Molecular Microbiology (1996) **19**(3), 639–647

MicroCorrespondence

Monofunctional biosynthetic peptidoglycan transglycosylases

Sir,

The high-molecular-weight penicillin-binding proteins (high-M_r PBPs) catalyse the final stages of bacterial peptidoglycan synthesis and are the physiological targets of the β-lactam group of antibiotics (Spratt and Cromie, 1988, *Rev Inf Dis* **10:** 699–711; Ghuysen, 1991, *Annu Rev Microbiol* **45:** 33–67). The carboxy-terminal domains of these enzymes catalyse the penicillin-sensitive peptidoglycan transpeptidation (TPase) reaction. The amino-terminal domains of the class A high-M_r PBPs (those that are homologous to *Escherichia coli* PBP1A and PBP1B) catalyse a penicillin-insensitive peptidoglycan transglycosylase (TGase) reaction; the function of this domain in the class B high-M_r PBPs (those that are homologous to *E. coli* PBP2 and PBP3) is uncertain (Ghuysen, 1991, ibid).

It has been proposed that several species, including *Staphylococcus aureus* and *E. coli*, possess a peptidoglycan TGase that is not associated with a TPase domain and which is therefore not a PBP (Hara and Suzuki, 1984, *FEBS Lett* **168:** 155–160; Park and Matsuhashi, 1984, *J Bacteriol* **157:** 538–544). The best-documented report is that of Hara and Suzuki (1984, ibid.), who described a protein of $M_r = 34\,000$ from *E. coli* that could synthesise linear glycan strands from lipid-linked precursors. This type of TGase activity appears to have been largely ignored since the initial reports. We have recently identified open reading frames (ORFs) in *Neisseria gonorrhoeae* (and *Neisseria meningitidis*), *Klebsiella pneumoniae*, *E. coli* and *Haemophilus influenzae* that appear to encode biosynthetic peptidoglycan TGases that are not PBPs, and which may correspond to this activity.

The *K. pneumoniae* and *N. gonorrhoeae* putative TGase genes were identified within regions that were being sequenced for other reasons in our laboratories. In both cases, comparisons of the products of translation, in all reading frames, against the protein databases, using BLASTX, gave very highly significant matches against the amino-terminal regions of class A high-M_r PBPs (using the *K. pneumoniae* TGase, the number of matches expected by chance to be equal to, or greater than, those observed ranged from 2.6×10^{-36} to 4.4×10^{-17}). No significant matches to any other classes of proteins (including class B high-M_r PBPs or lytic transglycosylases) were obtained.

The *K. pneumoniae* protein is encoded by an ORF that is downstream of the *rpoN* operon. A similar gene (unassigned, ORF242; EMBL/GenBank/DDJB Nucelotide Sequence Data Libraries accession number U18997) is found downstream of the *E. coli rpoN* operon. Although, whereas in *K. pneumoniae* ORF242 is immediately adjacent to the last gene of the operon (*ptsO*; accession number Z50803), in *E. coli* another ORF (ORF210) of unknown function separates *ptsO* and ORF242. In both species the putative TGase gene is transcribed in the opposite direction to the *rpoN* operon and appears to be translationally coupled to an upstream gene described in *E. coli* as 'sigma cross-reacting protein 27A' (accession number D13188).

The *N. gonorrhoeae* (and *N. meningitidis*) protein is encoded by an ORF immediately downstream of the shikimate dehydrogenase gene (*aroE*). These genes are transcribed in the same direction with only two base pairs between the end of *aroE* and the start of the ORF. A search of the *H. influenzae* genome sequence (accession number L42023) revealed a homologous TGase gene (unassigned ORF, HI0831). The products of all of these genes were of similar size (about 240 residues) and none of them were associated with a TPase domain. We propose the genetic nomenclature *mtgA* (monofunctional TGase) for this class of gene. The sequences of the *K. pneumoniae* and *N. gonorrhoeae mtgA* genes have been submitted to the EMBL database as Z54198 and L47159, respectively.

Figure 1 shows the alignment of the putative TGases with themselves and with the TGase domain of PBP1A of *E. coli*. Of the 19 residues that are completely conserved in all class A high-M_r PBPs (Popham and Setlow, 1995, *J Bacteriol* **177:** 327–335), 15 are conserved in the four monofunctional TGases, and two are conservative replacements (lysine for arginine and serine for threonine). The carboxy-terminus of these proteins corresponds to a position in PBP1A that is approx. 115 residues before the active-site serine (the residue in the TPase domain that is acylated by penicillin) and presumably marks the boundary between the TGase and TPase domains of this bifunctional high-M_r PBP.

High-M_r PBPs are translocated across, and retained in, the cytoplasmic membrane by a hydrophobic amino-terminal, non-cleaved, signal-like sequence (Edelman *et al.*, 1987, *Mol Microbiol* **1:** 101–106). This feature is present in the putative TGases, suggesting they are also translocated across the cytoplasmic membrane. The TGase from *H. influenzae* appears to lack this feature,

```
Eco PBP1A   MKFVKYFLILAVCCILLGAGSIYGLYRYIEPQLPDVATLKDVRLQIPMQIYSADGELIAQ  60
Eco TGase   MSKSRLTVFSFVRRFLLRLMVVLAVFWGGGIALFSVAPVPFSAVMVERQVSAWLHGNFRY  60
Kpn TGase   MTFRFSARCRLIKRFLLRLLLACAVLWGGGVALFSIVPVPFSAVMLERQLGAWLSGNFHY  60
Hin TGase   ------------------------------------PIPFSAYMVQQKIANLLQGDFRY  >23
Ngo TGase       MFRIVKWLIALPVGIFIFFNAYVYGNIITYRAVA-PHRTAFMSMRMKQFEQEGRDV  55
                                                 "

Eco PBP1A   YGEKRRIPVTLDQIPPEMVKAFIATEDSRFYEHHGVDPVGIFRAASVALFSGHASQGAST  120
Eco TGase   VAHSDWVS--MDQISPWMGLAVIAAEDQKFPEHWGFDVASIEKALAHNERNENRIRGAST  118
Kpn TGase   IAHSDWVG--MDEISPWMGLAVIAAEDQKFPEHWGFDVPAIEKALAHNERNENRIRGAST  118
Hin TGase   QIQYNWVS--LENISPNIQLAVISSEDQRFLEHLGFDFEAIQRAIRYNEKSNKGIRGAST  >81
Ngo TGase   ALDYRWVP--YNRISTNLKKALIASEDVRFAGHGGFDWDGIQNAIRRNRNSGEVKAGGST  113
                 ""      *"     * *  **  *  * *"*   *  *   "        * **

Eco PBP1A   ITQQLARNFFLSPERTLMRKIKEVFLAIRIEQLLTKDEILELYLNKIYLGYRAYGVGAAA  180
Eco TGase   ISQQTAKNLFLWDGRSWVRKGLEAGLTLGIETVWSKKRILTVYLNIAEFGDGVFGVEAAA  178
Kpn TGase   LSQQTAKNLFLWDGRSWLRKGLEAGLTVGIETVWSKKRILTVYLNIAEFGEGTFGVEAAS  178
Hin TGase   ISQQTAKNLMLWHGQNWLRKGLEVPATMLLELTWSKKRILEVYLNIAEFGNGIFGVEAAS  >141
Ngo TGase   ISQQLAKNLFLNESRNYLRKGEEAAITAMMEAVTDKNRIFELYLNSIEWHYGVFGAEAAS  173
             "** *"*" *       **" *   "   *    * "*   ***  "   " "* "**

Eco PBP1A   QVYFGKTVDQLTLNEMAVIAGLPKAPSTFNPLYSMDRAVARRNVVLSRMLDEGYITQQQFDQTR... 244
Eco TGase   QRYFHKPASKLTRSEAALLAAVLPNPLRFKVSSPSGYVRSRQAWILRQMYQLGGEPFMQQHQLD    242
Kpn TGase   QRYFHKPASRLTAAEAALLAAVLPNPIRFRADAPSGYIRSRQAWILRQLGGEGFMRANQLH      239
Hin TGase   RYYFKKSAKNLSQNEAALLAAVLPNPIIYKVNKPSLLVRKKQTWILRQMGNLGTEYLSHL      >201
Ngo TGase   RYFYKKPAADLTKQQAAKLTALVPAPLYYADHPKSKRLRNKTNIVLRRMGSAELPESDTD      233
                 * "  *    "* " "  " *        "   "       *"
```

Fig. 1. Alignment of putative monofunctional TGases with the TGase domain of class A high-M_r PBPs. Amino acid identity between all five sequences is shown with asterisks; additional identities between the four monofunctional TGases are shown with double apostrophes. The doubly underlined amino acids in the PBP1A sequence are conserved in the TGase domain of all class A high-M_r PBPs. The run of hydrophobic amino acids that is believed to function as an amino-terminal non-cleaved signal-like sequence is underlined. Eco, *E. coli*; Kpn, *K. pneumoniae*; Hin, *H. influenzae*; Ngo, *N. gonorrhoeae*.

but examination of the sequence indicates that the reading frame extends upstream of the proposed initiation codon, and that there is a frame-shift error within the region encoding the amino-terminal membrane anchor. The probable start of this ORF is the ATG codon at position 880 963; a single base-pair deletion within the immediate downstream sequence would then give an amino-terminal sequence with the characteristics of a membrane anchor.

The striking homology of these proteins to the TGase domains of the class A high-M_r PBPs (Fig. 1) suggests that they are members of a new class of monofunctional biosynthetic peptidoglycan transglycosylases. It seems likely that the *mtgA* genes encode the TGases that are not PBPs, which have been identified biochemically. The deduced M_r of 27 324 for the *E. coli mtgA* product is similar to that reported for the *E. coli* TGase described by Hara and Suzuki (1984, ibid). Clearly, the ability of the purified putative TGases to synthesise linear glycan chains from lipid-linked precursors needs to be established, and these experiments are in progress. Similarly, the function of monofunctional TGases in peptidoglycan biosynthesis needs to be addressed.

Monofunctional TGases may be useful targets for novel antibiotics and for structural studies. Attempts to obtain crystals of water-soluble forms of the bifunctional class A high-M_r PBPs, or of their TGase domains, have been unsuccessful. The removal of the amino-terminal non-cleaved signal-like peptide of a monofunctional TGase, or its replacement with an authentic signal peptide, may result in a stable water-soluble TGase that is more amenable to crystallisation and structural studies than those derived from the high-M_r bifunctional class A high-M_r PBPs.

This work was supported by The Wellcome Trust and the Medical Research Council. M.J.M. acknowledges support from the BBSRC through a Grant-in-Aid to the Nitrogen Fixation Laboratory. B.G.S. is a Wellcome Trust Principal Research Fellow.

Brian G. Spratt,[1]* Jiaji Zhou,[1] Mark Taylor[2] and Michael J. Merrick[2]
[1]Molecular Microbiology Group, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.
[2]Nitrogen Fixation Laboratory, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK.
**For correspondence. E-mail b.g.spratt@sussex.ac.uk; Tel. (01273) 678309; Fax (01273) 678433.*

Received 27 September, 1995; accepted 16 October, 1995.

A multi-layered sensory system controls yeast glycolytic gene expression

Sir,

The investigation of mechanisms which enable living cells to sense the absence or presence of glucose in their surroundings is an important research topic. Microorganisms and cells of specialized tissues of higher eukaryotic organisms have developed regulatory systems to cope with rapidly changing nutritional supplies. Generally, these systems are composed of a sensing unit, a signal-transduction component and specific target mechanisms which adapt metabolism, growth characteristics and development.

In Gram-negative bacteria, glucose sensing is connected to transport-associated phosphorylation of glucose by the phosphotransferase system. In Gram-positive bacteria, glucose catabolism generates metabolic intermediates such as fructose-1,6-bisphosphate or gluconate-6-phosphate which allosterically activate a protein kinase III, triggering catabolite repression (Saier *et al.*, 1995, *Trends Biochem Sci* **20:** 267–271). In yeast cells, different signal-transduction pathways are involved in the regulation of a wide variety of processes directed towards the efficient utilization of glucose, but there is almost a complete lack of knowledge concerning the actual glucose-sensing mechanisms (Thevelein and Hohmann, 1995, *Trends Biochem Sci* **20:** 3–10). In mammalian pancreatic β-cells, glucokinase controls the flux of glucose into the cells, thereby functioning as the glucose sensor that couples changes in the extracellular glucose concentration to insulin secretion. The actual glucose-sensing process is connected to the metabolism of glucose (Efrat *et al.*, 1994, *Trends Biochem Sci* **19:** 535–538).

Apparently, there is an enormous degree of diversity in the mechanisms used by different organisms to sense glucose. Furthermore, as seems to be the case in yeast, individual organisms may have developed different glucose-sensing mechanisms for triggering different response pathways. Recently, we provided evidence for the existence of a complex sensory system in the yeast *Saccharomyces cerevisiae* which is involved in the induction of glycolytic gene expression in the presence of glucose (Boles *et al.*,1993, *Yeast* **9:** 761–770; Boles and Zimmermann, 1993, *Arch Microbiol* **160:** 324–328; Müller *et al.*, 1995, *J Bacteriol* **177:** 4517–4519).

Addition of fermentable sugars like glucose to yeast cells growing on other carbon sources triggers a wide variety of regulatory processes. The best-studied pathways, which are involved in the generation of the glucose-induced regulatory phenomena, are the glucose-repression pathway, which triggers repression of genes that are used to metabolize other carbon sources, and the Ras-adenylate cyclase pathway, which seems to be involved in triggering the rapid switch from a gluconeogenic to a fermentative mode of metabolism (Ronne, 1995, *Trends Genet* **11:** 12–17; Thevelein, 1994, *Yeast* **10:** 1753–1790). However, in both cases, the actual glucose-sensing mechanisms are not known.

Moreover, the switch to a fermentative mode of metabolism in the presence of glucose also includes the induction of genes coding for enzymes of the second part of glycolysis. The mechanisms of this glucose-induced gene expression are not yet known. However, using a set of yeast strains deleted for various genes coding for enzymes involved in sugar catabolism, we demonstrated that the extent to which glucose has to be metabolized in order to obtain full expression of the glycolytic enzymes enolase II, pyruvate kinase, pyruvate decarboxylase and alcohol dehydrogenase I is different for the different enzymes (Boles *et al.*,1993, ibid.; Boles and Zimmermann,

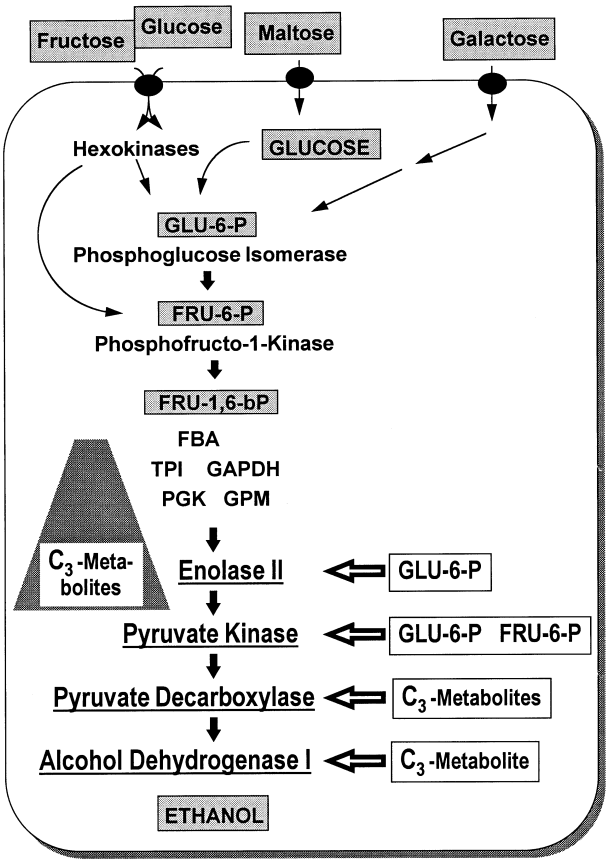


Fig. 1. Scheme of the proposed model for yeast glycolytic gene induction triggered by formation of different glycolytic metabolites.

1993, ibid.; Müller *et al*., 1995, ibid.). Based on our data, we have proposed the existence of distinct metabolic intermediates or even combinations of intermediates that trigger the differential induction of the glycolytic enzymes.

Even more intriguing, enzymes operating at an earlier step in glycolysis are induced by metabolic signals originating at an early stage of the glycolytic pathway, while the enzymes operating in the conversion of pyruvate to ethanol are induced by signals originating at the end of the glycolytic pathway. In this way, metabolism of glucose to glucose-6-phosphate is necessary and sufficient to trigger induction of enolase II, while full induction of pyruvate kinase requires the formation of glucose-6-phosphate as well as of fructose-6-phosphate. On the other hand, an increase in the concentrations of all three-carbon glycolytic intermediates leads to the induction of pyruvate decarboxylase, suggesting that a specific complement of metabolic intermediates serve as the actual signal. Finally, accumulation of one specific metabolic intermediate at the late stages of glycolysis is sufficient to trigger induction of alcohol dehydrogenase I.

The existence of such a hierarchical order of different metabolic signals offers the cell a sophistical means to react properly to the presence of different sugars and sugar concentrations under a variety of environmental conditions. Therefore, we would not be surprised if similar regulatory systems are operating in other organisms.

Eckhard Boles,* Susanne Müller and Friedrich K. Zimmermann
Institut für Mikrobiologie und Genetik, Technische Hochschule Darmstadt, Schnittspahnstrasse 10, D-64287 Darmstadt, Germany.
**For correspondence. E-mail boles@bio1.bio.th-darmstadt.de; Tel. (6151) 164476; Fax (6151) 164808.*
Received 12 September, 1995; revised 21 September, 1995; accepted 25 September, 1995.

Nomenclature of the genes encoding IHF

Sir,

We wish to change the names of the genes that encode the two subunits of Integration Host Factor (IHF) of *Escherichia coli*. The official names have been *himA* for the alpha subunit, and *himD* for the beta subunit. However, *hip*, the name of the first mutations affecting IHF-beta, is a widely used alternative to *himD*. The new names will be *ihfA* and *ihfB*, for the genes encoding IHF-alpha and IHF-beta, respectively. We have several reasons for making this change. First, it will remove the present confusion over the name of the gene for IHF-beta. Secondly, it respects the history of the field, because biochemical studies of IHF protein preceded, and to a large extent inspired, the isolation of virtually all of the existing *ihf* mutants. Thirdly, the new names are appropriately mnemonic.

We appreciate that IHF probably has functions in *E. coli* that are more central to the life of the cell than is prophage integration, and that neither the old nor the new names are suggestive of these (as yet largely unknown) functions. Nevertheless, in view of the extensive biochemical and physical characterization of this protein, and its wide distribution in prokaryotes, it is extremely unlikely that the name IHF will be abandoned as new functions are found.

Robert A. Weisberg,[1]* Martin Freundlich,[2] David Friedman,[3] Jeffrey Gardner,[4] Nora Goosen,[5] Howard Nash,[6] Amos Oppenheim[7] and Josette Rouvière-Yaniv[8]
[1]*Building 6B, Room 308, National Institutes of Health, Bethesda, Maryland 20892-0278 USA.*
[2]*Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, New York 11794, USA.*
[3]*Department of Microbiology, University of Michigan, Ann Arbor, Michigan 48109, USA.*
[4]*Department of Microbiology, University of Illinois, Urbana, Illinois 61801, USA.*
[5]*Laboratory of Molecular Genetics, Leiden Institute of Chemistry, PO Box 9502, 2300 RA Leiden, The Netherlands.*
[6]*Building 36, Room 1B08, National Institutes of Health, Bethesda, Maryland 20892, USA.*
[7]*Department of Molecular Genetics, Hadassah Medical School, POB 12272, Jerusalem 91120, Israel.*
[8]*Institut de Biologie, Physico-Chemique, 13 rue P. et M. Curie, 75005 Paris, France.*
**For correspondence. E-mail wia@cu.nih.gov; Tel. (301) 496 3555; Fax (301) 496 0243.*
Received 12 December, 1995; accepted 18 December, 1995.

The Gram-negative cell envelope 'springs' to life: coiled-coil *trans*-envelope proteins

Sir,

The *Escherichia coli* haemolysin D protein (HlyD) and its prokaryotic homologues in Gram-negative bacteria form a subfamily of MFPs (membrane fusion proteins), each having a single transmembrane domain and a large external, periplasmic domain (Dinh *et al*., 1994, *J Bacteriol* **176:** 3825–3831). Such proteins function in a variety of secretion and excretion processes, energized by an inner membrane partner (ABC protein HlyB with HlyD), to translocate a wide variety of compounds, from drugs to large polypeptides, directly to the culture medium, bypassing the periplasm. Precisely how such MFPs span the inner and outer membranes and function in *trans*-envelope transport is not understood, although in the case of the HlyD family, whose members promote protein secretion, it is clear that there is a third essential element of the transporter, namely the outer membrane protein TolC (and its homologues) (Létoffé *et al*., 1993, *J Bacteriol* **175:** 7321–7328). In addition, in the case of distant homologues like AcrA, which excrete drugs (Ma *et al*., 1995, *Mol Microbiol* **16:** 45–55), it has been argued that some form of tight contact with the outer membrane is essential to ensure that drug molecules are ejected directly to the medium, avoiding easy re-entry across the inner membrane.

We have detected the presence of an unexpected structural motif in HlyD and its closest relatives that may provide an important clue to their mechanisms of action. HlyD is specifically involved in the secretion of α-haemolysin, a 110 kDa polypeptide. At least 12 other homologues with a similar function have been identified in Gram-negative bacteria (Blight *et al*., 1994, *Curr Opp Biotech* **5:** 468–474). The HlyD polypeptide can be divided into four readily distinguishable regions: a 58-amino-acid N-terminal cytoplasmic domain, a 20-amino-acid inner membrane anchor, a 20 kDa central region predicted to be almost entirely α-helical, and an approximately 15 kDa, C-terminal β-strand, loop domain.

We have recently discovered, using the MACSTRIPE (Version1.3.1) program (Dr A. Knight, MIT) and the coiled-coil prediction algorithm of Lupas *et al*. (1991, *Science* **252:** 1162–1164), that HlyD (478 amino acids, aa) is predicted with high probability to contain coiled-coil structures (moving window = 28). The predicted coiled-coil is located within the central α-helical region of HlyD (aa 181–319) at position 249–299 (probability > 0.90). This domain contains a central region of 41 amino acids (253–293) with a coiled-coil probability greater than 0.99. This is considered to be a highly significant prediction since the shortest stable coiled-coil motifs are between 28 to 35 amino acids (Lupas *et al*., 1991). A shorter (28 aa) second coiled coil is predicted between residues 125 and 152 (probability 0.82), upstream of the central helical region but within a region rich in predicted α-helical structures.

Of the seven closest relatives of HlyD (all involved in protein secretion) for which sequence data are available, five are predicted to form coiled-coil structures with similar probability and at a similar position (see Fig. 1). Of the remaining two proteins, CvaA has two coiled-coil regions also predicted with lower probability (0.84 and 0.81) and CyaD has one coiled-coil region predicted with lower probability (0.88). CyaD and CvaA are in fact the most divergent of this group in overall sequence, and colicin V, the transport substrate for CvaA displays a unique structure and secretion signal compared with other proteins secreted by this pathway. This, in turn, may affect the requirement for coiled coils. We have also analysed the sequences of three other members of the MFP family, AcrA, NolF and EmrA, all of which transport small non-peptide molecules, each one functioning with a different type of inner membrane partner unrelated to HlyB. Predicted coiled-coil motifs of 40–50 aa were also detected towards the linear centre of these molecules, with probabilities of 0.96, 0.83 and 0.92, respectively. Although less clear than for HlyD, these molecules may also form coiled-coil structures important to their role as *trans*-envelope transporters.

The significance of predicted coiled-coil probabilities lower than 0.90 for prokaryotic monotopic membrane proteins is difficult to assess in the light of the data set used to compute the prediction algorithm. Lupas *et al*. (1991) used three major coiled-coil protein families (keratins, tropomyosins, and myosins) to develop an algorithm based upon the Gaussian distributions of coiled-coil motifs in GenBank or the coiled-coil protein data set. Application of the algorithm to other sequences revealed six prokaryotic coiled-coil proteins. However, the coiled-coil protein data set was not reconstructed with these additional prokaryotic sequences. Therefore, a possible bias may exist towards eukaryotic sequences, potentially diminishing probabilities within prokaryotic proteins.

In view of the above discussion, it is interesting to note that RNase E has been described as the *E. coli* 'myosin-like' protein (Casarégola *et al*., 1992, *J Mol Biol* **228:** 30–40) because of its cross-reaction with yeast (MYO1) monoclonal antibody and, in turn, the cross-reaction of RNase E specific antibody with smooth muscle and yeast myosins. Furthermore, the polynucleotide phosphorylase (PNPase) of *Photorhabdus* sp. has been reported to have amino acid sequence homology to the human myosin heavy chain (Clarke and Dowds, 1994, *J Bacteriol* **176:** 3775–3784). Both *E. coli* RNase E and PNPase are predicted to have regions of 32 (window = 28) and 28 (window = 21) amino acids with

Fig 1. Alignment of the amino acid sequences for coiled-coil motifs for members of the HlyD-like family. Sequences are numbered from internal positions as indicated to the left of the alignment. Residues within regions having coiled-coil probabilities of > 0.80 and > 0.99 are indicated by grey and black boxes, respectively. The lower sequence ('Predprot') indicates the secondary-structure prediction for HlyD according to the PREDICT PROTEIN program of Burkard and Sanders (1993), *J Mol Biol* **232:** 584–599. The proteins within the alignment include; AprE (*Pseudomonas aeruginosa)*, PrtE (*Erwinia chrysanthemi*), LktD (*Pasteurella haemolytica*), AppD (*Actinobacillus pleuropneumoniae*), HlyD (*Escherichia coli*), LktD (*Actinobacillus actinomycetemcomitans*), CyaD (*Bordetella pertussis*), and CvaA (*Escherichia coli*).

greater than 0.80 and 0.96 coiled-coil probabilities, respectively. Moreover, these proteins have been shown to interact by co-immunoprecipitation and co-purification (Carpousis *et al*., 1994, *Cell* **76:** 889–900; Py *et al*., 1994, *Mol Microbiol* **14:** 717–729). It therefore seems likely that despite apparently low (0.80–0.90) coiled-coil probabilities of a magnitude similar to those of the HlyD and MFP families, complexes between prokaryotic proteins, possibly formed through potential coiled-coil domains, have been identified.

Coiled-coil structures, normally associated with extended filaments of cytoskeletal proteins, are apparently extremely rare in prokaryotes (Hurme *et al*., 1994, *J Biol Chem* **269:** 10675–10682). An interesting and abundant prokaryotic coiled-coil protein is the *E. coli* lipoprotein (Lpp), which links the peptidoglycan to the outer membrane. The presence of such structures in HlyD and its homologues raises very interesting possibilities in relation to their function. HlyD is apparently expressed in several fold excess over its membrane partner, HlyB (an ABC membrane ATPase). It has been suggested that HlyB, like other ABC ATPases, functions as a dimer and we propose, therefore, that HlyD stably assembles as a higher order multimer (dependent upon HlyB and TolC; our manuscript in preparation). The presumed presence of short coiled-coil motifs in HlyD may therefore provide two potential structural features.

(i) Many extended filaments may associate to form a major part of the side walls of a transport chamber.

(ii) A 'spring-like' structure may be capable of limited

expansion or compaction to accommodate different steps in haemolysin translocation.

Fluctuations in the volume of the periplasm, in response to environmental changes, would be expected to alter the distance between the inner and outer membranes. Therefore a *trans*-envelope structure would be logically required to respond to such variations in order to maintain integrity.

From our analysis, the 'spring' structure appears to be consistently predicted to be within a component of the Gram-negative *trans*-envelope transporters. We propose that the ability of such complexes to accommodate volumetric changes in the periplasm is implicit for function, and potential coiled-coil structures might be required for the necessary internal conformational changes.

Andréa Pimenta, Mark Blight, David Clarke and I. Barry Holland*
Institut de Génétique et Microbiologie, URA 1354, CNRS–Bâtiment 409, Université de Paris XI, 91405 Orsay Cedex, France.
**For correspondence. E-mail Holland@igmors.u-psud.fr; Tel. (1) 69 41 66 99; Fax (1) 69 41 78 08.*
Received 15 August, 1995; accepted 1 September, 1995.

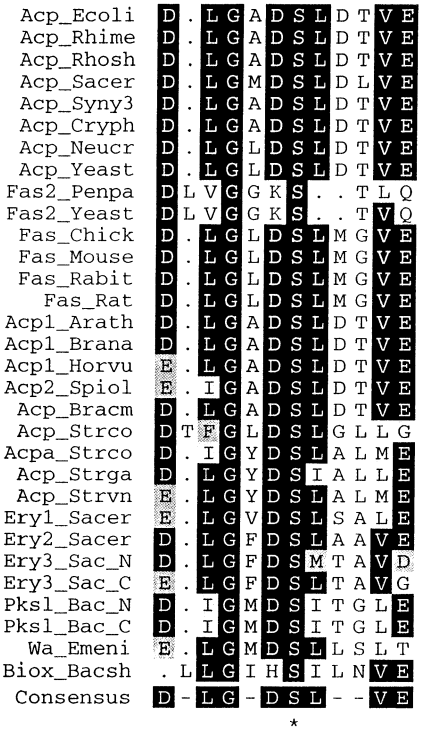
To be free or not: the fate of pimelate in *Bacillus sphaericus* and in *Escherichia coli*

Sir,

In a recent publication (Sanyal *et al*., 1994, *J Am Chem Soc* **116:** 2637–2638) hypothetical metabolic steps in the cellular biosynthesis of biotin were discussed. We wish to point out some additional facts which indicate that the conclusions reached by these authors should be modified slightly.

The biosynthetic pathway leading to pimeloyl-CoA, an important precursor of biotin or vitamin H, has not yet been elucidated. More than 30 years ago, *in vivo* biotin labelling experiments led F. Lynen (Lezius *et al*., 1963, *Biochem Z* **336:** 510–525) to propose a mechanism similar to that catalysed by fatty acid synthase, but with successive condensations of three malonyl-CoA units. The addition of pimelate has been shown to enhance biotin production in certain microorganisms. Indeed, a pimeloyl-CoA synthase has been purified and characterized in *Bacillus sphaericus* (Ploux *et al*., 1992, *Biochem J* **287:** 685–690). The corresponding genetic determinant, *bioW*, has been cloned and shown to be located in an operon together with the *bioF* and *bioX* genes (Gloeckler *et al*., 1990, *Gene* **87:** 63–70). Sequence analyses of the *bioX* gene revealed a consensus phosphopantetheine attachment site (Gloeckler *et al*., 1990, *Gene* **87:** 63–70). More detailed comparison (see Fig. 1A) with members of the acyl carrier protein family (ACP) from fatty acid and polyketide complexes fully confirm this initial observation. This leads us to propose that *bioX* encodes a specific ACP involved in pimeloyl-CoA synthesis. This ACP may interact directly with the fatty acid synthase complex of *B. sphaericus*. Malonyl-CoA extender units could then be directed towards pimeloyl-CoA synthesis, at a level depending on the cellular concentration of BIOX protein and/or its competition with the well-known fatty acid ACP. An unusual feature of BIOX is its deduced length of 166 amino acids, which is much larger than the average length (80 amino acids) of ACPs. A search for consensus sequences linked to other functions shared by the fatty acid and polyketide synthases (Fernandez-Moreno *et al*., 1992, *J Biol Chem* **267:** 19278–19290) did not reveal any significant match other than the identified ACP signature. An additional activity carried by this new ACP would be to restrict the acyl transfer to a starter malonyl-CoA unit and to limit the condensation steps up to two extender units only, possibly acting like a chain-length determining factor identified in polyketide synthases (McDaniel *et al*., 1994, *Proc Natl Acad Sci USA* **91:** 11542–11546). Furthermore, genetic constructions and functional analysis of hybrid polyketide synthases containing heterologous ACPs have revealed subtle modifications in the nature of the end-products (Khosla *et al*., 1993, *J Bacteriol* **175:** 2197–2204). A fatty acid thioesterase would release free pimelate before its condensation with Coenzyme A, this reaction being catalysed by the *bioW* gene product. At this stage, however, we cannot exclude the possibility that all or part of the other activities, i.e. malonyl-CoA acyltransferase, β-ketoacyl synthase, β-ketoacyl reductase, β-hydroxyacyl dehydratase and enoyl reductase, are different from those of the classical fatty acid biosynthetic complex(es). In *Escherichia coli*, extensive screenings for *bio* auxotrophic mutants identified only two loci *bioC* and *bioH* as being likely to encode proteins specifically involved in pimeloyl-CoA synthesis (Barker and Campbell, 1980, *J Bacteriol* **143:** 789–800). Strikingly, pimelate does not enhance biotin productivity by this host, although *E. coli* cells are clearly permeable to this compound (Ploux *et al*., 1992, *Biochem J* **287:** 685–690). Moreover, the *bioC* and *bioH* sequences (Otsuka *et al*., 1988, *J Biol Chem* **263:** 19577–19580; O'Regan *et al*., 1989, *Nucl Acids Res* **17:** 800) do not seem to be related to either *bioW* or *bioX*, in contrast with the other *bio* genes products which show an

A



B

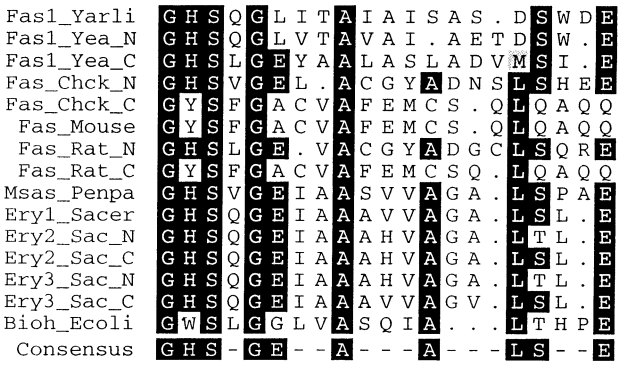


Fig. 1. Alignments from the results of a motif search in the SwissProt database (Release 29.0, 6/94) using programs of the GCG program package (Version 8, 9/1994, Genetics Computer Group, Madison, Wisconsin). Alignments were generated from the output of a MOTIFS search with the program PILEUP and then refined with LINEUP.

A. ACP motif 4′-phosphopantheine-binding site (active serine labelled with *) containing the [(D,E)(X){0,1}(F,I,L,V)GX(D,H,K)-S(I,L,M){0,1}XX{1,2}(L,M,V)(D,E,G,Q,T)] consensus sequence found in different ACPs. Prokaryotes: Acp_Ecoli, P02901 (31–42); Acp_Rhime, P19372 (31–42); Acp_Rhosh, P12784 (31–42); Acp_Sacer, P11830 (34–45); Acp_Syny3, P20804 (33–44). Plant chloroplasts: Acp_Cryph, P29189 (34–45). Fungi: Acp_Neucr, P11943 (86–97); Acp_Yeast, P32463 (77–88); Fas2_Penpa, P15368 (167–178); Fas2_Yeast, P19097 (174–185). Higher Eukaryotes: Fas_Chick, P12276 (2079–2090); Fas_Mouse, P19096 (479–490); Fas_Rabit, P07855 (33–44); Fas_Rat, P12785 (2146–2157). Plants: Acp1_Arath, P11829 (88–99); Acp1_Brana, P10352 (85–96); Acp1_Horvu, P02902 (99–110); Acp2_Spiol, P23235 (82–93); Acp_Bracm, P07088 (85–96). Prokaryotic polyketide synthases: Acp_Strco, P23153 (41–52); Acpa_Strco, Q02054 (37–48); Acp_Strga, P12884 (36–47); Acp_Strvn, P12885 (36–47); Ery1_Sacer, Q03131 (1916–1927); Ery2_Sacer, Q03132 (3443–3454); Ery3_Sac_N, Q03133 (1419–1430); Ery3_Sac_C, Q03133 (2847–2858); Pksl_Bac_N, Q05470 (2518–2529); Pksl_Bac_C, Q05470 (2659–2670); Wa_Emeni, Q03149 (1799–1810). BIOX_Bacsh: P22821 (68–79).

B. Thioesterase/acyl transferase (active serine labelled with *) motifs containing the consensus sequence [G(H,W,Y)S(F,L,Q,V)G(x){2,3}A(X){8,11}(E,Q)]. Fungi: Fas1_Yarli, P34229 (274–294); Fas1_Yea_N, P07149 (272–292); Fas1_Yea_C, P07149 (1806–1826). Higher Eukaryotes: Fas_Chck_N, P12276 (504–524); Fas_Chck_C, P12276 (2233–2253); Fas_Mouse, P19096 (633–653); Fas_Rat_N, P12785 (579–599); Fas_-Rat_C, P12785 (2300–2320). Fungal polyketide synthases: Msas_Penpa, P22367 (651–671). Prokaryotic polyketide synthases: Ery1_Sacer, Q03131 (2596–2616); Ery2_Sac_N, Q03132 (649–669); Ery2_Sac_C, Q03132 (2103–2123); Ery3_Sac_N, Q03133 (640–660); Ery3_Sac_C, Q03133 (2105–2125). BIOH_Ecoli: P13001 (80–100).

average of 50% amino acid identity (including the conservative replacements) among very different species (Gloeckler *et al.*, 1990, *Gene* **87:** 63–70; I. Kuhn *et al.*, unpublished results). This led Sanyal *et al.* to speculate that a *bioX* equivalent must exist in *E. coli*, awaiting further characterization.

A detailed analysis has indeed not allowed us to identify a putative phosphopantetheine attachment site in BIOC protein. In contrast to the model of Sanyal *et al.*, we propose that the *bioC* gene product catalyses the stepwise condensation of the malonyl-CoA starter group by addition of acetate units from malonyl groups in a similar way to chalcone synthase (Lanz *et al.*, 1991, *J Biol Chem* **266:** 9971–9976), which does not possess the 4′-phosphopantetheine group residue characteristic of ACP function. Again, the cyclic reductive and dehydration steps could be shared with (one of) the cellular fatty acid synthase complexes. We have identified a consensus sequence in BIOH protein around an active serine residue which is characteristic of acyltransferase and thioesterase proteins (Fig. 1B). A possible role for the *bioH* gene product would be to transfer pimeloyl from the active cysteinyl residue of BIOC directly to Coenzyme A, preventing the accumulation of free pimelate inside the cells. To the best of our knowledge, this is the first time that enzymatic functions have been proposed for the *bioC* and *bioH* gene products. Medium/high-chain classical fatty acids represent probably the major end-products synthesized by the fatty

acids complexes, unlike pimelate/pimeloyl-CoA. If our model is correct, re-routing of the metabolic flux towards pimelate/pimeloyl-CoA would represent a major challenge for the biotechnology industry in terms of production of biotin by biological means. In the case of *B. sphearicus*, a thermosensitive mutant specifically altered in the classical ACP would allow a transient increase in the metabolic flow via the specific pimeloyl-ACP at restrictive temperatures. The concentration of pimeloyl-ACP could also be increased by the use of a tightly regulated expression system, provided that no enzymatic inhibition by vitamer end-products were to act at this level.

The support of Yves Boulanger throughout this study was much appreciated. We are also indebted to Peter Philippsen for his generous help.

Yves Lemoine,[1*] Achim Wach[2] and Jean-Marc Jeltsch[1]
[1]*Biotechnology Unit, Ecole Supérieure de Biotechnologie de Strasbourg, Université Louis Pasteur, Parc d'Innovation, Boulevard Sébastien Brant, 67400 Illkirch, France.*
[2]*Institute for Applied Microbiology, Biocentrum, Universität Basel, Klingelbergstraβe 70, CH 4056 Basel, Switzerland.*
**For correspondence. E-mail lemoine@esbs.u-strasbg.fr; Tel (33) 88655337; Fax (33) 88655249.*

Received 14 September, 1995; revised 25 September, 1995; accepted 28 September, 1995.