Identification of potential $\sigma^N$-dependent promoters in bacterial genomes

The alternative form of bacterial RNA polymerase containing sigma factor $\sigma^N$ ($\sigma^N$, RpoN) recognizes promoters with the consensus DNA sequence YTGGCACGrNNN-TGTCW (1). The promoter-bound $\sigma^N$ holoenzyme is subject to positive control by enhancer-binding activator proteins (3). We have previously predicted several new potential $\sigma^N$-dependent promoters from genome sequence data using the World Wide Web-based program seqscan (13, 14; http://www.bmb.psu.edu/seqscan/). Now we have searched the complete genomes of several bacteria using a new program, promscan (http://www.promscan.uklinux.net/). The results are summarized in Table 1. It is important to remember that these predictions remain hypothetical until tested by experiment. Notwithstanding the limitations of prediction, a number of interesting results and hypotheses emerge.

Escherichia coli and Bacillus subtilis have considerably larger genomes than most of those sequenced and are rather distantly related to each other by phylogeny. Therefore the repertoire of $\sigma^N$-dependent promoters in the B. subtilis genome is of some interest for comparison with what we know about enteric bacteria. In addition to the four previously known or predicted $\sigma^N$-binding sites in B. subtilis (bkd, levD, rocA and rocD), candidates were found upstream of acoA, bglC, ytgB and yweB. Therefore, apart from acetoin metabolism (in Ralstonia eutropha, acoX is under the control of a $\sigma^N$-dependent promoter), there is no apparent overlap of $\sigma^N$-dependent functions between B. subtilis and the Gram-negative bacteria.

We have previously predicted $\sigma^N$-dependent promoters for the chlamydial genes Cpn0725 and CT652.1 (13). Now we have found several more. Two of these are linked to ORFs encoding hypothetical proteins of unknown function. Another is found upstream of a helicase in Chlamydia muridarum and another is upstream of a gene potentially encoding a protein with a TPR (tetratricopeptide repeat) motif. Matthews & Timms (9) recently confirmed that this latter promoter is active in vivo by a primer-extension assay. TPR-containing proteins are often associated with multiprotein complexes and may be involved in the functioning of chaperones, cell-cycle, transcription and protein-transport complexes (2). Matthews et al. (10) propose that $\sigma^N$ may be involved in the conversion from reticulate bodies (RB) to elementary bodies (EB) in chlamydias and thus it may be that these genes have some role in that developmental process.

We have previously predicted $\sigma^N$-dependent promoters for rpoS in Borrelia burgdorferi and the TpFI gene in Treponema pallidum (3). Here we identify an additional potential promoter upstream of pfoR/S, a gene with some similarity to pfoR encoding a protein regulating perfringolysin in Clostridium perfringens (11).

Several strains of Neisseria species, including those whose whole genomes have been sequenced, contain an rpoN-like sequence (RLS). The RLS appears to have arisen from an ancestral rpoN gene which underwent a deletion of an essential part of the gene so that these organisms lack a functional $\sigma^N$ (8). Carrick et al. (11) went on to demonstrate that Neisseria gonorrhoeae contains the remnants of a two-component regulatory system that probably once controlled $\sigma^N$-dependent transcription of the pilin gene pilE. Our searches detected three potential $\sigma^N$-dependent promoters. Presumably an ancestral Neisseria species lost $\sigma^N$ relatively recently and these sites may represent the remnants of genuine promoters.

Mutants of Vibrio cholerae lacking a functional $\sigma^N$ are non-motile and auxotrophic for glutamine. In V. cholerae $\sigma^N$ is required for the transcription of several flagellar genes. It also appears to have some other role in colonization of the host that is separate from its role in motility (7). Here we predict several new potential $\sigma^N$-dependent promoters, found on both chromosomes, and thus suggest some candidates that may be involved in colonization. $\sigma^N$-dependent regulation of hae-molysin co-regulated protein (encoded by hcp) has previously been suggested (15). Among the gene products possibly regulated by our new proposed promoters are a histidine kinase sensor, an Fe–S-centre binding protein, a di-carboxylate-binding protein and hypothetical proteins of unknown function.

The phytopathogen Xylella fastidiosa, like E. coli and V. cholerae, belongs to the gamma proteobacteria. Like some strains of Pseudomonas species, in X. fastidiosa $\sigma^N$ potentially regulates biogenesis of the fimbriae which are involved in several stages of patho-
genesis. The gene encoding a regulatory protein similar to MarR, which in E. coli represses a multi-drug resistance operon (5), has a potential $\sigma^N$-dependent promoter. A third potential $\sigma^N$-dependent promoter in X. fastidiosus is found upstream of XF0924, a homologue of smf. In E. coli, smf encodes polypeptide deformylase, an enzyme essential for protein synthesis.

Flagellar motility plays an important role in many bacteria including several species of Campylobacter and the closely related Helicobacter pylori. Kinsella et al. (6) showed that the fglE gene, encoding a flagellar hook protein, is under the control of $\sigma^N$ in Campylobacter coli. Consistent with this, we recovered the potential promoter upstream of fglE2 in Campylobacter jejuni. We also found potential promoters for a flagellar gene flaG, Cj1026c encoding a putative lipoprotein, and rpmE encoding a ribosomal protein. The function of the lipoprotein is unknown. If the rpmE promoter turns out to be genuine, then this may represent a novel link between regulation at the levels of transcription and translation.

In H. pylori $\sigma^N$ is implicated in the regulation of five operons involved in flagellar synthesis (12). In H. pylori there is a single $\sigma^N$-dependent transcriptional activator, FlgR, which has been identified as the master regulator of several basal body and hook genes (12). We also predict two additional promoters in H. pylori 26695. One of these lies upstream of an apparent operon containing HP1154 (of unknown function) and murG (involved in peptidoglycan synthesis). The other lies upstream of pfr (encoding a ferritin) and serB (synthesis of the amino acid serine). Therefore, it may be that these functions also form part of a regulon with several flagellar genes in this organism. 

In conclusion, we have developed tools for rapidly scanning microbial genome data to identify new potential $\sigma^N$-binding sites. Although most of these predictions remain hypothetical at present, we hope that they will help to stimulate fruitful research. A database of the complete data and the source code for the PROMSCAN software can be found at http://www.promscan.uklinux.net and at http://homepage. virgin. net/d. studholme. We intend to update this database as new complete genome sequences become available.
Acknowledgements

We are grateful to Enrique Morett for providing the compilation of known promoters and for useful discussions. We thank Sarah Matthews for sharing data before publication.

David J. Studholme¹, Martin Buck² and B. Tracy Nixon³

¹Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK.
²Department of Biology, Sir Alexander Fleming Building, Imperial College of Science, Technology and Medicine, London SW7 2AZ, UK.
³Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802-4500, USA.

Author for correspondence: David J. Studholme. Tel: +44 23 8059 4324. Fax: +44 23 8059 4459.

e-mail: djs@soton.ac.uk