identical and ordered pores (Reed & Breck, 1956; Olson, 1970). MS-mediated hetero-epitaxic nucleation can produce distinct forms of protein crystals that are not available from spontaneous nucleation, thereby enhancing the variety of crystals in crystallization screening.

2. Materials and methods

2.1. Proteins

Six proteins were used in this experiment, three from *Pyrococcus horikoshii* OT3 (PH1033, 11492 and 70020) and three from *Thermus thermophilus* HB8 (TTHB142, 00005 and 70134), chosen to cover a wide range of molecular size and a variety of crystallization conditions. Two uncharacterized proteins (PH1033 and 11492) from *P. horikoshii* OT3 share 32% sequence identity with the UPF0310 protein from

Sulfolobus solfataricus and 80% sequence identity with the acetyl-CoA synthetase β subunit from *Thermococcus kodakarensis*. However, the *BLAST* search (Altschul *et al.*, 1997) for these two proteins in the PDB revealed no significant homologues. The glycerol kinase from *T. thermophilus* HB8 (TTHB142) shares 44% sequence identity with the glycerol kinase from *Escherichia coli* (PDB code 1bu6; Feese *et al.*, 1998). The structure of *E. coli* glycerol kinase is dimeric in crystal form. The UDP-*N*-acetylglucosamine pyrophosphorylase (protein 00005) from *T. thermophilus* HB8 shares 42% sequence identity with the *N*-acetylglucosamine-1-phosphate uridyltransferase from *Streptococcus pneumoniae* (PDB code 1hm9; trimeric in crystal form; Sulzenbacher *et al.*, 2001). A dynamic light-scattering experiment showed a monomeric state for PH1033 and 11492, and a dimeric state for TTHB142 in solution. Proteins 70020 and 70134 are the mutants of the diphthine synthase from *P. horikoshii* OT3 (PDB code for the wild-type protein 1wng; dimeric in crystal form and in solution; Kishishita *et al.*, 2008) and the α-ribazole-5′-phosphate phosphatase from *T. thermophilus* HB8 (PDB code for wild-type protein 1v7q; monomeric in crystal form and in solution), respectively.

2.2. Expression and purification

The purified recombinant protein samples for crystallization were mainly supplied from RIKEN Genomic Science Center and Structurome Research Group (Yokoyama *et al.*, 2000). The six proteins were cloned using PCR from genomic DNA of *P. horikoshii* OT3 (PH1033, gene *ID1443356*; 11492, gene *ID1442635*; 70020, gene *ID1443058*) and *T. thermophilus* HB8 (TTHB142, gene *ID3169247*; 00005, gene *ID3168468*; 70134, gene *ID3169579*). Each plasmid encoding PH1033 (residues 1–145; 17.4 kDa), TTHB142 (residues 1–495; 53.3 kDa), 00005 (residues 1–456; 48.9 kDa), 11492 (residues 1–238; 26.8 kDa), 70020 (residues 1–265; N142M mutation; 29.3 kDa), 70134 (residues 1–177; P127H mutation; 19.6 kDa) was digested with *Nde*I and *Bgl*II and the fragment was inserted into the expression vector pET11a (Novagen) linearized with *Nde*I and *Bam*HI. The recombinant pET11a plasmids encode the full-length wild-type proteins without any N- or C-terminal modifications.

E. coli BL21 Codon Plus (DE3)-RIL cells for PH1033, 11492 and 70020, *E. coli* BL21 (DE3) cells for TTHB142, 00005 and 70134 were transformed with the recombinant plasmids and grown at 310 K in Luria–Bertani medium containing 50 μg ml^{-1} ampicillin for 20 h. The cells were harvested by centrifugation at 4500*g* for 5 min at 277 K and suspended in: 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 5 m*M* 2-mercaptoethanol for protein 00005; 20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 5 m*M* 2-mercaptoethanol, 1 m*M* phenylmethylsulfonyl fluoride for the other proteins. The suspension was disrupted by sonication and heated at: 363 K, 13 min for PH1033, 11492 and 70020; 343 K, 18 min for protein 00005; 343 K, 13 min for TTHB142; 343 K, 11.5 min for protein 70134. The cell debris and denaturated proteins were removed by centrifugation at: 20 000*g*, 90 min for protein 00005; 20 000*g*, 30 min for the other proteins. The supernatant solution was used as the crude extract for purification.

The crude extract was desalted with a HiPrep 26/10 desalting column (Amersham Biosciences) and applied onto a Super Q Toyopearl 650M (Tosoh) column equilibrated with 20 m*M* Tris–HCl pH 8.0 (buffer *A*). Each protein was eluted with a linear gradient of 0–0.5 *M* NaCl for PH1033; 0–0.4 *M* NaCl for protein 11492; 0–0.3 *M* NaCl for TTHB142, 00005, 70020 and 70134. For protein 00005, the fraction containing target protein was then sequentially applied onto eight columns: (1) a Resource-Q column (Amersham Biosciences) equilibrated with 20 m*M* MES–NaOH pH 6.0 (buffer *B*) and eluted with a linear gradient of 0–0.3 *M* NaCl in buffer *A*; (2) a Bio-Scale CHT-20-I column (Bio-Rad) equilibrated with 10 m*M* potassium phosphate pH 7.0 and eluted with a linear gradient of 10–200 m*M* potassium phosphate pH 7.0; (3) a Resouce PHE1 column (Amersham Biosciences) equilibrated with 50 m*M* potassium phosphate pH 7.0 and eluted with a linear gradient of 0.75–0 *M* ammonium sulfate; (4) a HiTrap Heparin column (Pharmacia) equilibrated with buffer *B* containing 0.06 *M* NaCl and eluted with a linear gradient of 0.06–1.0 *M* NaCl; (5) a HiTrap Blue column (Pharmacia) equilibrated with buffer *B* containing 0.06 *M* NaCl and eluted with a linear gradient of 0–2.0 *M* NaCl in buffer *B*; (6) a Mono Q HR 10/10 column (Pharmacia) equilibrated with buffer *A* and eluted with a linear gradient of 0–0.3 *M* NaCl in buffer *A*; (7) a Resource-S1 column (Amersham Biosciences) equilibrated with buffer *B* and eluted with a linear gradient of 0–0.3 *M* NaCl in buffer *B*; (8) Resource-S6 column (Amersham Biosciences) equilibrated with buffer *B* and eluted with a linear gradient of 0–0.3 *M* NaCl in buffer *B*. After the buffer replacement with buffer *B* for all six proteins, the fraction containing target protein was subjected to a Resource-Q column (Amersham Biosciences) equilibrated with buffer *B* and eluted with a linear gradient of 0–0.5 *M* NaCl for PH1033 and 0–0.3 *M* NaCl for the other proteins. Each sample for PH1033, TTHB142, 11492 and 70020 was then applied onto a Bio-Scale CHT-10-I column (for PH1033 and TTHB142) and a CHT-20-I column (Bio-Rad; for proteins 11492 and 70020) both equilibrated with 10 m*M* potassium phosphate pH 7.0 and eluted with a linear gradient of the same phosphate buffer: 10–500 m*M* for PH1033, 11492 and 70020; 10–100 m*M* for TTHB142. The samples were concentrated by ultrafiltration (Vivaspin; 10 kDa cutoff) and loaded onto a HiLoad 16/60 Superdex 75 prep-grade column (Amersham Biosciences; for PH1033 and 70134) and a HiLoad 16/60 Superdex 200 prep-grade column (Amersham Biosciences; for TTHB142, 00005, 11492 and 70020) both equilibrated with buffer *A* containing 0.2 *M* NaCl. The homogeneity and identity of the purified samples were assessed by SDS–PAGE (Laemmli, 1970) and N-terminal sequence analysis. Finally, the purified protein solutions in buffer *A* with 0.2 *M* NaCl were stored at 203 K after the concentration by ultrafiltration to 21.4 mg ml^{-1} for PH1033, 14.6 mg ml^{-1} for TTHB142, 44.8 mg ml^{-1} for protein 00005, 46.7 mg ml^{-1} for protein 11492, 10.8 mg ml^{-1} for protein 70020 and 9.7 mg ml^{-1} for protein 70134.

Table 1
Data-collection and structure refinement statistics.

Values in parentheses are for the outermost shell.

Protein ID	PH1033 (condition II)	PH1033 (condition II)	TTHB142	70020	70134
MS	none	MS 5A	MS 5A	MS 5A	MS 5A
Space group	$P3_121$	$P3_221$	$P2_12_12_1$	$P6_322$	$P4_12_12$
Unit-cell parameters (Å)	$a = b = 95.0$, $c = 44.5$	$a = b = 39.9$, $c = 168.5$	$a = 68.0$, $b = 90.5$, $c = 183.0$	$a = b = 115.8$, $c = 114.4$	$a = b = 43.4$, $c = 155.3$
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000	1.5418
Resolution range (Å)	40–2.0 (2.07–2.00)	40–1.35 (1.40–1.35)	30–2.8 (2.90–2.80)	30–3.2 (3.31–3.20)	20–1.7 (1.76–1.70)
No. of unique reflections	15961 (1567)	35514 (3455)	28303 (2766)	7866 (756)	17232 (1658)
Redundancy	9.6 (9.4)	9.9 (9.6)	4.8 (3.6)	16.4 (17.2)	14.6 (14.6)
Completeness (%)	100 (99.9)	100(100)	99.4 (98.7)	99.9 (100.0)	99.7 (99.0)
R_{sym}† (%)	6.5 (55.5)	12.2 (63.6)	8.8 (32.5)	10.9 (58.7)	5.6 (38.7)
$\langle I/\sigma(I)\rangle$	9.9 (4.2)	6.7(3.2)	9.2 (4.1)	9.3 (4.2)	9.3 (6.0)
Refinement					
Resolution range (Å)	39.2–2.0	32.0–1.35	29.8–2.8		
R_{cryst}/R_{free} (%)	24.4/25.1	22.8/23.4	21.3/26.6		
R.m.s. deviation					
Bond lengths (Å)	0.008	0.008	0.009		
Bond angles (°)	1.4	1.5	1.5		

† $R_{sym} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean intensity of multiple $I_i(hkl)$ observations of Bijvoet-equivalent indices.

2.3. Crystallization without MS

Crystallization of the six proteins (PH1033, TTHB142, 00005, 11492, 70020 and 70134) without MS was performed using the conventional oil-microbatch method (Chayen *et al.*, 1990) using Nunc HLA plates (Nalge Nunc International). Each crystallization drop was prepared by mixing equal volumes of the purified protein solution and the precipitant solution. The crystallization drop was overlaid with mixed silicone/paraffin oils, allowing slow evaporation of water in the drop, and stored at a constant temperature. The initial crystallization conditions were established utilizing the automated crystallization system TERA operated at 291 K with a 1 µl drop scale and with a 1:1 mixture of silicone/paraffin oils (Sugahara & Miyano, 2002) using the precipitant solution set as follows: Crystal Screening I, II and Cryo (Hampton Research) for TTHB142; original screening set of TERA with 144 initial conditions for the other proteins. Further optimization of the initial conditions was performed as follows: manual optimization for PH1033; automatic optimization implemented in the TERA system for proteins 00005, 11492 and 70020 (for wild-type). Crystallization was performed manually at 295 K with a 2 µl drop scale and with paraffin oil for covering the drops. The final conditions for the precipitant solution were: 22.5%(*w*/*v*) PEG 4000, 0.1 *M* citrate buffer pH 5.7 for PH1033 (condition I); 22.5%(*w*/*v*) PEG 4000, 15%(*w*/*v*) glycerol, 0.1 *M* citrate buffer pH 5.7 for PH1033 (condition II); 25.5%(*w*/*v*) PEG 8000, 0.17 *M* ammonium sulfate, 15%(*v*/*v*) glycerol, 0.085 *M* cacodylate–NaOH pH 6.5 for TTHB142; 1.23 *M* ammonium sulfate, 10%(*v*/*v*) dioxane, 0.1 *M* Tris–HCl pH 7.6 for protein 00005; 22.5%(*w*/*v*) PEG 4000, 0.1 *M* citrate buffer pH 5.9 for protein 11492; 3.85 *M* sodium formate, 0.1 *M* acetate–NaOH pH 5.5 for protein 70020 (for wild-type); 2.75 *M* NaCl, 0.1 *M* citrate buffer pH 4.8 for protein 70134 (for wild-type). The mercury derivative crystals of PH1033 were prepared by cocrystallization with 10 m*M* $Hg(CH_2COO)_2$ under the same conditions as used for the native (condition I) crystals.

2.4. Crystallization with MS

Crystallization with MS was performed using the conventional oil-microbatch method except with the presence of MS. MS types 3A, 4A, 5A and 13X were used for PH1033, 00005, 11492 and 70020, and MS type 5A was used for TTHB142 and 70134. Each MS was broken using a mortar into small pieces (size 0.2–0.7 mm) and placed inside the Nunc HLA plate well. Crystallization was performed manually at 295 K with a 2 µl drop scale and with paraffin oil for covering the drops. The morphology of the crystals was observed under a low-power microscope to compare those grown with and without MS in parallel. The crystallization experiments were repeated at least 12 times for each condition in order to confirm the reproducibility of the results.

2.5. Data collection and data processing

The protein crystals were flash-cooled by placing them in a cryostat nitrogen gas stream at 100 K after a brief soaking with cryoprotectant solutions: oil-based versatile cryoprotectant (Sugahara & Kunishima, 2006) for PH1033 native (condition I); 20%(*w*/*v*) glycerol in the precipitant solution for PH1033 Hg derivative (condition I); the precipitant solution for TTHB142 with and without MS; 30%(*w*/*v*) glycerol in the precipitant solution for protein 00005 without MS; 30%(*w*/*v*) glycerol in the precipitant solution for protein 11492 with MS 5A; 30%(*w*/*v*) glycerol in the precipitant solution for protein 70020 with MS 5A; water-based versatile cryoprotectant (Sugahara & Kunishima, 2006) for protein 70134 with MS 5A. The PH1033 (condition II) crystals were directly mounted in a cryoloop from the crystallization drop and flash-cooled at 100 K in a nitrogen gas stream. Complete diffraction data sets were collected using an in-house Rigaku R-AXIS VII image-plate detector with Cu $K\alpha$ radiation and a Rigaku R-AXIS V image-plate detector and a Rigaku JUPITER 210 CCD detector with synchrotron radiation at BL26B1 of SPring-8

(Ueno *et al.*, 2006). All data were processed with *HKL2000* (Otwinowski & Minor, 1997).

2.6. Structure determination of protein PH1033

The native (condition I) crystals belonged to the space group $P3_121$ and contained one chain of PH1033 in the asymmetric unit. The mercury derivative crystals were nearly isomorphous to the native (condition I) crystals. The structure was solved by a mercury MAD (multiple anomalous dispersion) phasing at 3.1 Å resolution (Supplementary Table 1[1]). The heavy-atom positions were determined using the program *SOLVE* (Terwilliger & Berendzen, 1999). The MAD phase was improved by density modification using the program *RESOLVE* (Terwilliger, 2000). The structures from the native (condition II) crystals with and without MS were determined by molecular replacement using the program *MOLREP* (Vagin & Teplyakov, 1997) with the native (condition I) coordinates as a search model.

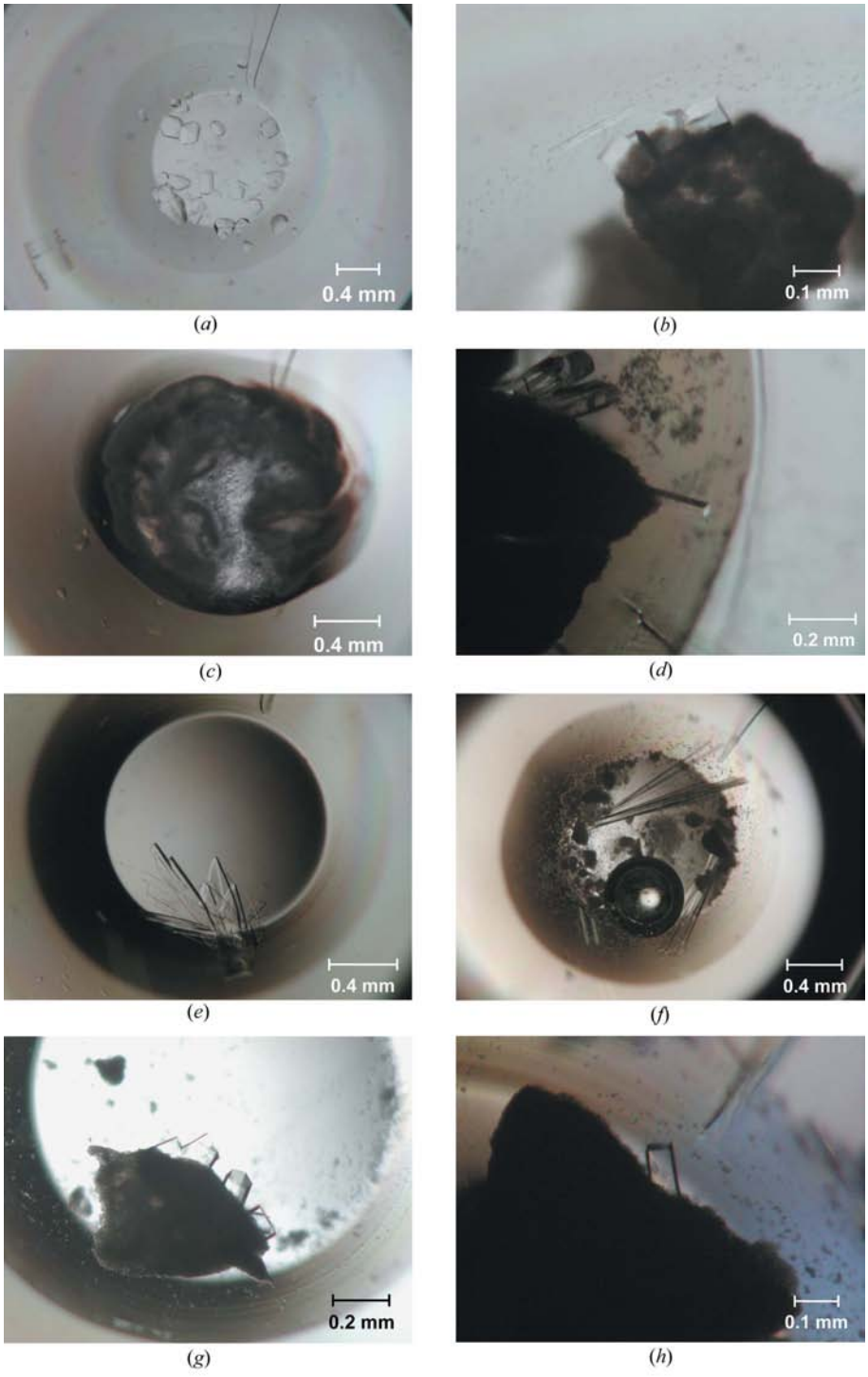

Figure 2
Photographs of protein crystals. (*a*) Protein PH1033 (condition II), (*b*) protein PH1033 (condition II) from MS 5A, (*c*) protein 11492, (*d*) protein 11492 from MS 5A, (*e*) protein TTHB142, (*f*) protein TTHB142 from MS 5A, (*g*) protein 70020 from MS 5A and (*h*) protein 70134 from MS 5A.

2.7. Structure determination of TTHB142

The TTHB142 crystals contained two chains of TTHB142 in the asymmetric unit. Positioning of the two TTHB142 molecules in the asymmetric unit was carried out using the molecular replacement method as implemented in the program *MOLREP*. Chain O (residues 3–499) of glycerol kinase from *E. coli* (sequence identity 44%; PDB code 1bu6) was used as a search model.

2.8. Refinement

Manual model revision was performed using *QUANTA2000* software (Accelrys Inc.). The program *CNS* (Brünger *et al.*, 1998) was used for the structure refinement and the electron-density map calculation. Each cycle of refinement with bulk-solvent and overall anisotropic *B*-factor corrections consisted of rigid-body refinement, simulated annealing incorporating the slow-cool protocol, positional refinement and *B*-factor refinement (individual or group). Several cycles of model revision and refinement yielded the final models. Stereochemical validation using the program *PROCHECK* (Laskowski *et al.*, 1993) revealed that of all the non-glycine residues, 87.9% and 12.1% for the PH1033 condition I structure,

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BW5223). Services for accessing this material are described at the back of the journal.

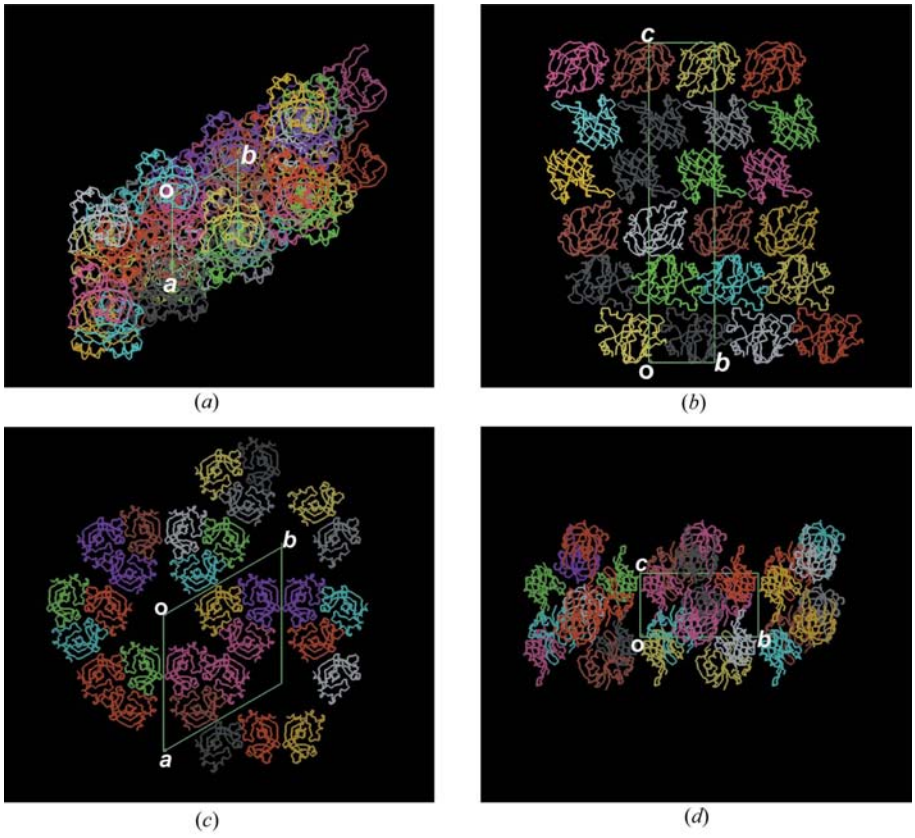

Figure 3
Packing of PH1033 crystals in the $P3_221$ cell [condition II from MS 5A; (*a*) and (*b*)] and the $P3_121$ cell [condition II; (*c*) and (*d*)]. C^{α}-trace models of PH1033 molecules are shown. The boundary of the unit cell is boxed. (*a*) The *ab* plane projection. (*b*) The *bc* plane projection. Six layers of PH1033 molecules are observed. (*c*) The *ab* plane projection. (*d*) The *bc* plane projection. Drawn in *QUANTA2000*.

90.3% and 9.7% for the PH1033 condition II structure, 91.3% and 8.7% for the PH1033 condition II structure from MS 5A and 85.0% and 15.0% for the TTHB142 structure lie within the most favoured and additionally allowed regions of the Ramachandran plot, respectively. Statistics of the data collection and refinement are shown in Table 1.

2.9. Crystallization of salts with MS 5A

To clarify the nucleation point of ammonium sulfate crystal on the MS surface, various salt concentrations, 1.23 and 1.0 *M*, 750, 500, 250, 100 and 50 m*M*, were examined in the presence of MS 5A. Firstly, MS 5A was placed inside the crystallization plate well. Each crystallization drop was prepared by adding 2.0 µl ammonium sulfate aqueous solution in the well. All drops were covered with paraffin oil, and stored at 295 K.

3. Results

3.1. Properties of molecular sieves

Nucleant-mediated crystallization of six proteins was examined using four MS materials (types 3A, 4A, 5A and 13X) as nucleants. The MS material was chosen based on the properties of known heterogeneous nucleants: (1) mineral or synthetic materials with ordered two-dimensional arrays such as crystalline inorganic compounds (McPherson & Shlichta, 1988); (2) porous substances (Chayen *et al.*, 2001, 2006); (3) cation-containing substrates, because the interfacial concentration of the calcium ion has been used to enhance the heterogeneous nucleation of $CaCO_3$ crystals at a Langmuir monolayer (Mann *et al.*, 1988); (4) commercially available, inexpensive and nonhazardous materials. Unlike the previously used porous silicon (Chayen *et al.*, 2001) and mesoporous gel-glass (Chayen *et al.*, 2006) that have non-uniform pore sizes, the MS used in this research is an aluminosilicate crystalline material with identical and ordered pores: 3A, 4A, 5A and 13X with pore sizes of 3, 4, 5 and 13 Å, respectively (Reed & Breck, 1956; Olson, 1970). These pore sizes are much smaller than the molecular sizes of the proteins examined in this study: 46 × 44 × 32 Å for protein PH1033 (17.4 kDa monomer); 106 × 60 × 56 Å for protein TTHB142 (106.7 kDa dimer); 57 × 54 × 36 Å for protein 11492 (26.8 kDa monomer); 64 × 65 × 65 Å for protein 70020 (58.6 kDa dimer); 47 × 39 × 28 Å for protein 70134 (19.6 kDa monomer). Additionally, each MS material noncovalently binds specific metal ions with regular coordination architecture: K^+ for MS 3A, Na^+ for MS 4A, Ca^{2+} for MS 5A and Na^+ for MS 13X. Therefore, the possible interactions between the MS and protein molecule are polar or nonpolar interactions mediated by the silicate matrix or the bound ion on the MS surface.

3.2. MS-mediated protein crystallization

3.2.1. PH1033. The protein PH1033 was crystallized with and without MS. Under conventional conditions without MS, the protein crystallized after 1–2 weeks (Fig. 2*a*). Crystals of protein PH1033 could not be obtained using MS 3A, 4A and 13X; in the presence of MS 5A, however, hexagonal crystals took 3–4 weeks to appear on the material's surface (Fig. 2*b*). X-ray diffraction data for the crystals of protein PH1033 were collected (Table 1). The crystals obtained without MS belong to the space group $P3_121$, with unit-cell parameters of $a = b = 95.0$, $c = 44.5$ Å and diffract X-rays to 2.0 Å resolution. The crystals obtained with MS 5A belong to the space group $P3_221$, with unit-cell parameters of $a = b = 39.9$, $c = 168.5$ Å. The crystals diffract X-rays to 1.35 Å resolution. Thus, the presence of MS 5A converted the crystal form of protein PH1033, which resulted in a substantial improvement in diffraction quality. The crystal structure of the $P3_121$ form was firstly determined by the MAD method (Figs. 3*c* and 3*d*; Table

1; Supplementary Table 1 and that of the $P3_221$ form was subsequently determined by the molecular replacement method (Figs. 3*a* and 3*b*; Table 1).

3.2.2. Protein 11492. Without MS, the protein 11492 yielded many microcrystals after 2 weeks (Fig. 2*c*). With MS 5A, protein crystals were obtained after 1 week (Fig. 2*d*). In contrast to the small and unuseable crystals obtained without MS, large crystals grew on the MS 5A surface and were of a suitable size for X-ray diffraction studies, suggesting that MS 5A may be useful for obtaining crystals of much better quality. As for protein PH1033, crystals of protein 11492 could not be obtained in the trials with the other MS materials. The diffraction experiment on 11492 protein crystals from MS 5A was performed using an in-house X-ray setup. Unfortunately, the crystals did not diffract X-rays, indicating that MS-mediated crystals are not always of high X-ray quality.

3.2.3. TTHB142. The availability of TTHB142 is limited because of the poor expression level of this protein. From the results on the proteins PH1033 and 11492, MS 5A was selected as a nucleant. Without MS, plate-shaped crystals were observed in drops incubated for 2–3 weeks (Fig. 2*e*). In contrast, rod-shaped crystals were obtained in the presence of MS 5A after 1 week (Fig. 2*f*). The crystals from the conventional method diffract X-rays to around 7 Å resolution and belong to the space group *P*2, with unit-cell parameters of $a = 154$, $b = 59$, $c = 194$ Å, $\beta = 94^\circ$. By contrast, the crystals obtained with MS 5A belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters of $a = 98.0$, $b = 90.5$, $c = 183.0$ Å, and the resolution of X-ray diffraction was remarkably improved from 7 to 2.8 Å (Table 1). The crystal structure of the orthorhombic form was determined by the molecular replacement method (Fig. 4).

3.2.4. Proteins 70020 and 70134. Crystallization of the mutant proteins 70020 and 70134 was performed with and without MS. The wild-type protein for the mutant 70020 crystallized under conditions of 3.85 *M* sodium formate and 0.1 *M* acetate–NaOH pH 5.5; however, the mutant protein 70020 did not crystallize under similar conditions. Although a large-scale crystallization screening of protein 70020 was performed using the automatic crystallization system TERA (Sugahara & Miyano, 2002), protein crystals could not be obtained. Crystallization of the wild-type protein in the presence of MS 3A, 4A, 5A and 13X yielded hexagonal crystals after 2–3 d on all the MS surfaces (Fig. 2*g*). The diffraction data set at 3.2 Å resolution was collected (Table 1).

Similarly, crystals of the mutant protein 70134 could not be obtained using the same crystallization conditions as for its wild-type protein. In the case of protein 70134, large-scale crystallization screening could not be performed because of the poor expression level of this mutant. Therefore, crystallization with MS 5A only was attempted. The mutant protein 70134 was successfully crystallized on the MS 5A surface, and block-shaped crystals were obtained after 3 d (Fig. 2*h*). The crystals belong to the space group $P4_12_12$ and diffract X-rays to 1.7 Å resolution (Table 1). These results demonstrate that the MS materials can effectively promote the crystallization of mutant proteins.

3.2.5. Protein 00005. Crystallization of the protein 00005 without MS produced hexagonal crystals after 3–4 weeks. In the crystallization trials with MS 3A, 4A, 5A and 13X, protein crystals were not observed; instead, ammonium sulfate crystals grew on all MS surfaces after 2–3 d (Fig. 5). It is probable that the high salt concentration of 1.23 *M* ammonium sulfate induced the nucleation of salt crystals on the MS surface, indicating that this method may not be adapted to crystallization conditions where ammonium sulfate is present in high concentrations. On the other hand, although crystals of TTHB142 were formed in the presence of 0.17 *M* ammonium sulfate (the final concentration was 85 m*M*), only protein crystals formed on the MS surface, indicating that the use of ammonium sulfate as an additive at low concentrations is possible. In the crystallization of the salts with MS 5A (see §2), we observed salt crystals after 1 d in all drops except for that of 50 m*M* ammonium sulfate concentration. This crystallization experiment showed that a salt concentration higher than 100 m*M* induced crystal nucleation on the MS surface. The crystallization of lithium sulfate, magnesium sulfate, ammonium acetate, ammonium chloride and ammonium formate aqueous solutions at the concentrations of 1.23 *M* and 615 m*M* with MS 5A was performed in order to elucidate further salt crystallization on the MS surface. Two samples at

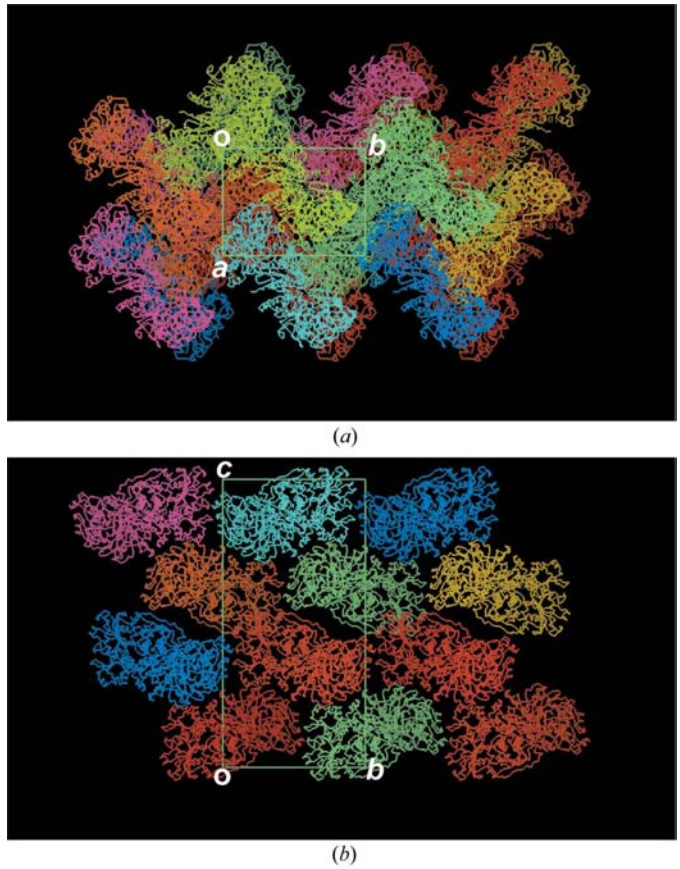

Figure 4
Packing of TTHB142 crystal in the $P2_12_12_1$ cell. C^α-trace models are shown. The boxed lines are the projection of the unit cell. (*a*) The *ab* plane projection. (*b*) The *bc* plane projection. Four layers of the TTHB142 molecules are observed. Drawn in *QUANTA2000*.

both the concentrations crystallized after 1–3 d for lithium sulfate and 1 week for magnesium sulfate. The results from the crystallization revealed that sulfate compounds would induce nucleation on the MS surface in general. The crystals of proteins 70020 and 70134 could be obtained in spite of their crystallization conditions at high salt concentrations (3.85 *M* sodium formate and 2.75 *M* NaCl, respectively), suggesting that this phenomenon of MS may be specific to sulfate. It is reported that the affinity of acid molecules to the Al_2O_3 surface increases in the following order: formate, chloride, carbonate < acetate < sulfate < salicylate < fumarate < maleate < malonate < oxalate, fluoride < citrate (Žutić & Werner, 1984).

3.3. Evidence for hetero-epitaxic growth

In order to confirm the hetero-epitaxic growth of protein crystals using MS, a crystal-packing analysis was performed. Interestingly, comparison of the two solved crystal structures of PH1033 reveals a totally different crystal packing, where the $P3_221$ crystal grown with MS 5A (Figs. 3*a* and 3*b*) has a lower solvent content of 44.2% compared with 62.8% of the $P3_121$ crystal grown without MS (Figs. 3*c* and 3*d*), suggesting that the more intimate crystal packing in the $P3_221$ form is one of the reasons for the crystal improvement by MS. Furthermore, the packing of PH1033 and TTHB142 crystals from MS 5A revealed a layer-like structure (Figs. 3*b* and 4*b*) which is typically observed in the protein crystal structures reported from hetero-epitaxic growth (Edwards *et al.*, 1994; Hemming *et al.*, 1995). Six and four layers in the unit cell of PH1033 and TTHB142 crystals, respectively, are clearly observed along the *ab* plane of the crystals. Assuming that the 3D crystals are composed of a stack of the respective 2D array, hetero-epitaxic crystal growth occurs along the *c* axis. In contrast, the packing of PH1033 crystals without MS 5A shows no layer-like structure in the crystal lattice (Figs. 3*c* and 3*d*). These results from the crystals from two proteins suggest that the MS-mediated 3D crystals are formed by hetero-epitaxic growth on the MS surface.

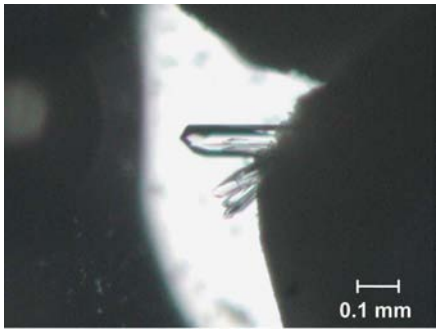

Figure 5
Photograph of MS-mediated ammonium sulfate crystals. Ammonium sulfate crystals were formed on the surface of MS 13X during the crystallization of protein 00005.

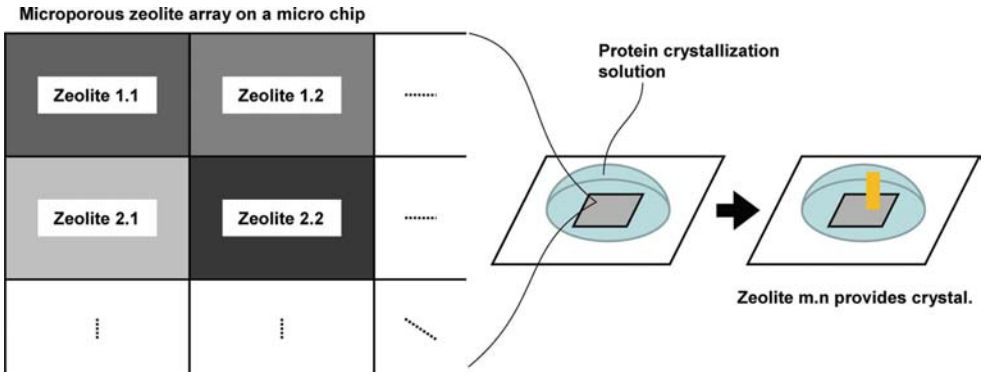

Figure 6
The future application using an array of microporous zeolite.

4. Discussion

Two decades have passed since the first report in 1988 of a hetero-epitaxic nucleant for protein crystallization (McPherson & Shlichta, 1988). To the best of our knowledge, MS is the first candidate for a universal hetero-epitaxic nucleant. The reasons for MS's versatility in hetero-epitaxic protein crystallization are intriguing. From its structural regularity, MS can be regarded as a crystal. Interestingly, the phenomenon observed in the present work is reminiscent of the ice recognition of antifreeze protein (AFP) where the protein molecules bind to the ice surface to prevent the growth of ice crystal. The crystal structures of four types of AFPs have been determined to date, and various recognition modes proposed: type I comprising a single α-helix (Sicheri & Yang, 1995); type II homologous to Ca^{2+}-dependent lectin (Gronwald *et al.*, 1998); type III with a unique compact α/β fold (Jia *et al.*, 1996); and type IV β-helix type AFP (Liou *et al.*, 2000). For instance, in the type II Ca^{2+}-bound AFP, it has been suggested that four residues on a loop region and the Ca^{2+} ion recognize specific water molecules on the {1 0−1 0} plane of the ice crystal (Liu *et al.*, 2007). This work on AFPs demonstrates that a large variety of proteins with different folds can interact specifically with crystal surfaces. Although hydrogen bonds dominate the interactions between AFP and ice crystals, nonpolar interactions in terms of shape complementarity may contribute substantially to the stabilization of the complex.

By analogy, it is likely that a large variety of proteins interact with MS surfaces in similar ways, which may be relevant to the versatility of MS-mediated protein crystallization. From the present crystallization experiments, all MS do not necessarily facilitate the formation of crystal nuclei, indicating that MS-mediated crystallization seems to depend critically upon the differences in size of pores or bound metal ions. Interestingly, most of the MS-mediated crystals were obtained only when MS 5A with 5 Å pores and bound

Ca^{2+} was used as a nucleant. The clearly higher success rate is likely to reflect the recognition of MS 5A surface by the general regular structures of proteins such as α-helix and β-sheet with structural pitches of 5.4 Å and 4.5 Å, respectively, or the generally high occurrence frequency of Ca^{2+} recognition on MS 5A surface by protein molecules; both these factors have been observed in the ice recognition of AFPs (Sicheri & Yang, 1995; Liou *et al.*, 2000; Liu *et al.*, 2007). The carboxylate–calcium interaction (Mann *et al.*, 1988) demonstrated that stearic acid at a Langmuir monolayer can enhance the heterogeneous nucleation of $CaCO_3$ crystals. The results indicated that Ca binding to the carboxylate headgroups preceded nucleation at the monolayer surface through electrostatic and stereochemical interaction at the inorganic–organic interface. Furthermore, Addadi & Weiner (1985) have reported the effect of aspartic-acid-rich β-pleated-sheet proteins on the crystallization of calcite, in which the formation of a regular array of calcium-bound aspartate side-chains emerging from the planar protein β-sheets is available for the electrostatic, structural and stereochemical requirements to direct calcite nucleation. Because the structure of MS 5A is characterized by its ordered surface calcium array, the MS 5A-mediated regulation of crystal nucleation may be accomplished based on a specific interaction between the carboxylate moieties of the protein and the calcium array of the MS surface. These possibilities require further investigation.

The MS-mediated crystallization is found to improve the crystal quality in many cases: new crystal forms with better resolutions for the proteins PH1033 and TTHB142; larger crystals for the protein 11492; MS-dependent crystal formation for the mutant proteins 70020 and 70134. Importantly, MS has a tendency to facilitate selectively new crystal forms which are not available spontaneously from solution, indicating that MS can be regarded as a crystallization 'catalyst' with crystal-form specificity in protein crystallization. Although protein crystallization is not a chemical reaction, MS can be considered as a type of catalyst because it lowers the activation energy for the nucleation of protein crystals. The concept of 'crystallization catalyst' is already accepted in the crystallization of small molecules such as metals in glass crystallization (Stookey, 1959). With regard to the applicability of this technique, a clear advantage is its simplicity as all that is required is the addition of MS to the conventional formulation of crystallization. Therefore, this method could be most effective in the crystallization of proteins with certain modifications such as mutation and selenomethionyl substitution (Hendrickson *et al.*, 1990) where the crystallization condition for the original protein is available. For difficult proteins, it may be worth carrying out a sparse matrix (Jancarik & Kim, 1991) screening of crystallization conditions with and without MS in parallel to increase the chances of obtaining crystals. Notably, the hetero-epitaxic growth suggested in this study may be suitable for the crystallization of membrane proteins, because membrane proteins tend to form 2D crystals and have better potential for crystallization at their hydrophilic molecular surfaces *i.e.* 2D crystal surfaces. In fact, successful methods for the crystallization of membrane proteins using the lipidic cubic phase (Landau & Rosenbusch, 1996; Pebay-Peyroula *et al.*, 1997; Nollert *et al.*, 2001; Paas *et al.*, 2003) and the lipidic sponge phase (Cherezov *et al.*, 2006) are based on the homo-epitaxic growth of 3D crystals from 2D crystals formed in lipid layers. Therefore, this technology may provide a new paradigm in the crystallization of difficult targets with poor crystallizability, where the success rate of crystallization screening is ultimately enhanced using an array of various types of microporous zeolite (Fig. 6). The continued search for MS-like materials that promote the nucleation of protein crystals will help in securing more efficient control of protein crystallization.

MS was the main contributor to this study and wrote the paper; YA, YM and YK established the crystallization conditions for the proteins TTHB142, 70020/70134 and 11492, respectively; NK supervised this work and co-wrote the paper. We thank the beamline staff for assistance during data collection at BL26B1 of SPring-8. The purified protein solutions from *Pyrococcus horikoshii* OT3 and *Thermus thermophilus* HB8 were obtained from RIKEN Genomic Science Center and the Structurome Research Group. This work was supported by the National Project on Protein Structural and Functional Analyses, funded by the MEXT of Japan.

References

Adachi, H., Takano, K., Matsumura, H., Niino, A., Ishizu, T., Inoue, T., Mori, Y. & Sasaki, T. (2004). *Jpn J. Appl. Phys.* **43**, L76–L78.
Addadi, L. & Weiner, S. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 4110–4114.
Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chayen, N. E. & Saridakis, E. (2001). *J. Cryst. Growth*, **232**, 262–264.
Chayen, N. E., Saridakis, E., El-Bahar, R. & Nemirovsky, Y. (2001). *J. Mol. Biol.* **312**, 591–595.
Chayen, N. E., Saridakis, E. & Sear, R. P. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 597–601.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Cherezov, V., Clogston, J., Papiz, M. Z. & Caffrey, M. (2006). *J. Mol. Biol.* **357**, 1605–1618.
Edwards, A. M., Darst, S. A., Hemming, S. A., Li, Y. & Kornberg, R. D. (1994). *Nature Struct. Biol.* **1**, 195–197.
Falini, G., Fermani, S., Conforti, G. & Ripamonti, A. (2002). *Acta Cryst.* D**58**, 1649–1652.
Feese, M. D., Faber, H. R., Bystrom, C. E., Pettigrew, D. W. & Remington, S. J. (1998). *Structure*, **6**, 1407–1418.
Gronwald, W., Loewen, M. C., Lix, B., Daugulis, A. J., Sönnichsen, F. D., Davies, P. L. & Sykes, B. D. (1998). *Biochemistry*, **37**, 4712–4721.
Hemming, S. A., Bochkarev, A., Darst, S. A., Kornberg, R. D., Ala, P., Yang, D. S. & Edwards, A. M. (1995). *J. Mol. Biol.* **246**, 308–316.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hiraki, M. *et al.* (2006). *Acta Cryst.* D**62**, 1058–1065.
Hui, R. & Edwards, A. (2003). *J. Struct. Biol.* **142**, 154–161.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Jia, Z., DeLuca, C. I., Chao, H. & Davies, P. L. (1996). *Nature* (*London*), **384**, 285–288.

Kishishita, S., Shimizu, K., Murayama, K., Terada, T., Shirouzu, M., Yokoyama, S. & Kunishima, N. (2008). *Acta Cryst.* D**64**, 397–406.
Laemmli, U. K. (1970). *Nature* (*London*), **227**, 680–685.
Landau, E. M. & Rosenbusch, J. P. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14532–14535.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leung, C. J., Nall, B. T. & Brayer, G. D. (1989). *J. Mol. Biol.* **206**, 783–785.
Liou, Y-C., Tocilj, A., Davies, P. L. & Jia, Z. (2000). *Nature* (*London*), **406**, 322–324.
Liu, Y., Li, Z., Lin, Q., Kosinski, J., Seetharaman, J., Bujnicki, J. M., Sivaraman, J. & Hew, C. L. (2007). *PLoS ONE*, **6**, e548.
McPherson, A. & Shlichta, P. (1988). *Science*, **239**, 385–387.
Mann, S., Heywood, B. R., Rajam, S. & Birchall, J. D. (1988). *Nature* (*London*), **334**, 692–695.
Miyatake, H., Kim, S.-H., Motegi, I., Matsuzaki, H., Kitahara, H., Higuchi, A. & Miki, K. (2005). *Acta Cryst.* D**61**, 658–663.
Nollert, P., Qiu, H., Caffrey, M., Rosenbusch, J. P. & Landau, E. M. (2001). *FEBS Lett.* **504**, 179–186.
Olson, D. H. (1970). *J. Phys. Chem.* **74**, 2758–2764.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paas, Y., Cartaud, J., Recouvreur, M., Grailhe, R., Dufresne, V., Pebay-Peyroula, E., Landau, E. M. & Changeux, J. P. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 11309–11314.
Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. (1997). *Science*, **277**, 1676–1681.
Punzi, J. S., Luft, J. & Cody, V. (1991). *J. Appl. Cryst.* **24**, 406–408.
Ray, W. J.Jr & Bracker, C. E. (1986). *J. Cryst. Growth*, **76**, 562–576.
Reed, T. B. & Breck, D. W. (1956). *J. Am. Chem. Soc.* **78**, 5972–5977.
Shah, A. K., Liu, Z.-J., Stewart, P. D., Schubot, F. D., Rose, J. P., Newton, M. G. & Wang, B.-C. (2005). *Acta Cryst.* D**61**, 123–129.
Sicheri, F. & Yang, D. S. C. (1995). *Nature* (*London*), **375**, 427–431.
Stevens, R. C. (2000). *Curr. Opin. Struct. Biol.* **10**, 558–563.
Stookey, S. D. (1959). *Ind. Eng. Chem.* **51**, 805–808.
Sugahara, M. & Kunishima, N. (2006). *Acta Cryst.* D**62**, 520–526.
Sugahara, M. & Miyano, M. (2002). *Tanpakushitsu Kakusan Koso*, **47**, 1026–1032.
Sulzenbacher, G., Gal, L., Peneff, C., Fassy, F. & Bourne, Y. (2001). *J.Biol.Chem.* **276**, 11844–11851.
Sulzenbacher, G. *et al.* (2002). *Acta Cryst.* D**58**, 2109–2115.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. (2006). *J. Struct. Funct. Genomics*, **7**, 15–22.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Watanabe, N., Murai, H. & Tanaka, I. (2002). *Acta Cryst.* D**58**, 1527–1530.
Wood, S. P., Janes, R. W., Sweeney, E. & Palmer, R. A. (1992). *J. Cryst. Growth*, **122**, 204–207.
Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, 943–945.
Žutić, V. & Werner, S. (1984). *Geochim. Cosmochim. Acta*, **48**, 1493–1503.