Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 7411–7413, August 1993
Biochemistry

Cloning and characterization of a cDNA encoding glyoxal oxidase, a H_2O_2-producing enzyme from the lignin-degrading basidiomycete *Phanerochaete chrysosporium*

(hydrogen peroxide/peroxidase/fungus/galactose oxidase)

PHILIP J. KERSTEN*† AND DANIEL CULLEN*‡

*Institute for Microbial and Biochemical Technology, Forest Products Laboratory, U.S. Department of Agriculture, Madison, WI 53705; and †Biotechnology Center and ‡Bacteriology Department, University of Wisconsin, Madison, WI 53706

Communicated by T. Kent Kirk, May 18, 1993 (received for review March 8, 1993)

ABSTRACT **Glyoxal oxidase is produced by ligninolytic cultures of the white-rot fungus *Phanerochaete chrysosporium* and is a source of the extracellular H_2O_2 that is required by ligninolytic peroxidases. We report here the cloning and characterization of glx-1c cDNA, which encodes glyoxal oxidase. The deduced mature protein has 537 amino acids, a molecular size of 57 kDa, and a pI of 5.1. Five potential N-glycosylation sites are present. The predicted N-terminal sequence is identical to the experimentally determined sequence of purified enzyme and is preceded by a leader peptide of 22 amino acids. The sequence of glx-1c lacks significant homology with known sequences. Specific comparisons were made between the glx-1c translated sequence and that of galactose oxidase from *Dactylium dendroides* because of previously observed catalytic similarities of the enzyme. Although no significant homology is observed, in both cases extensive ß-sheet regions are predicted from the primary sequences. Glyoxal oxidase activity correlates with transcript levels and is also coordinate with the lignin peroxidases in nutrient nitrogen-starved cultures.**

Considerable progress has been made in recent years on the biochemistry and molecular genetics of lignin biodegradation. The major degraders of lignin are the basidiomycetous fungi, of which the white-rot fungus *Phanerochaete chrysosporium* is the species most studied for its lignin-degrading ability. *P. chrysosporium* has been shown to produce three classes of extracellular enzymes under ligninolytic (secondary metabolic) conditions: lignin peroxidases, manganese peroxidases, and glyoxal oxidases (GLOX) (for review see refs. 1 and 2).

The presumed role of GLOX is to supply the H_2O_2 required as oxidant in the catalytic cycles of the extracellular peroxidases (3). GLOX is nonspecific; substrates include formaldehyde, acetaldehyde, glycolaldehyde, glyoxal, glyoxylic acid, dihydroxyacetone, glyceraldehyde, and methylglyoxal. GLOX is expressed when P. chrysosporium is grown on glucose or xylose, the major sugar components of lignocellulosics. However, the physiological substrates for GLOX are not these growth carbon compounds but rather intermediary metabolites including glyoxal and methylglyoxal. Thus, the organism produces H_2O_2 without producing an assortment of extracellular sugar oxidases. In addition, products of ligninolysis, including glycolaldehyde (4), are also substrates for GLOX; this suggests that once the oxidase system is "primed" with substrates derived from carbohydrates, the system can, in part, be perpetuated by the action of lignin peroxidase on lignin itself.

GLOX has been purified to homogeneity, and biochemical properties have been determined (5); it is a glycoprotein, has a molecular size of 68 kDa (by SDS/PAGE), has two isozymic forms of pI 4.7 and 4.9, and, if purified, is activated by Cu^{2+}. Furthermore, the activity of glyoxal oxidase is modulated in response to peroxidase systems in vitro, suggesting a possible mechanism for control of the extracellular oxidase–peroxidase systems in vivo. The physiochemical basis for this extracellular regulation has not been elucidated.

Considerable progress has been made on the molecular genetics of P. chrysosporium peroxidase gene families (6, 7), although nothing has been reported concerning the number, structure, genomic organization, or regulation of the gene encoding GLOX. As a first step toward elucidating the molecular genetics of GLOX, we report here the cloning and sequence of a cDNA clone encoding GLOX.[§]

MATERIALS AND METHODS

Organisms. *P. chrysosporium* strain BKM-F-1767 (ATCC 24725) was grown in nitrogen-limited shake cultures, which were supplemented with additives on day 2 as described earlier (5). *Escherichia coli* strain DH5α was the host for plasmid pBluescript (Stratagene). *E. coli* Y1090 was used for transfections with λgt11.

Enzyme Assay. GLOX activity was determined as described previously using methylglyoxal as the oxidase substrate and phenol red as the substrate for horseradish peroxidase in the coupled-enzyme assay (3).

Antibody Production and Purification. Purified GLOX (5) was used for the production of antiserum in rabbits. Polyclonal antibodies were partially purified with Econo-Pac serum IgG purification columns (Bio-Rad).

N-Terminal Analysis of GLOX. N-terminal sequence was determined at the University of Wisconsin Biotechnology Center, Madison. Two determinations were made with no contradictory assignments. The sequence is APGWRFDLKPNLSGIVALEAIVVXsslvvi (uppercase letters indicate high confidence; lowercase letters indicate moderate confidence). Enzyme for these determinations was purified essentially as for antibody production (addressed previously), except that fractionation on a phenyl-Sepharose CL-4B (Sigma) column [0.8 M–0M $(NH_4)_2SO_4$ gradient] preceded that on the DEAE Biogel column.

cDNA Library. RNA was purified from day 8 nitrogen-starved cultures (8), and cDNAs were synthesized from poly(A)$^+$ RNA by oligo(dT) priming. *Eco*RI adapters were ligated to cDNAs, and a library was constructed in lgt11 (Promega). Approximately 10^5 clones were screened with

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: glx-1c, a cDNA encoding glyoxal oxidase; GLOX, glyoxal oxidase.

[§]The sequence reported in this paper has been deposited in the GenBank data base (accession no. L18991).

antibody against glyoxal oxidase with alkaline phosphatase conjugant (Promega) for detection.

DNA Sequencing and Analysis. A full-length λgt11 clone designated glx-1c was subcloned into Bluescript KSII+ (Stratagene) and into M13mp18 and M13mp19 (New England Biolabs). Templates were sequenced by the dideoxynucleotide chain-termination method (9) using Sequenase (United States Biochemical). Both strands were sequenced. The databases European Molecular Biology Laboratory/GenBank (Release 74), National Biomedical Research Foundation/Protein Identification Resource (Release 34), and SwissProt (Release 23) were searched for GLOX sequence similarities by the method of Wilbur and Lipman (10) using DNASTAR software (Madison, WI).

RNA Analysis. Total RNA was purified from ligninolytic shake flasks (5) on days 2, 4, 6, and 8. Formaldehyde gels were probed with a full-length cDNA (1.3 kb) encoding lignin peroxidase isozyme H8 and a 1.2-kb *Kpn* I fragment of the glx-1c cDNA. High-stringency hybridizations were used, although closely related transcripts may have contributed to the hybridization signal obtained for the lignin peroxidase probe.

RESULTS

Nucleotide Sequence and Analysis of glx-1c. The cDNA library constructed in λgt11 was screened with antibody raised against GLOX. Approximately 0.03% of the clones were immunologically positive. Additional plaque lifts were made from the library and screened for lignin peroxidase and manganese peroxidase clones using antibodies against the corresponding enzymes; the frequency of positives was ≈0.11% and 0.06%, respectively. Of ≈20 immunologically positive GLOX clones analyzed, four extended beyond the translational start codons.

The sequence of one full-length cDNA, glx-1c, and the predicted translation product are shown in Fig. 1. The experimentally determined N terminus (30 amino acids with one indeterminate amino acid at position 24) matches without contradiction the deduced amino acid sequence beginning with alanine (position l), thus providing the most conclusive evidence that the clone encodes GLOX. The coding region of 559 amino acids indicates that the mature peptide is preceded by a signal peptide of 22 amino acids, as would be expected for an extracellular protein. Characteristic of signal sequences, the N terminus has a net positive charge, and the core predominantly consists of hydrophobic amino acids (leucine, alanine, and valine). However, the proteolytic processing at the acidic aspartate (−1 position) of the C terminus is atypical (11, 12).

The molecular size of 57 kDa (537 amino acids) for the predicted mature peptide is in reasonable agreement with the experimental determination of 68 kDa (5). The molecular size difference presumably represents the carbohydrate content of the glycoprotein; five potential N-glycosylation sites (Asn-Xaa-Ser/Thr) are indicated at Asn^{11}, Asn^{24}, Asn^{78}, Asn^{209}, and Asn^{399}. Glycosylation at Asn^{11} seems unlikely because N-terminal sequence analysis unambiguously identified the amino acid. However, the failure to identify the N-terminal amino acid residue at position 24 might have been due to the glycosylation of asparagine.

A pI of 5.1 for the mature peptide is predicted from the cDNA sequence, ignoring any possible secondary ionic effects, such as contributions of the carbohydrate component. This pI is also in reasonable agreement with the experimental observations of pI 4.7 and 4.9 for the isozymes.

As with many other fungal genes (13), the canonical eukaryotic polyadenylylation signal AATAAA is not observed in glx-1c. However, the sequence 5′-GCATATGAAATATC-TGT-3′ 10 bp upstream of the poly(A) tail in glx-1c is highly homologous to 3′-untranslated regions in lignin peroxidase

```
AGACTTACGATGTTGTCGCTGCTAGCCGTAGTCTCACTCGCCGCCGCCACACTCGCTGCT  60
         MetLeuSerLeuLeuAlaValValSerLeuAlaAlaAlaThrLeuAlaAla  -6
         -22
                 ↓beginning of mature peptide
CCAGCTGCCAGCGATGCACCAGGCTGGCGGTTCGACCTGAAGCCCAACCTTTCCGGGATC  120
ProAlaAlaSerAspAlaProGlyTrpArgPheAspLeuLysProAsnLeuSerGlyIle  15

GTGGCTCTCGAGGCGATCGTCGTGAACAGCTCACTGGTGGTCATATTCGACCGCGCCACC  180
ValAlaLeuGluAlaIleValValAsnSerSerLeuValValIlePheAspArgAlaThr  35

GGCGATCAGCCTCTCAAGATCAACGGCGAATCGACCTGGGGTGCGCTCTGGGATCTCGAT  240
GlyAspGlnProLeuLysIleAsnGlyGluSerThrTrpGlyAlaLeuTrpAspLeuAsp  55

ACGAGCACTGTCCGCCCGCTTTCGGTCCTGACGGACTCCTTCTGCGCCAGCGGTGCGCTG  300
ThrSerThrValArgProLeuSerValLeuThrAspSerPheCysAlaSerGlyAlaLeu  75

CTCAGCAATGGCACTATGGTCAGCATGGGAGGAACCCCTGGCGGTACTGGAGGGGATGTT  360
LeuSerAsnGlyThrMetValSerMetGlyGlyThrProGlyGlyThrGlyGlyAspVal  95

GCTGCTCCTCCTGGCAACCAGGCGATCCGGATCTTCGAGCCCTGCGCCTCGCCTTCGGGA  420
AlaAlaProProGlyAsnGlnAlaIleArgIlePheGluProCysAlaSerProSerGly  115

GATGGCTGCACGCTCTTCGAAGACCCTGCGACTGTTCATCTCCTTGAGGAGCGGTGGTAC  480
AspGlyCysThrLeuPheGluAspProAlaThrValHisLeuLeuGluGluArgTrpTyr  135

CCATCGTCCGTGCGCATTTTCGATGGCAGTCTCATGATCATTGGTGGTTCGCATGTCCTC  540
ProSerSerValArgIlePheAspGlySerLeuMetIleIleGlyGlySerHisValLeu  155

ACACCGTTCTACAACGTCGATCCAGCCAACTCCTTCGAGTTCTTCCCAAGCAAGGAGCAG  600
ThrProPheTyrAsnValAspProAlaAsnSerPheGluPhePheProSerLysGluGln  175

ACGCCCAGGCCTTCAGCCTTCTTGGAACGCTCGCTGCCAGCGAATCTTTTCCCACGGGCT  660
ThrProArgProSerAlaPheLeuGluArgSerLeuProAlaAsnLeuPheProArgAla  195

TTCGCCTTGCCAGACGGAACCGTCTTTATCGTCGCTAATAACCAGTCTATCATCTACGAT  720
PheAlaLeuProAspGlyThrValPheIleValAlaAsnAsnGlnSerIleIleTyrAsp  215

ATCGAGAAGAATACTGAGACCATCCTCCCGGATATCCCCAACGGTGTTCGTGTCACGAAC  780
IleGluLysAsnThrGluThrIleLeuProAspIleProAsnGlyValArgValThrAsn  235

CCCATCGATGGAAGCGCCATCCTTCTCCCGTTGTCGCCCCCTGACTTCATTCCCGAGGTT  840
ProIleAspGlySerAlaIleLeuLeuProLeuSerProProAspPheIleProGluVal  255

CTCGTCTGTGGAGGCTCGACTGCGGACACCTCGCTCCCGTCCACCAGCTTGTCCTCCCAA  900
LeuValCysGlyGlySerThrAlaAspThrSerLeuProSerThrSerLeuSerSerGln  275

CATCCCGCTACCAGCCAATGCAGCCGGATCAAGCTCACCCCCGAAGGTATCAAGGCCGGC  960
HisProAlaThrSerGlnCysSerArgIleLysLeuThrProGluGlyIleLysAlaGly  295

TGGCAGGTCGAACATATGCTTGAGGCGCGCATGATGCCTGAGCTCGTGCACGTCCCGAAC  1020
TrpGlnValGluHisMetLeuGluAlaArgMetMetProGluLeuValHisValProAsn  315

GGCCAGATCCTCATCACCAACGGCGCTGGTACAGGCTTTGCCGCACTGTCTGCCGTCGCG  1080
GlyGlnIleLeuIleThrAsnGlyAlaGlyThrGlyPheAlaAlaLeuSerAlaValAla  335

GACCCCGTCGGCAACTCGAACGCGGACCACCCCGTGCTCACACCTTCGCTCTACACGCCC  1140
AspProValGlyAsnSerAsnAlaAspHisProValLeuThrProSerLeuTyrThrPro  355

GACGCACCTCTTGGAAAGCGCATCAGCAACGCGGGCATGCCGACGACGACGATTCCGCGC  1200
AspAlaProLeuGlyLysArgIleSerAsnAlaGlyMetProThrThrThrIleProArg  375

ATGTACCACTCGACCGTGACGCTCACGCAGCAGGGCAACTTCTTCATCGGGGGCAACAAC  1260
MetTyrHisSerThrValThrLeuThrGlnGlnGlyAsnPhePheIleGlyGlyAsnAsn  395

CCGAACATGAACTTCACGCCGCCTGGCACACCCGGGATCAAGTTCCCCAGCGAGCTGCGC  1320
ProAsnMetAsnPheThrProProGlyThrProGlyIleLysPheProSerGluLeuArg  415

ATCGAGACCCTCGACCCGCCGTTCATGTTCAGGAGCCGCCCCGCGCTTCTCACGATGCCT  1380
IleGluThrLeuAspProProPheMetPheArgSerArgProAlaLeuLeuThrMetPro  435

GAGAAGCTCAAGTTCGGCCAGAAGGTGACCGTGCCTATCACGATCCCGAGCGACCTGAAG  1440
GluLysLeuLysPheGlyGlnLysValThrValProIleThrIleProSerAspLeuLys  455

GCGAGCAAGGTCCAAGTCGCCCTGATGGACCTCGGCTTTTCGAGCCACGCCTTCCACTCC  1500
AlaSerLysValGlnValAlaLeuMetAspLeuGlyPheSerSerHisAlaPheHisSer  475

AGTGCGCGCCTCGTCTTCATGGAGTCGTCCATCTCGGCGGACCGCAAGTCGCTCACCTTC  1560
SerAlaArgLeuValPheMetGluSerSerIleSerAlaAspArgLysSerLeuThrPhe  495

ACGGCTCCGCCCAACGGACGCGTCTTCCCACCGGGCCCGGCTGTCGTTTTCCTCACGATT  1620
ThrAlaProProAsnGlyArgValPheProProGlyProAlaValValPheLeuThrIle  515

GACGACGTGACAAGTCCGGGAGAGAGAGTCATGATGGGCAGTGGCAACCCTCCGCCGACC  1680
AspAspValThrSerProGlyGluArgValMetMetGlySerGlyAsnProProProThr  535

CTGGAGTAAATATGGAGTGTTCGCGGGTACCTCGGCGCAGACCTTGTCTTGACTGTTGTG  1740
LeuGluTer

GTGCGGTATGTACGTTGTTTTTTGTTGTGAGCTCGAAAGTTGCATATGAAATATCTGTTT  1800


GGGCGAAAAAAAAAA
```

FIG. 1. Nucleotide and deduced amino acid sequence of glx-1c. The beginning of the mature protein is indicated, and experimentally determined amino acids are underlined. The 3′-untranslated region conserved in certain lignin peroxidase genes is double underlined.

clones λML1 (GCTTATGAAATATCGGT) and CLG4 (ACGAAATA) (14, 15). Similar to other highly expressed genes, including *P. chrysosporium* peroxidases, substantial codon bias is observed in glx-1c. Specifically, codons ending with guanine or cytosine are heavily favored.

RNA Analysis. Cultures of *Phanerochaete* were grown under ligninolytic conditions (9, and RNA was extracted for

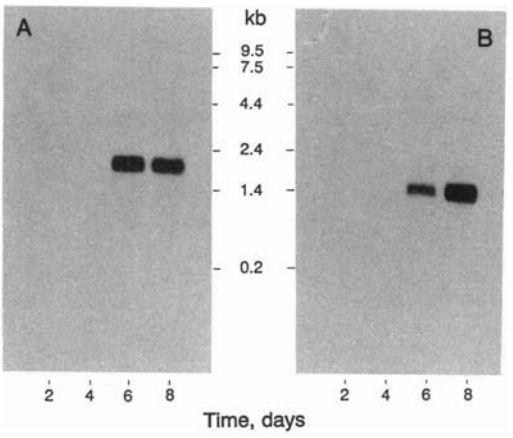

FIG. 2. Autoradiograph of RNA blot hybridized with glx-1c cDNA (A) and lip cDNA (*B*). Total RNA was purified from ligninolytic shake flasks (5) on days 2, 4, 6, and 8. When harvested, glyoxal oxidase activities were as follows: day 2, <1 nmol per min· ml; day 4, <1 nmol per min· ml; day 6, 7 nmol per mind; day 8, 26 nmol per min· ml. Approximately 20 μg of total RNA was size-fractionated on formaldehyde gels and blotted to Hybond-N (Amersham), according to the manufacturer's recommendations. For probes, a full-length cDNA (1.3 kb) encoding lignin peroxidase isozyme H8 and a 1.2-kb *Kpn* I fragment of the glx-1c cDNA were nick-translated, and 3 × 10^6 dpm/ml were used in the hybridization buffer. High-stringency hybridizations [60°C at 0.1× standard saline phosphate/EDTA (SSPE) (1× SSPE is 0.18 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA)] were used. Blots were exposed to Kodak XAR-5 film with amplifying screens for 4 hr (A) or 7 hr (B).

RNA analysis (Fig. 2). For comparison, probes for transcripts corresponding to both GLOX and lignin peroxidase isozyme H8 were used. The time course for glyoxal oxidase activity correlates with transcript levels and is also coordinate with the lignin peroxidases in nutrient nitrogen-starved cultures. We have also recently demonstrated the presence of GLOX transcript in carbon-starved cultures (16).

DISCUSSION

We report a cDNA nucleotide sequence that encodes glyoxal oxidase. Evidence that glx-1c codes glyoxal oxidase includes (*i*) positive immunological identification of the clone in a lgt11 expression library with antibodies specific for glyoxal oxidase, (*ii*) 100% identity of the experimental N-terminal amino acid sequence with the deduced sequence, (*iii*) coincidence of transcript expression with enzyme activity in cultures, (*iv*) presence of a signal sequence for the extracellular enzyme, and (*v*) reasonable agreement of experimental and predicted physical characteristics (M_r and pI) of the protein.

The distinctive characteristics of glyoxal oxidase were demonstrated within the physiological context of the lignin-degrading system of *P. chrysosporium* (3, 5). However, there were some similarities with the extracellular fungal enzyme galactose oxidase. The enzymes have substrates in common (although galactose is not oxidized by GLOX), and both appear to have inactive forms that can be reactivated by peroxidase systems.

Structural details of galactose oxidase were recently elucidated by x-ray crystallography (17), and the gene encoding this enzyme was cloned and sequenced (18). Galactose oxidase has an unusual copper active site with tyrosine and histidine ligands; one ligand, Tyr^{272}, is linked to the sulfur of Cys^{228} by a thioether bond. It is proposed that this structure, in conjunction with a stacking Trp^{290}, is a built-in secondary cofactor important in the redox activity of the enzyme. Another striking characteristic of galactose oxidase is the extensive ß structure of the three structural domains of the monomeric protein.

Overall, there is no significant sequence homology between glx-1c and the galactose oxidase gene. However, Gamier–Robsen analysis (19) predicts extensive ß-sheet secondary structure in GLOX. Furthermore, the number and distribution of tyrosine, histidine, and cysteine residues in glx-1c could provide for a similar catalytic mechanism. Two adjacent copper ligands of galactose oxidase, Tyr^{495} and His^{496}, might correspond to Tyr^{377} and His^{378} of GLOX. Residues surrounding these pairs are also similar—i.e., Arg-Val-Tyr-His-Ser in galactose oxidase compared with Arg–Met-Tyr-His-Ser in GLOX.

The precise role of GLOX in ligninolysis and the full implications of its interaction with extracellular peroxidases remain to be established. In this and another recent report (16), we demonstrated that GLOX expression is coordinate with lignin peroxidase and manganese peroxidase. In addition to the obvious relationship of supply and demand of H_2O_2, the extracellular modulation of glyoxal oxidase activity by interaction with peroxidase systems could also have important implications in coordinating extracellular metabolism of lignin and carbohydrates and controlling levels of reactive oxygen species. Ultimately, the catalytic mechanism and interactions of GLOX with peroxidases will be elucidated through crystal structures, site-specific mutagenesis of active sites, and gene disruptions. The cloning and sequencing of GLOX provide an important step toward these ends.

This work was supported, in part, by U.S. Department of Agriculture Competitive Research Grant 88-33521-4089 to D.C. and P.J.K., Department of Energy Grant DE-FG02-87ER13712 to D.C. and T. K. Kirk, and by a Biopulping Consortium involving industry, the University of Wisconsin Biotechnology Center, and the Forest Products Laboratory.

1. Kirk, T. K. & Farrell, R. L. (1987) *Annu. Rev. Microbiol.* **41,** 465–505.
2. Cullen, D. & Kersten, P. J. (1987) in Applied Molecular Genetics of Filamentous Fungi , ed~. Kinghorn, J. R. & Turner, G. (Chapman & Hall, New York), pp. 100–131.
3. Kersten, P. J. & Kirk, T. K. (1987) *J. Bacteriol.* **169,** 2195–2201.
4. Tien, M. & Kirk, T. K. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2280–2284.
5. Kersten, P. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2936–2940.
6. Alic, M. & Gold, M. H. (1991) in *More Gene Manipulations in Fungi*, eds. Bennett, J. & Lasure, L. (Academic, New York), pp. 320–335.
7. Pease, E. A. & Tien, M. (1991) in *Biocatalysts for Industry*, ed. Dordick, J. S. (Plenum, New York), pp. 115–135.
8. Timberlake, W. E. & Barnard, E. C. (1981) *Cell* **26,** 26–37.
9. Sanger, F., Nicklen, S. & Coulson, R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
10. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726-730.
11. von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
12. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683-4690.
13. Gurr, S. J., Unkles, S. E. & Kinghorn, J. R. (1987) in Gene Structure in Eukaryotic Microbes, ed. Kinghorn, J. R. (IRL, Oxford), pp. 93–139.
14. Tien, M. & Tu, C. (1987) Nature (London) 326,520–523.
15. deBoer, H. A., Zhang, Y., Collins, C. & Reddy, C. A. (1987) Gene **60,** 93–102.
16. Stewart, P., Kersten, P., Vanden Wymelenberg, A., Gaskell, J. & Cullen, D. (1992) *J. Bacteriol.* **174,** 5036–5042.
17. Ito, N., Philips, S. E. V., Stevens, C., Ogel, Z. B.. McPherson, M. J., Keen, J. N., Yadav, K. D. S. & Knowles, P. F. (1991) *Nature (London)* **350,** 87–90.
18. McPherson, M. J., Ogels, Z. B., Stevens, C., Yadav, K. D. S., Keen, J. N. & Knowles, P. F. (1992) *J. Biol. Chem.* **267,** 8146–8152.
19. Gamier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.