Volume 258, number 2, 297–300 FEB 07899 December 1989

An ice nucleation active gene of *Erwinia ananas*

Sequence similarity to those of *Pseudomonas* species and regions required for ice nucleation activity

Keiko Abe, Satoshi Watabe, Yasufumi Emori*, Michiko Watanabe and Soichi Arai

Department of Agricultural Chemistry, The University of Tokyo, Yayoi-cho and **Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, Honkomagome, Bunkyo-ku, Tokyo 113, Japan*

Received 17 October 1989

The ice nucleation active gene, *ina*A, of *Erwinia ananas* IN-10 has been sequenced This gene encodes a protein composed of 1322 amino acid residues The *ina*A protein contains a 1120-residue segment consisting of 70 repeats of closely related 16 amino acid motifs (R-domain), which is flanked by N- and C-terminal sequences (N- and C-domains, respectively) Its primary structure is similar to, but not identical with, those of *Pseudomonas ina*W and *ina*Z gene products By truncating the *ina*A gene to various extents, it was found that deletion of the C-domain resulted in complete loss of the ice nucleation activity, whereas removal of the N-domain led to a moderate decrease in the activity Complete loss of the activity was also observed when the N-domain plus a large part of the R-domain were deleted It is suggested that the C-domain is required for the assembly of *ina*A protein to form a functional ice nucleus

Ice nucleation gene, *Erwinia ananas*, *ina*A, Repetition sequence, *ina*Z

1. INTRODUCTION

Some bacterial strains belonging to the genera of *Pseudomonas* [1], *Xanthomonas* [2] and *Erwinia* [3] act as strong ice nuclei and promote freezing of supercooled water. The structural genes responsible for the ice nucleation active (ina) phenotype have been isolated from *Pseudomonas syringae* and *P. fluorescence* (called *ina*W and *ina*Z, respectively) and characterized [4,5]. However, *ina* genes from the other genera have not yet been examined in detail. We have recently succeeded in cloning an *ina* gene, termed *ina*A, from *Erwinia ananas* IN-10 [6], which causes frost injury to tea plants and other crops in Japan [7]. Here we report the complete nucleotide sequence of the *ina*A gene and compare the deduced primary structure of *ina*A protein with those of *ina*W and *ina*Z gene products. The results of an attempt to identify the regions of *ina*A protein required for the ice nucleation activity are also reported.

2. MATERIALS AND METHODS

E ananas IN-10 isolated from the gemmisphere of tea plants in the field of the Shizuoka Agricultural Experiment Station, Japan [7] was a kind gift It was cultuered with a Pseudomonas F medium [6] at 22°C for 2 days. *Escherichia coli* MM294 was used as a host for plasmid pUC18 derivatives Total DNA was extracted from grown cells and partially digested with *Sau*3AI The DNA fragments obtained were inserted into a λEMBL3 phage vector to construct a genomic DNA library Isolation of *ina*A of *E. ananas* IN-10 was conducted as described previously [6] A plasmid carrying *ina*A gene, pINA6S13, was used for DNA sequencing analysis and truncation experiments pINA6S13 was digested with appropriate restriction enzymes and subcloned into the multicloning site of the plasmid vector, pUC18 pINA6S13 was also treated with exonuclease III from both ends of the insert, digested with mung bean nuclease, and circularized with T4 DNA ligase by self-ligation reaction The nucleotide sequence was determined for both strands of *ina*A gene by modified dideoxy chain termination method [8]

Ice nucleation activity was measured by gradually cooling to −5°C a 2-ml portion of an aqueous suspension of appropriately diluted *E ananas* cells or *E coli* cells carrying pINA6S13 or its truncates Five independent measurements were carried out, and ice nucleation activity was represented as the cell number necessary to make the bulk water freeze in all probability Ice nucleation activity of *E ananas* IN-10 used as a control was observed at 10^3 cells in 2 ml water under these conditions of measurement

3. RESULTS AND DISCUSSION

The complete nucleotide sequence of the *ina*A gene contains an open reading frame coding for 1322 amino acid residues (M_r = 131 093) (fig.1). The open reading frame is preceded by putative −35 and −10 sequences for transcription initiation and a ribosome-binding (SD) sequence for translation initiation (fig.1). The deduced primary structure of *ina*A protein has a long segment spanning Ala^{162} through Ile^{1281}, in which closely related 16 amino acid motifs are repeated 70

Correspondence address: S Arai, Department of Agricultural Chemistry, University of Tokyo, Yayoi-cho, Tokyo, Japan

Published by Elsevier Science Publishers B V (Biomedical Division)
00145793/89/$3 50

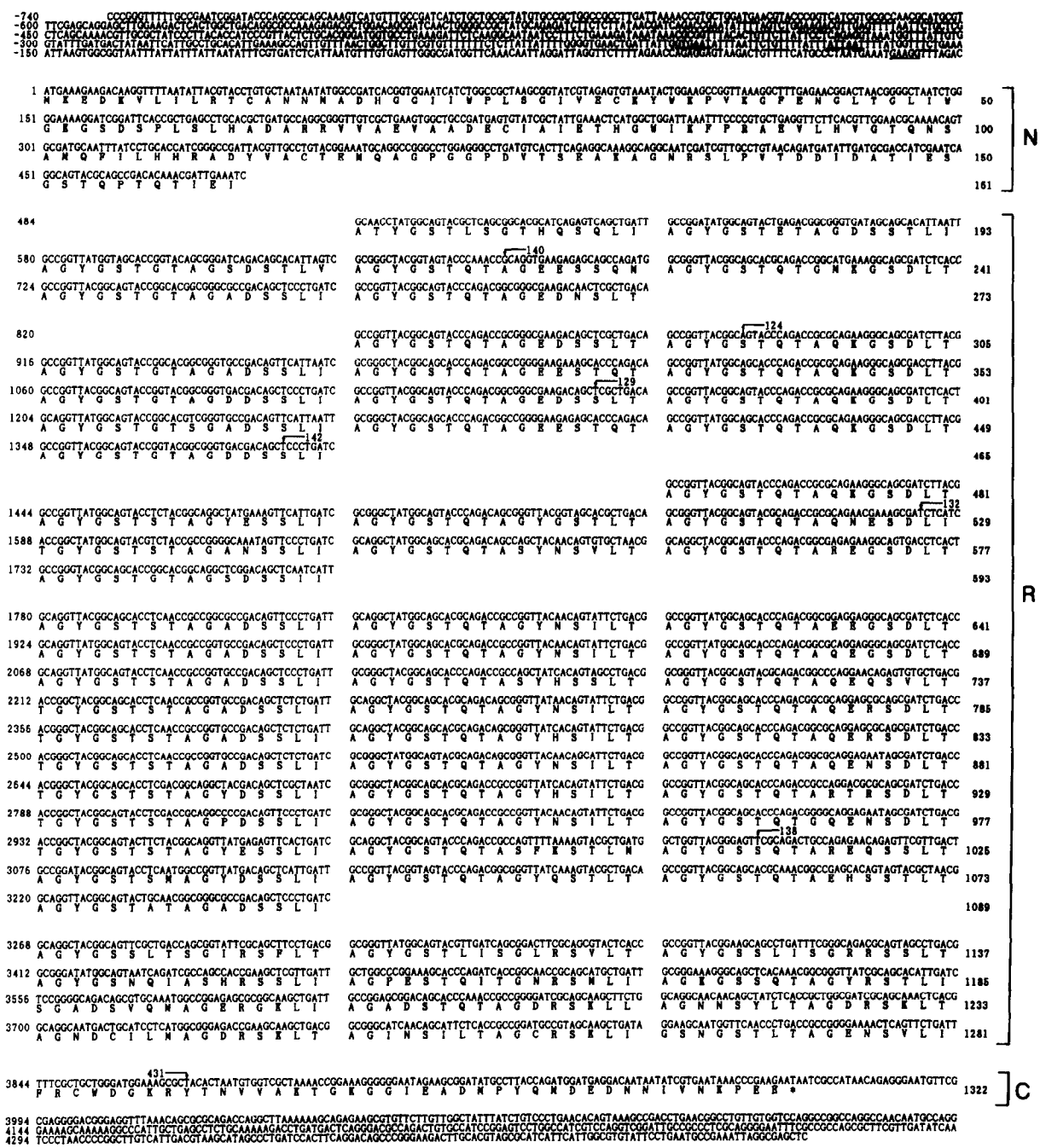

Fig 1 Nucleotide sequence of *ina*A gene and predicted amino acid sequence for *ina*A product Both the nucleotide and amino acid sequences are numbered from the initiation codon The −35 and −10 promoter sequences and a putative SD sequence are underlined As a 16-residue periodicity is present 70 times from residue 162 to residue 1281, the graphic layout is chosen to emphasize it The mutants obtained by truncation from NH_2- or COOH terminal in fig 3 are indicated by arrows in the sequence

times. This repetitive segment, termed R-domain, is flanked by a 161-residue N-terminal sequence (N-domain) and a 41-residue C-terminal sequence (C-domain). From their sequence similarities, the 16 amino acid motifs in the R-domain can be classified into 3 groups, which are separately shown in the left-hand, middle and right-hand columns in fig.1. The occurrence of a closely similar R-domain and the classification of the 16 amino acid motifs into 3 groups have also been reported for *ina*W and *ina*Z proteins [4,5]. The R-domain of *ina*A protein is, however, different from those of *ina*W and *ina*Z proteins in several respects. Thus, in the R-domain of *ina*A protein the motifs are repeated 70 times, whereas only 62 and 60 repeats are found in *ina*W and *ina*Z proteins, respectively. As shown in fig.2, the consensus sequences of the 3 groups of 16 amino acid motifs are also slightly different among the 3 proteins. Both the N- and C-

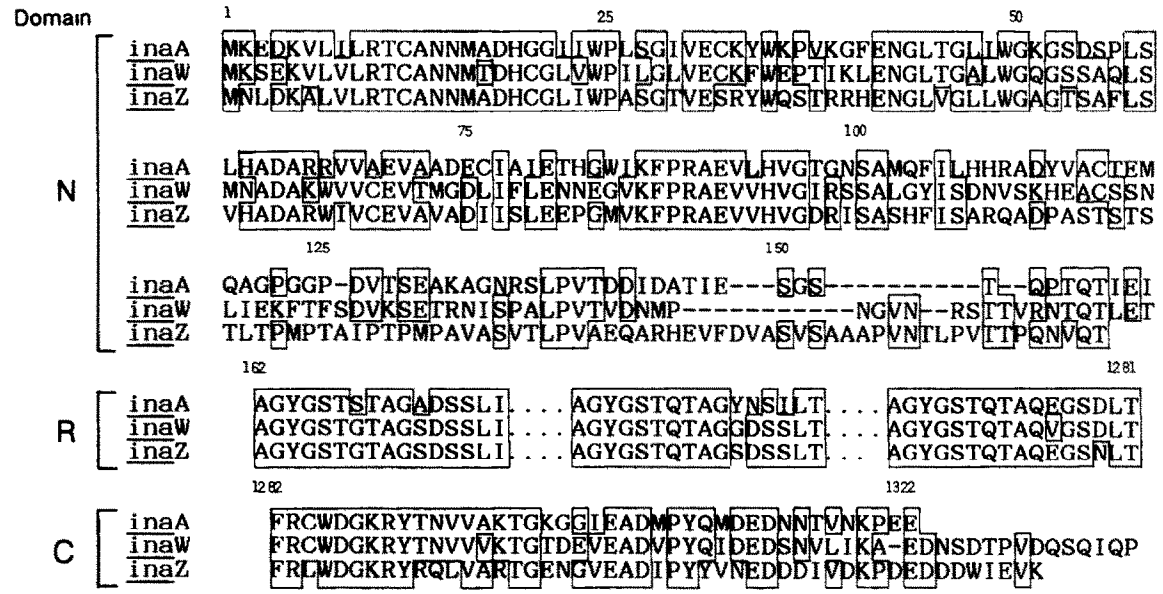

Fig 2 Amino acid sequence homology of *ina*A gene product with *ina*W and *ina*Z gene products Identical amino acids in the sequences are boxed In R-domain 3, only typical amino acid repeat motifs are represented Gaps are introduced to maximize the sequence homology

domains of *ina*A protein exhibit fairly high sequence similarities to the corresponding domains of *ina*W and *ina*Z proteins, but there are definite differences among the 3 proteins (fig.2). Particularly, the C-domain of *ina*A protein significantly shorter than those of *ina*W and *ina*Z proteins. The high sequence similarities among *ina*A, *ina*W and *ina*Z proteins suggest that these 3 genes have been derived from a common ancestor. The sequence data, however, also suggest that the *ina*A gene of *E. ananas* is distantly related to the other two genes of *Pseudomonas* species.

Since these 3 *ina* proteins, though different from one another, all consist of the N-, R- and C-domains, all the 3 domains seem to be important for the ice nucleation activity. To test this possibility, we constructed variously truncated mutants of the *ina*A gene, as schematically illustrated in fig.3. (The precise site of cleavage is shown in fig.1.) The truncated genomic DNAs, inserted into an appropriate expression vector, were introduced into *E. coli*, and the ice nucleation activities of cells harboring these mutants were examined. As can be seen in fig.3, the activity of a mutant in which the N-domain

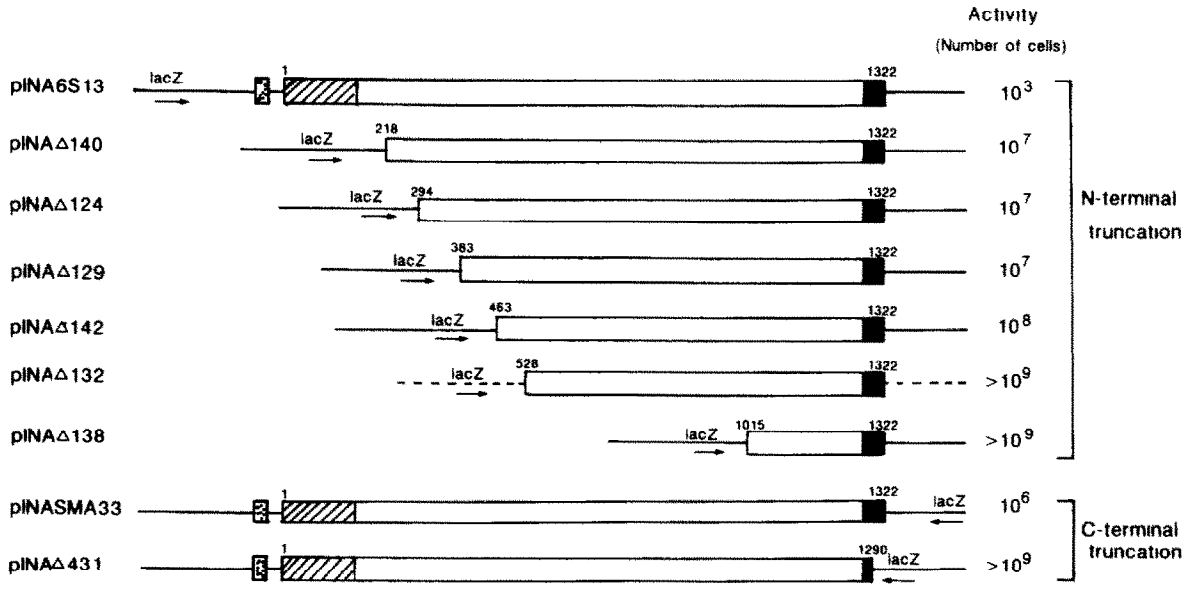

Fig.3 Truncation mutants of the *ina*A gene and their ice nucleation activity. The direction of transcription from the *lacZ'* promoter is shown by arrows Numbers show the positions of amino acids in *ina*A product A plasmid indicated by dashed line is out of frame Columns are as follows ⊠, *ina*A gene promoter, ▨, NH_2-terminal domain; □, repeat domain, ■, COOH-terminal domain Ice nucleation activity was represented by the number of cells necessary to make the bulk water (2 ml) freeze at −5°C

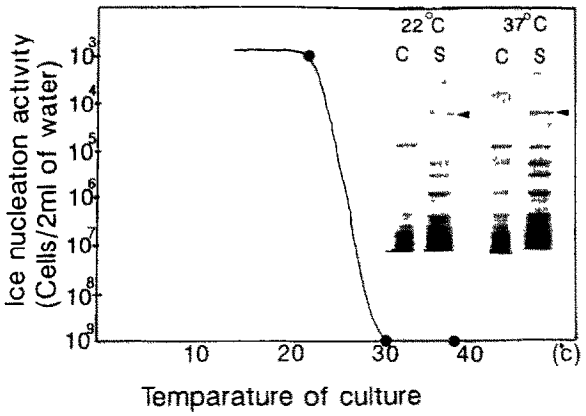

Fig 4 Temperature dependency of ice nucleation activity of *E.coli* carrying pINA6S13 The *E coli* cells were incubated for 48 h at 22, 30 and 37°C SDS-PAGE pictures of the cells carrying p6S13(S) or those carrying pUC18(C) are inserted Arrows show 130 kDa protein as a *ina*A gene product

had been deleted (pINAΔ140) was lower than that of the wild-type by 4 orders of magnitude. Further removal of the R-domain sequence from the N-terminal caused a further decrease in the activity (pINAΔ142), and the activity was completely lost when the N-domain plus a large part of the R-domain were deleted (pINAΔ138). An interesting finding was that deletion of a short (32-residue) C-terminal sequence resulted in complete loss of the activity, even though both the N- and R-domain were intact in this mutant (pINAΔ431). These results indicate that the C-domain is essential for the ice nucleation activity, although the other two domains are also more or less involved in the activity. The indispensability of the C-domain may be explained by assuming its involvement in self-assembly of *ina*A protein or in conjugation of the *ina* protein with other kind of molecules. It has been shown that *ina*W and *ina*Z proteins do not exist as monomers but as polymers or conjugates with other molecules that are precipitable by centrifugation at $100000 \times g$ as observed for *E. ananas* cells [9] and recombinant *E. coli* cells [10]. Moreover, the ice nucleation activity of *E. ananas* has been reported to be temperature-sensitive [11]. In this study, we could confirm the temperature-sensitivity of the ice nucleation activity of *E. coli* cells carrying the *ina*A gene (fig.4), even though the synthesis of *ina*A protein was not temperature-dependent (fig.4, inset). These observations suggest that functional ice nuclei can be formed by self-assembly of *ina*A protein or its conjugation with other molecules that exist universally in bacterial cells, and that this temperature-sensitive process depends on the C-domain of the protein as in the case of self-assembly process of collagen molecules [12].

REFERENCES

[1] Schnell, R C and Vali, C (1972) Nature 236, 163–165
[2] Maki, L R , Galyon, E L , Chang-chien, M and Caldwell, D R (1974) Appl Microbiol 28, 456–460
[3] Lindow, S E , Arny, D C and Upper, C D (1978) Phytopathology 68, 523–527
[4] Warren, G , Corotto, L and Wolber, P (1986) Nucleic Acids Res 14, 8047–8060
[5] Green, R L and Warren, G J (1985) Nature 317, 645–648
[6] Arai, S., Abe, K , Watabe, S , Emori, Y and Watanabe, M (1989) FEMS Lett in press
[7] Goto, M., Huang, B L , Makino, T , Goto, T and Inaba, T (1988) Ann Phytopathol Soc Jap 54, 189–197
[8] Sanger, F , Coulson, A.R , Barrell, B G , Smith, A J H and Roe, B A (1980) J Mol Biol 143, 161–173
[9] Watanabe, M and Arai, S (1987) Agric Biol Chem 51, 557–563
[10] Wolber, P K , Deininger, C A , Southerworth, M W , Vandekerckhove, J , Montagu, M V and Warren, G J (1986) Proc Natl Acad Sci USA 83, 7256–7260
[11] Arai, S and Watanabe, M (1986) Agric Biol Chem 50, 169–175
[12] Superti-Furga, A , Gugler, E , Gitzelmann, R and Steinmann, B (1988) J Biol Chem 263, 6226–6232