

TERTIARY STRUCTURE OF PROTEINS INVOLVED IN THE BACILLUS SUBTILIS CELL DIFFERENTIATION PROCESS

Imrich Barák*, Katarína Muchová, Patrik Florek and Zuzana Chromiková

*Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, 845 51 Bratislava 45, Slovakia; phone: ++421 2 5930 7418, fax: ++421 2 5930 7416, e-mail: imrich.barak@savba.sk, * Corresponding author*

Keywords:

crystallography, protein structure, sporulation, *Bacillus subtilis*

Abstract

The number of proteins with functions restricted to sporulation in *Bacillus subtilis* is proposed to be about 200. There is an enormous amount of genetic, biochemical and molecular biology data about this simplest cell differentiation process. However, at the present, the number of known structures of sporulation specific proteins is about ten. This article gives a short overview about recent progress in crystallization of key regulators of sporulation.

Bacillus subtilis as an example of primary cellular differentiation

Bacillus subtilis is a model organism for the study of one of the simplest cell differentiation processes, called sporulation. Thus this nonpathogenic soil bacterium is of great importance for basic research in general and most likely it will be the first organism in which the cell differentiation process will be understood at the molecular level. The study of this process has advanced to solving the structure of key protein regulators. This review focuses on the progress in protein crystallography oriented toward the understanding of sporulation mechanisms, that have been achieved in the last couple of years. The life cycle of *B. subtilis* comprises of three different processes: vegetative growth, sporulation and germination - the spore outgrowth (Fig. 1A). During vegetative growth, the cell grows and divides in the middle, giving rise to identical progeny.

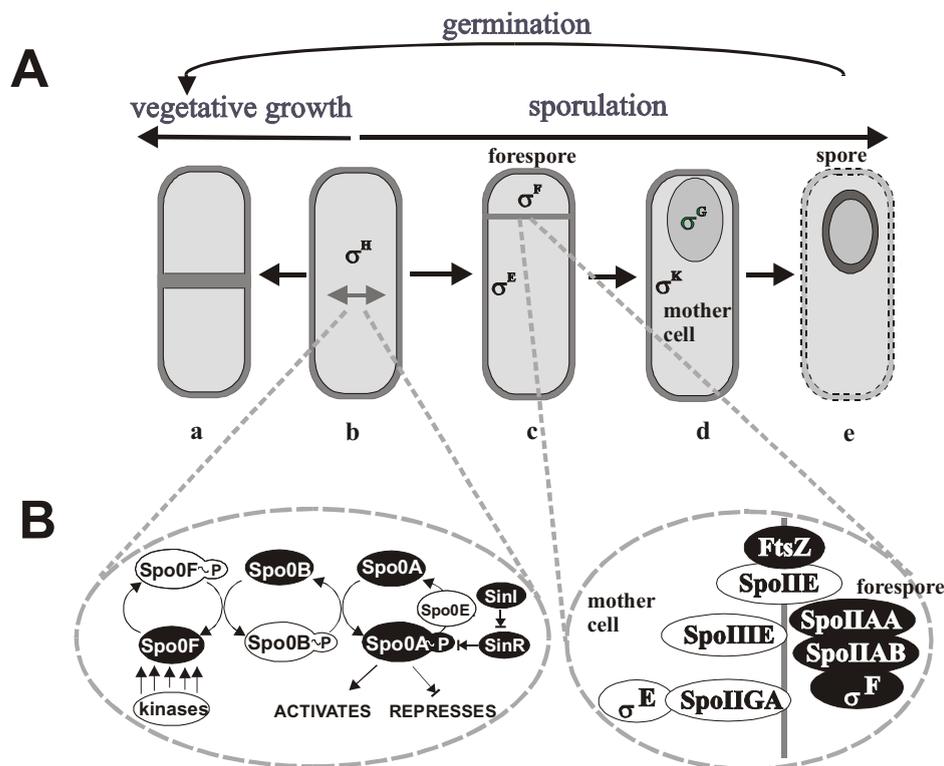


Fig. 1A. Life cycle of *Bacillus subtilis*: (a) In nutritional rich medium *B. subtilis* grows by longitudinal growth and divides in the middle. This process is called vegetative growth. (b) In the case of nutritional deprivation, *B. subtilis* might enter the sporulation process. (c) The first clear morphological step in this process is the formation of an asymmetric septum. (d-e) The small forespore develops into a very resistant spore and the mother cell lyses. The spore can germinate when living conditions improve. The process of sporulation is regulated by the depicted five sigma factors (σ^H , σ^F , σ^E , σ^G and σ^K), the last four of which are compartment specific, activated either in the forespore or in the mother cell.

Fig. 1B. The proteins involved in the two most important steps of sporulation - the initiation of sporulation (first dashed oval) and activation of σ^F and σ^E (second dashed oval). The proteins with solved tertiary structure are shown in filled ovals and some other protein structures that would be most important to know are in empty ovals.

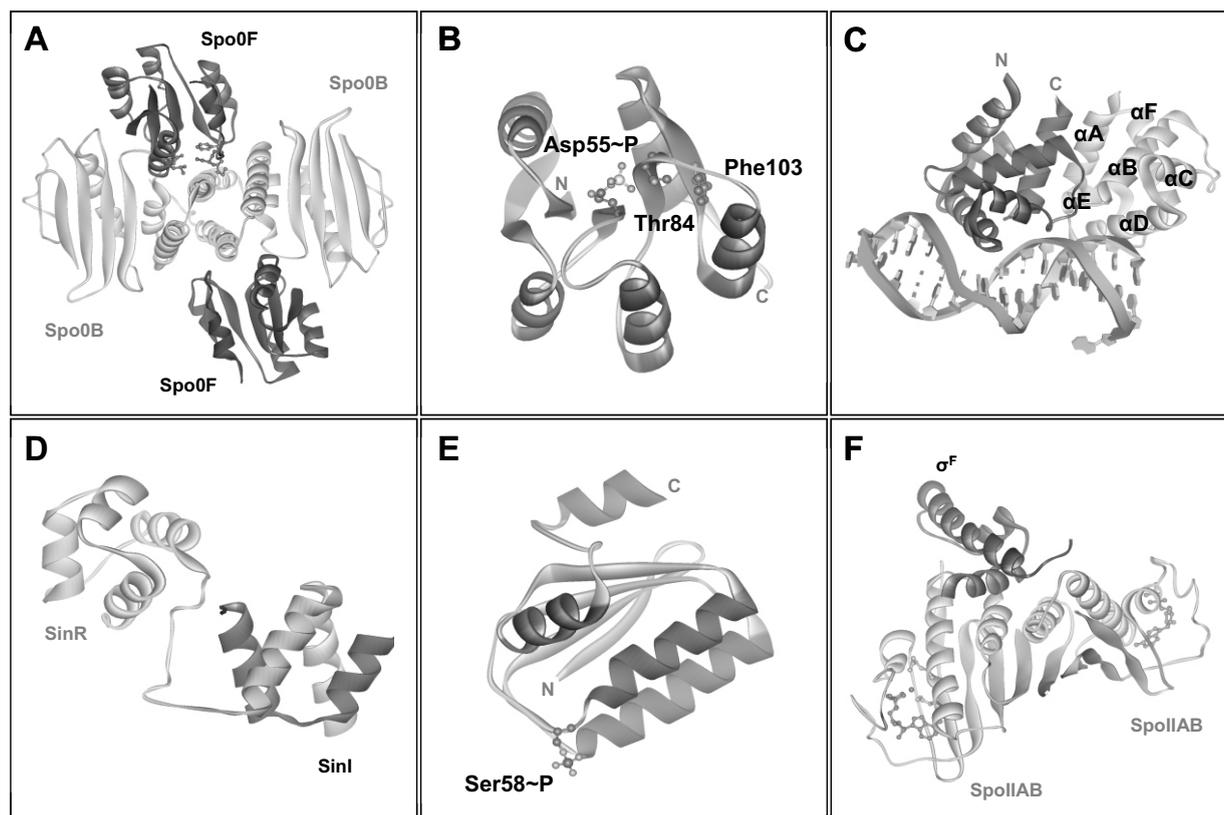


Fig. 2. Crystal structure of some sporulation specific proteins. A. Complex of two Spo0F monomers with Spo0B dimer; aminoacid residues of Spo0F active site are visualised on upper Spo0F molecule (PDB ID: 1F51). B. Active site of N-Spo0A; phosphorylated Asp55 and aminoacid residues, which change their conformation during phosphorylation are shown (PDB ID: 1QMP). C. Complex of two C-terminal domain of Spo0A in complex with DNA; helices α A- F are depicted (PDB ID: 1LQ1). D. Structure of SinR-SinI complex (PDB ID: 1B0N). E. Phosphorylated form of SpoIIAA with site of phosphorylation pointing to solvent (PDB ID: 1H4X). F. Complex of SpoIIAB dimer with part of σ^F monomer (PDB ID: 1L0O).

Initiation of sporulation

Vegetative cells are able to monitor the number of cells in the population and the amount of nutritional substances in the environment. In case of nutrition deprivation, the bacterial cell can enter the process of spore development. There are many proteins involved in the initiation stage of sporulation that are important for proper chromosome segregation, cytokinesis, cell length and cell shape determination. Some of these proteins play a crucial role also during sporulation. The initiation of sporulation depends on a complex series of external and internal signals and is stringently controlled by a network of regulatory proteins of the phosphorelay. Recently, the structure of some proteins involved in this adaptive response were solved and questions partially answered regarding interactions between the components of this regulatory network.

This phosphorelay is an expanded version of a two-component signal transduction system. The first component, a sensor kinase (up to five sporulation sensor kinases Kin A-E have been identified [1]), is autophosphorylated by ATP on its histidine residue in response to environmental changes. The phosphoryl group is subsequently transferred to the aspartyl residue of the response regulator Spo0F. Then phosphotransferase Spo0B transfers phosphate from Spo0F to the final response regulator, transcriptional activator and repressor Spo0A [2]. Spo0F is a typical single domain response regulator that has an α -

structure with five α -helices arranged around a central α -sheet consisting of five parallel α -strands [3] (Fig. 2A). The active site comprising three aspartates Asp10, Asp11 and Asp54, and the threonine Thr82 and lysine Lys104 residues is situated as a small pocket at the carboxy-terminal end of the α -sheet. The site of phosphorylation, Asp54 is at the bottom of the pocket and is accessible to solvent. The overall structure and active site geometry of Spo0F is remarkably similar to other regulatory domains such as CheY [4], Fix J [5] or Spo0A [6].

The phosphate accumulated on Spo0F is immediately delivered to Spo0A by the action of Spo0B. Spo0B is a phosphotransferase with some functional and structural similarities with histidine kinases [7]. It is phosphorylated on its histidine residue and forms a dimer. The monomer is made up of two domains, an amino-terminal α -helical hairpin domain and a carboxy-terminal domain with an α -fold. The crystal structure reveals that dimer is formed by the association of the hairpin helical domains from two molecules to form a four-helix bundle with the site of phosphorylation His30 protruding into the solvent. There are two active sites per dimer. The crystal structure of the complex between Spo0F and Spo0B [8] revealed how the regulatory domain and phosphotransfer domain interact together (Fig. 2A). The cocrystal contained two Spo0F molecules per Spo0B dimer. The 1-helix of each Spo0F molecule and the five α -loops on the top of the molecule

interact with the α -helix of Spo0B. This interaction aligns the histidine of Spo0B with the aspartate of Spo0F in correct configuration and distance for phosphotransfer. The hydrophobic surface of Spo0F comprised of hydrophobic residues, Ile15, Leu18, Pro105, Phe106 and Ile108, is involved in this interaction. Since the structure of Spo0B is very similar to the structure of histidine sensor kinases, the Spo0B-Spo0F structure should be viewed as a model for response regulator-sensor kinase interactions and these six highly conserved residues may play a key role in this interaction.

Spo0A master control for entry into sporulation

The final acceptor of phosphate migration via Spo0F and Spo0B is Spo0A, the response regulator and key control element in the decision to sporulate [2]. Spo0A consists of two domains, the N-terminal phosphoacceptor (receiver) domain and the C-terminal DNA-binding (effector) domain [9] connected by a flexible linker peptide. As the DNA-binding domain is active after removal of the N-terminal domain, this receiver domain is thought to inhibit the effector domain and this autoinhibition is overcome by conformational changes following phosphorylation of the N-terminal domain. The crystal structure of N-Spo0A was determined in both phosphorylated and unphosphorylated forms [6, 10]. The overall fold of N-Spo0A perfectly matches the structures of the regulatory domains of other response regulators. The structure of N-Spo0A~P revealed that phosphorylation of aspartate 55 leads to conformational changes at the active site (Fig. 2B). The reorientation of the Thr84 hydroxyl group towards the phosphoryl group is accompanied by movement of a side chain of Phe103 from a solvent exposed position into the protein core. Since these active-site residues are highly conserved among response regulators and similar changes were observed in the structures of other activated response regulators [5, 11], this "aromatic switch" could represent a general mechanism of signal propagation through the protein following aspartate phosphorylation. Moreover, mutational and biochemical studies suggest that Phe103 could be involved in dimer formation of Spo0A after phosphorylation (unpublished data).

Recently, structures of the C-terminal domain of Spo0A [12] and C-Spo0A in complex with DNA [13] also were solved. The C-Spo0A (Fig. 2C) forms a large helical assembly comprising of six α -helices (A-F) connected by short peptides. The structure contains a classical helix-turn-helix (helices C and D) as a DNA binding motif. Helices A, B and F form a three-helical bundle located atop the HTH and helix E protrudes from the bundle opposite the HTH. The residues involved in contact with σ^A subunit of RNA polymerase in the transcription initiation complex are situated at the opposite end of the DNA recognition helix and form in contrast with the rest of the structure a flexible and mobile segment. The crystal structure of C-Spo0A in complex with its target DNA (Fig. 2C) revealed that two molecules form a dimer upon binding to the tandem binding sites on the DNA. The molecules are bound in head to tail arrangement such that the C-terminus of one molecule (helix F) interacts with the N-terminus of another (helix B).

Helix F of one molecule and helices A and B of the second molecule form a C-Spo0A dimer interface. Such dimerization should stabilize the protein-DNA complex. Also analytical ultracentrifugation and gel filtration experiments revealed dimerization of entire Spo0A after phosphorylation [14]. These results indicate that dimerization after phosphorylation is mediated principally by the receiver domain probably followed by dimerization of the effector domain bound on DNA. Obtained structural and biochemical data provide an important framework for further interpretation of the mutational data of Spo0A and thus understanding of the biological function of Spo0A at the molecular level.

Other proteins involved in initiation of sporulation

To ensure that spore formation is the only way which enables the bacterium is able to overcome unfavourable environmental conditions, the entry into sporulation is precisely controlled not only by the above mentioned components of phosphorelay, but also by other proteins. Among these, an important role belongs to the sporulation inhibitor SinR and its antagonist SinI [15]. SinR is a tetrameric DNA binding protein that controls sporulation directly through repression of *spo0A* and stage II sporulation genes, *spoIIA*, *spoIIIE* and *spoIIIG*. For sporulation to proceed, the activity of SinR must be switched off. This is brought about by the action of SinI which forms SinI-SinR complex that is unable to bind to DNA and therefore the repressive effects of SinR on transcription are relieved. The crystal structure of this heterodimer [16] revealed an α -helical assembly of two domains of approximately equal size, an oligomerisation domain and a DNA-binding domain. The oligomerisation domain is formed by four α -helices, two from the C-terminal residues of SinR and two from the central residues of SinI (Fig. 2D). Recently, structure of another protein, Obg, a Spo0B-associated GTP binding protein, involved in the sporulation pathway, was solved [17]. This structure revealed the unique architecture of a GTP binding protein and together with biochemical analysis suggests a potential role for this protein in the life cycle of *B. subtilis*.

Proteins involved in the activation of first compartment specific sigma factor - σ^F

The first clear morphological feature of sporulation is the formation of an asymmetric septum (Fig. 1A) that bisects the bacterial cell into two unequally sized compartments, the larger mother cell and the smaller forespore. The proper positioning of this sporulation septum is dependent on Spo0A activity, which mediates the assembly of division proteins near the cell pole instead of at mid-cell. Such positional switch could be partially triggered through the activity of the sporulation specific protein SpoIIIE, also a component of the sporulation septum, that is expressed as a result of Spo0A activity [18, 19, 20, 21]. The central domain of SpoIIIE is involved in oligomerization of the protein and it is responsible for interaction with FtsZ, the protein essential for cell division [22, 23]. SpoIIIE protein can dephosphorylate SpoIIAA-P and thus it plays a crucial

role in activation of the first cell-type specific factor σ^F (Fig. 1B) [24, 25]. The crystal structure of phosphorylated and unphosphorylated form of SpoIIAA was solved from *Bacillus sphaericus* [26]. This single domain globular protein consists of a central sheet of five strands, and four

helices (Fig. 2E). On one face, the β -pleated sheet is embedded with a pair of α -helices, while the other face is exposed to the solvent. The phosphorylation site, dephosphorylation of which is important for activation of σ^F , is located at the N-terminus of the helix 2 so that the phosphoryl group points into the solvent. Comparisons between unphosphorylated and phosphorylated SpoIIAA had shown, that structural changes accompanying phosphorylation are only slight. Extended hydrophobic surfaces together with flexibility in helix 3 and the following loop indicate that this protein often undergoes a disorder to order transition and therefore has a disposition for forming stable protein-protein interactions [26]. Also the structure of the anti- σ^F factor SpoIIAB in the complex with σ^F of *Bacillus stearothermophilus* was resolved [27]. SpoIIAB- σ^F complex consists of a SpoIIAB dimer bound to one σ^F molecule so that σ^F binding directly to the dimer interface makes contact with both SpoIIAB monomers (Fig. 2F). The whole structure of σ^F is comprised of three compactly folded domains connected by flexible linkers [27]. σ^F is kept inactive in a complex with SpoIIAB. In the presence of unphosphorylated SpoIIAA molecules, SpoIIAB binds to SpoIIAA and active σ^F is released. However, SpoIIAB kinase phosphorylates SpoIIAA, and the SpoIIAB-SpoIIAA complex decays, releasing so the anti- σ^F factor ready to block another σ^F . SpoIIAB has the ATP binding Bergerat fold [28], found in histidine kinases and ATP-ases of the GHKL superfamily. This consists of an α/β -sandwich with a four-stranded antiparallel β -sheet and three α -helices. In the SpoIIAB monomer 1 extends the four-stranded β -sheet. Dimerization interactions occur among 1 and 1 of each SpoIIAB monomers. The ATP-binding site is represented by Asn50, which plays a key role in chelating Mg^{2+} , critical for ATP-ase or kinase activity of the GHKL superfamily members [27]. The catalytic site is formed by Glu46, which functions as a catalytic base in the kinase reaction [29]. SpoIIAA phosphorylated by SpoIIAB requires the phosphatase activity of SpoIIE to become active again.

After activation of σ^F , which follows only after completion of the sporulation septum, other compartment specific factors are sequentially activated. This leads to differentiation of gene expression. In the course of the sporulation process, the mother cell engulfs the prespore and the process of sporulation culminates with programmed death of the mother cell. The survivor is the forespore, which matures into a very resistant spore able to outlive hostile envi-

ronments. The life cycle is completed by the process of germination, which is initiated when the free spore returns to conditions favorable for life.

Future perspectives

Especially interesting for our understanding of this mechanism of the gene expression asymmetry would be the detailed study of sporulation septa formation by solving the crystal structure of key proteins of this cell division process. Although the asymmetric cell division that occurs during sporulation differs from the cell division during vegetative growth, both processes use fundamentally the same protein machinery [30, 31, 32]. SpoIIE protein is one of the most interesting candidates for crystallographic studies. This large (92 kDa) protein consists of three domains: N-terminal region containing 10 membrane-spanning helices [33, 34] followed by central domain, possibly involved in intermolecular interaction [22], and the C-terminal domain, which is specific protein phosphatase involved in activation of the first forespore sigma factor [25]. Another very interesting and highly studied sporulation specific protein from *B. subtilis* is DNA translocase SpoIIIE [35, 36]. Recently it has been shown that this protein also cooperates in membrane fusion during the spore engulfment [37]. SpoIIIE protein consists of a N-terminal membrane bound domain mediating its localization to the division septum and cytosolic C-terminal domain capable of tracking along DNA in the presence of ATP.

DivIVA protein is an important regulator of cell division during vegetative growth [38] and it also has been proposed to have a crucial role in the sporulation process by anchoring the chromosomes to the cell poles prior to asymmetric division [39]. It was shown that in its native state DivIVA oligomerizes [40], and has a structural similarity to myosin and other proteins having α -helical coiled-coil structure [41].

This review is an attempt to summarize what is known from crystallographic studies about the structure of some of the approximately 200 sporulation specific proteins of *B. subtilis*. The structures of only a small number of these proteins are known due to problems associated with the crystallization of the proteins. Structures for many of the interesting candidates remain a challenge. Among them are the membrane bound proteins and proteins with very flexible domains.

Acknowledgements

The work in IB's laboratory is supported by grant 2/1004/21 from the Slovak Academy of Sciences and Wellcome Trust Grant 066732/Z/01/Z.

**References**

1. M. Jiang, W. Shao, M. Perego & J. A. Hoch, *Mol. Microbiol.* **38** (2000) 535-542.
2. D. Burbulys, K. A. Trach & J. A. Hoch, *Cell* **64** (1991) 545-552.
3. Madhusudan, J. Zapf, J. M. Whiteley, J. A. Hoch, N. H. Xuong & K. I. Varughese, *Structure* **4** (1996) 679-690.
4. A. M. Stock, J. M. Mottonen, J. B. Stock & C. E. Schutt, *Nature* **337** (1989) 745-749.
5. C. Birck, L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn & J. P. Samama, *Structure* **7** (1999) 1505-1515.
6. R. J. Lewis, J. A. Brannigan, K. Muchová, I. Barák & A. J. Wilkinson, *J. Mol. Biol.* **294** (1999) 9-15.
7. K. I. Varughese, Madhusudan, X. Z. Zhou, J. M. Whiteley & J. A. Hoch, *Mol. Cell.* **2** (1998) 485-493.
8. J. Zapf, U. Sen, Madhusudan, J. A. Hoch & K. I. Varughese, *Structure* **8** (2000) 851-862.
9. J. K. Grimsley, R. B. Tjalkens, M. A. Strauch, T. H. Bird, G. B. Spiegelman, Z. Hostomsky, J. M. Whiteley & J. A. Hoch, *J. Biol. Chem.* **269** (1994) 16977-16982.
10. R. J. Lewis, K. Muchová, J. A. Brannigan, I. Barák, G. Leonard & A. J. Wilkinson, *J. Mol. Biol.* **297** (2000) 757-770.
11. H. S. Cho, S. Y. Lee, D. Yan, X. Pan, J. S. Parkinson, S. Kustu, D. E. Wemmer & J. C. Pelton, *J. Mol. Biol.* **297** (2000) 543-551.
12. R. J. Lewis, S. Krzywda, J. A. Brannigan, J. P. Tukenburg, K. Muchová, E. J. Dodson, I. Barák & A. J. Wilkinson, *Mol. Microbiol.* **38** (2000) 198-212.
13. H. Zhao, T. Msadek, J. Zapf, Madhusudan, J. A. Hoch & K. I. Varughese, *Structure* **10** (2002) 1041-1050.
14. R. J. Lewis, D. J. Scott, J. A. Brannigan, J. C. Ladds, M. A. Cervin, G. B. Spiegelman, J. G. Hoggett, I. Barák & A. J. Wilkinson, *J. Mol. Biol.* **316** (2002) 235-245.
15. I. Mandic-Mulec, L. Doukhan & I. Smith, *J. Bacteriol.* **177** (1995) 4619-4627.
16. R. J. Lewis, J. A. Brannigan, W. A. Offen, I. Smith & A. J. Wilkinson, *J. Mol. Biol.* **283** (1998) 907-912.
17. J. Buglino, V. Shen, P. Hakimian & C. D. Lima, *Structure* **10** (2002) 1581-1592.
18. I. Barák & P. Youngman, *J. Bacteriol.* **178** (1996) 4984-4989.
19. A. Feucht, T. Magnin, M. D. Yudkin & J. Errington, *Genes Develop.* **10** (1996) 794-803.
20. A. Khvorova, L. Zhang, M. L. Higgins & P. J. Piggot, *J. Bacteriol.* **180** (1998) 1256-1260.
21. K. York, T. J. Kenney, S. Satola, C. P. Moran, H. Poth & P. Youngman, *J. Bacteriol.* **174** (1992) 2648-2658.
22. I. Lucet, A. Feucht, M. D. Yudkin & J. Errington, *EMBO J.* **19** (2000) 1467-1475.
23. P. Prepiak, Z. Chromiková & I. Barák, *Folia Microbiol. (Praha)* **46** (2001) 292-296.
24. L. Duncan, S. Alper. & R. Losick, *Curr. Opin. Genet. Dev.* **4** (1994) 630-636.
25. L. Duncan, S. Alper, F. Arigoni, R. Losick & P. Stragier, *Science* **270** (1995) 641-644.
26. P. R. Seavers, R. J. Lewis, J. A. Brannigan, K. H. G. Verschuere, G. N. Murshudov & A. J. Wilkinson, *Structure* **9** (2001) 614-615.
27. E. A. Campbell, S. Masuda, J. L. Sun, O. Muzzin, C. A. Olson, S. Wang & S. A. Darst, *Cell* **108** (2002) 795-807.
28. A. Bergerat, *Nature* **386** (1997) 414-417.
29. S. G. Park and M. D. Yudkin, *Gene* **194** (1997) 25-33.
30. J. Errington, *Curr. Opin. Microbiol.* **4** (2001) 660-666.
31. B. Beall and J. Lutkenhaus, *Genes Dev.* **5** (1991) 447-455.
32. K. Ireton, N. W. Gunther & A. D. Grossman, *J. Bacteriol.* **176** (1994) 5320-5329.
33. I. Barák, J. Behari, G. Olmedo, P. Guzmán, D. P. Brown, E. Castro, D. Walker, J. Westpheling & P. Youngman, *Mol. Microbiol.* **19** (1996) 1047-60.
34. F. Arigoni, A. M. Guérout-Fleury, I. Barák & P. Stragier, *Mol. Microbiol.* **31** (1999) 1407-1715.
35. L. Wu & J. Errington, *Science* **264** (1994) 572-575.
36. J. Bath, L. J. Wu, J. Errington and J. C. Wang, *Science* **290** (2000) 995-997.
37. M. D. Sharp & K. Pogliano, *Proc. Natl. Acad. Sci. USA* **96** (1999) 14553-14558.
38. D. H. Edwards and J. Errington, *Mol. Microbiol.* **24** (1997) 905-915.
39. H. B. Thomaides, M. Freeman, M. E. Karoui & J. Errington, *Genes Dev.* **15** (2001) 1662-1673.
40. K. Muchová, E. Kutejová, D. J. Scott, J. A. Brannigan, R. J. Lewis, A. J. Wilkinson & I. Barák, *Microbiol.* **148** (2002) 807-813.
41. D. H. Edwards, H. B. Thomaides & J. Errington, *EMBO J.* **19** (2000) 2719-2727.

