Molecular Microbiology (1999) **31**(1), 393–395

MicroCorrespondence

RpoN-dependent transcription of *rpoH*?

Sir,

Kallipolitis and Valentin-Hansen (1998, *Mol Microbiol* **29**: 1091–1099) recently reported that transcription of *rpoH* (which encodes the heat shock sigma factor in *Escherichia coli*) is controlled by the cAMP–CRP/CytR complex. This adds to the known complexity of the regulation of *rpoH* transcription, which involves at least four promoters (P1, P3, P4 and P5) and two sigma factors, sigma-E (RpoE) and sigma-70 (RpoD). However, it seems likely that the situation is yet more complex, as *rpoH* is probably also transcribed by the core enzyme in association with yet another sigma factor, sigma-54 (RpoN or sigma-N), a sigma factor structurally unrelated to the sigma-70 family (Merrick, 1993, *Mol Microbiol* **10**: 903–909).

While searching for promoters dependent on RpoN, I discovered an excellent match to the RpoN promoter consensus sequence (TGGCAC-N5-TTGCa/t; Merrick, 1993, *Mol Microbiol* **10**: 903–909; Buck and Cannon, 1989, *Nucleic Acids Res* **17**: 2597–2612) upstream of the *E. coli rpoH* coding region (Fig. 1). The putative RpoN binding site begins 10 bp downstream of the P5 start site (the most downstream of the known promoters), and the RpoN-dependent transcriptional start site (P6) can be predicted to occur 31 bases upstream of the ATG start codon. The sequence and position of the putative RpoN binding site is well conserved in the regions upstream of *rpoH* in *Citrobacter freundii* and *Enterobacter cloacae* (Nakahigashi *et al*., 1995, *Nucleic Acids Res* **23**: 4383–4390), adding credence to its likely biological significance (Fig. 2).

As the regulation of RpoN-dependent genes is governed by activator proteins, it is difficult to predict under what circumstances the RpoN-dependent promoter might be active (Merrick, 1993, *Mol Microbiol* **10**: 903–909). However, for those with the interest and experimental expertise, it should be relatively straightforward to show experimentally that RpoN binds to the predicted site. If these predictions are confirmed, then a previously unsuspected link between two important regulatory networks, the RpoH regulon and the RpoN regulon, will be established.

Acknowledgements

I thank Martin Buck for helpful discussion, and Gordon Dougan and the Wellcome Trust for sponsoring my Research Leave Fellowship.

Mark Pallen
Molecular Pathogenesis Research Group, Department of Medical Microbiology, St Bartholomew's and the Royal London School of Medicine and Dentistry, London EC1A 7BE, UK.
Received 11 August 1998; accepted 8 September 1998.

```
        -90                 -70                 -50                 -30                 -10         +1
         .                   .                   .                   .                   .          .
TCTGATAAAACAGTGAATGATAACCTCGTTGCTCTTAAGCTCTGGCACAGTTGTTGCTACCACTGAAGCGCCAGAAGATATCGATTGAGGAGGATTTGAATG
           |      |             |         TGGCACNNNNNTTGCW          |                                |
           P3     P4            P5        RpoN-binding site         P6                               start codon
```

Fig. 1. The proximal part of the *rpoH* promoter region. The proven transcriptional start sites from promoters P3 P4 and P5 and the predicted start site P6 are marked by vertical lines. The RpoN binding site is underlined. Numbering is set with +1 as the first base of the *rpoH* coding region. The start codon is indicated in bold. Sequence from Kallipolitis and Valentin-Hansen, 1998, *Mol Microbiol* **29**: 1091–1099.

```
Consensus                TGGCACNNNNNTTGCW
Escherichia coli         TGGCACAGTTGTTGCT <--41bp--> ATG start codon
Citrobacter freundii     TGGCATGGTTGTTGCT <--47bp--> ATG start codon
Enterobacter cloacae     TGGCATGGTTGTTGCC <--44bp--> ATG start codon
```

Fig. 2. Putative RpoN binding sites upstream of the *rpoH* coding regions in related organisms. Consensus from Buck and Cannon, 1989, *Nucleic Acids Res* **17**: 2597–2612, cited by Merrick, 1993, *Mol Microbiol* **10**: 903–909. Sequences from Nakahigashi *et al*., 1995, *Nucleic Acids Res* **23**: 4383–4390.

The multidrug efflux protein NorM is a prototype of a new family of transporters

Sir,

Four families of transporters have previously been described that contain multidrug efflux systems: the major facilitator superfamily (MFS), the small multidrug resistance (SMR) family, the resistance/nodulation/cell division (RND) family and the ATP-binding cassette (ABC) superfamily. The MFS, SMR and RND families consist of secondary transporters, typically energized by the proton-motive force (pmf), and the ABC superfamily comprises ATP-dependent transporters (Paulsen *et al*., 1996, *Microbiol Rev* **60:** 575–608). Recently, two new multidrug efflux proteins have been identified, NorM from *Vibrio parahaemolyticus*, and a homologue in *E. coli*, YdhE (Morita *et al*., 1998, *Antimicrob Agents Chemother* **42:** 1778–1782). These two proteins mediate resistance to a range of cationic dyes, aminoglycosides and fluoroquinolones, probably via a pmf-dependent efflux mechanism. NorM and YdhE have 12 predicted transmembrane segments (TMS) and on this basis have been suggested to be members of the MFS (Morita *et al*., 1998, ibid.).

However, our analyses have demonstrated that these proteins do not share significant sequence similarity with any member of the MFS and do not exhibit any of the signature sequences specific to the 18 families of the MFS identified by Pao *et al*., 1998 (*Microbiol Mol Biol Rev*, **62:** 1–34). We have found that NorM and YdhE are members of a previously unidentified family, which contains more than 30 proteins, including representatives from all three kingdoms of life (Eukarya, Archaea and Eubacteria). Statistical analyses evaluated by binary comparisons revealed significant amino acid sequence similarity between NorM and all of the other members of the family ($9 \leq Z \leq 82$; Lipman and Pearson, 1985, *Science* **227:** 1435–1441). The proteins in this family range in size from 363 to 1141 residues, and hydropathy analyses revealed that they characteristically possess 12 putative TMS (Fig. 1A). Multiple sequence alignments indicated that their most highly conserved regions are located in the vicinity of TMS 5 and 6, and near the terminus of TMS 8 (Fig. 1B).

Extensive phylogenetic studies of more than 70 families of transport proteins (Paulsen *et al*., 1998, *J Mol Biol* **277:** 573–592; Saier, 1998, *Adv Microbiol Physiol*, **40**: 81–136) has revealed that transporter substrate specificity typically correlates with phylogeny, and hence such analyses provide a credible foundation for making functional predictions. Phylogenetic analysis of this family revealed the presence of three distinct clusters (clusters 1, 2 and 3 in Fig. 1C), implying that these three clusters may differ functionally in some respect.

The first of these clusters includes the bacterial multidrug efflux proteins NorM and YdhE, as well as hypothetical proteins from *Haemophilus*, *Bacillus* and *Synechocystis*. The branching pattern of these proteins parallels the lineage of their respective organisms, indicating that they are likely to be orthologues, and may consequently all be drug/multidrug efflux proteins. The constituents of the second cluster are exclusively eukaryotic proteins, from either fungi or plants. The one functionally characterized member of this cluster, the yeast Erc1 protein, confers resistance to the methionine analogue ethionine (Shiomi *et al*., 1991, *J. Ferment Bioeng* **71:** 211–215). Although the mechanism of resistance mediated by Erc1 has not been established, it is plausible that Erc1 and the other members of this cluster are eukaryotic transporters.

The third cluster includes the DinF proteins from *E. coli* and *Streptococcus pneumoniae* and homologues of these proteins from Eubacteria and Archaea. In contrast to cluster 1, the branching pattern of the proteins within this cluster does not correlate with their respective organismal phylogeny, indicating that this cluster includes a number of paralogous proteins. The functions of the two DinF proteins are unknown, but expression of both of them has been demonstrated to be DNA damage-inducible (Mortier-Barrière *et al*., 1998, *Mol Microbiol* **27:** 159–170; Thoms and Wackernagel, 1987, *J Bacteriol* **169:** 1731–1736). It is tempting to speculate they may be stress-induced efflux proteins of unknown specificity; whether this is a common feature of proteins in this cluster remains to be explored.

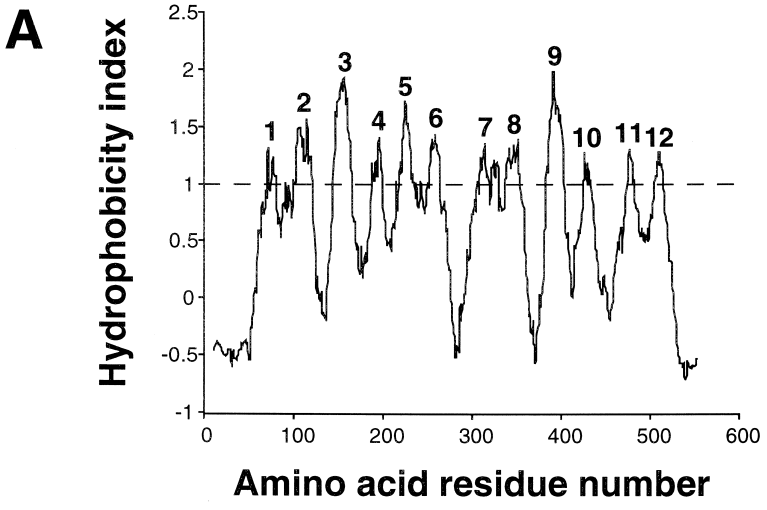
Based on the currently available sequences, the recently described NorM multidrug efflux protein is manifestly not a member of the MFS, but instead is the prototype of a new family of transporters, which we have termed the MATE (multidrug and toxic compound extrusion) family. This represents a significant new finding in the field of multidrug efflux, as for several years it has been thought that there were only four transporter families that included multidrug efflux systems. The inclusion of *E. coli* DinF and its homologues in this family provides an exciting new insight into the function of this SOS-response protein, potentially implicating it in the efflux of DNA damage-inducing compounds. The MATE family appears to represent a family of transport proteins responsible for protecting cells from drugs and other toxic compounds, although the physiological role of most of its 32 members remains to be elucidated.

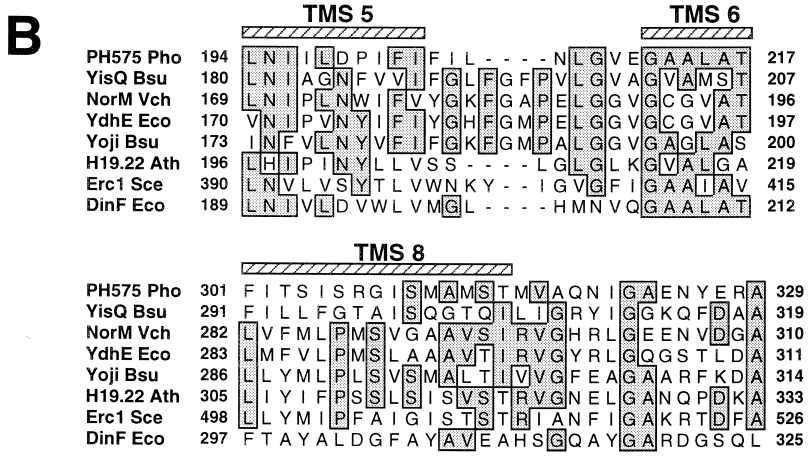
Melissa H. Brown, Ian T. Paulsen and Ronald A. Skurray*

School of Biological Sciences, University of Sydney, New South Wales 2006, Australia.

**For correspondence. E-mail skurray@bio.usyd.edu.au; Tel. (2) 9351 2376; Fax (2) 9351 4771.*

Received 29 August, 1998; revised 10 September, 1998; accepted 14 September, 1998.





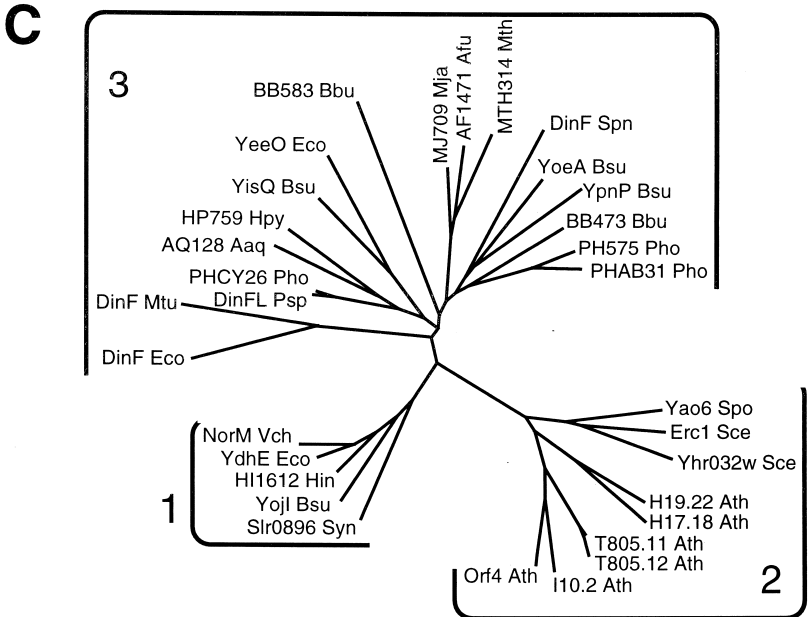


Fig. 1 A. Average hydropathy plot for NorM and its homologues, based on the multiple sequence alignment used to construct the tree shown in C below. The plot was derived using a sliding window of 21 residues and the hydrophobicity scale of Kyte and Doolittle, 1982 (*J. Mol Biol* 157: 105–132). The locations of the 12 predicted TMS are indicated.
B. Multiple sequence alignment of the most conserved regions of NorM with representative homologues. The protein designations used list the protein name and a three-letter organism abbreviation. The hatched boxes represent the predicted positions of TMS. Amino acid numbering is shown at both ends of the sequence presented. Identical residues are indicated by the boxed shading.
C. Unrooted phylogenetic tree for NorM and its 31 homologues, the three clusters are delineated and numbered. The tree was prepared using the PHYLIP package (Felsenstein, 1989, *Cladistics* 5: 164–166). The protein abbreviations used are as described in B above.