Polymorphism of the yeast pyruvate carboxylase 2 gene and protein: effects on protein biotinylation

Dale L. VAL,* Anne CHAPMAN-SMITH,*† Michelle E. WALKER,* John E. CRONAN, Jr.† and John C. WALLACE*

* Department of Biochemistry, University of Adelaide, Adelaide, South Australia 5005, Australia

and † Departments of Microbiology and Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A.

In *Saccharomyces cerevisiae* there are two isoenzymes of pyruvate carboxylase (Pyc) encoded by separate genes designated *PYC1* and *PYC2*. We report the isolation and sequencing of a *PYC2* gene, and the localization of both genes on the physical map of *S. cerevisiae*. Comparison with the previously reported sequence [Stucka, Dequin, Salmon and Gancedo (1991) Mol. Gen. Genet. 229, 307–315] revealed significant differences within the open reading frame. The most notable difference was near the 3' end, where we found a single base deletion reducing the open reading frame by 15 bases. We have confirmed the C-terminus of Pyc2 encoded by the gene isolated here by expressing and purifying an

86-amino-acid biotin-domain peptide. In addition, we investigated the effects of the two changes in the Pyc2 biotin domain (K1155R substitution and Q1178P/five-amino-acid extension) on the extent of biotinylation *in vivo* by *Escherichia coli* biotin ligase, and compared the biotinylation of peptides containing these changes with that of two different-length Pyc1 biotin-domain peptides. The K1155R substitution had very little effect on biotinylation, but the five-amino-acid C-terminal extension to Pyc2 and the N-terminal extension to Pyc1 both improved biotinylation *in vivo*.

INTRODUCTION

Pyruvate carboxylase (Pyc) (EC 6.4.1.1) belongs to the family of biotin-dependent carboxylases, enzymes that catalyse the carboxylation of certain acceptor molecules using a covalently attached biotin moiety as a mobile carboxyl carrier [1,2]. Pyc catalyses the two-step carboxylation of pyruvate to oxaloacetate, a reaction of importance both for gluconeogenesis and the replenishment of tricarboxylic acid-cycle intermediates. Sequencing of the yeast *PYC* gene (*PYC1*) and partial proteolysis of the purified enzyme showed that the enzyme has a multidomain structure, composed of an N-terminal ATP-binding domain, a central pyruvate-binding domain and a C-terminal biotinyl-carrier domain [3].

Sequence comparisons between a range of biotin enzymes has revealed that the biotinyl group is attached to a specific lysine residue usually 34 or 35 residues from the C-terminus [4,5], and that the sequence immediately surrounding the attachment site shows a high degree of identity between a wide range of enzymes and species [3,4,6,7]. Experiments with enzymes from prokaryotes have shown that biotin domains exist as relatively isolated and stable regions [8] of 75–80 residues [5,7], and amino acids 30–40 residues either side of the attachment site are known to be important for biotinylation [5,8,9]. Biotin ligase enzymes from various sources have been found to recognize and biotinylate acceptor proteins from very different species both *in vivo* [5] and *in vitro* [10], suggesting an essentially common tertiary structure.

Interestingly, biotin domains have some sequence similarity to lipoyl domains [3] which are found in pyruvate dehydrogenase (PDH) and the H-protein of the glycine-cleavage system. The lipoyl cofactor, much like biotin, functions as a swinging arm [11] carrying substrate between active sites, and it too is covalently

attached to a lysine residue similarly located within the domain [12,13]. The three-dimensional structure of the PDH lipoyl domains from *Bacillus stearothermophilus* [14] and *Escherichia coli* [15] have been determined by NMR, and the domain from the pea leaf H-protein by X-ray crystallography [16]. Brocklehurst and Perham [17] have proposed a three-dimensional structure for the biotin domain of yeast *PYC1*, predicting that it contains the same fold as the PDH lipoyl domains. However, as yet there have been no direct structural or functional studies on a biotin domain from pyruvate carboxylase.

After the sequencing of *PYC1*, we [18] and others [19] reported the existence and chromosomal localization of a second *PYC* gene (*PYC2*) in yeast. Analysis of both single null mutant strains (*pyc1* and *pyc2*), and the partially purified enzymes, led Stucka and co-workers [19] to conclude that there were no notable differences between the two isoenzymes. In contrast, we do find significant differences in the activities of the two Pyc isoenzymes [18], as revealed for example in the growth phenotypes of the single null mutants [20]. One explanation of this disparity could be the existence of different polymorphic forms of the *PYC2* and/or *PYC1* genes.

In this paper we report the independent isolation and sequencing of a yeast *PYC2* gene with a significant number of differences in the open reading frame (ORF) when compared with the gene isolated by Stucka et al. [19], the most notable being a single base deletion resulting in translation termination five amino acids early. We present data confirming the predicted C-terminus of the Pyc2 protein expressed from the gene sequenced here. In addition, we report the relative extent of biotinylation *in vivo* of the biotin domains encoded by the different forms of *PYC2* with respect to the sequence differences found in this region.

Abbreviations used: Pyc, pyruvate carboxylase; PDH, pyruvate dehydrogenase; BCCP, biotin carboxyl carrier protein; ACC, acetyl-CoA carboxylase; IPTG, isopropyl β -D-thiogalactopyranoside; ORF, open reading frame.

† To whom correspondence should be addressed.

The nucleotide sequences reported in this paper have been submitted to the GenBank/EMBL Data Bank under accession numbers BankIt14449 and U35647.

EXPERIMENTAL

Strains and materials

The pEMBL Ye23 *Sau* 3A genomic library from *Saccharomyces cerevisiae* strain S288C was obtained from Dr. C. P. Morris, the λ EMBL3 library was obtained as previously outlined [3], and prime λ -clone grid filters were obtained from the laboratory of Dr. M. V. Olson. BL21 Δ DES, used as a host for expression studies, is an *E. coli* B strain ($F^- ompT r_B^- m_B^-$) lysogenized with a prophage (λ DES) which expresses T7 RNA polymerase on induction with isopropyl β -D-thiogalactopyranoside (IPTG) [21].

Nitrocellulose was obtained from Schleicher & Schuell; [^{3} H]biotin (35 Ci/mmol) was from Amersham; [α - ^{32} P]ATP (3000 mCi/mmol), DNA polymerase I (Klenow fragment) and T4 DNA ligase were purchased from Bresatec Ltd. (Adelaide, Australia); T7 DNA polymerase (unmodified) was from New England Biolabs; *Taq* DNA polymerase (Pyrostate) was obtained from Molecular Genetic Resources; restriction enzymes and other DNA-modification enzymes were from Pharmacia-LKB.

Oligonucleotides were purchased from Bresatec. The restriction sites in the oligonucleotides are indicated by underlining, and the mutagenic changes are shown by bold lettering. The sequences of the oligonucleotides are as follows: O3671, 5'-CTGTTGCTAAACCCATGGCTGATGTCCAC-3'; O6612, 5'-CTACCCCCATCCCCAAAAAAGTAA-3'; O6613, 5'-GTTTTCATTAGGGATGGTGAAAGTG-3'; pSeq, 5'-CTAGATAAAATATACGCC-3'; PCR*Bam*HI, 5'-GAATGG-ATCCATGGTAAGTG-3'; O5011, 5'-CGAATTTTTC-TCATGACACCATTCAAATC-3'; O5012, 5'-CATGCATGT-CTGCCATGGATTAGTAAC-3'.

DNA manipulations and sequencing

Transformations, isolation of plasmid DNA, restriction enzyme digests, agarose-gel electrophoresis, and library screening were all carried out using standard methods [22]. Hybridization of prime λ -clone grid filters was performed by methods supplied by Dr. L. Riles from the laboratory of Dr. M. V. Olson.

PCR was performed using 100 ng of oligonucleotide and 10 ng of template DNA in 20 μ l of 1 $\times$ Pyrostate buffer (supplied by Molecular Genetic Resources), containing 1.25 mM MgCl₂ and 200 μ M dNTPs. Twenty cycles were performed for each reaction with the cycle conditions being: denaturation, 95 °C for 2 min; annealing, 60 °C for 1.5 min; and synthesis, 72 °C for 3 min.

Sequencing was performed using denatured double-stranded plasmid DNA and Sequenase (Pharmacia), by the dideoxy-chain-termination method as outlined in the Pharmacia Sequenase Version II Kit. Overlapping sequencing clones were generated by a combination of nested deletions (Pharmacia Exo III Nested Deletion Kit) and subcloning.

Preparation of the Pyc biotin-domain expression constructs

Subcloning and subsequent expression of each of the Pyc2 biotin-domain peptides required the introduction of an initiating methionine (K1095M) and *Nco*I cloning site at +3281 in *PYC2*. This was performed by site-directed mutagenesis using a modification (unmodified T7 DNA polymerase instead of Klenow) of the method of Kunkel et al. [23] with the oligonucleotide O3671 and clone p58 (see the Results section). The construct required to express the Pyc2 biotin-domain peptide corresponding to the *PYC2* DNA sequence reported here (c86K) was then obtained by subcloning the 3' end of *PYC2* as a 900 bp *Nco*I-*Xho*I fragment into the pET-16b expression system (Novagen).

The Q1178P/VIFTR extension and the K1155R mutation were both introduced into *PYC2* (clone p58) by PCR with the

Figure 1 PCR mutagenesis strategy

Short arrows (not to scale) indicate the position of the mutagenic/PCR primers, and the dashed arrow indicates the size of the full-length PCR product obtained using the external O3671 and PCR*Bam*HI mutagenic primers. The solid lines show the size and position of the PCR fragments that were combined in the PCR reactions used to prepare the various mutant full-length products, and the scale below shows the nucleotide positions relative to the first ATG in *PYC2*. The schematic diagram shows the position, length and important features of the biotin domain expressed from the *PYC2* sequence reported here.

oligonucleotides O6612 and O6613 respectively, as follows: the 460 bp 3' *PYC2* fragment containing the O6613 mutation was obtained by PCR using the external primers (O3671 and PCR*Bam*HI) with a mixture of the wild-type 349 bp and mutant 209 bp overlapping fragments shown in Figure 1. The 460 bp fragment thus obtained was then blunt-ended with Klenow, digested with *Nco*I and *Bam*HI, and subcloned into pET-16b. Resultant clones were screened by DNA sequencing to check for the correct (mutant) construct required to express the c86R peptide. The construct required to express the peptide with the O6612 mutation (c91K) was obtained as outlined above using the O6612 mutant 275 bp fragment (Figure 1) in place of the mutant 209 bp fragment, and the construct expressing the peptide with both mutations (c91R) was similarly obtained by PCR with the external primers from a mixture of the two 460 bp single mutant fragments followed by subcloning into pET-16b.

Introduction of the restriction sites and initiating methionine (where required) for the subcloning and subsequent expression of the c104 and c85 *Pyc1* peptides was performed using site-directed mutagenesis as previously outlined. To express the c104 peptide a *Bsp*HI site was introduced at the native ATG located at +3221 in *PYC1* using the oligonucleotide O5011. The 3' end of *PYC1* containing this O5011 mutation was then subcloned into pET-16b as a 980 bp *Bsp*HI-*Bam*HI fragment. To construct the plasmid for expressing the c85 peptide, an ATG (K1094M) and *Nco*I site were introduced into *PYC1* at +3278 using the oligonucleotide O5012. The 3' end of *PYC1* containing this O5012 mutation was then subcloned into pET-16b as a 920 bp *Nco*I-*Bam*HI fragment.

Expression of the Pyc biotin-domain constructs

Expression of each of the Pyc biotin-domain constructs was performed in a derivative of strain BL21 Δ DES which had previously been transformed with pCY216 [24], a plasmid that expresses *E. coli* biotin ligase on induction with arabinose. Bacterial growth, induction conditions and preparation of cell lysates were as described previously [24].

Biotinylation and purification of the C-terminal peptide from Pyc2

The cells obtained from a 250 ml fermentation after induction with 0.25 mM IPTG were harvested, lysed, biotinylated *in vitro* and labelled with [³H]biotin as outlined by Chapman-Smith et al. [24]. The resultant biotinylated lysate was diluted 1:3 with buffer A (20 mM Bistris, pH 6.1, 0.1 mM EDTA), and run on a Q-Sepharose column (12 cm × 2.6 cm). The Pyc2 biotin-domain peptide was eluted at 2.5 ml/min with a 0–1.5 M NaCl gradient in buffer A, over 320 min. Fractions of volume 10 ml were collected and those containing ³H were pooled and dialysed against 2 mM ammonium acetate. The peptide was concentrated on a 1 ml Resource-Q column (Pharmacia) employing a scaled-down version of the Q-Sepharose protocol and run on a 300 ml Sephacryl S-200 gel-filtration column in 50 mM potassium phosphate, pH 7.0, containing 0.15 M NaCl, 0.2 mM dithioerythritol and 0.1 mM EDTA. Fractions containing the Pyc2 biotin-domain peptide (³H counts) were dialysed and subjected to C4 reverse-phase HPLC using a 1%/min acetonitrile gradient in 0.082% trifluoroacetic acid. The resultant peptide peaks from each run were pooled and further purified by HPLC with a 0.1%/min acetonitrile gradient.

Protein methods

The methods for SDS/PAGE (performed under reducing conditions on Tris/tricine 12% polyacrylamide gels) and protein detection by staining with Coomassie Blue were essentially as described by Schagger and von Jagow [25], and biotinylated proteins were detected by Western blotting as described by Lim et al. [26]. Protein concentrations for SDS/PAGE were determined by Bradford assays [27], whereas concentrations of samples for N-terminal sequencing or MS were determined by either the microbiuret assay [28] or HPLC peak integration [29] depending on the origin of the sample. N-Terminal sequence analysis was performed by automated Edman degradation in an Applied Biosystems 470A Sequencer, and molecular-mass determination by MS was performed by electrospray ionization using a VG Quattro mass spectrometer. Quantification of Coomassie Blue-stained gels and Western blots was performed by laser densitometry with a Molecular Dynamics model 300A densitometer using the ImageQuant software. Before laser densitometry the nitrocellulose Western-blot filters were soaked in paraffin oil to make them translucent.

RESULTS

Isolation and sequencing of the *PYC2* gene

A 3.54 kb clone encompassing the 5' end of *PYC2* was isolated from a *S. cerevisiae* pEMBL Ye23 *Sau3A* genomic library by hybridization to a 5' 2.4 kb *Bgl*II fragment from *PYC1* containing the *PYC1* promoter and 1.87 kb of 5' coding region [18]. The *PYC2* clone obtained was selected on the basis of its strong hybridization to the *PYC1* probe, while having a different restriction pattern. This fragment was then subcloned into KS+ Bluescript (Stratagene) and the complete sequence of both strands was determined. This revealed that the fragment that had been isolated extended between the *Sau3A* sites at –675 (Figure 2) and +2866 (sites inclusive) relative to the first ATG. Subsequent sequencing of an overlapping 3' clone (described below) revealed that the 3'-terminal *Sau3A* site was contained within a *Bgl*II site.

Repeated attempts to isolate the 3' end of the *PYC2* gene from the same library using various *PYC* probes were unsuccessful. Instead, a clone containing the 3' end of the gene (p58) was isolated by screening a λEMBL3 genomic library with the 367 bp *PYC2* *Eco*RI fragment. The 3' end of this clone was then

subcloned as a 2.3 kb *Bam*HI–*Sal*I fragment into pHSG396 [30] and sequenced from both directions.

In total, we have sequenced the 4.3 kb of *PYC2* DNA containing the entire 3.54 kb ORF, 675 bp of 5' non-coding, and 97 bp of 3' non-coding sequence extending to the *Dra*I site at +3641 (Figure 2). The ORF encodes a protein of 1180 amino acids in length (Figures 2 and 3). Apart from the extreme non-coding ends of the DNA fragment, sequencing was performed on both strands. In all cases considerable care was taken to ensure that the sequence recorded was correct so that any differences observed between our data and the reported sequence could not be attributed to sequencing errors on our part.

Sequence comparisons between the various *PYC* isoforms

Comparison between our data and the reported sequence data for the *PYC2* gene

Comparing the *PYC2* non-coding regions, we found that in the 772 bp of non-coding DNA sequenced, there were a total of 16 differences giving a sequence identity of 97.9% (average of one sequence difference every 48 residues) between our data and the published sequence [19]. These included four insertions, three deletions and nine transversions. However, there were no sequence differences observed within either the incomplete TATAT sequence at –65 or the TATTTA sequence at –290. Furthermore, the putative binding sites for the regulatory factors ABF1/GF1 (–492), GRF2/REB1 (–496) and the CT block at –493 [19] all remained unchanged.

Comparison of the two ORFs revealed 36 sequence differences in a total of 3.54 kb, equivalent to one sequence difference every 98 bp or a sequence identity of approximately 99%. Of the 36 sequence differences, 35 were transversions (Table 1). Of the 35 transversions 25 (71%) occurred in the third or 'wobble' position. Consequently only three of these resulted in an amino acid change in the predicted protein sequence. The majority of the remaining ten transversions, being in either the first or second codon positions, resulted in changes in the predicted protein sequence giving a total of 12 amino acid differences due to transversions (Table 1 and Figure 3).

The most 3' of the sequence differences in the ORF was a deletion at +3533, which results in a frame shift, changing P-1178 into a glutamine (Q) and generating a stop codon 15 bp earlier (Figure 4a). Hence the protein predicted from the published sequence (1185 amino acids) has a five-amino-acid (VIFTR) extension at the C-terminus relative to that predicted on the basis of our sequence data (Figures 3 and 4a). The sequence of the *PYC2* non-coding strand from the gene isolated by us, at position +3533, is presented in Figure 4(b).

Comparisons between *PYC2* and *PYC1* and their protein products

The *PYC2* gene we have sequenced has an ORF 6 bp longer than the ORF of *PYC1* with an identity of 85.4%. At the protein level Pyc2 has 92.6% identity with Pyc1, having one additional amino acid at each of the N- and C-termini and 92 internal amino acid sequence differences (Figure 3).

Further chromosomal localization of the *PYC2* and *PYC1* genes

Previously we used the technique of hybridization (with a *PYC1* probe) of chromosome blots obtained from agarose gels containing yeast chromosomes separated out by pulse-field electrophoresis in order to localize the *PYC1* gene to chromosome VII, and *PYC2* to chromosome II [18]. To localize each of the *PYC* genes further we hybridized gene-specific probes to a set of prime λ-clone grid filters (kindly supplied by Dr. M. V. Olson) con-

Figure 2 Nucleotide and inferred protein sequence for the *PYC2* gene

The top sequence refers to the nucleotide sequence of the coding strand; the lower sequence refers to the inferred protein sequence, and the numbers on the right correspond to the positions for each of these respective sequences relative to the first ATG. Underlining indicates the positions of the flanking *Sau3A* and *Dral* restriction sites. Asterisk indicates stop codon.

taining an ordered set of λ -clones of known location on the physical map of *S. cerevisiae* [31].

The *PYC2* gene

The grid filters were hybridized with a 527 bp *PvuII*-*BglII* promoter fragment of *PYC2* (gene-specific), which was found to bind to filter 2 position D5 corresponding to clone 5564 from the

right arm of chromosome II (results not shown). The flanking clones that partially overlap with the proximal and distal ends of clone 5564 are clones 6471 (known to contain the *CDC28* gene) and 6734 respectively. The *PYC2* promoter fragment did not hybridize to clone 6471, but did hybridize to the DNA at position M7 on a supplementary filter corresponding to clone 6734. Hence these results suggest that the *PYC2* gene is contained within the small region of DNA in common to clones 5564 and

Figure 3 Comparison between the inferred protein sequence from our *PYC2* gene and the published sequences for Pyc2 and Pyc1

The primary sequence presented is that obtained from the *PYC2* gene reported here, and the residue positions are indicated by the numbers on the right. The letters connected to the primary sequence by open boxes indicate the corresponding residue present in Pyc1 [3], and those in the shaded boxes indicate the residues present in the published sequence for Pyc2 [19]. The letters in the shaded box directly following the terminus of the primary sequence indicate the residues in the five-amino-acid extension present in the published Pyc2 sequence. Asterisk indicates stop codon.

Table 1 Nucleotide and amino acid sequence differences in the coding region of *PYC2*

The first base listed in the change column refers to the base present in the (+) strand of the gene reported here, while the second base refers to the sequence of Stucka et al. [19]. The symbols listed in the amino acid change column refer to the amino acids encoded by the corresponding nucleotide sequences.

Nucleotide position	Nucleotide change	Codon position	Amino acid change	Amino acid no.
+43	A → T	1	S → C	15
+138	C → T	3		
+156	T → C	3		
+177	C → T	3		
+372	C → T	3		
+396	C → A	3	D → E	132
+408	C → T	3		
+411	A → G	3		
+465	T → C	3		
+714	C → A	3	N → K	238
+804	G → C	3	L → F	268
+1068	G → T	3		
+1320	T → G	3		
+1380	A → T	3		
+1431	G → A	3		
+1637	C → G	2	S → C	546
+1785	T → G	3		
+1815	T → G	3		
+1914	A → G	3		
+1925	A → C	2	N → T	642
+1926	C → T	3	N → T	
+2013	T → C	3		
+2145	C → T	3		
+2311	G → A	1	G → S	771
+2317	G → C	1	A → R	773
+2318	C → G	2	A → R	773
+2472	T → C	3		
+2490	C → T	3		
+2491	T → A	1	W → R	831
+2515	T → C	1	S → P	839
+2541	A → C	3		
+3001	T → A	1	Y → N	1001
+3372	A → G	3		
+3453	C → T	3		
+3464	A → G	2	K → R	1155

6734 on the right arm of chromosome II distal to the *CDC28* gene.

In the same year that we reported the chromosomal localization of the *PYC* genes by pulse-field gel electrophoresis, Stucka et al. [19] later reported the same findings, also using this method. In addition, they reported that *PYC2* was located just distal to the *DUR1,2* loci, i.e. on a cosmid clone (c411) which overlaps the distal end of another clone that they had earlier found to contain the *DUR1,2* loci and a *tRNA^{Glu3}* gene. Feldman et al. [32] have since published the complete DNA sequence of chromosome II assembled from the sequence data contributed by a consortium of 28 European laboratories, including the published sequence of *PYC2* [19]. According to their map, the *PYC2* gene is approximately 15 kb distal to the *DUR1,2* loci. The *PYC2* localization determined by hybridization to the 'Olson filters' reported here coincides exactly with the location revealed by DNA sequencing.

The *PYC1* gene

As the *PYC1* gene had not previously been localized to a specific region of chromosome VII, the same grid filters were used to

Figure 4 The single base insertion near the C-terminus of the published *PYC2* gene

(a) Schematic diagram showing the effect of the single base insertion on the protein sequence. The protein sequence in bold presented above corresponds to the nucleotide sequence for the gene reported here, and the protein sequence below the central nucleotide sequence (+ strand) shows the effect of the additional cytosine present at +3533 in the sequence reported by Stucka et al. [19]. The position of the additional cytosine is shown by the arrow. (b) Photograph of the DNA sequence of the negative strand at +3533. The arrow indicates the site at which the published sequence has an additional G residue. Asterisk indicates stop codon.

Figure 5 Biotinylation *in vivo* of the Pyc2 and Pyc1 biotin-domain peptides

Schematic diagram of the Pyc biotin-domain peptides used in the study of biotinylation *in vivo*. The numbers above each group of peptides indicate the amino acid positions in the native Pyc proteins, and the numbers below indicate the positions relative to the N-terminus of each expressed peptide. The N- and C-terminal residues and the key internal residues distinguishing the various peptide constructs are shown in capital lettering, and the shadowed capitals indicate the residues in Pyc2 predicted by the published sequence [19]. The dotted line indicates the position of the residues in the Pyc sequences that correspond to first and last 'structured' residues in the three-dimensional structure determined for the biotin domain of the BCCP subunit from *E. coli*. ACC (S. M. Brocklehurst, Y. N. Kalia, J. D. Kulman, A. Chapman-Smith, J. E. Cronan, Jr., J. C. Wallace and R. N. Perham, unpublished work).

localize this gene further. When the grid filters were hybridized to a 655 bp *HindIII*-*NdeI* promoter fragment from *PYC1*, the fragment hybridized to filter 2 position P4 and filter 3 position

Figure 6 Purification of the Pyc2 biotin-domain peptide

Duplicate protein samples were prepared from an aliquot taken at each stage of the purification of the C-terminal Pyc2 peptide. Samples were analysed by running duplicate Tris/tricine SDS/12% polyacrylamide gels. One gel (a) was Coomassie Blue stained for total proteins, and the other (b) used for the detection of biotinylated proteins by Western blotting and avidin alkaline detection as outlined in the Experimental section. Lys, induced lysate; QS, Q-Sepharose step; GF, gel-filtration (Sephacryl-S200) step; P1 and P2, HPLC peaks 1 and 2; BCCP, the purified and biotinylated 87-amino-acid C-terminal biotin-domain peptide from the BCCP of *E. coli* ACC [24]; Mw, molecular-mass markers.

K21 (results not shown), corresponding to the overlapping λ clones 6619 and 7086 respectively. Therefore the *PYC1* gene is contained within the region of DNA common to both of these clones, and as the *RAD6* gene has previously been localized to clone 6619, the *PYC1* gene is either at, or very close to, the *RAD6* locus.

Expression and biotinylation of the Pyc2 biotin-domain peptide

By expressing a Pyc2 C-terminal peptide containing the biotin domain we were able to: (a) purify the domain, determine its molecular mass by MS, and thus confirm the nature of the C-terminus of the enzyme expressed from our *PYC2* gene; (b) determine whether the biotin domain from a yeast enzyme such as Pyc2 can be biotinylated by *E. coli* biotin ligase.

As the *E. coli* T7 pET expression system (pET-16B) gives good levels of expression of the biotin domain from the biotin carboxyl carrier protein (BCCP) subunit of *E. coli* acetyl-CoA carboxylase (ACC) [24], we engineered a construct to express the Pyc2 biotin domain in this system. The C-terminus of Pyc2 containing the entire biotin domain (c86K; Figure 5) was expressed as an 86-amino-acid recombinant peptide from an engineered methionine residue (K1095M) to the native C-terminus. Expression was induced with IPTG in the presence of excess biotin (10 μ M) and biotin ligase (*E. coli* *birA* gene product), with the excess biotin ligase being supplied from the *araBAD* promoter on the plasmid pCY216 in the presence of arabinose [24]. Production of the Pyc2 biotin-domain peptide in the induced lysate was observed by SDS/PAGE (Lys in Figure 6a), and avidin alkaline phosphatase

detection after Western transfer (Lys in Figure 6b) clearly demonstrated that the yeast Pyc2 biotin domain was indeed biotinylated *in vivo* by *E. coli* biotin ligase.

When expressing the BCCP subunit of *E. coli* ACC using the same system, we found that significant amounts of the expressed peptide remained as apoprotein even in the presence of excess biotin and biotin ligase, apparently because of insufficient ATP being available for the ligase reaction [24]. Hence, by completing the biotinylation of the Pyc2 biotin domain *in vitro* (with excess ATP) we were able to label the peptide with [^{3} H]biotin, which facilitated its detection during purification.

Purification, N-terminal sequencing and MS analysis of the Pyc2 biotin-domain peptide

Small-scale purification of the Pyc2 biotin-domain peptide was performed in order to obtain sufficient material for N-terminal sequencing and molecular-mass determination by electrospray MS. The Pyc2 peptide was eluted from the Q-Sepharose anion-exchange column at 0.25 M NaCl. Analysis of the eluate by SDS/PAGE (QS in Figure 6a) and Western blotting (QS in Figure 6b) showed that the Q-Sepharose step had removed all the other low-molecular-mass proteins (below about 17 kDa), and the majority of the larger proteins as well. After the gel-filtration step, the Pyc2 biotin-domain peptide was found to be reasonably pure (GF in Figure 6a). Pure material was ultimately obtained as a single peak on C4 reverse-phase HPLC (P1 in Figures 6a and 6b), with a minor peak containing a small quantity of the peptide eluted at a higher acetonitrile concentration (P2 in Figure 6).

N-Terminal sequencing of the first ten residues of the purified material confirmed that the N-terminus of the expressed Pyc2 biotin-domain peptide was in agreement with the predicted protein sequence (results not shown), with the expected removal of the N-terminal methionine by *E. coli* methionyl-aminopeptidase [33]. Furthermore, electrospray MS of the pure peptide gave a molecular mass of 9234.9 Da, which is within 0.01 % of the predicted value of 9235.7 Da for the c86K peptide, and 620.6 atomic mass units below the predicted value for the c91R peptide (Figure 5). Taken together, these data provide convincing evidence for the amino acid sequence predicted for the C-terminus of Pyc2 (Figure 2), and hence the DNA sequence determined for this region of the *PYC2* gene isolated here (Figure 2).

Comparative *in vivo* biotinylation of various Pyc biotin-domain peptides

We have engineered constructs to express the four Pyc2 biotin-domain peptides covering all possible combinations of the two sequence differences found in this region of the *PYC2* DNA sequence (Figure 3). In addition, we engineered constructs to express two different-length Pyc1 biotin-domain peptides. A schematic diagram of each of these peptides is presented in Figure 5.

Induced and uninduced lysates were prepared from cells containing each of the constructs essentially as previously described [24], in the presence of excess biotin and *E. coli* biotin ligase. Samples of each lysate were subjected to SDS/PAGE in duplicate; one gel was Coomassie Blue stained and the other was used for Western blotting and detection of biotinylated proteins with an avidin alkaline phosphatase probe. The extent of biotinylation of each peptide *in vivo* was estimated by determining the intensity of the biotinylated band detected with avidin alkaline phosphatase by laser densitometry. This value was then divided by the intensity of the Coomassie Blue-stained band of expressed protein, also quantified by laser densitometry. To ensure that

Table 2 Comparative *in vivo* biotinylation of the Pyc peptides

Lysates were prepared (see the Experimental section) from cells expressing each of the Pyc peptides at an appropriate level for measuring the degree of biotinylation *in vivo*. Samples of each lysate were run on Tris/tricine 12% polyacrylamide gels in duplicate with one gel Coomassie Blue stained for total protein and the other used for Western blotting and avidin alkaline phosphatase detection. These data represent the average obtained from a number of such gels and Western blots, by quantifying and comparing the amount of Coomassie Blue-stained protein with the amount of biotinylated material detected on the Western blots as described in the text.

		Relative biotinylation <i>in vivo</i> (arbitrary units)
Pyc2	c86K	1.0
	c86R	1.2
	c91K	2.4
	c91R	4.0
Pyc1	c85	1.5
	c104	9.2

meaningful comparisons were made, the constructs being compared were all analysed on the same gels and Western blots, and the experiment was repeated a number of times. Comparison of induced and uninduced lysates showed very little endogenous *E. coli* protein co-migrating with the expressed peptides (results not shown), and within the concentration range investigated, the individual domains and their biotinylated and unbiotinylated forms appear to stain equally well with Coomassie Blue. The only endogenous biotinylated protein in *E. coli* is the BCCP subunit of ACC which has a molecular mass of 16.7 kDa [7] and thus is clearly distinguishable from the ~10 kDa peptides expressed in the Tris/tricine gel system [24].

Experiments with BCCP-87 [24] and the Pyc2 c86K peptide showed that with T7-expressed biotin domains some of the biotin-domain material remains unbiotinylated unless the reaction is forced to completion *in vitro* by the addition of excess ATP, even when the peptides are expressed in the presence of excess biotin and *E. coli* biotin ligase expressed from the strong *araB* promoter. However, analysis of samples from a number of the Pyc constructs expressed in the same system revealed that the proportion of peptide biotinylated *in vivo* for a given peptide remained relatively constant with increasing expression provided that the amount of expressed peptide was less than 20% of the total cellular protein (results not shown). Consequently, calculations of the extent of biotinylation *in vivo* for each of the peptides were made using lysates from cultures in which the cells were expressing a quantity of peptide within this range. Thus, whereas the level of expression of the compared domains was not necessarily the same, the extent of biotinylation of a given peptide was not detectably affected by the level of expression under the conditions used.

The comparative extent of biotinylation *in vivo* for the various Pyc peptides calculated from a number of such gels and Western blots is shown in Table 2. Comparing biotinylation of the four Pyc2 constructs, it appears that the peptide containing the biotin-domain sequence from the gene reported here (c86K) is the most poorly biotinylated. Both of the sequence differences in the biotin domain corresponding to the DNA sequence reported by Stucka et al. [19] improved biotinylation of the Pyc2 biotin domain *in vivo*. The K1155R change in c86R produced a slight increase in biotinylation on its own, but the Q1178P/VIFTR extension in c91K produced a 2-fold increase. When both of these changes were combined in the one peptide (c91R; the

peptide corresponding to the DNA sequence of Stucka et al. [19]), the resultant peptide had a 4-fold increase in biotinylation. Comparison of the extent of biotinylation of the c85 and c104 Pyc1 peptides suggests that the 19-amino-acid N-terminal extension present in c104 produced a 6-fold increase in biotinylation.

DISCUSSION

Moss'e and co-workers [34] recently published the compilation of a sequence database for *S. cerevisiae* containing 1001 reported nucleotide sequences for 818 non-allelic genes. Although most of the genes had only been sequenced once, there were 149 genes for which more than one sequence had been determined. Of these, 30 genes showed differences at the nucleotide level within the ORF, but were totally identical at the predicted amino acid level (silent changes). They concluded that these silent changes were probably due to natural polymorphisms. However, by virtue of their frequency, it could be expected that at least some of the non-silent changes observed with the other genes could also be due to polymorphisms rather than sequencing errors. Furthermore, there are some examples in the literature of genes from different laboratory strains of *S. cerevisiae* having polymorphic variation at the amino acid level, e.g. *URA3* [35] and *POP2* [36]. In the case of the *PYC2* gene, we found that 61% (22 of 36; Table 1) of the nucleotide differences in the coding region were silent mutations, consistent with their being due to natural polymorphisms.

Considering the predicted amino acid substitutions in the coding region, only two of the 13 are strictly conservative. The majority of the 11 non-conservative amino acid changes either occur infrequently with amino acids in equivalent environments or are unlikely or avoided on the basis of residue change allowing the same global protein topology [37]. Consequently, two such Pyc2 enzyme variants might well be expected to have different activities, leading to differences in growth phenotype between the respective *pyc1/PYC2* null mutant strains. Hence the sequence differences may account for the reported differences in Pyc2 specific activity (relative to Pyc1), and the differences in aspartate dependence of the *pyc1/PYC2* single null mutants [18,19].

Although all known biotin enzymes have methionine residues flanking the biotinylated lysine, these methionine residues are not required for biotinylation. Instead, mutagenesis of these residues affects the ability of the biotin moiety to participate in the carboxyl-transfer reactions [38,39]. Similarly, changing either of the flanking residues in the analogous DKA sequence present in all lipoyl enzymes was shown to affect the reductive acylation of the lipoyl moiety, rather than lipoylation itself [40]. Changing the DKA in the PDH lipoyl domain to MKM did not result in biotinylation instead of lipoylation, but displacing the K from its position at the tip of the β -hairpin loop by one residue in either direction destroyed lipoylation [40]. Furthermore, the G70N mutation in the bovine H-protein, which is located at the first residue of the β -strand (no. 6) adjacent to the C-terminus of the protein [16], is known to destroy the lipoylation of this domain [41]. These results suggest that, in addition to the exposed position of the recipient lysine at the top of the β -hairpin loop, other features of the domains are significant for correct interaction with the respective ligase enzyme.

By expressing Pyc2 biotin domains we have been able to assess the effects of the two differences in this domain on its ability to be biotinylated by *E. coli* biotin ligase *in vivo*. Although expression of the biotin domains in the presence of yeast biotin ligase would be ideal, biotin ligase enzymes are well known to biotinylate acceptor proteins from very different species [5,10]. Thus we

consider that the results presented here using a heterologous biotin ligase are indicative of real variation in the ability of the different peptides to act as biotin acceptors. Moreover, as biotin domains have been shown to be stable independently folding domains [5,8,42], a sequence change that affects the biotinylation of biotin-domain peptides, in this case the K1155R substitution, may have a similar effect on biotinylation of the whole protein, and hence may influence the relative specific activities of the different isoforms *in vivo*.

In addition, both the five-amino-acid extension at the C-terminus of Pyc2 and the 19-amino-acid N-terminal extension of Pyc1 produced a noticeable increase in biotinylated peptide. This was unexpected, as the C-terminus of Pyc2 expressed from the gene isolated here already extends some ten residues past the C-terminus of the biotin-domain core structure, and the 19-amino-acid extension of Pyc1 was to an N-terminus already eight residues upstream of the beginning of the structure ([17]; S. M. Brocklehurst, Y. N. Kalia, J. D. Kulman, A. Chapman-Smith, J. E. Cronan, Jr., J. C. Wallace and R. N. Perham, unpublished work; see Figure 5). However, the N- and C-termini of the Pyc1 (and presumably Pyc2) biotin domain are adjacent in the predicted three-dimensional structure on the opposite side of the molecule to the biotinylated lysine [17], as is the case with the recently solved NMR structure of the biotin domain from the BCCP subunit of *E. coli* ACC (S. M. Brocklehurst, Y. N. Kalia, J. D. Kulman, A. Chapman-Smith, J. E. Cronan, Jr., J. C. Wallace and R. N. Perham, unpublished work). Although it is possible that the effect of these extensions on biotinylation reflects the inherent stability of the isolated domains rather than their interaction with biotin ligase *per se*, it is clear that the C- and N-termini contribute to the efficient biotinylation of biotin-domain peptides. A definitive understanding of the role of N- and C-terminal extensions to the Pyc peptides in relation to the interaction with biotin ligase requires determination of the structure of the yeast Pyc biotin domain, together with purification of unbiotinylated biotin-domain peptides and examination of their stability and the kinetics of the ligase interaction *in vitro*.

This work was supported by an Australian Research Council Grant A09131713 to J. C. W. D. L. V. was the recipient of an Australian Postgraduate Research Award. The pEMBL Ye23 *Sau* 3A genomic library from *S. cerevisiae* strain S288C was obtained from Dr. C. P. Morris, Department of Chemical Pathology, Adelaide Womens and Childrens Hospital, Adelaide, SA, Australia. The λ -prime grid filters were supplied by Dr. Linda Riles formerly of the laboratory of Dr. M. V. Olson, Department of Molecular Biotechnology, University of Washington School of Medicine, Seattle, WA 98195, U.S.A. Electrospray MS was performed by members of the Mass Spectroscopy Laboratory, School of Chemical Sciences, University of Illinois. The VG Quattro mass spectrometer was purchased in part with a grant from the Division of Research Resources, National Institutes of Health (RR 07141).

REFERENCES

- 1 Attwood, P. V. and Keech, D. B. (1984) *Curr. Top. Cell. Regul.* **23**, 1–55
- 2 Knowles, J. R. (1989) *Annu. Rev. Biochem.* **58**, 195–221
- 3 Lim, F., Morris, C. P., Occhiodoro, F. and Wallace, J. C. (1988) *J. Biol. Chem.* **263**, 11493–11497
- 4 Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C. and Wood, H. G. (1988) *J. Biol. Chem.* **263**, 6461–6464
- 5 Cronan, J. E., Jr. (1990) *J. Biol. Chem.* **265**, 10327–10333
- 6 Kondo, H., Shiratsuchi, K., Yoshimoto, Y. et al. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9730–9733
- 7 Li, S.-J. and Cronan, J. E., Jr. (1992) *J. Biol. Chem.* **267**, 855–863
- 8 Reed, K. E. and Cronan, J. E., Jr. (1991) *J. Biol. Chem.* **266**, 11425–11428
- 9 Murtif, V. L. and Samols, D. (1987) *J. Biol. Chem.* **262**, 11813–11816
- 10 McAllister, H. C. and Coon, M. J. (1966) *J. Biol. Chem.* **241**, 2855–2861
- 11 Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46
- 12 Reed, L. J. and Hackett, M. L. (1990) *J. Biol. Chem.* **265**, 8971–8974
- 13 Perham, R. N. (1991) *Biochemistry* **30**, 8501–8512
- 14 Dardel, F., Davis, A. L., Laue, E. D. and Perham, R. N. (1993) *J. Mol. Biol.* **229**, 1037–1048
- 15 Green, J. F. D., Laue, E. D., Perham, R. N., Ali, S. T. and Guest, J. R. (1995) *J. Mol. Biol.* **248**, 328–343
- 16 Pares, S., Cohen-Addad, C., Sieker, L., Neuberger, M. and Douce, R. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4850–4853
- 17 Brocklehurst, S. M. and Perham, R. N. (1993) *Protein Sci.* **2**, 626–639
- 18 Walker, M. E., Val, D. L., Rohde, M., Devenish, R. J. and Wallace, J. C. (1991) *Biochem. Biophys. Res. Commun.* **176**, 1210–1217
- 19 Stucka, R., Dequin, S., Salmon, J. and Gancedo, C. (1991) *Mol. Gen. Genet.* **229**, 307–315
- 20 Brewster, N. K., Val, D. L., Walker, M. E. and Wallace, J. C. (1994) *Arch. Biochem. Biophys.* **310**, 1–10
- 21 Studier, F. W. and Moffat, B. A. (1986) *J. Mol. Biol.* **189**, 113–130
- 22 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 23 Kunkel, T. A., Roberts, J. D. and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
- 24 Chapman-Smith, A., Turner, D. L., Cronan, J. E., Jr. and Wallace, J. C. (1994) *Biochem. J.* **302**, 831–837
- 25 Schagger, H. and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
- 26 Lim, F., Rhode, M., Morris, C. P. and Wallace, J. C. (1987) *Arch. Biochem. Biophys.* **258**, 259–264
- 27 Bradford, M. (1976) *Anal. Biochem.* **72**, 248–254
- 28 Munkres, K. D. and Richards, F. M. (1965) *Arch. Biochem. Biophys.* **109**, 466–478
- 29 Buck, M. A., Olah, T. A., Weitzmann, C. J. and Cooperman, B. S. (1989) *Anal. Biochem.* **182**, 295–299
- 30 Takeshita, S., Sato, M., Tabo, M., Masahashi, W. and Hashimoto-Gotoh, T. (1987) *Gene* **61**, 63–74
- 31 Riles, L., Dutchik, J. E., Baktha, A. et al. (1993) *Genetics* **134**, 81–150
- 32 Feldman, H., Aigle, M., Aljinovic, G. et al. (1994) *EMBO J.* **13**, 5795–5809
- 33 Hirel, P.-H., Schmitter, D. P., Fayat, G. and Blanquet, S. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8247–8251
- 34 Moss'e, M.-O., Linder, P., Lasowska, J. and Slonimski, P. P. (1993) *Curr. Genet.* **23**, 66–91
- 35 Rose, M., Grisafi, P. and Boltstein, D. (1984) *Gene* **29**, 113–124
- 36 Sakai, A., Chibazakura, T., Shimizu, Y. and Hishinuma, F. (1992) *Nucleic Acids Res.* **20**, 6227–6233
- 37 Bordo, D. and Argos, P. (1991) *J. Mol. Biol.* **217**, 721–729
- 38 Kondo, H., Uno, S., Komizo, Y. and Sunamoto, J. (1984) *Int. J. Pept. Protein Res.* **23**, 559–564
- 39 Shenoy, B. C., Xie, Y., Park, V. L. et al. (1992) *J. Biol. Chem.* **267**, 18407–18412
- 40 Wallis, N. G. and Perham, R. N. (1994) *J. Mol. Biol.* **236**, 209–216
- 41 Fujiwara, K., Okamura-Ikeda, K. and Motokawa, Y. (1994) *FEBS Lett.* **293**, 115–118
- 42 Fall, R. R. and Vagelos, P. R. (1973) *J. Biol. Chem.* **248**, 2078–2088