# Selectivity of post-translational modification in biotinylated proteins: the carboxy carrier protein of the acetyl-CoA carboxylase of *Escherichia coli*

Pedro RECHE, Yi-Li LI, Christopher FULLER, Knut EICHHORN[1] and Richard N. PERHAM[2]

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

Biotin-dependent enzymes contain a biotinyl-lysine residue in a conserved sequence motif, MKM, located in a surface hairpin turn in one of the two $\beta$-sheets that make up the domain. A subgene encoding the 82-residue C-terminal biotinyl domain from the biotin carboxy carrier protein of acetyl-CoA carboxylase from *Escherichia coli* as a fusion protein with glutathione S-transferase was created and over-expressed in *E. coli*. The biotinyl domain was readily released by cleavage with thrombin. Five mutant domains were created in which the conserved MKM motif was systematically replaced: by MAK and KAM, in which the target lysine is moved one place; by KKM and MKK, in which a second potential site for biotinylation is introduced; and by DKA, the motif found in the correspondingly conserved site of lipoylation in the structurally related lipoyl domains of 2-oxo acid dehydrogenase multienzyme complexes. No biotinylation of the MAK or KAM mutants was observed *in vivo* or by purified biotinyl protein ligase *in vitro*; in the KKM and MKK mutants, only one lysine residue, presumed to be that in its native position in the hairpin turn, was found to be biotinylated *in vivo* and *in vitro*. The DKA mutant was not biotinylated *in vivo*, but was partly lipoylated and octanoylated. It was also a poor substrate for lipoylation *in vitro* catalysed by the *E. coli* lipoyl protein ligase encoded by the *lplA* gene. The flanking sequence in the MKM motif is important, but not crucial, and appears to have been conserved in part to be compatible with the subsequent carboxylation reactions of biotin-dependent enzymes. The DKA motif, displayed in the hairpin loop, is sufficient to address lipoylation in *E. coli* but probably by a pathway different from that mediated by the *lplA*-dependent ligase. The recognition of the structurally homologous lipoyl and biotinyl domains by the appropriate ligase evidently has a major structural component to it, notably the positioning of the target lysine residue in the exposed hairpin loop, but there appear to be additional recognition sites elsewhere on the domains.

## INTRODUCTION

Biotin carboxy carrier protein (BCCP) is one of the three functional subunits in the acetyl-CoA carboxylase of *Escherichia coli*. This enzyme catalyses the biotin-dependent carboxylation of acetyl-CoA to malonyl-CoA, the first committed step in the biosynthesis of fatty acids [1]. BCCP is a homodimer [2], each polypeptide chain containing 156 amino acid residues, as deduced from the DNA sequence of the corresponding gene [3]. The biotinyl group is attached post-translationally by the action of a biotinyl protein ligase (BPL, also known as biotin holoenzyme synthetase), which catalyses the formation of an amide linkage between the carboxy group of the biotin and the $N^6$-amino group of a specific lysine residue in BCCP. The biotinyl-lysine (biocytin) residue is located in the C-terminal region of the BCCP protein, in an independently folded domain of about 80 amino acids [4–6], the structure of which has been established independently by means of X-ray crystallography [7] and, in solution, by means of NMR spectroscopy (E. L. Roberts, R. W. Broadhurst, M. J. Howard, A. Chapman-Smith, J. E. Cronan, J. C. Wallace and R. N. Perham, unpublished work).

The structure of the biotinyl domain is that of a flattened $\beta$-barrel, comprising two 4-stranded $\beta$-sheets, with the N- and C-terminal residues close together in space in one sheet and the biotinyl-lysine residue located in a highly conserved MKM sequence in an exposed $\beta$-turn in the other sheet (Figure 1). The biotinyl-lysine residue is thus well placed to act as a swinging arm in the mechanism of acetyl-CoA carboxylase, becoming carboxylated at the active site of the biotin carboxylase subunit and passing on the carboxy group at the active site of the carboxytransferase subunit. In *E. coli*, BCCP is the only biotinylated protein, but a comparable domain with a similar function in the catalytic mechanism is present in all biotin-dependent enzymes [8,9]. As predicted previously and despite minimal amino acid sequence similarity [10], the structure of the biotinyl domain closely resembles that of the lipoyl domain in the dihydrolipoyl acyltransferase component of 2-oxo acid dehydrogenase multienzyme complexes [11,12]. The lipoyl domain also consists of two 4-stranded $\beta$-sheets, with the lipoylated lysine residue prominently displayed in a highly conserved DKA sequence at the tip of the single $\beta$-turn in the $\beta$-sheet opposite the N- and C-terminal residues [13–17]. In these enzymes too, the modified (lipoylated) lysine residue functions as a swinging arm ferrying substrate between the three active sites in the enzyme complex.

In *E. coli*, BPL is encoded by the *birA* gene [18,19], which has been sequenced [20]. The X-ray crystal structure of BPL, which also acts as a repressor of the biotin operon, has been solved [21]. The catalytic mechanism of the enzyme involves a two-step process: activation of the carboxy group of biotin by reaction

Abbreviations used: BCCP, biotin carboxy carrier protein; BPL, biotinyl protein ligase; DTT, dithiothreitol; ES, electrospray; GST, glutathione S-transferase; IPTG, isopropyl 1-thio-$\beta$-galactopyranoside; LPL, lipoyl protein ligase.

[1] Present address: Wellcome CRC Institute of Cancer and Developmental Biology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, U.K.

[2] To whom correspondence should be addressed.


**Figure 1 Three-dimensional structure of the biotinyl domain from the BCCP of *E. coli* acetyl-CoA carboxylase**

The two 4-stranded $\beta$-sheets are indicated in dark and light-shading, and the MKM motif in the hairpin loop housing the biotinyl-lysine residue (Lys-50) is indicated. The structure is taken from the coordinates of Athappilly and Hendrickson [7] and the Figure was constructed using the program MOLSCRIPT [7a].

with ATP to form biotin-5′-AMP and elimination of pyrophosphate, followed by transfer of the biotinyl group to the target lysine residue of BCCP and release of AMP. The biotin-5′-AMP, bound to BPL, acts as the co-repressor of the biotin operon, which creates an elegant model for the regulation of biotin biosynthesis [22].

A similar system exists for the post-translational modification of lipoyl domains. In *E. coli* two genes (*lplA* and *lipB*) have been postulated to encode independent pathways for the attachment of the lipoyl group to the apo-protein [23]. The *lplA* gene has been cloned and over-expressed [24] and the product, lipoyl protein ligase (LPL), catalyses a reaction corresponding to that of BPL, but with lipoyl-5′-AMP as the reaction intermediate. It utilizes exogenous D-lipoic acid as its preferred substrate but can use octanoic acid and L-lipoic acid [25], corresponding to the B form of the enzyme described previously [26]. In contrast, the product of the *lipB* gene relies on an endogenous source of lipoic acid supplied by the *lipA*-dependent biosynthetic route [23] and is likely to be responsible for most of the aberrant octanoylation of lipoyl domains that occurs under conditions of lipoic acid deficiency [27,28].

Two important questions present themselves in the light of the above findings. How do the BPL and LPL enzymes distinguish biotinyl and lipoyl apo-domains as substrates for biotinylation and lipoylation, respectively? And how are the target lysine residues in the biotinyl and lipoyl domains selected for post-translational modification? We have shown previously that the positioning of the target lysine residue at the tip of the $\beta$-turn in the lipoyl domain, rather than the conserved DKA sequence in which it is found, plays a dominant role in the targetting process *in vivo* [29]. In this paper, we describe experiments *in vivo* and *in vitro* which demonstrate that placing the target lysine residue at the tip of the $\beta$-turn in the biotinyl domain is likewise essential for its post-translational modification by BPL and that the highly conserved MKM sequence is of lesser importance. Importing the DKA motif from the $\beta$-turn of the lipoyl domain in place of the MKM motif in the biotinyl domain renders the latter ineffective as a substrate for BPL but confers on it the ability to become lipoylated, albeit poorly. The results throw new light on the specificities of BPL and LPL and the pattern of differing recognition signals displayed by lipoyl and biotinyl domains to ensure correct post-translational modification, which is essential to their biological activity.

## MATERIALS AND METHODS

### Materials

All chemicals used in the preparation of buffers and solutions were of reagent grade. Bacteriological media were purchased from Oxoid, Unipath Laboratories. *E. coli* host strain XL1-Blue was from Stratagene. *E. coli* host strains BL21(DE3) and BL21(DE3)plysS and the expression vector pET11c were from Novagen. Plasmid pGEX-2T was purchased from Pharmacia Biotech. DL-Lipoic acid, octanoic acid, *d*-biotin and PMSF were purchased from Sigma, ATP was from Boehringer Mannheim, and isopropyl 1-thio-$\beta$-galactopyranoside (IPTG) and dithiothreitol (DTT) were from Melford Laboratories Ltd. Restriction endonucleases and DNA-modifying enzymes were from Pharmacia Biotech; Pfu DNA polymerase was purchased from Stratagene Ltd. Oligonucleotide primers were synthesized by Mr. M. Weldon in the Protein and Nucleic Acid Facility, Department of Biochemistry, University of Cambridge, Cambridge, U.K.

### DNA manipulations

Standard protocols for molecular biology were used as described elsewhere [30]. DNA fragments were purified after agarose gel electrophoresis using a Geneclean® kit (BIO101). The Quiagen plasmid kit and the Wizard™ DNA purification kit from Promega were used for plasmid isolation. Automated DNA sequence analysis was carried out by Mr. J. Lester and Mrs. K. Pennock (Department of Biochemistry, University of Cambridge, Cambridge, U.K.) using the Prism™ Ready Reaction DyeDeoxy™ Terminator and the Catalyst 800 Molecular Biology Labstation, as described by the manufacturer (Applied Biosystems Inc.). Sequence data were analysed by the seqEd™ analysis software (Applied Biosystems Inc.). PCRs were carried out under mineral oil in a volume of 100 $\mu$l. The reaction mixtures contained 200 $\mu$M dNTPs, 2 mM $MgCl_2$, 100 pmol each of sense and anti-sense primers, DNA template (amount as stated) and 2.5 units of Pfu DNA polymerase. PCR parameters were: 5 min at 94 °C (1 cycle); 1 min at 94 °C, 1 min 30 s at 55 °C, 2 min at 72 °C (25 cycles); and 10 min at 72 °C (1 cycle).

### Construction of the expression system for the biotinyl domaian

The region of the gene encoding the biotinyl domain of *E. coli* BCCP was amplified by PCR using 250 ng of *E. coli* genomic DNA as template and the oligonucleotides Dbccp and Rbccp as primers. Dbccp, 5′-AGCGGATCCGCAGCGGAAATCAGTGGTCACATC, primes on the coding strand and contains a *Bam*HI cleavage site (underlined); Rbccp, 5′-GTTCGAATTC **TTA**CTCGATGACGACCAGCGGCTC, primes on the non-coding strand and carries an *Eco*RI site (underlined) and a stop codon (in bold). After digestion of the purified PCR product with *Bam*HI and *Eco*RI, the resulting 255 bp DNA fragment was

**Table 1 Oligonucleotides employed in the mutagenesis of the biotinyl domain sub-gene**

The oligonucleotide sequences encoding the biocytin motif are indicated in bold.

| Mutant | Oligonucleotide | |
|---|---|---|
| MAK | senMAK | 5′ GAAGCC**ATGGCAAAG**ATGAACCAGAT |
| | antMAK | 5′ GTTCAT**CTTTGCCAT**GGCTTCAACGA |
| KAM | senKAM | 5′ GAAGCC**AAGGCAATG**ATGAACCAG |
| | antKAM | 5′ GTTCAT**CATTGCCTT**GGCTTCAACGA |
| DKA | senDKA | 5′ GAAGCC**GATAAAGCG**ATGAACCAG |
| | antDKA | 5′ GTTCAT**CGCTTTATC**GGCTTCAACGA |
| KKM | senKKM | 5′ GTTGAAGCC**AAGAAAATG**ATGAAC |
| | antKKM | 5′ CAT**CATTTTCTT**GGCTTCAACGA |
| MKK | senMKK | 5′ GCC**ATGAAAAAG**ATGAACCAGAT |
| | antMKK | 5′ CTGGTTCAT**CTTTTTCAT**GGCTTC |

cloned into pGEX-2T, to express the biotinyl domain as a fusion protein with glutathione S-transferase (GST). The resulting plasmid (pGwtBCCP) was propagated in *E. coli* XL1-Blue cells, and the DNA sequence encoding the entire biotinyl domain was checked on both strands to verify its identity.

### Directed mutagenesis of biotinyl domains by overlap extension PCR

The mutant biotinyl domains MAK, KAM, MKK, KKM and DKA were generated by overlap extension PCR [31]. For each mutant, three PCR fragments were obtained: the upstream PCR fragment, the downstream PCR fragment and the overlapping PCR extension fragment. In addition, a pair of sense and anti-sense oligonucleotides overlapping on their 3′-ends (20–25 bp) were required for each mutant (Table 1). These oligonucleotides primed in the region to be modified and encoded the appropriate changes of amino acid residue (Table 1). Upstream PCR fragments were obtained using the Dbccp oligonucleotide (described above) and the anti-sense oligonucletides listed in Table 1 as primers. Downstream PCR fragments were obtained using the sense oligonucleotides (Table 1) and the Rbccp oligonucleotide (described above) as primers. Downstream and upstream PCR reactions were carried out using 25 ng of pGwtBCCP plasmid DNA as template. Overlapping extension PCRs to generate the entire mutated sub-genes were performed using 20 ng of the appropriate purified downstream and upstream PCR fragments as template, and the oligonucleotides Dbccp and Rbccp as primers. The overlapping fragments were purified after agarose gel electrophoresis, digested with *Bam*HI and *Eco*RI, and then cloned into pGEX-2T cut with the same enzymes. In this way five plasmids were generated (pGkamBCCP, pGmakBCCP, pGkkmBCCP, pGmkkBCCP and pGdkaBCCP), each encoding a mutated biotinyl domain. The plasmids were propagated in *E. coli* XL1-Blue cells and the DNA sequences of the sub-genes were all verified.

### Expression and purification of recombinant wild-type and mutant biotinyl domains

*E. coli* BL21(DE3) cells transformed with the plasmids bearing the wild-type and mutated sub-genes were grown in LB medium [30] containing 100 $\mu$g/ml ampicillin and 10 mg/l *d*-biotin, and the cells were induced at an $A_{600}$ of 0.7–0.8 with IPTG (final concentration 1 mM). The cultures were grown for a further 3–6 h and harvested by centrifugation. The cells were washed in PBS [30], resuspended in 10 ml of PBS containing 1 mM PMSF, disrupted ($131 \times 10^6$ Pa) in a French press (SLM-AMINCO) and the cell debris removed by centrifugation (15000 ***g*** for 1 h at 4 °C). Samples were then loaded on to a Pharmacia Glutathione 4B Sepharose® affinity column (1.6 cm × 10 cm; flow rate 0.4 ml/min) equilibrated with PBS. The column was washed with the same buffer until a stable base line ($A_{280}$) was reached, and 100 units of thrombin (Calbiochem-Novabiochem Corporation) in 5 ml of PBS were then loaded on to the column (flow rate 1 ml/min). The column was incubated at 37 °C for 2 h and the recombinant biotinyl domain released by thrombin cleavage was then eluted from the column with PBS (100 ml) at a flow rate of 1 ml/min. Fractions (2 ml) containing the protein domain were concentrated using Centriprep-3 concentrators (Amicon) and loaded (in 2 ml batches) on to a Pharmacia Superdex™ 30 HR 16/60 gel filtration column (flow rate 1 ml/min) previously equilibrated with 20 mM potassium phosphate buffer, pH 7.0. Fractions containing the protein domain were pooled, concentrated and stored at −80 °C for future use.

### Anion-exchange chromatography of wild-type and mutant biotinyl domains

Further purification of the recombinant domains was carried out by anion-exchange chromatography on a Pharmacia Resource™ Q HR 16/10 column. Protein domains were loaded on to the column (flow rate 5 ml/min) previously equilibrated with 20 mM potassium phosphate (buffer A), pH 7.5, for purification of the wild-type biotinyl domain; pH 7.0 for the DKA mutant; and pH 8.0 for the KKM and MKK mutants. Biotinylated and unbiotinylated forms of the recombinant domains were resolved by applying a linear gradient (80 ml) of 0.1–0.2 M NaCl at a flow rate of 5 ml/min. Fractions (1 ml) of interest were pooled, dialysed against buffer A, pH 7.0, concentrated and stored at −80 °C.

### Expression and purification of BPL

BPL was purified from *E. coli* BL21(DE3)plysS cells transformed with the plasmid pTbpl (P. Reche and R. N. Perham, unpublished work). pTbpl was derived from the plasmid pMBR10 [32] and carries the gene encoding the BPL of *E. coli* under the T7lac promoter in pET11c. Cells were grown in LB medium containing 100 $\mu$g/ml ampicillin and at an $A_{600}$ of 0.6–0.8 were induced with IPTG (final concentration 1 mM). Cells were harvested by centrifugation and resuspended in 50 mM Tris/HCl, pH 8.2, containing 100 mM KCl, 1 mM DTT, 0.1 mM EDTA and 5% glycerol (buffer B). The cells were disrupted in a French press and the cell-free extract was applied to a Pharmacia Hiload™ Q HR 16/10 anion-exchange column (flow rate 2.5 ml/min). The recombinant protein was eluted with a linear gradient (70 ml) of 0–1 M KCl in buffer B at a flow rate of 2.5 ml/min. Fractions (2.5 ml) containing the protein of interest were pooled, concentrated (using a Centriprep-10 from Amicon) and loaded on to a Pharmacia Superdex™ 75 HR 16/60 filtration column (flow rate 1 ml/min) equilibrated with 50 mM potassium phosphate, pH 6.8, containing 100 mM KCl, 1 mM DTT, 0.1 mM EDTA and 5% glycerol (buffer C). Fractions (1 ml) containing the recombinant BPL were pooled and applied to a Pharmacia Hiload™ S HR 16/10 cation-exchange column (flow rate 2.5 ml/min). BPL was eluted with a linear gradient (70 ml) of 0–1.0 M of KCl in buffer C at a flow rate of 2.5 ml/min. Fractions (1 ml) containing the recombinant enzyme were pooled, dialysed against buffer C, concentrated and stored at −80 °C.

### Purification of LPL

*E. coli* BL21(DE3) cells transformed with the plasmid TM202 [24] were grown in LB medium containing 100 $\mu$g/ml ampicillin and at an $A_{600}$ of 0.6–0.7 were induced with IPTG (final concentration 1 mM) for a further 6 h. The cells were then harvested by centrifugation and resuspended in 10 ml of 50 mM Tris/HCl, pH 7.5, containing 1 mM PMSF, 1 mM EDTA and 10 % glycerol (buffer D). The cells were harvested again by centrifugation and disrupted in a French press. The cell-free extract was applied to a Pharmacia Hiload™ Q HR 16/10 column (flow rate 2.5 ml/min), and the recombinant protein was eluted by applying a linear gradient (100 ml) of NaCl (0–1 M) in buffer A. Fractions (2.5 ml) containing the ligase were concentrated to 1 ml, using a PM10 membrane in an Amicon stirred cell, and applied to a Pharmacia Superdex™ 75 HR 16/60 gel filtration column (flow rate 1 ml/min) equilibrated with buffer D. Fractions of interest (1 ml) were pooled, concentrated and stored at −80 °C.

### Expression and purification of E2p lipoyl domain

A recombinant lipoyl domain from *E. coli* E2p was purified using the expression plasmid pET11cE2p (E. L. Roberts and R. N. Perham, unpublished work). This plasmid is derived from the plasmid pGS233 [33] and carries the sub-gene encoding the innermost of the three lipoyl domains of *E. coli* E2p [34] under the control of the T7lac promoter in pET11c. *E. coli* BL21(DE3) cells transformed with pET11cE2p were grown in LB medium containing 100 $\mu$g/ml ampicillin and at an $A_{600}$ of 0.7–0.8 were induced for expression with IPTG (final concentration 1 mM) for 2 h. Recombinant protein was then purified from the cell-free extract by fractional precipitation with $(NH_4)_2SO_4$ (30–80 % saturation), followed by gel filtration on a Pharmacia Superdex™ 75 column and ion-exchange chromatography on a Pharmacia Resource™ Q column, essentially as described elsewhere [28].

### Lipoylation and biotinylation reactions *in vitro*

Biotinylation reactions were carried out in 15 $\mu$l mixtures containing 50 mM Tris/HCl, pH 8.3, 4 mM $MgCl_2$, 2.6 mM ATP (pH 7.0), 200 $\mu$M *d*-biotin and approximately 3 $\mu$g of the substrate protein. Lipoylation reactions were carried out in 15 $\mu$l mixtures containing 30 mM potassium phosphate buffer, pH 7.5, 1.5 mM ATP (pH 7.0), 1.5 mM $MgCl_2$ and 200 $\mu$M lipoic acid (pH 7.0) and approximately 3 $\mu$g of the substrate protein. Reactions were started by the addition of the relevant ligase and were incubated at 22–24 °C for 30 min, unless otherwise stated. After incubation, reactions were subjected to non-denaturing PAGE. The amounts of modified and unmodified domains were determined densitometrically from scanned pictures of the relevant gels using the program Phoretix 1D Advanced (v 3.0).

### Determination of molecular mass

Molecular masses of recombinant domains were estimated by electrospray (ES)-MS using a VG BioQ quadrapole mass spectrometer and myoglobin as the calibration standard. Proteins for ES-MS were prepared in aqueous solution (salt concentration $< 2$ mM) at a concentration of approx. 5 $\mu$M. Samples were diluted in acetonitrile (50 % final concentration) and formic acid (1 % final concentration). Theoretical molecular masses were calculated from the amino acid sequences using the *MW/pI* computing tool from the ExPasy Database.

### General techniques for protein analysis and purification

Protein purifications were carried out on a Pharmacia FPLC™ at 4 °C (unless otherwise stated) and column fractions were analysed by SDS/PAGE (12.5 and 20 % acrylamide) using the Pharmacia PhastSystem™. The concentrations of pure proteins were estimated by amino acid analysis [35]. Biotinylated and lipoylated forms of the domains were separated by means of non-denaturing PAGE (20 % acrylamide/1 % bis-acrylamide as separating gel; 5 % acrylamide as stacking gel) using a Tris/glycine buffer [36] lacking SDS. All gels were stained with Coomassie Brilliant Blue R-250.

## RESULTS

### Cloning, expression and purification of the wild-type BCCP biotinyl domain

A plasmid, pGwtBCCP, was constructed that encodes a fusion protein with GST at the N-terminus, followed by a thrombin cleavage site and an 82-residue peptide corresponding to the biotinyl domain of the BCCP component of *E. coli* ACC at the C-terminus. As judged by means of SDS/PAGE and Coomassie Blue staining (results not shown), the fusion protein represented 20–30 % of the total soluble protein in *E. coli* BL21(DE3) cells after IPTG induction. The biotinyl domain was purified by means of glutathione–Sepharose® affinity chromatography, thrombin cleavage of the retained fusion protein, followed by filtration of the released biotinyl domain on a Superdex™ 30 column to remove contaminating material (see the Materials and Methods section). The purified domain was found to be at least 98 % pure, as judged from Coomassie-stained 20 % SDS/PAGE gels (results not shown). The average yield of domain was 3 mg/l of *E. coli* cell culture. N-terminal sequence analysis revealed the expected Gly-Ser sequence generated by the thrombin cleavage, indicating a polypeptide chain of 84 residues in total.

### Construction of the mutant biotinyl domains MAK, KAM, KKM, MKK and DKA

The sub-gene encoding the biotinyl domain in pGwtBCCP was subjected to site-directed mutagenesis (see the Materials and Methods section) in the region encoding the highly conserved MKM sequence motif (residues M49, K50 and M51) that embodies the biotinyl-lysine residue (Figure 1). Mutant biotinyl domains with the altered sequence motifs MAK, KAM, KKM, MKK and DKA were generated. The KAM and MAK mutants present the target lysine residue for biotinylation moved by one position (to the N- and C-terminal side, respectively) in the $\beta$-turn, whereas the MKK and KKM mutants offer two possible target lysine residues in adjacent positions. In the DKA mutant, the target lysine residue is situated in the conserved DKA motif found in the structurally related lipoyl domains [13].

### Post-translational modification of the wild-type and mutant biotinyl domains *in vivo*

Wild-type and mutant biotinyl domains were purified from *E. coli* BL21(DE3) cells after induction with IPTG in media supplemented with 10 mg/l *d*-biotin (see the Materials and Methods section). The purified domains were analysed by means of non-denaturing 20 % PAGE, where the biotinylated form of the domain can be distinguished from the apo-form because it carries one fewer positive charge and therefore migrates more rapidly towards the anode (Figure 2). In addition, the proteins were analysed by ES-MS. The molecular mass of the wild-type

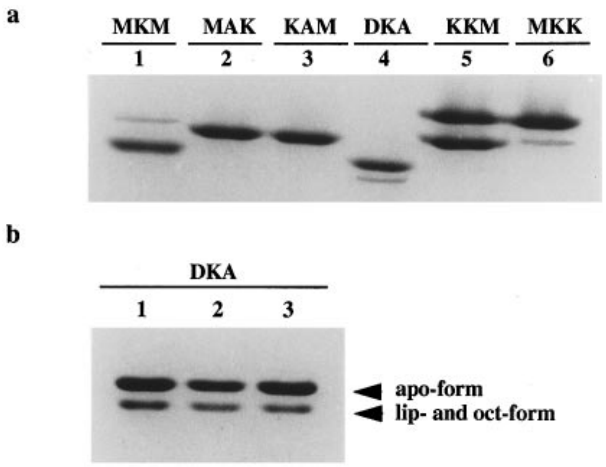


**Figure 2 Post-translational modification of the wild-type and mutant biotinyl domains *in vivo***

Sub-genes encoding the protein domains were expressed after induction with IPTG of *E. coli* Bl21(DE3) cells transformed with the relevant plasmids and grown in the presence of 10 mg/l *d*-biotin. Approximately 3 $\mu$g of each of the purified domains was subjected to non-denaturing PAGE (20% gels) and stained with Coomassie Blue. (**a**) Lane 1, wild-type biotinyl domain (MKM); lane 2, MAK mutant; lane 3, KAM mutant; lane 4, DKA mutant; lane 5, KKM mutant; lane 6, MKK mutant. The domains were purified from cells 3 h after induction. (**b**) DKA mutant domain purified from *E. coli* cells 6 h after induction and growth in the presence of: lane 1, 10 mg/l *d*-biotin; lane 2, 10 mg/l DL-lipoic acid; lane 3, 10 mg/l octanoic acid. In each lane, the upper band represents the apo-form of the domain. Abbreviations: lip, lipoyl; oct, octanoyl.

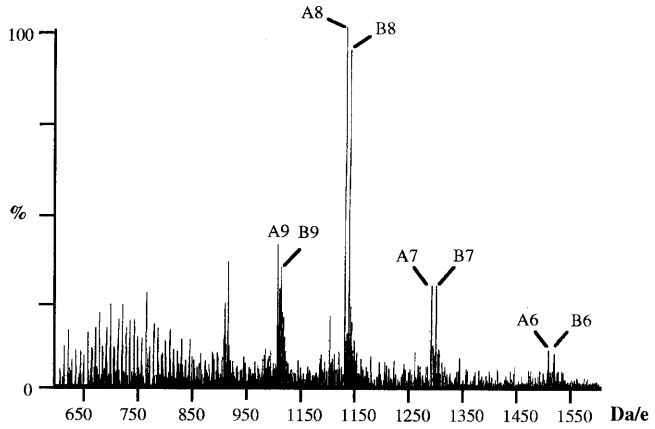


**Figure 3 ES mass spectrum of the DKA mutant biotinyl domain**

The post-translationally modified form of the DKA mutant domain (faster-migrating protein band in non-denaturing PAGE) was purified and analysed by ES-MS. A-series of peaks, octanoylated DKA domain, mass 9030 $\pm$ 2 Da (predicted mass 9028 Da); B-series, lipoylated DKA domain, mass 9090 $\pm$ 1 Da (predicted mass 9090 Da). The integers by each peak indicate the number of positive charges borne by the species.

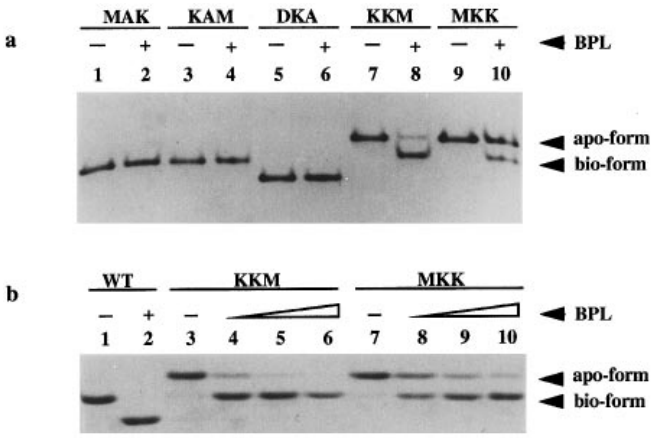


**Figure 4 Biotinylation of the biotinyl domains with BPL *in vitro***

Samples (approx. 22 $\mu$M) of purified apo-biotinyl domain were incubated for 30 min with BPL under the conditions described in the Materials and methods section and then analysed by means of non-denaturing PAGE (20% gels). The gels were stained with Coomassie Blue. (**a**) The mutant domains were incubated with BPL (concentration 0.7 $\mu$M); MAK mutant without BPL (lane 1) and with BPL (lane 2); KAM mutant without BPL (lane 3) and with BPL (lane 4); DKA mutant reaction mixtures without BPL (lane 5) and with BPL (lane 6); KKM mutant without BPL (lane 7) and with BPL (lane 8); MKK mutant without BPL (lane 9) and with BPL (lane 10). (**b**) Wild-type (MKM) domain incubated without BPL (lane 1) and with 0.2 $\mu$M BPL (lane 2); KKM mutant incubated without BPL (lane 3), with 0.5 $\mu$M BPL (lane 4), with 0.6 $\mu$M BPL (lane 5), and with 0.7 $\mu$M BPL (lane 6); MKK mutant incubated without BPL (lane 7), with 1.4 $\mu$M BPL (lane 8), with 2.8 $\mu$M BPL (lane 9) and with 5.6 $\mu$M BPL (lane 10). In each lane, the symbol + or — represents incubation with or without BPL respectively. Abbreviation: bio, biotinyl.

biotinyl domain was found to be 9204 $\pm$ 0.5 Da, 226 Da more than that calculated from the amino acid sequence (8978 Da) and corresponding exactly to that required for the addition of a biotinyl group [37]. The extent of biotinylation of the wild-type domain was over 90%, as judged by non-denaturing PAGE (Figure 2a). Thus the wild-type biotinyl domain appears to be almost fully biotinylated *in vivo* under the conditions used. The molecular mass of the putative apo-form was also determined by means of ES-MS after it was separated from the modified form by means of anion-exchange chromatography and, as expected, its molecular mass (8978 $\pm$ 0.8 Da) was that calculated from the amino acid sequence.

The MAK and KAM mutant domains, when subjected to non-denaturing PAGE, displayed a single protein band (Figure 2a). Their measured molecular masses (MAK 8918 $\pm$ 0.6 Da, KAM 8918 $\pm$ 0.7 Da) corresponded exactly with those of the apo-forms, as calculated from the amino acid sequences (8918 Da), which indicates that they were not biotinylated *in vivo*. In contrast, the KKM mutant displayed two protein bands of approximately the same intensity in non-denaturing PAGE (Figure 2a). It also revealed two components on ES-MS, whose molecular masses were those expected for the apo-protein (measured mass 8975 $\pm$ 2 Da, expected mass 8975 Da) and the singly biotinylated form of the domain (measured mass 9202 $\pm$ 1 Da, expected mass 9201 Da). Thus, under the experimental conditions used (*E. coli* cells were harvested 3 h after induction in a medium supplemented with biotin), 50% of the mutant was biotinylated *in vivo*. We could not detect any doubly biotinylated form of the KKM mutant by means of non-denaturing PAGE or ES-MS. Likewise, post-translational modification of the MKK mutant was found to occur *in vivo*, but to be less efficient than that of the KKM mutant. In non-denaturing PAGE (Figure 2a) the protein band corresponding to the modified form represented only 10% of the total domain and its molecular mass (9201 $\pm$ 1 Da), as determined by ES-MS, corresponded with that of the singly biotinylated form of the protein (9201 Da). Again, no doubly biotinylated form was detected.

The purified DKA mutant domain also contained two components when subjected to non-denaturing PAGE (Figure 2a), that corresponding to the apo-domain (upper band) representing 90% of the protein. These two components were purified and

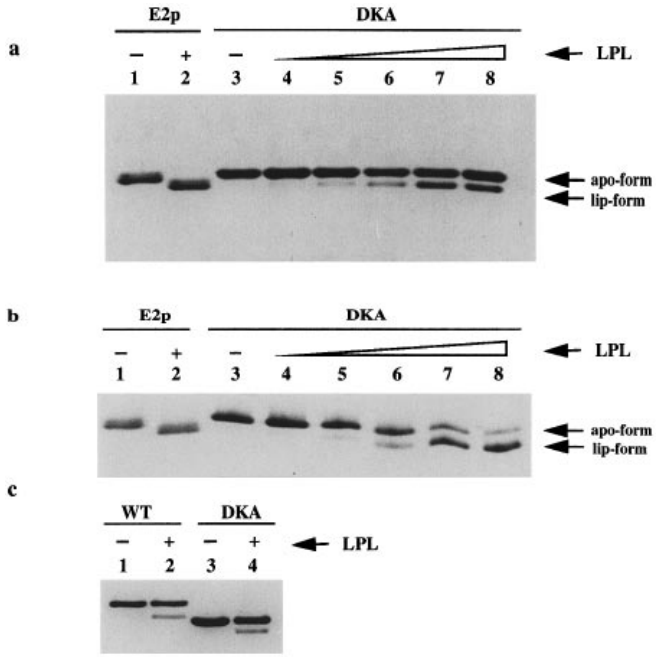


**Figure 5 Lipoylation of the DKA mutant biotinyl domain by LPL *in vitro***

Samples (approx. 3 $\mu$g) of purified apo-DKA mutant domain were incubated with LPL under the conditions described in the Materials and Methods section and then analysed by means of non-denaturing PAGE (20% gels). The gels were stained with Coomassie Blue. (**a**) All incubations were for 30 min. Control samples of the innermost of the three lipoyl domains from *E. coli* E2p were incubated without LPL (lane 1) and with 1.8 $\mu$M LPL (lane 2). The DKA biotinyl domain was incubated without LPL (lane 3), with 1.8 $\mu$M LPL (lane 4), with 3.6 $\mu$M LPL (lane 5), with 7.2 $\mu$M LPL (lane 6), with 18 $\mu$M LPL (lane 7) and with 36 $\mu$M LPL (lane 8). (**b**) Same as (**a**), except the incubations were extended to 12 h. (**c**) Protein domains were incubated for 30 min with LPL at a concentration of 18 $\mu$M. The wild-type biotinyl domain was incubated without LPL (lane 1) and with LPL (lane 2); the DKA mutant was incubated without LPL (lane 3) and with LPL (lane 4). In each lane, the symbol + or − represents incubation with or without LPL respectively. Abbreviations: lip, lipoyl; WT, wild-type.

analysed by ES-MS. The molecular mass of the upper protein band (8902 $\pm$ 1 Da) was that expected for the apo-form of the mutant (8902 Da). The mass spectrum of the lower protein band (Figure 3) displayed two products, with molecular masses of 9090 Da and 9030 Da. These mass differences correspond closely with those predicted if the protein had been lipoylated (+188 Da) and octanoylated (+126 Da) respectively. The molar ratio between the putative lipoylated and octanoylated forms of the DKA mutant was approximately 1:1 and no biotinylated form of the domain was detectable. The ratio of modified to unmodified form of the DKA mutant did not change significantly when the expression of the relevant mutant was induced in the presence of lipoic acid or octanoic acid instead of *d*-biotin (Figure 2b), nor did the proportion of lipoylated to octanoylated form, as judged by ES-MS, under the same conditions (results not shown). However, increasing the time of expression from 3 h (Figure 2a) to 6 h (Figure 2b) was accompanied by an increase of the modified protein band from approx. 10% to 20%.

### Biotinylation of the biotinyl domains *in vitro*

Biotinylation of the apo-biotinyl domains was carried out *in vitro* by incubating them with purified BPL under the conditions described in the Materials and Methods section. The extent of biotinylation was analysed by means of non-denaturing PAGE (Figure 4a) and the results were found to be fully consistent with those obtained *in vivo*. Thus, the MAK or KAM mutants were not biotinylated; the KKM and MKK mutants were biotinylated but neither of them appeared to be doubly biotinylated; and the DKA mutant was not biotinylated at all (even if the incubation time was extended to 12 h or more). The concentration of BPL used in these experiments was 0.7 $\mu$M, which in preliminary experiments had been found to be sufficient to achieve the complete biotinylation of the wild-type biotinyl domain in 5 min (data not shown). We also studied the biotinylation of the KKM and MKK mutant domains by varying the concentration of BPL in the reaction mixture. Under identical conditions the concentration of ligase needed to biotinylate the KKM and MKK mutants was found to be 2- and 20-fold respectively that required to achieve complete modification of the wild-type biotinyl domain (Figure 4b).

### Lipoylation of the biotinyl domains *in vitro*

In view of the lipoylation and octanoylation of the DKA biotinyl domain observed *in vivo*, we attempted to lipoylate the domain *in vitro* using the LPL. The concentration of LPL was varied under the conditions described in the Materials and Methods section and the reaction mixtures were analysed by means of non-denaturing PAGE. The DKA mutant was observed not to be lipoylated in 30 min at concentrations of LPL up to 1.8 $\mu$M (Figure 5a), nor did it become detectably lipoylated if the incubation time was increased to 12 h (Figure 5b). Under identical conditions, the lipoyl domain of *E. coli* E2p was fully lipoylated in less than 5 min (Figure 5). However, when the concentration of LPL was increased 10-fold or more, lipoylation of the DKA biotinyl domain was significant (Figure 5, lanes 7 and 8), but an incubation of more than 12 h was required to bring it to completion (results not shown). This large amount of LPL was also found to be capable of catalysing the lipoylation of the wild-type biotinyl domain to a similar extent (Figure 5c). In all cases the identities of the products were verified by ES-MS.

## DISCUSSION

The biotin cofactor of *E. coli* acetyl-CoA carboxylase is attached covalently by amide linkage to the $N^6$-amino group of a unique lysine residue of the BCCP component. The attachment is catalysed by BPL as a post-translational modification, and the ligase selects a unique target lysine residue located in a conserved amino acid sequence motif MKM near the tip of a tight $\beta$-turn in the structure of the biotinyl domain (Figure 1). Li and Cronan [3] were the first to show that fusion of the C-terminal 84 amino acids of BCCP to $\beta$-galactosidase led to the production of biotinylated $\beta$-galactosidase molecules *in vivo*. However, the expression level of this fusion protein was low. Chapman-Smith et al. [5] have also reported the expression of a sub-gene in *E. coli* encoding the C-terminal 87 residues of the *E. coli* BCCP, generating a partly biotinylated protein *in vivo*. As pointed out by these authors, the production of the peptide, where expression was under the control of the T7 promoter of pET, appeared to have saturated the capacity of the intracellular BPL. The GST–biotinyl domain fusion that we have used allows the preparation of a biotinyl domain that has been biotinylated to a large extent (90% or more) *in vivo* under the conditions used (supplementing the growth medium with 10 mg/l *d*-biotin). A small amount of apo-domain (Figure 2) could be separated from the biotinylated form by ion-exchange chromatography.

The presentation of the target lysine residue in the structure of the biotinyl domain is evidently of crucial importance for the

biotinylation reaction. Thus, the KAM and MAK mutant domains, with the lysine residue moved one position to the N- and C-terminal side respectively, were not biotinylated either *in vivo* (Figure 2) or *in vitro* (Figure 4) and the KKM or MKK mutant domains, with two potential sites for biotinylation, were found to be biotinylated at only one site. Given the results with the KAM and MAK mutants, it is reasonable to suppose that this is the lysine residue in its normal position, though this has not been formally proved by amino acid sequence analysis. Thus, as with lipoylation of the lipoyl domain from 2-oxo acid dehydrogenase multienzyme complexes where a full sequence analysis was carried out [29], it clearly is essential for the target lysine residue to be properly positioned in the hairpin loop on the surface of the biotinyl domain for post-translational modification to occur.

However, the biotinylation of the KKM and MKK mutant domains, although it occurred, was significantly impaired. Under identical conditions *in vivo*, the modification of the KKM mutant reached 50% of that of the wild-type domain and that of the MKK mutant only 10%. The DKA mutant domain was not biotinylated at all. Similar results were obtained when the biotinylation of the various domains by purified BPL was examined *in vitro* (Figure 4). Thus it would appear that the methionine residues flanking the target lysine form part of the surface of the biotinyl domain that is recognized by BPL. None the less, although the MKM motif for the biotinylated lysine residue is highly conserved in all biotinylated proteins [8], it is not sufficient to ensure biotinylation. This is clear from the observation that importing the MKM motif into the lipoyl domain of *Bacillus stearothermophilus* pyruvate dehydrogenase complex leaves it capable of being lipoylated but does not render it a substrate for biotinylation *in vivo* in *E. coli* [29].

Replacement of the methionine residues flanking the target lysine with leucine on the biotinyl domain of the $\alpha$-subunit from human propionyl-CoA carboxylase [38] and the 1.3 S subunit of *Pseudomonas shermanii* transcarboxylase [39] has demonstrated that the methionine residues are not essential for correct biotinylation of these proteins *in vivo* by *E. coli* BPL. However, in the *P. shermanii* transcarboxylase, further analysis showed the subsequent carboxylation of the biotin moiety to be defective [40]. Taking these data together with our own results, it appears that methionine residues flanking the target lysine residue may have been evolutionarily selected in order to be compatible with the interaction first with the BPL and subsequently with the relevant carboxylase. Further investigation of the effects of systematic exchange of the methionine residues on the biotinylation and carboxylation reactions will be necessary to understand these protein–protein interactions and their needs for compatibility in more detail.

Previous studies of the lipoylation of the lipoyl domain from the pyruvate dehydrogenase multienzyme complex of *B. stearothermophilus* in *E. coli* indicated that the conserved aspartic acid and alanine residues surrounding the target lysine residue (the DKA motif) are not essential for lipoylation, but that the correct positioning of the lysine at the tip of the tight $\beta$-turn in the domain is crucial [29]. The DKA motif is replaced by DKV in the lipoyl domains of the 2-oxoglutarate dehydrogenase complexes of *E. coli* [15,41] and *Azobacter vinelandii* [16], by AKA in the N-terminal lipoyl domain of the pyruvate dehydrogenase complex of *A. vinelandii* [17], and by VKA in the lipoylated H-protein of the glycine decarboxylase complex of ox [42], pea leaf [43,44] and *E. coli* [45]. None the less, all these domains can be lipoylated *in vivo* in *E. coli*. Thus, we were surprised to find that the DKA mutant biotinyl domain was lipoylated and octanoylated *in vivo*, albeit at a low level (Figures 2 and 3). Octanoylation of lipoyl domains as an aberrant modification *in vivo*, under conditions of lipoic acid deficiency [27] or where over-expression of the lipoyl domain swamps the lipoylating machinery of the *E. coli* cell, has been observed before [28,42].

It would appear that the DKA motif, when displayed at the right position in a structurally related protein, is sufficient to direct lipoylation of the target lysine residue *in vivo*. This is in contrast with the lack of lipoylation of the lipoyl domain of the bovine branched-chain 2-oxo acid dehydrogenase complex, which also exhibits the DKA motif, expressed in *E. coli* [46]. It is significant that the lipoylation of the DKA biotinyl domain catalysed by the purified *E. coli* LPL *in vitro*, although it took place, was very slow. Indeed, it was not very different from the lipoylation of the wild-type biotinyl domain under forcing conditions (stoichiometric amounts of domain and ligase) *in vitro* (Figure 5). This may indicate that the post-translational modification of the DKA mutant domain *in vivo* is catalysed by an enzyme different from the purified LPL used here, the product of the *lplA* gene [24]. That enzyme could be the *lipB*-dependent LPL, which is believed to be responsible for most of the octanoylation occurring *in vivo* [23]. However, it cannot be discounted that the lipoylation of the DKA mutant observed *in vivo* could also be due, at least in part, to the product of the *lplA* gene under the conditions of high level expression of the domain.

Overall, it is apparent that lipoylation and biotinylation of the relevant protein domains do not rely simply on the recognition of unique and well-conserved sequence motifs. Instead, the recognition process between the domain and its cognate ligase has a major structural component to it, notably the positioning of the target lysine residue in the exposed hairpin loop in one of the two $\beta$-sheets that make up the domains, together with additional recognition sites elsewhere on the domains. This emerging position has been somewhat clouded recently by the observation that a 13-residue peptide tag can be added to proteins that appears to target biotinylation by BPL to a lysine residue embedded in it [47,48]. Whether this peptide has any persistent structure in solution, and if so whether it resembles the $\beta$-turn that houses the biotinylated lysine residue in native proteins, is unknown. How these recognition processes function at the molecular level remains an important problem to be solved in detail.

We thank the Federation of European Biochemical Societies for a Fellowship awarded to P.R., the European Union for a Marie Curie Fellowship awarded to Y.-L.L., and the BBSRC for a research grant awarded to R.N.P. The core facilities of the Cambridge Centre for Molecular Recognition are supported by the BBSRC and The Wellcome Trust. We are grateful to Dr. J. E. Cronan for the gift of plasmids pMBR10 and TM202 encoding BPL and LPL respectively.

## REFERENCES

1 Wakil, S. J., Stoops, J. K. and Joshi, V. C. (1983) Annu. Rev. Biochem. **52**, 537–579
2 Fall, R. R. and Vagelos, P. R. (1972) J. Biol. Chem. **247**, 8005–8015
3 Li, S.-J. and Cronan, Jr., J. E. (1992) J. Biol. Chem. **267**, 855–863
4 Reed, K. E. and Cronan, Jr., J. E. (1991) J. Biol. Chem. **266**, 11425–11428
5 Chapman-Smith, A., Turner, D. L., Cronan, Jr., J. E., Morris, T. W. and Wallace, J. C. (1994) Biochem. J. **302**, 881–887
6 Neonortas, E. and Beckett, D. (1996) J. Biol. Chem. **271**, 7559–7567
7 Athappilly, F. K. and Hendrickson, W. A. (1995) Structure **3**, 1407–1419
7a Kraulis, P. J. (1991) J. Appl. Crystallogr. **24**, 946–950
8 Knowles, J. R. (1989) Annu. Rev. Biochem. **58**, 195–221
9 Salmons, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C. and Wood, H. G. (1988) J. Biol. Chem. **263**, 6461–6464
10 Brocklehurst, S. M. and Perham, R. N. (1993) Protein Sci. **2**, 626–639
11 Patel, M. S. and Roche, T. E. (1990) FASEB J. **4**, 3224–3233
12 Perham, R. N. (1991) Biochemistry **30**, 8501–8512
13 Dardel, F., Davis, A. L., Laue, E. D. and Perham, R. N. (1993) J. Mol. Biol. **229**, 1037–1048

14 Green, J. D. F., Laue, E. D., Perham, R. N., Ali, S. T. and Guest, J. R. (1995) J. Mol. Biol. **248**, 328–343
15 Ricaud, P. M., Howard, M. J., Roberts, E. L., Broadhurst, R. W. and Perham, R. N. (1996) J. Mol. Biol. **264**, 179–190
16 Berg, A., Vervoort, J. and de Kok, A. (1996) J. Mol. Biol. **261**, 432–442
17 Berg, A., Vervoort, J. and de Kok, A. (1997) Eur. J. Biochem. **244**, 352–360
18 Barker, D. F. and Campbell, A. M. (1981) J. Mol. Biol. **146**, 451–467
19 Barker, D. F. and Campbell, A. M. (1981) J. Mol. Biol. **146**, 469–492
20 Howard, P. K., Shaw, J. and Otsuka, A. J. (1985) Gene **35**, 321–331
21 Wilson, K. P., Shewchuk, L. M., Brennan, R. G., Otsuka, A. J. and Matthews, B. W. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 9257–9261
22 Cronan, Jr., J. E. (1989) Cell **58**, 427–429
23 Morris, T. W., Reed, K. E. and Cronan, Jr., J. E. (1995) J. Bacteriol. **177**, 1–10.
24 Morris, T. W., Reed, K. E. and Cronan, Jr., J. E. (1994) J. Biol. Chem. **269**, 16091–16100
25 Green, D. E., Morris, T. W., Green, J., Cronan, Jr., J. E. and Guest, J. R. (1995) Biochem. J. **309**, 853–862
26 Brookfield, D. E., Green, J., Ali, S. T., Machado, R. S. and Guest, J. R. (1991) FEBS Lett. **295**, 13–16
27 Ali, S. T., Moir, A. J. G., Ashton, P. R., Engel, P. C. and Guest, J. R. (1990) Mol. Microbiol. **4**, 943–950
28 Dardel, F., Packman, L. C. and Perham, R. N. (1990) FEBS Lett. **264**, 206–210
29 Wallis, N. G. and Perham, R. N. (1994) J. Mol. Biol. **236**, 209–216
30 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY
31 Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R. (1989) Gene **77**, 51–59
32 Buoncristiani, M. R. and Otsuka, A. J. (1988) J. Biol. Chem. **263**, 1013–1016
33 Stephens, P. E., Darlison, M. G., Lewis, H. M. and Guest, J. R. (1983) Eur. J. Biochem. **133**, 481–489
34 Packman, L. C., Hale, G. and Perham, R. N. (1984) EMBO J. **3**, 1315–1319
35 Packman, L. C., Borges, A. and Perham, R. N. (1988) Biochem. J. **252**, 79–86
36 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
37 Krishna, R. G. and Wold, F. (1993) in Methods in Protein Sequence Analysis (Imahori, K. and Sakiyama, F., eds.), pp. 167–172, Plenum Press, New York
38 Leon-Del-Rio, A. and Gravel, R. A. (1994) J. Biol. Chem. **269**, 22964–22968
39 Shenoy, B. C., Paranjape, S., Murtif, V. L., Kumar, G. K., Salmons, D. and Wood, H. G. (1988) FASEB J. **2**, 2505–2511
40 Shenoy, B. C., Salmons, D. and Kumar, G. K. (1993) Arch. Biochem. Biophys. **304**, 359–366
41 Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K. and Guest, J. R. (1984) Eur. J. Biochem. **141**, 361–374
42 Fujiwara, K., Okamura-Ikeda, K. and Motokawa, Y. (1992) J. Biol. Chem. **267**, 20011–20016
43 Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M. and Douce, R. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 4850–4853
44 Macherel, D., Bourguignon, J., Forest, E., Faure, M., Cohen-Addad, C. and Douce, R. (1996) Eur. J. Biochem. **236**, 27–33
45 Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K. and Motokawa, Y. (1993) Eur. J. Biochem. **216**, 539–548
46 Griffin, T. A., Wynn, R. M. and Chuang, D. T. (1990) J. Biol. Chem. **265**, 12104–12110
47 Schatz, P. J. (1993) Bio/Technology **11**, 1138–1143
48 Tsao, K.-L., DeBarbieri, B., Michel, H. and Waugh, D. S. (1996) Gene **169**, 59–64

Received 13 August 1997/24 September 1997; accepted 3 October 1997