Commentary

Comment on 'Duplicated sporulation genes in bacteria' by J. Errington, P. Fort and J. Mandelstam (FEBS Letters 188 (1985) 184–188)

Patrick Stragier

Institut de Microbiologie, Bât 409, Université Paris-Sud, 91405 Orsay Cedex, France

Received 14 October 1985

We recently presented a model for the structural organization of bacterial sigma factors based on a thorough comparison of available sequences [1]. The major sigma factor of *Escherichia coli*, σ^{70}, its *Bacillus subtilis* vegetative counterpart, σ^{43}, the *E. coli* heat shock regulatory protein encoded by the *htp*R gene and the product of the *B. subtilis* sporulation-essential *spo*IIG gene, share common characteristics: a highly conserved internal domain proposed to interact with core RNA polymerase and a carboxy-terminal region postulated to be involved in promoter recognition. In their recent article, Errington et al. [2] identify in the previously published sequence of the *B. subtilis spo*IIAC gene [3] a region exhibiting homology with known sigma factors. This region corresponds to the internal domain conserved in bacterial sigma factors [1] which was also the first one to be noticed in the SpoIIG sequence [4]. However, no homology can be found in the carboxy-terminal region of the *spo*IIAC product with the postulated DNA-binding domain of the other sigma factors, a serious discrepancy with our model.

Therefore we undertook a careful examination of the published *spo*IIAC sequence [3]. This study revealed that shifting the reading frame in the distal part of the *spo*IIAC sequence by addition of one extra nucleotide at two different locations extended the *spo*IIAC coding sequence by 60 amino acids and restored the missing homologous domain. In fact the homology created with the carboxy-terminal part of the *spo*IIG product was so striking that it could barely be the result of chance. These proposed alterations of the original *spo*IIAC sequence have now been experimentally confirmed (see Note Added in Proof in [2]). The corrected SpoIIAC sequence is shown in fig.1, aligned with the 4 previously studied sigma factors.

Addition of a new sequence introduces minor modifications to our previous alignment [1] as it allows a better estimation of the homologous domains: a highly conserved 65-amino-acid internal region (called A) flanked by a more variable 33-amino-acid domain (called B) and a 59-amino-acid carboxy-terminal region (called C) containing the characteristic α-helix–β-turn–α-helix structure of DNA-binding proteins. The alignment of the less conserved regions found between domains A-B and C has also been determined. It contains a second, albeit less typical, DNA-binding structure that is also present in the SpoIIAC sequence.

The corrected SpoIIAC sequence corroborates perfectly the hypothesis presented in the paper by Errington et al. [2]: the new version is much closer to the SpoIIG sequence and makes it very likely that the genes encoding these two proteins have arisen by duplication of an ancestral gene, itself evolutionarily related to the *rpo*D gene, encoding the major vegetative sigma factor, σ^{43}. On the other hand, the SpoIIAC sequence fits with the general scheme proposed in our original article [1], thus strengthening our hypothesis. However, we

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

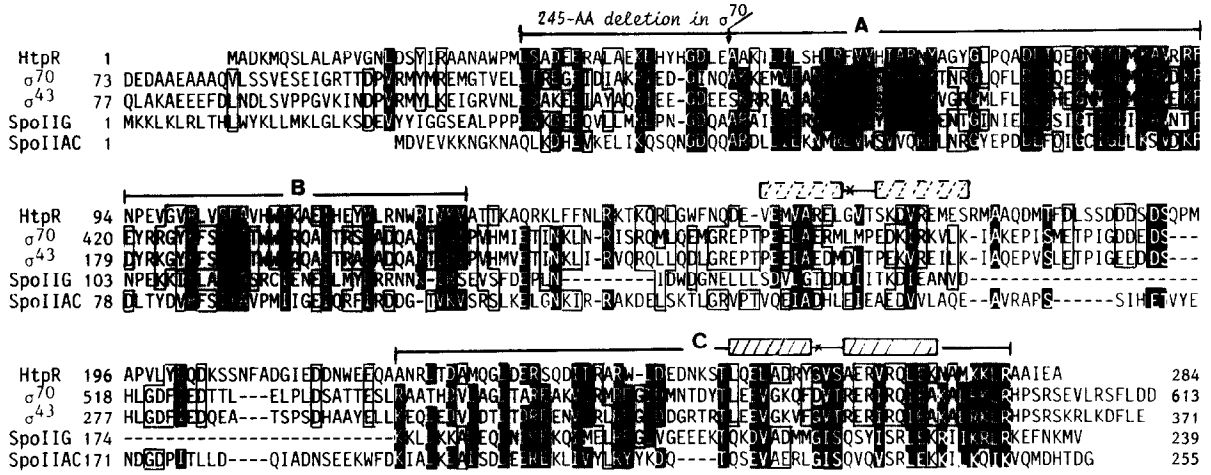

Fig.1. Alignment of the amino acid sequences of the *E. coli* HtpR and σ^{70} proteins with the *B. subtilis* σ^{43}, SpoIIG and SpoIIAC proteins. A 245-amino-acid deletion corresponding to residues 130–374 has been introduced in the σ^{70} sequence at the position shown by the arrow. The one-letter amino acid notation is used. Chemically similar amino acids are defined as the following groups: D and E; K and R; S and T; F and Y; I, L and V. Gaps have been introduced to maximize the alignments and are shown by hyphens. White letters on a black background indicate positions occupied by identical or chemically similar amino acids in 4 or 5 proteins. Positions at which 3 residues are identical or similar are boxed. The homologous regions discussed in the text are indicated by brackets (A, B and C). The locations of potential α-helix–β-turn–α-helix structures typical of DNA-binding proteins are schematized.

suggested that interaction with the core RNA polymerase (mediated by region A) could be modulated by additional association with specific amino acids present in region B. In 9 positions, pairs of identical or chemically similar amino acids could be related either to the major sigma factors, σ^{70} and σ^{43}, or to their alternate competitors, HtpR and SpoIIG. The SpoIIAC product is related to the minor sigma factors in only 2 of these positions and to the major sigma factors in 3 of these positions which precludes any extension of our 'displacement model'. Obviously, more biochemical and genetic data are needed to explain how alternate sigma factors can displace their cognate major sigma factors from core RNA polymerase.

The structural organization conserved in the *spo*IIAC product strongly suggests that it is an authentic sigma factor. According to its predicted molecular mass, 29372 Da, it could be identical to minor sigma factors previously described in *B. subtilis*, σ^{28}, σ^{29} or σ^{32}. Transcription of the *spo*IIA operon starts about 1 h after the initiation of sporulation and is greatly diminished in a *spo*OA mutant [5]. In contrast, σ^{28} and σ^{32} are present during vegetative growth and can readily be purified from *spo*OA cells [6,7], making it very unlikely that they could be encoded by the *spo*IIAC gene. σ^{29} is a sporulation-specific sigma factor and is absent from *spo*OA bacteria [8], thus appearing to be the best candidate for the *spo*IIAC product. From the same arguments it was previously proposed that the *spo*IIG gene could encode σ^{29} [1]. It has actually been shown that the *spo*IIG product is recognized by a monoclonal antibody directed against σ^{29} and that the *spo*IIG41 mutation leads to the synthesis of a truncated form of immunoreacting σ^{29} [9]. Taken together these data suggest that the *spo*IIAC gene should encode an as yet undescribed minor sigma factor. Alternatively, the so-called σ^{29} factor could be a mixture of two polypeptides, the products of the *spo*IIG and *spo*IIAC genes, with similar biochemical properties but distinct promoter recognition specificities. In any case, the existence of two different sporulation sigma factors reinforces the 'cascade model' of Losick and Pero [10]. Elucidation of their specific role should help in understanding of the regulatory network involved in the early stages of sporulation.

REFERENCES

[1] Stragier, P., Parsot, C. and Bouvier, J. (1985) FEBS Lett. 187, 11–15.
[2] Errington, J., Fort, P. and Mandelstam, J. (1985) FEBS Lett. 188, 184–188.
[3] Fort, P. and Piggot, P.J. (1984) J. Gen. Microbiol. 130, 2147–2153.
[4] Stragier, P., Bouvier, J., Bonamy, C. and Szulmajster, J. (1984) Nature 312, 376–378.
[5] Savva, D. and Mandelstam, J. (1985) in: The Molecular Biology of Microbial Differentiation (Setlow, P. and Hoch, J. eds) pp.55–59, American Society for Microbiology, Washington, DC.
[6] Gilman, M.Z. and Chamberlin, M.J. (1983) Cell 35, 285–293.
[7] Johnson, W.C., Moran, C.P. and Losick, R. (1983) Nature 302, 800–804.
[8] Haldenwang, W.G., Lang, N. and Losick, R. (1981) Cell 23, 615–624.
[9] Trempy, J.E., Bonamy, C., Szulmajster, J. and Haldenwang, W.G. (1985) Proc. Natl. Acad. Sci. USA 82, 4189–4192.
[10] Losick, R. and Pero, J. (1981) Cell 25, 582–584.