A Bacillus cereus member of the SNF2 family

Toril Lindbæk and Anne-Brit Kolstø

Author for correspondence: Anne-Brit Kolstø. Tel: +47 22 95 84 60. Fax: +47 22 69 41 30. e-mail: a.b.kolsto@farmasi.uio.no

Biotechnology Centre of Oslo and School of Pharmacy, University of Oslo, Pb 1125 Blindern, 0317 Oslo, Norway

The complete sequence of a Bacillus cereus member of the SNF2 family of putative helicases showed conservation of all seven motifs typical of this family. Bcsnf predicted a protein of 1064 aa where the conserved SNF2 domain was located at the carboxy terminus, whereas the 633 amino-terminal aa showed no homology to any protein in the databases. A putative transcriptional start was identified by primer extension, indicating that Bcsnf is not a part of a larger operon. No phenotypical changes were observed after insertional inactivation of Bcsnf. The completely sequenced genomes of Mycoplasma genitalium and Haemophilus influenzae contain one ORF each with similarity to the SNF2 family: MG018 and H10616, respectively. A phylogenetic tree of the SNF2 family showed that BcSNF and MG018 were most closely related, and appeared closer to the eukaryotic members of the SNF2 family than to the two other bacterial members of the family, HepA from Escherichia coli and H10616.

Keywords: SNF2 family, Bacillus cereus

We have previously described the sequencing of a 1-9 kb EcoRI fragment from Bacillus cereus ATCC 10987 carrying an ORF which encodes a protein of 635 aa with homology to the SNF2 family of putative helicases (Kolsta et al., 1993). To clone the entire gene, three additional libraries, MspI, AccI and HindIII, were constructed and screened. Preparation, cloning and analysis of DNA was carried out essentially as described by Sambrook et al. (1989). The libraries were constructed in pUC18 and pUC19 using Escherichia coli BK2118 (Evensen & Seeberg, 1982) as host. Three overlapping clones that contained 3-9 kb upstream and 1-35 kb downstream of the original EcoRI fragment were isolated and sequenced (Fig. 1). Sequence analysis was performed with the GCG sequence analysis programs (Devereux et al., 1984). Translation of the nucleotide sequence revealed two ORFs: orf-1 of 981 nt (327 aa) and Bcsnf of 3207 nt (1069 aa). Bcsnf has previously been mapped to the 840 kb NotI fragment of the NotI physical map of B. cereus ATCC 10987 (Carlson et al., 1992). The gene is located on an internal Ascl fragment, E', of 250 kb on this NotI fragment (Carlson & Kolstø, 1994).

Sequence analysis of BcSNF

The deduced amino acid sequences were compared to protein sequences from SWISS-PROT release 32.0 (12/95), by fasta searches (Pearson & Lipman, 1988).

**Abbreviations:** SNF, sucrose non-fermenting; BcSNF, Bacillus cereus SNF2.

The EMBL accession numbers for the sequence reported in this paper are X98455 (B. cereus ATCC 10987) and X98626 (B. cereus ATCC 14579).

**Fig. 1.** Restriction map of the sequenced region of 5-2 kb with two ORFs, orf-1 and Bcsnf. Bc203, Bc203.1, Bc203.2 and Bc203.4 are individual clones identified by screening of genomic libraries. The SNF2 carboxy-terminal domain of BcSNF and the Bc203 homologue from B. cereus ATCC 14579, TS203 are indicated.
BcSNF showed significant homology to the SNF2 family of proteins involved in transcriptional control, DNA repair or maintenance of chromosome stability. Proteins sharing amino acid sequence homology with the yeast transcriptional activator SNF2 have been grouped into a protein family as a part of the DEAD/H superfamily of helicases (Gorbalenya & Koonin, 1993; Bork & Koonin, 1993). These proteins appear to contain motifs characteristic of RNA- and DNA-dependent helicases and other nucleic-acid-dependent ATPases (Gorbalenya et al., 1989), and are therefore suggested to possess a helicase function. All seven motifs conserved within the family are present in the amino acid sequence of BcSNF. The overall identity to the SNF2 family is 25–30% for the 431 carboxy-terminal aa, while the sequence identity within the conserved motifs and additional regions was between 20–82%.

The number of proteins assigned to the SNF2 family has increased rapidly over the last few years, but HepA from E. coli has been the only completely sequenced bacterial gene encoding a member of the family (Lewis et al., 1992). A phylogenetic tree of the SNF2-related proteins was recently constructed by Eisen et al. (1995), where they suggest that the proteins of this family may be divided into seven subfamilies with distinct sequences and function; HepA was determined to be the deepest-branching member of the SNF2 family.

When searching the Mycoplasma genitalium genome database (Fraser et al., 1995), an SNF2-related sequence of 440 aa, MG018, showed 34% identity to BcSNF in the SNF2 domain. MG018 was later extended at the amino terminus by 126 aa (Venter, 1996) that show no significant homology to BcSNF or any protein in the databases. The SNF2-like sequences of B. cereus and M. genitalium were included in the phylogenetic tree constructed by Eisen et al. (1995) (Fig. 2). The two sequences clustered as a novel subfamily more closely related to the other eukaryotic members of the SNF2 family than to HepA from E. coli which grouped together with an SNF2-related sequence, HI0616, from the Haemophilus influenzae database (Fleischmann et al., 1995). The homology between BcSNF and MG018 is
Bacterial member of the SNF2 family

limited to the SNF2 domain, while HepA and the H. influenzae sequence are conserved throughout their length.

Primer-extension analysis of the transcript of Bcsnf was performed on total RNA isolated from B. cereus during exponential growth, using an oligonucleotide complementary to the DNA sequence at position 1599–1617. The likely transcriptional start site was identified as guanine position 1242, located 229 bp upstream of the putative translation start site, indicating that TAGAAA (G in position -10) is the putative RNA-polymerase-binding site. This indicates a -10 box that shows some homology to the consensus sequence of the Bacillus subtilis $\alpha$-binding site, TATAAT (Moran, 1993). However, no sequence homologous to the consensus sequence of the RNA polymerase recognition site (TTGACA) is found in the -35 region.

Conservation of Bcsnf among B. cereus strains

We have cloned the Bc203 homologue TS203 (Fig. 1) from the type strain B. cereus ATCC 14579. This EcoRI fragment was 1616 bp, where the upstream EcoRI site was conserved between ATCC strains 10987 and 14579. The identity between the EcoRI fragments from the two strains was 90% for the nucleotide sequences and 93% for the amino acid sequences.

Chromosomal DNAs from five additional B. cereus strains and one Bacillus thuringiensis strain were digested with EcoRI, HincII and HindIII. A Southern blot of the DNA separated on agarose was hybridized with Bc203 (Fig. 1) as a probe. This shows the presence of Bcsnf in all the tested strains and a 3.6 kb HindII fragment was conserved in all seven strains. The size of the HindIII fragment that hybridized to Bc203 was identical in four strains and the EcoRI fragment in three of the strains.

Insertional inactivation of Bcsnf

To gain information about the function of the B. cereus member of the SNF2 family we wanted to do an insertional inactivation of the gene. Protoplast transformation or electroporation of strain ATCC 10987 with the commonly used pC194 gave a very low efficiency, even when pC194 was isolated from B. subtilis 168 rather than from E. coli. However, successful conjugation between E. coli and Bacillus anthracis has been reported (Pezard et al., 1991; Trieu-Cuot et al., 1991; Sirard et al., 1994). The conjugation turned out to be successful only in the type strain B. cereus ATCC 14579. The Bc203 homologue from B. cereus ATCC 14579, TS203 (Fig. 1), where the XcmI-ClaI fragment was replaced by a spectinomycin-resistance cassette, was cloned into the conjugative-suicide vector pAT113, and used in transconjugal transfer between E. coli JM83(pRK24) and B. cereus. Hybridization verified that the chromosomal XcmI-ClaI fragment had been replaced by the spectinomycin-resistance cassette. The mutant did not show any changes in growth rate, UV-resistance or spore heat-resistance compared to the wild-type.

Sequence analysis of ORF-1

A second ORF, orf-1, lies 105 bp upstream of Bcsnf. The amino acid sequence of ORF-1 showed 53% identity to a hypothetical sensory transduction protein of B. subtilis, YCBB, which is implicated in resistance to protonophore compounds (Quirk et al., 1994).

The amino-terminal part of the two Bacillus proteins showed about 30% identity to a group of prokaryotic response regulator proteins (reviewed by Stock et al., 1989), including SpoOA from B. subtilis, B. thuringiensis and B. anthracis. ORF-1 possesses the two conserved aspartates and the conserved lysine characteristic of the response regulator family. The homology between the two Bacillus proteins ORF-1 and YCBB, and the other members of the response regulator family is limited to about 130 amino-terminal aa. The similarity between ORF-1 and SpoOA is within the amino-terminal region, but it is interesting to observe that a region of six aa that is almost perfectly conserved is suggested to be part of the second helix of a 20 aa region predicted to be the helix-turn-helix DNA-binding motif of SpoOA (Brown et al., 1994).

A putative translational start site of orf-1 is located at DNA sequence position 421, TTG (Leu). No ATG in the B. cereus sequence corresponds to the translational start site of YCBB or SpoOA, but 30% of B. subtilis proteins are known to be initiated by the start codon UUG (Leu) or GUG (Val) instead of AUG (Met) (Hager & Rabinowitz, 1985).

ACKNOWLEDGEMENTS

This work was supported by grant from the Norwegian Research Council to A.-B.K. We thank Agnes Fouet, Institut Pasteur, Paris, for helpful suggestions and for reading the manuscript, and Jonathan Eisen, Stanford University, CA, USA, for generating the phylogenetic tree.

REFERENCES


