Volume 308, number 3, 271–276 FEBS 11472 August 1992

Trimethylamine dehydrogenase of bacterium W_3A_1

Molecular cloning, sequence determination and over-expression of the gene

Geoffrey Boyd[a], F. Scott Mathews[a], Leonard C. Packman[b] and Nigel S. Scrutton[b]

[a]*Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO 63110, USA* and [b]*SERC Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK*

Received 26 June 1992

The gene encoding trimethylamine dehydrogenase (EC 1.5.99.7) from bacterium W_3A_1 has been cloned. Using the polymerase chain reaction a 530 bp DNA fragment encoding a distal part of the gene was amplified. Using this fragment of DNA as a probe, a clone was then isolated as a 4.5 kb *Bam*HI fragment and shown to encode residues 34 to 729 of trimethylamine dehydrogenase. The polymerase chain reaction was used also to isolate the DNA encoding the missing N-terminal part of the gene. The complete open reading frame contained 2,190 base pairs coding for the processed protein of 729 amino acids which lacks the N-terminal methionine residue. The high-level expression of the gene in *Escherichia coli* was achieved by the construction of an expression vector derived from the plasmid pKK223-3. The cloning and sequence analysis described here complete the partial assignment of the amino acid sequence derived from chemical sequencing [1] and will now permit the refinement of the crystallographic structure of trimethylamine dehydrogenase and also a detailed investigation of the mechanism and properties of the enzyme by protein engineering.

Trimethylamine dehydrogenase; Bacterium W_3A_1; Iron-sulphur flavoprotein

1. INTRODUCTION

Trimethylamine dehydrogenase (EC 1.5.99.7) catalyses the oxidative *N*-demethylation of trimethylamine to yield dimethylamine and formaldehyde [2]. The enzyme is one of several proteins that are induced to high levels in the bacterium W_3A_1 or related organisms when grown on trimethylamine as the sole carbon source [3]. The structure of the homodimeric enzyme is solved at 2.4 Å resolution [4] revealing the overall chain fold of the enzyme. The enzyme is comprised of three domains, the N-terminal $\beta_8\alpha_8$ barrel domain linked covalently to the 6-position of the FMN cofactor by a cysteine residue [5,6], a medium domain and a small domain. The latter two domains bear a striking resemblance to the FAD and NADPH-binding domains of glutathione reductase [4] but dinucleotide binding to this region of the protein has not been demonstrated. The binding of ADP, however, has been observed in crystallographic studies suggesting that this region of the enzyme represents the vestigial remains of an FAD-binding domain [7]. Crystallographic studies of the enzyme in three oxidation states and in the presence of substrate and the inhibitor tetramethylammonium chloride have also been carried out [8]. The amino acid sequence of trimethylamine dehydrogenase has been deduced from the X-ray diffraction pattern obtained at 2.4 Å resolution. This sequence has been compared to a chemically determined sequence that represents approximately 80% of the enzyme subunit [1]. Here we report the cloning, nucleotide sequence determination and high-level expression of the gene (*tmd*) encoding trimethylamine dehydrogenase. This has allowed us to deduce the sequence of amino acid residues inferred from the X-ray diffraction data which were not confirmed by chemical sequencing and also to revise twelve assignments previously obtained by chemical sequencing. The sequence of the full protein will now enable the refinement of the crystal structure of the enzyme and will also permit a detailed study of the enzyme by protein engineering.

Abbreviations: IPTG, isopropyl-β-D-1-thiogalactopyranoside; PCR, polymerase chain reaction; TMADH, trimethylamine dehydrogenase.

Correspondence address: N.S. Scrutton, SERC Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK. Fax: (44) (223) 333345.

2. EXPERIMENTAL

2.1. *Materials*

Complex bacteriological media were from Difco laboratories, and all media were prepared as described in Maniatis et al. [9]. [^{35}S]dATP-αS triethylammonium salt (400 Ci/mmol) for DNA sequencing, [^{32}P]ATP-γP triethylammonium salt (3,000 Ci/mmol), [^{32}P]dCTP-αP triethylammonium salt (3,000 Ci/mmol) and 'Mega prime' random prime labeling kit were supplied by Amersham International. The

restriction enzymes *Bam*HI, *Eco*R1, *Hin*dIII, *Kpn*I, *Nsi*I, *Pst*I, *Pvu*I and *Sma*I, the DNA modification enzymes T4 DNA ligase, T4 polynucleotide kinase and DNA polymerase (Klenow fragment) and the plasmid pKK223-3 and nucleotides were purchased from Pharmacia LKB Biotechnology. *Taq* DNA polymerase was supplied by Promega and calf intestinal alkaline phosphatase and terminal transferase were from Boehringer Mannheim GmbH. *E. coli* strain JM109 [*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17, *sup*E44, *rel*A1, Δ(*lac-pro*AB), F'*tra*D36, *pro*AB, *lacI*ᵠΔM15] and DH5α [F' *rec* A Δ(*lac* ZYA-*arg* F) U169 *hsd* R17 *thi*-1 *gyr* A66 *sup* E44 *end* A1 *rel* A1 Φ80d *lac* Z Δ M15] were purchased from Promega. Bacterium W_3A_1 was a kind gift of Dr. W. McEntire and was maintained on minimal base E Media [10].

2.2. *Methods used for protein analysis*

Trimethylamine dehydrogenase for N-terminal sequence analysis was purified as described previously [2]. N-Terminal sequence analysis was kindly performed by Dr. J. McCormick (Department of Biochemistry, University of Cambridge) on an Applied Biosystems Model 477A sequencer. Polyacrylamide gels were run in the presence of sodium dodecyl sulphate [11] using the Phast System marketed by Pharmacia LKB Biotechnology. Gels were run according to the manufacturers instructions and stained with Coomassie brilliant blue R250. Cell-free extracts of *E. coli* strains were prepared and analysed by SDS-PAGE as described previously [12].

2.3. *Cloning and sequence determination of the tmd gene*

Molecular cloning techniques were adopted from [9]. Chromosomal DNA of bacterium W_3A_1 was prepared by the method of Marmur [13]. The polymerase chain reaction was used to isolate a 530 bp portion (nucleotides 1470 to 2,000) of the *tmd* gene from genomic DNA of bacterium W_3A_1. The primers (Fig. 1) were synthesised according to known chemically determined polypeptide sequence ([1]; Fig. 1) and the conditions used for amplification were: cycle 1, denaturation 94°C (60 s), annealing 46°C (80 s), extension 72°C (80 s); cycles 2–25, denaturation 95°C (40 s), annealing 48°C (60 s), extension 72°C (80 s) Following isolation by polyacrylamide gel electrophoresis, the identity of the amplified 530 bp fragment was confirmed by its hybridisation to an oligonucleotide (Fig. 1) that corresponds to the chemical sequence of trimethylamine dehydrogenase (residues 537 to 542). Hybridisations were carried out in dried agarose gels according to the procedure of Meinkoth and Wahl [14]. The amplified PCR amplified fragment was subsequently ligated into *Sma*I-cut pUC 18 which had been modified by terminal transferase. The PCR fragment was subsequently released from the pUC 18 vector by digestion with *Bam*HI and *Kpn*I, isolated by polyacrylamide gel electrophoresis, labelled by random priming and used to hybridise to *Bam*HI-digested W_3A_1 chromosomal DNA which had been separated by agarose gel electrophoresis. Hybridisation occurred to fragments of DNA of approximate size 4.5 kb. Separated chromosomal DNA in the range 4 to 5 kb was eluted from the agarose gel and ligated into *Bam*HI-cut pUC 18. The screening of 600 colonies by colony transfer and hybridisation [9] using the PCR fragment as a probe identified two identical clones (designated pUC TM) that contained part of the gene (*tmd*) encoding trimethylamine dehydrogenase.

The *tmd* gene (nucleotides 105 to 2,190) in plasmid pUC TM was sequenced by the dideoxy- chain termination method [15,16] with the T7 sequencing kit (Pharmacia LKB Biotechnology) and by fluorescent sequencing on a Pharmacia LKB A.L.F. sequencer. The plasmid pUC TM was sonicated and fragments of length 400 to 700 bp were isolated by polyacrylamide gel electrophoresis, end-repaired and cloned into M13mp18 previously restricted with the enzyme *Sma*I. Fragments for sequencing were also generated by digesting the plasmid pUC TM with the restriction endonuclease *Pst*I. These fragments were likewise ligated into M13mp18 restricted with *Sma*I. Nucleotide sequence was determined using the M13 Universal Primer to initiate DNA fragment synthesis. Gaps in the sequence were filled using specific oligonucleotide primers designed to prime near the sequence to be determined in the plasmid pUC TM. The nucleotide sequence was determined on both strands (Fig. 2) and a consensus sequence was derived using the

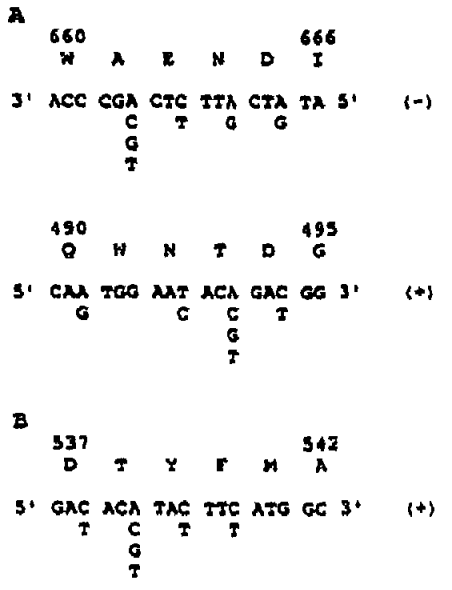

Fig. 1. (A) The degenerate primers used in PCR amplification of the distal portion of the *tmd* gene and the peptide sequences within TMADH to which they correspond. (B) The oligonucleotide used to verify the identity of the PCR product generated using the primers in Fig. 1A . Numbers indicate the amino acid residue locations of the peptide fragments within TMADH. (+), coding strand; (−), non-coding strand.

software package Assembly Align marketed by the IBI software company, New Haven, CT, USA.

The missing N-terminal portion of the *tmd* gene in plasmid pUC TM was amplified by the polymerase chain reaction from genomic DNA of bacterium W_3A_1 which had previously been digested partially with the restriction enzyme *Pst*I and size-fractionated (5 to 10 kb) by sucrose gradient ultracentrifugation. One of the PCR primers was designed according to the revised N-terminal protein sequence of trimethylamine dehydrogenase (see section 3; Fig. 3) and the second primer was made as a section of the non-coding strand of the *tmd* gene (nucleotides 405 to 422). Amplification was for 25 cycles using the following parameters: denaturation 93°C (2 min), annealing 60°C (2 min), extension 72°C (3 min). The amplified product was purified by polyacrylamide gel electrophoresis, end-repaired and ligated into *Sma*I-cut M13mp18 to yield the recombinant bacteriophage M13Nterm. The nucleotide sequence of the inserted fragment was determined on both strands using the T7 sequencing kits as described above.

2.4 *Construction of an over-expressing clone (pKTM) for trimethylamine dehydrogenase*

The high-level expression of the *tmd* gene was achieved by inserting the reconstructed *tmd* gene into the plasmid expression vector pKK223-3. The reconstruction of the gene was achieved as detailed below (Fig. 4). The distal part of the gene was released from plasmid pUC TM by digestion with *Bam*HI/*Nsi*I and ligated into M13mp18 previously restricted with *Bam*HI and *Pst*I. This new construct was designated M13TM. The N-terminal portion of the gene was amplified by the polymerase chain reaction using the bacteriophage M13Nterm as template. The cycling parameters were as decribed for the amplification of the N-terminal portion of the gene from genomic DNA of bacterium W_3A_1. The first PCR primer was modified so as to encode a unique *Eco*R1 cleavage site positioned immediately before the initiating methionine codon of the *tmd* gene. The second primer was identical to that previously used to isolate the 5' end of the gene from

genomic DNA (section 2.3). The amplified product was isolated by polyacrylamide gel electrophoresis, restricted with *Eco*RI and *Bam*HI and ligated into *Eco*RI/*Bam*HI-cut M13TM. The new construct carrying the reconstructed *tmd* gene was designated M13TM2. The region of the *tmd* gene encoded by the inserted PCR product was re-sequenced to check that no spurious changes in DNA sequence had occurred during the amplification procedure. The *tmd* gene was then released from M13TM2 by complete restriction with *Eco*RI and partial restriction with *Hin*dIII. Following its isolation by polyacrylamide gel electrophoresis this fragment was ligated into pKK223-3 previously cut with *Eco*RI and *Hin*dIII to produce plasmid pKTM. The constructs were transformed into *E. coli* strain JM109 and recombinants were isolated by selection on ampicillin/2×YT agar plates and identified by restriction analysis of DNA prepared by the rapid boiling procedure [9].

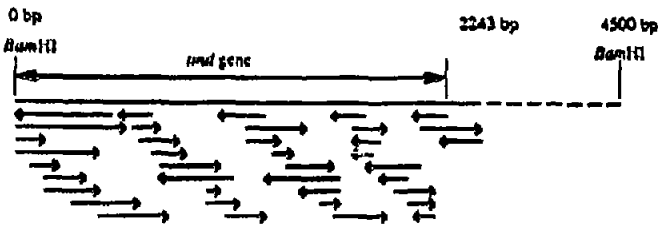

Fig. 2. Strategy for sequencing the *tmd* gene (corresponding to amino acid residues 34 to 729) as part of the 4.5 kbp *Bam*HI fragment contained in the plasmid pUC TM.

3. RESULTS AND DISCUSSION

Based on the known partial chemical sequence [1], the polymerase chain reaction was used to isolate a 530 bp DNA fragment encoding a portion of the *tmd* gene. This fragment was subsequently used as a probe to isolate the *tmd* gene encoding amino acid residues 34 to 729 from a genomic library constructed by the ligation of *Bam*HI-cut W_3A_1 genomic DNA to *Bam*HI-digested pUC 18. The missing portion of the gene encoding the N-terminus of the protein was isolated by the polymerase chain reaction using two independent amplifications and the nucleotide sequence of both amplified products was shown to be identical. For the purpose of clarity, the numbering scheme adopted here is that used by Barber et al. [1]. The initiating methionine residue is therefore referred to as position 0 even though it was observed in about 25% of the material subjected to N-terminal sequence analysis of the protein. The DNA sequence of the PCR fragment was found to be consistent with the sequence of the first thirty amino acid residues of trimethylamine dehydrogenase determined by N-terminal sequence analysis of the intact enzyme (this work and [1]) except for residues 23 and 29 that could not be unequivocally assigned and with residues 30 to 33 determined previously by sequence analysis of a peptide fragment containing covalently linked FMN [6]. Residues 1 to 33 correspond to the sequence not present in the clone pUC TM. The complete DNA sequence and inferred amino acid sequence is presented in Fig. 5. The *tmd* gene of 2190 bp encodes a protein of 730 amino acids (including the N-terminal methionine residue which is absent when the protein is expressed in bacterium W_3A_1) with an apoenzyme subunit molecular mass of 81,623 Da . The calculated subunit molecular mass is in good agreement with the published values of approximately 83,000 Da. There is a putative stem-loop structure at the 3' end of the gene covering nucleotides 2205 to 2238 (Fig. 5) which may act as a transcription termination signal. The *tmd* gene represents the first gene reported to be cloned from bacterium W_3A_1 and is composed of 24.4% T, 23.1% C, 25.7 % A and 26.8% G . The ratio of the C/G content over the A/T content at the third position of amino acid codons is approximately unity indicating no preference for the type of base pairing at these positions. The codon usage is displayed in Table I. Comparisons of the consensus sequence of TMADH with sequences deposited in the protein and DNA sequence databases (Swissprot version 21 and Genbank) revealed no significant similarities except, as previously reported by Barber et al. [1], for the bile acid inducible operon protein of *Eubac-*

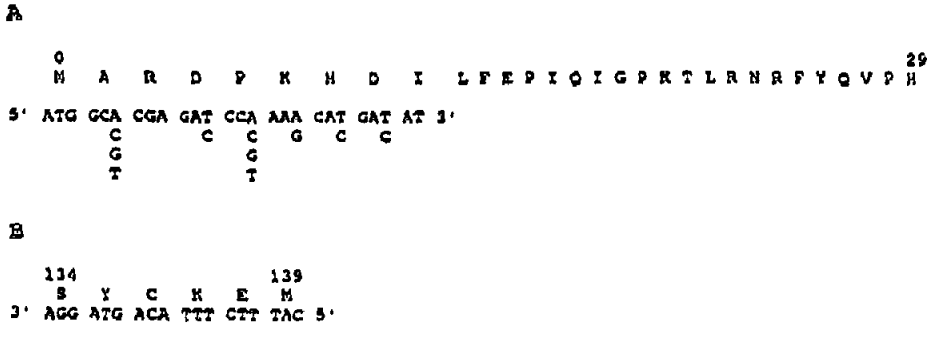

Fig. 3. (A) The revised N-terminal amino acid sequence of TMADH obtained by protein sequencing and the DNA sequence of the degenerate primer used for PCR amplification of this region of the gene. (B) Internal protein sequence of TMADH (residues 134 to 139) deduced from the nucleotide sequence of pUC TM and the antisense primer used for PCR amplification of the gene encoding the N-terminal portion of TMADH.

Table I

Codon usage table for the *tmd* gene

Phe	TTT	8	Ser	TCT	9	Tyr	TAT	13	Cys	TGT	8
Phe	TTC	11	Ser	TCC	10	Tyr	TAC	18	Cys	TGC	4
Leu	TTA	6	Ser	TCA	9	*	TAA	1	*	TGA	0
Leu	TTG	20	Ser	TCG	2	*	TAG	0	Trp	TGG	16
Leu	CTT	1	Pro	CCT	10	His	CAT	10	Arg	CGT	30
Leu	CTC	0	Pro	CCC	1	His	CAC	13	Arg	CGC	12
Leu	CTA	2	Pro	CCA	21	Gln	CAA	15	Arg	CGA	0
Leu	CTG	18	Pro	CCG	6	Gln	CAG	10	Arg	CGG	0
Ile	ATT	22	Thr	ACT	4	Asn	AAT	8	Ser	AGT	0
Ile	ATC	22	Thr	ACC	17	Asn	AAC	16	Ser	AGC	6
Ile	ATA	1	Thr	ACA	11	Lys	AAA	22	Arg	AGA	0
Met	ATG	16	Thr	ACG	8	Lys	AAG	18	Arg	AGG	0
Val	GTT	13	Ala	GCT	17	Asp	GAT	29	Gly	GGT	41
Val	GTC	7	Ala	GCC	11	Asp	GAC	15	Gly	GGC	20
Val	GTA	8	Ala	GCA	17	Glu	GAA	29	Gly	GGA	1
Val	GTG	16	Ala	GCG	16	Glu	GAG	24	Gly	GGG	3

The first nine codons of the gene are omitted from the analysis because they are derived from the sequence of the degenerate oligonucleotide primer used for PCR amplification of the 5' end of the *tmd* gene.

terium spp. and dimethylamine dehydrogenase of *Hyphomicrobium* X.

The protein sequence of trimethylamine dehydrogenase deduced from the *tmd* gene is, but for twelve assignments, consistent with that previously determined chemically [1]. These discrepancies no doubt reflect the difficulty in making confident assignments at certain positions during peptide sequencing. We were unable to amplify the 5' region of the gene by PCR using the published amino acid sequence for the N terminus of the protein. A re-determination of the N-terminal sequence of trimethylamine dehydrogenase revealed that the residue at position eight is isoleucine and not, as previously reported, tyrosine. The eighth position in the sequence is also encoded by the 3' end of the primer used for PCR amplification of the N-terminal portion of the *tmd* gene and therefore successful PCR amplifications were only made using a modified primer which encoded isoleucine at its 3' end. The amino acid composition of TMADH determined after acid hydrolysis has been published elsewhere [17]. The composition deduced from the gene sequence is in reasonable agreement with the hydrolysis data of Kasprzak et al. [17] when normalised to the same number of residues per dimer (1,458 residues; Table II) The found values mostly agree with the expected to within 4%, the generally accepted limits of accuracy of this technique. With a few residues, notably His, Trp Pro, Val, Ile and Cys the differences approach or exceed 10% but these residues are the most difficult to determine accurately; in particular Trp and Cys are subject to oxidative loss and Val and Ile are resistant to release by hydrolysis. The protein sequence presented here completes the partial assignment of the amino acid sequence of trimethylamine dehydrogenase derived previously from chemical sequencing [1] and will now facilitate the refinement of the crystallographic structure of the enzyme molecule.

The high-level expression of the *tmd* gene in *E. coli* was achieved by reconstructing the complete gene in a bacteriophage vector followed by the insertion of the gene into the plasmid expression vector pKK223–3 (Fig. 4). During gene reconstruction, the opportunity was taken to introduce by PCR a unique *Eco*RI restriction site immediately 5' of the *tmd* gene. As a result, the complete *tmd* gene was directionally subcloned as an

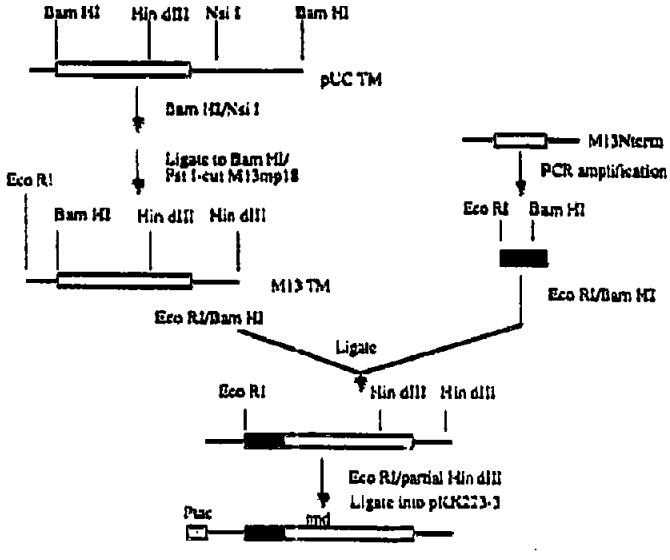

Fig. 4. Strategy for the isolation of an expression clone (pKTM) for TMADH.

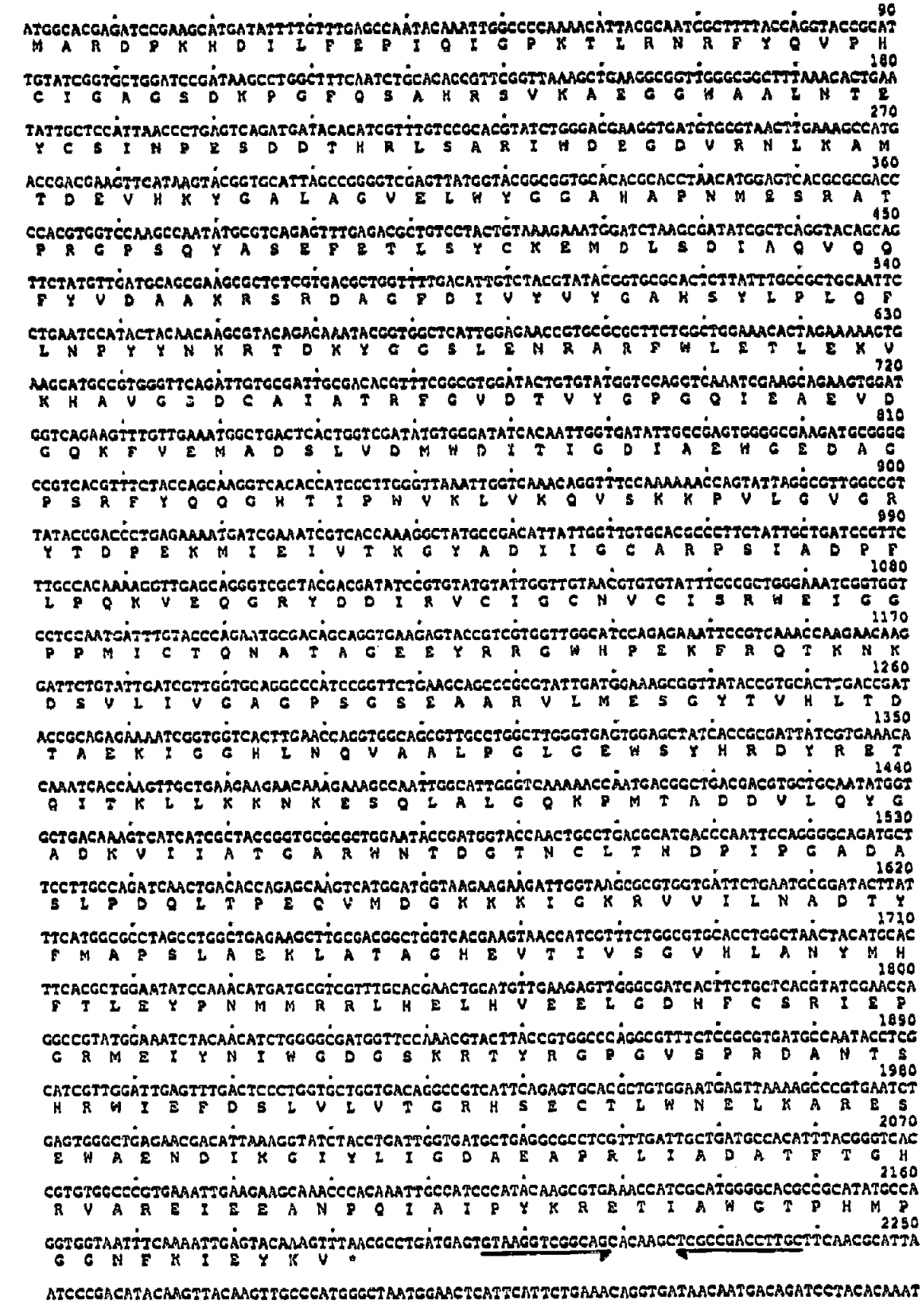

Fig. 5. DNA sequence of the trimethylamine dehydrogenase gene (*tmd*) of bacterium W_3A_1 and the deduced amino acid sequence. Directionally opposed arrows indicate the position of a potential stem-loop structure.

Table II

Comparison of the amino acid compositions derived from acid hydrolysis of trimethylamine dehydrogenase [17] with that deduced from the sequence of the *tmd* gene

Amino acid	Amino acid composition (mol residue/mol enzyme)	
	Hydrolysis	DNA
Asn + Asp	143.9	140
Thr	78.6	80
Ser	75.4	72
Gln + Glu	160.1	156
Pro	91.2	78
Gly	135.8	130
Ala	127.9	124
Val	76.6	88
Met	30.8	32
Ile	75.5	92
Leu	96.4	94
Tyr	61.5	62
Phe	38.5	38
His	43.7	48
Lys	79.8	82
Trp	35.2	32
Arg	90.0	86
Cys	16.7	24
Total	1,458	1,458

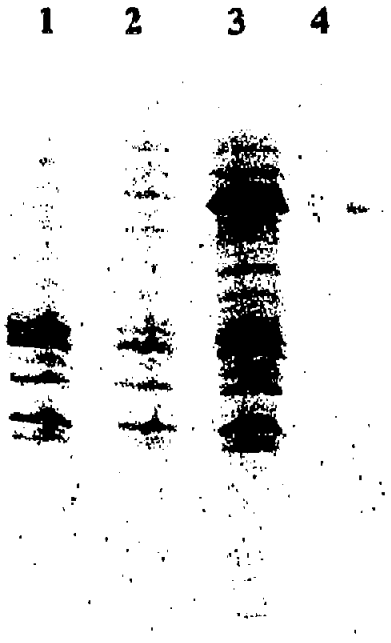

Fig. 6. 12.5% SDS-PAGE analysis of the expression of trimethylamine dehydrogenase in *E. coli* strain JM109 transformed with plasmid pKTM. Lane 1, cell extract of untransformed *E. coli* strain JM109; lane 2, cell extract of *E. coli* strain JM109 transformed with plasmid pKTM; lane 3, cell extract of *E. coli* strain JM109 transformed with plasmid pKTM and grown in the presence of 2 mM IPTG; lane 4, trimethylamine dehydrogenase purified from bacterium W_3A_1. Protein was visualised by staining with Coomassie brilliant blue R250.

*Eco*RI/*Hin*dIII fragment into the expression vector pKK223–3 where it came under the influence of the strong *tac* promoter and the plasmid derived ribosome-binding site which is positioned 9 bases from the initiating methionine codon of the *tmd* gene. SDS-PAGE analysis of *E. coli* strain JM109 carrying the recombinant pKTM expression clone which was grown in the presence of the inducer molecule IPTG revealed that the pKTM-transformed strain over-produces trimethylamine dehydrogenase to approximately 20% of the total cell protein (Fig. 6). By achieving the high-level expression of the gene encoding trimethylamine dehydrogenase together with the availability of the crystallographic structure [4] and the results of detailed spectroscopic studies of the enzyme [18,19] the way is now clear to initiate a detailed investigation of the mechanism and properties of the enzyme by protein engineering.

Acknowledgements: We thank Mr. J.B. Lester and Mr. M. Weldon for skilled technical assistance and Dr. J. McCormick for N-terminal sequence analysis. We also thank A.Y. Chistoserdov for valuable advice. This work was funded by the Royal Society (N.S.S.) and the U.S.P.H.S. (F.S.M). The support of the Wellcome Trust for oligonucleotide synthesis is gratefully acknowledged.

REFERENCES

[1] Barber, M.J., Neame, P.J., Lim, L.W., White, S. and Mathews, F.S. (1991) J. Biol. Chem. 267, 6611–6619.
[2] Steenkamp, D.J. and Mallinson, J. (1976) Biochim. Biophys. Acta 429, 705–719.
[3] Kasprzak, A.A. and Steenkamp, D.J. (1983) J. Bacteriol. 156, 348–353.
[4] Lim, L.W., Shamala, N., Mathews, F.S., Steenkamp, D.J., Hamlin, R. and Xuong, N. (1986) J. Biol. Chem. 261, 15140–15146.
[5] Steenkamp, D.J., McIntire, W. and Kenney, W.C. (1978) J. Biol. Chem. 253, 2818–2824.
[6] Kenney, W.C., McIntire, W., Steenkamp, D.J. and Benisek, W.F. (1978) FEBS Lett. 85, 137–139.
[7] Lim, L.W., Mathews, F.S. and Steenkamp, D.J. (1988) J. Biol. Chem. 263, 3075–3078.
[8] Bellamy, H.D., Lim, L.W., Mathews, F.S. and Dunham, W.R. (1989) J. Biol. Chem. 624, 11887–11892.
[9] Maniatis, T. Fritsch, E.F. and Sambrook, J. (1982) in Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[10] McIntire, W.S. (1990) in: Methods in Enzymology, Vol. 188, pp. 250–290.
[11] Laemmli, U.K. (1970) Nature 227, 680–685.
[12] Scrutton, N.S., Berry, A. and Perham, R.N. (1987) Biochem. J. 245, 875–880.
[13] Marmur, J. (1961) J. Mol. Biol. 3, 208–218.
[14] Meinkoth, J. and Wahl, G. (1978) Anal. Biochem. 138, 267–284.
[15] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J. H. and Roe, B.A. (1980) J. Mol. Biol. 143, 167–178.
[16] Biggin, M.D., Gibson, T.G. and Hong, G.F. (1983) J. Mol. Biol. 144, 103–116.
[17] Kasprzak, A.A., Papas, E.J. and Steenkamp, D.J. (1983) Biochem. J. 211, 535–541.
[18] Rohlfs, R.J. and Hille, R. (1991) J. Biol. Chem. 266, 15244–15252.
[19] Hazzard, J.T., McIntire, W.S. and Tollin, G. (1991) Biochemistry 30, 4559–4564.