Volume 224, number 2, 257-260 November 1987

Hypothesis

Two separable functional domains in the σ-subunit of RNA polymerase in *Bacillus subtilis*?

J. Errington

Microbiology Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, England

Received 31 July 1987

The σ-subunit of RNA polymerase is responsible for promoter recognition in prokaryotes [(1969) Nature 221, 43-46]. Alterations in the σ-subunit are thought to be involved in controlling 'global' changes in gene expression, such as those involved in differentiation in the spore-forming bacterium *Bacillus subtilis* [(1981) Cell 25, 582–584]. Stragier et al. [(1985) FEBS Lett. 195, 3–11] have proposed that σ-factors are composed of two domains: a C-terminal domain involved in promoter recognition and an N-terminal domain involved in interactions with RNA polymerase. We have sequenced another developmental gene from *B. subtilis*, *spoIIIC*, and the strong homology of its predicted product suggests that it too may be a σ-factor. However, the *spoIIIC* product is small and lacks completely the conserved N-terminal domain of the σ-subunits. I propose that the product of the *spoIIIC* gene may carry out the DNA-recognition functions of a σ-factor but that it probably requires an auxiliary factor to interact with core RNA polymerase.

RNA polymerase; Sigma factor; Sporulation; (*Bacillus subtilis*)

1. INTRODUCTION

A single type of RNA polymerase is responsible for transcription in prokaryotes. The catalytic core enzyme is composed of four large subunits: two α-subunits ($M_r \sim 40\,000$), β- ($M_r \sim 155\,000$) and β'- ($M_r \sim 160\,000$). Initiation of transcription is controlled by a fifth subunit, σ, which is not required for elongation and which dissociates from the complex soon after initiation. In *Bacillus subtilis*, in addition to the major form of σ found in vegetative cells (σ_{43}), a number of minor forms have been identified by their stimulatory effect on transcription from specific promoter sequences in vitro [4–7]. Others have been tentatively identified on the basis of strong homologies between their predicted amino acid sequences and those of known σ-subunits [8]. Although *B. subtilis* has by far the widest range of known σ forms at the moment, multiple forms of σ have been identified in other prokaryotes, notably *Escherichia coli* and *Klebsiella*, where they have been shown to control genes involved in the heat-shock response and in nitrogen regulation [9]. The major form of σ is highly conserved in *B. subtilis* and *E. coli* [10]. On the basis of a comparison of the sequences of these and related σ-factors, Stragier et al. [3] have recently proposed that there are two conserved functional domains in σ-factors. The C-terminal conserved domain contains the consensus α-helix-turn-α-helix motif, which is characteristic of a range of prokaryotic proteins capable of site-specific DNA-binding [11]. The presence of this structure prompted Stragier et al. [3] to propose

Correspondence address: J. Errington, Microbiology Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, England

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

that the C-terminal domain of the σ-factors is involved in promoter recognition. By elimination, Stragier et al. [3] also proposed that the conserved N-terminal domain might be involved in interactions with core RNA polymerase. More recently, a region of σ_{30} involved in recognition of the '−10' region of the *spoVG* promoter has been identified by the location of an amino acid alteration that suppresses a specific promoter-down mutation (Losick, R., personal communication). The conserved α-helix-turn-α-helix motif near the C-terminus of the σ-factors is thus now thought to be involved specifically in the recognition of '−35' promoter sequences.

It seems likely that development in *B. subtilis* is controlled, at least in part, by changes in the σ-subunit of RNA polymerase [2]. Many of the regulatory genes involved in this process have been identified by the sites of mutations (*spo*) that have a highly pleiotropic effect, often causing a complete block in development [12]. However, only three of the 23 or so *spo* genes that have been sequenced so far appear to encode σ-factors, on the basis of either sequence homology or function [13]. As part of a continuing effort to identify the functions and regulation of the *spo* genes, we have recently characterised, in terms of expression and nucleotide sequence, the *spoIIIC* gene ([14], Errington, J., Rong, S., Rosenkrantz, M.S. and Sonenshein, A.L., submitted). In this report I discuss the relationship between the predicted product of the *spoIIIC* gene and the family of σ-factors and its implications for σ-function.

2. RESULTS AND DISCUSSION

The predicted product of the *spoIIIC* gene (Errington, J., Rong, S., Rosenkrantz, M.S. and Sonenshein, A.L., submitted) was compared with the NIH Protein Sequence Database and a single

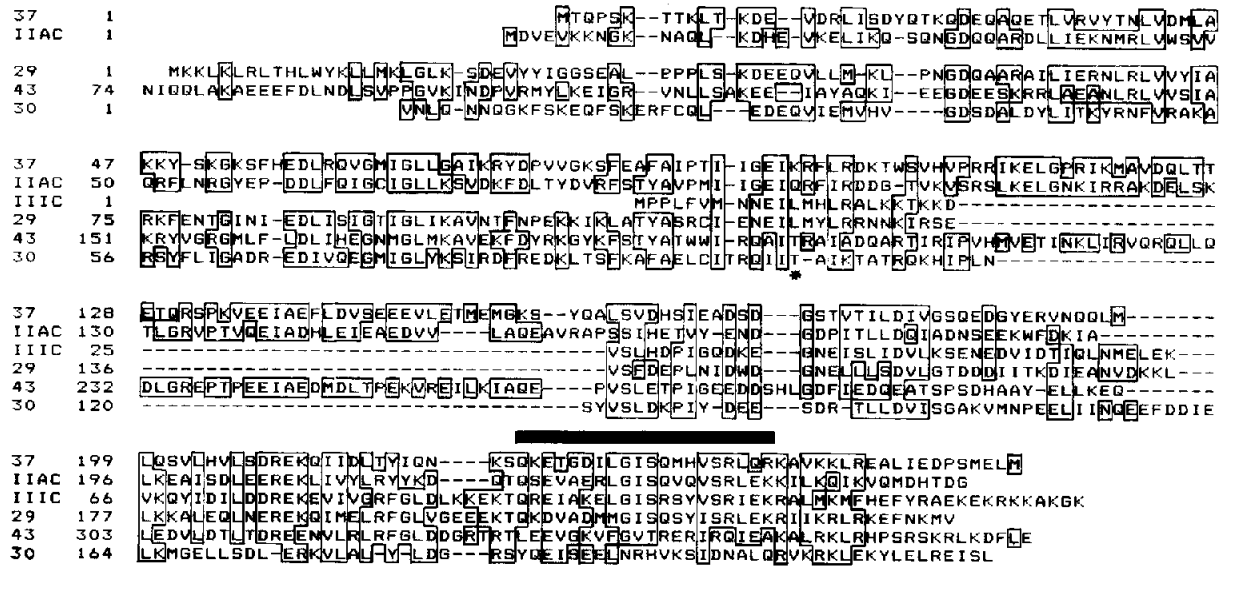

Fig.1. Alignment of six σ-like proteins from *B. subtilis*. The standard one-letter amino acid notation is used. Numbers at the beginning of each line refer to the first amino acid residue in that line. Gaps, indicated by hyphens, were introduced to maximise the alignment. A solid bar overlies the conserved α-helix-turn-α-helix motif characteristic of many prokaryotic DNA-binding proteins and thought to be involved in recognition of the −35 regions of promoters. The asterisk indicates the location of the threonine to isoleucine alteration that may indicate the location of a −10 promoter specificity determining region in σ_{30} (Losick, R., personal communication). Vertical lines of amino acids are boxed when the majority of members in a line are identical or represent conservative substitutions. Conservative substitutions are defined as those involving pairs of amino acids from the following groups: A and G; D and E; F and Y; I, L, M and V; K and R; N and Q; S and T. Sequence information was taken from the following references [8,10,19–21]. The σ_{30} sequence is that of *B. licheniformis* and was kindly provided by Dr I. Smith.

significant homology was obtained to the major form of σ-factor from *E. coli*. The major form of σ found in vegetative cells of *B. subtilis* and the other minor forms of σ in this organism were not included in the database but examination of their sequences indicated that the *spoIIIC* product was also homologous to them. A preliminary alignment was obtained by analysis of each pair of protein sequences in turn by a dot-matrix computer programme, DIAGON [15]. The programme ALIGN (the alignment score programme of the Protein Identification Resource) was then used to compare each pair of sequences with regard to the significance of the homology and to determine the introduction of gaps. The *spoIIIC* product showed significant homology to all of the sequences. For example, the mean alignment score for 100 random sequences of the same length and composition as the products of *spoIIIC* (all residues) and *spoIIG* (residues 113–239) was 816.04, with a standard deviation (SD) of 15.58. The score for the aligned proteins was 1019, 13.03 SDs above the mean score, indicating a highly significant similarity between the sequences ($P<10^{-6}$). It is clear from the alignment shown in fig.1 that the *spoIIIC* product is very closely related to the family of genes. It aligns particularly well with the σ_{29} sequence throughout its length, except for 6 residues at the N-terminus and 12 at the C-terminus. It was not necessary to introduce any gaps between residues in order to align this pair of sequences. We have previously commented upon the extensive gene duplication that is evident now that a number of developmental genes in *B. subtilis* have been characterised [8]. The structure of *spoIIIC* is highly suggestive of an origin by partial gene duplication, and *spoIIG* (encoding σ_{29}) seems the most likely candidate for an immediately ancestral gene. The *spoIIIC* sequence is, however, unusual in that it is much shorter than the other σ-subunits; of the two putative domains distinguished in this family of proteins (see above), it contains only the C-terminal segment. Nevertheless, the *spoIIIC* product must be functional, as mutations in it have very strong pleiotropic effects on sporulation [12]. It thus seems likely that the putative C-terminal region of the σ-subunits does represent a separate domain and it is presumably responsible for promoter recognition by site-specific DNA-binding. Note that the threonine to isoleucine alteration that confers an altered -10 promoter specificity on σ_{30} (see above and fig.1) lies within the region of homology to the predicted *spoIIIC* product, though rather near its N-terminus. The fact that the predicted product shows strong homology to the other σ-subunits throughout its length suggests that its function is specifically related to that of the σ-factors, rather than to the general class of DNA-binding proteins: activators and repressors.

One way to avoid the problem of core polymerase interaction by the rather truncated *spoIIIC* product, if that is the function of the missing N-terminal domain, would be to suppose that it interacts with a second protein capable of mediating this function. Support for the concept of separating the host σ-factors into two components is provided by bacteriophage SPO1 of *B. subtilis*, in which the 'late' genes are controlled by the products of two genes, *33* and *34*, that act synergistically with σ-like activity [16]. However, neither of these products resembles that of *spoIIIC* in overall structure.

On the assumption that the proposed model is correct, which other sporulation gene products might interact with that of *spoIIIC*? Although none of the sporulation genes that have been sequenced appear to have a suitable structure, mutations in several distinct developmental genes, *spoIVC*, *spoIVD* and *spoIVE*, give phenotypes that are indistinguishable from those produced by *spoIIIC* mutations [12], as might be expected if one or more of their gene products were to act in concert with that of *spoIIIC*. Clarification of the true relationships between the products of these genes must, however, await their characterisation in molecular terms but at least one of them, *spoIVC*, has been cloned by several groups [17,18]. The separation of σ-polypeptides into two domains capable of acting independently would further complicate the problem of developmental gene regulation in *B. subtilis* and may have important implications for studies of gene regulation in other prokaryotes.

ACKNOWLEDGEMENTS

I thank Drs R. Losick and I. Smith for communicating their results prior to publication, and Professor J. Mandelstam, Dr P. Butler, Dr M.D. Yudkin and Dr A.L. Sonenshein for helpful com-

ments and criticisms. I am grateful to the Royal Society for a Research Fellowship and to the SERC for funding.

REFERENCES

[1] Burgess, R.R., Travers, A.A., Dunn, J.J. and Bautz, E.K.F. (1969) Nature 221, 43–46.
[2] Losick, R. and Pero, J. (1981) Cell 25, 582–584.
[3] Stragier, P., Parsot, C. and Bouvier, J. (1986) FEBS Lett. 195, 9–11.
[4] Haldenwang, W.G. and Losick, R. (1980) Proc. Natl. Acad. Sci. USA 77, 7000–7004.
[5] Wiggs, J.L., Gillman, M.Z. and Chamberlin, M.J. (1981) Proc. Natl. Acad. Sci. USA 78, 2762–2766.
[6] Haldenwang, W.G., Lang, N. and Losick, R. (1981) Cell 23, 615–624.
[7] Carter, H.L.iii and Moran, C.P.jr (1986) Proc. Natl. Acad. Sci. USA 83, 9438–9442.
[8] Errington, J., Fort, P. and Mandelstam, J. (1985) FEBS Lett. 188, 184–188.
[9] Doi, R.H. and Wang, L.-F. (1986) Microb. Rev. 50, 227–243.
[10] Gitt, M.A., Wang, L.-F. and Doi, R.H. (1985) J. Biol. Chem. 260, 7178–7185.
[11] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–323.
[12] Piggot, P.J. and Coote, J.G. (1976) Bact. Rev. 40, 908–962.
[13] Mandelstam, J. and Errington, J. (1987) Microb. Sci. 4, 238–244.
[14] Turner, S.M., Errington, J. and Mandelstam, J. (1986) J. Gen. Microbiol. 132, 2995–3003.
[15] Staden, R. (1982) Nucleic Acids Res. 10, 2951–2961.
[16] Tjian, R. and Pero, J. (1976) Nature 262, 753–757.
[17] Fujita, M. and Kobayashi, Y. (1985) Mol. Gen. Genet. 199, 471–475.
[18] Errington, J. and Jones, D. (1987) J. Gen. Microbiol. 133, 493–502.
[19] Binnie, C., Lampe, M. and Losick, R. (1986) Proc. Natl. Acad. Sci. USA 83, 5943–5947.
[20] Fort, P. and Piggot, P.J. (1984) J. Gen. Microbiol. 130, 2147–2153.
[21] Stragier, P., Bouvier, J., Bonamy, C. and Szulmajster, J. (1984) Nature 312, 376–378.